PREFACE

This volume contains the full-length papers presented in the 2nd International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2017) that was held on June 15-17, 2017 on the Island of Rhodes, Greece.

UNCECOMP 2017 is a Thematic Conference of ECCOMAS and a Special Interest conference of the International Association for Structural Safety & Reliability (IASSAR).

The objective of UNCECOMP 2017 Conference is to act as a forum for the recent research efforts and progress towards analysis and design processes under uncertainty, with emphasis in multiscale analysis and design of complex systems. The aim of the conference is to bring together researchers seeking an interaction of stochastic methods and computational mechanics in order to obtain reliable predictions of the behavior of physical systems. The UNCECOMP conference serves as a forum for facilitating the exchange of ideas and as a platform for establishing links between research groups with complementary activities.

The UNCECOMP 2017 Conference is supported by the National Technical University of Athens (NTUA), the European Committee on Computational Solids and Structural Mechanics of ECCOMAS (ECCSM) and the Greek Association for Computational Mechanics (GRACM).

The editors of this volume would like to thank all authors for their contributions. Special thanks go to the colleagues who contributed to the organization of the Minisymposia and to the reviewers who, with their work, contributed to the scientific quality of this e-book.

M. Papadrakakis
National Technical University of Athens, Greece

V. Papadopoulos
National Technical University of Athens, Greece

G. Stefanou
Aristotle University of Thessaloniki, Greece
ACKNOWLEDGEMENTS

The conference organizers acknowledge the support towards the organization of the “2nd International Conference on Uncertainty Quantification in Computational Sciences and Engineering”, to the following organizations:

- European Community on Computational Methods in Applied Sciences (ECCOMAS)
- European Committee on Computational Solids and Structural Mechanics of ECCOMAS (ECCSM)
- International Association for Structural Safety & Reliability (IASSAR)
- School of Civil Engineering, National University of Athens (NTUA)
- Greek Association for Computational Mechanics (GRACM)
- Municipality of Rhodes

Plenary Speakers and Invited Session Organizers

We would also like to thank the Plenary and Semi-Plenary Speakers and the Minisymposia Organizers for their help in the setting up of a high standard Scientific Programme.

Plenary Speakers: Jie Li, Arvid Naess, Martin Ostoja-Starzewski, Christian Soize

Semi-Plenary Speakers: Sondipon Adhikari, Siu-Kui Au, Hermann G. Matthies, David Moens, Thiago Ritto, Alexandros Taflanidis

SUMMARY

Preface.................................................................................................................................................. iii

Acknowledgements.................................................................................................................................. iv

Contents.................................................................................................................................................. vii

Plenary Lectures .................................................................................................................................... 1

Minisymposia

MS 1: UNCERTAINTY QUANTIFICATION IN VIBRATION BASED MONITORING AND STRUCTURAL
DYNAMICS SIMULATIONS............................................................................................................................ 16
Organized by Eleni Chatzi, Costas Papadimitrou

MS 2: ENGINEERING ANALYSES WITH VAGUE AND IMPRECISE INFORMATION ............................................. 90
Organized by Edoardo Patelli, Bruno Sudret, Matteo Broggi, Michael Beer

MS 3: MULTISCALE ANALYSIS AND DESIGN OF RANDOM HETEROGENEOUS MEDIA ...................................... 111
Organized by George Stefanou, Dimitrios Savvas, Vissarion Papadopoulos

MS 4: ADVANCED SIMULATION-BASED APPROACHES TO UNCERTAINTY QUANTIFICATION AND
RELIABILITY ANALYSIS .......................................................................................................................... 130
Organized by Edoardo Patelli, Francisco Alejandro Diaz De la O, Siu-Kui Au, Michael Shields

MS 5: SURROGATE MODELS FOR UNCERTAINTY QUANTIFICATION, RELIABILITY ANALYSIS AND ROBUST
DESIGN ..................................................................................................................................................... 187
Organized by Stefano Marelli, Bruno Sudret, Sankaran Mahadevan, Alex Taflanidis

MS 6: SOFTWARE FOR UNCERTAINTY QUANTIFICATION .............................................................................. 212
Organized by Stefano Marelli, Edoardo Patelli, Brian M. Adams, Bruno Sudret

MS 7: NON-PROBABILISTIC APPROACHES FOR UNCERTAINTY REPRESENTATION AND ANALYSIS IN
ENGINEERING .......................................................................................................................................... 269
Organized by David Moens, Dirk Vandepitte, Michael Hanss, Alba Sofi

MS 8: BAYESIAN ANALYSIS OF NUMERICAL MODELS .................................................................................. 339
Organized by Iason Papaioannou, Daniel Straub, Costas Papadimitriou

MS 10: CURRENT TOPICS IN UNCERTAINTY CHARACTERIZATION ................................................................. 381
Organized by D. T. Hristopulos, I. C. Tsantili

MS 11: UNCERTAINTY QUANTIFICATION ACROSS MULTIPLE SCALES FOR SOLID MECHANICS ......................... 406
Organized by Savvas Triantafyllou, Eleni Chatzi, Manolis Chatzis

MS 12: COMPUTATIONAL STOCHASTIC MULTISCALE MODELLING .................................................................. 419
Organized by Paul Steinmann, Dmytro Pivovarov
Regular Sessions

RS 6: EFFICIENT STOCHASTIC SIMULATION METHODS .................................................................426
RS 12: UNCERTAINTY QUANTIFICATION ..................................................................................543
RS 16: STOCHASTIC FRACTURE AND DAMAGE .....................................................................605
RS 17: SYSTEM RELIABILITY ANALYSIS, DESIGN AND RISK ASSESSMENT ......................627
RS 21: STOCHASTIC OPTIMIZATION METHODS ...................................................................690
CONTENTS

Semi-plenary Lectures

PROBABILISTIC LEARNING ON MANIFOLD FOR OPTIMIZATION UNDER UNCERTAINTIES ........................................ 1
C. Soize, R. Ghanem

Minisymposia

MS 1: UNCERTAINTY QUANTIFICATION IN VIBRATION BASED MONITORING AND STRUCTURAL DYNAMICS SIMULATIONS

OBTAINING HIGH ACCURACY VIBRATION MEASUREMENTS WITH LOW-COST SENSORS USING BAYESIAN VIRTUAL SENSING ................................................................................................................................. 16
Jyrki Kullaa

FINITE ELEMENT MODEL DEVELOPED AND MODAL ANALYSIS OF LARGE SCALE STEAM TURBINE ROTOR: QUANTIFICATION OF UNCERTAINTIES AND MODEL UPDATING .............................................................. 32
Dimitrios Giagopoulos, Alexandros Arailopoulos, Ilias Zacharakis, Eleni Pipili

OPERATIONAL MODAL ANALYSIS OF BRODIE TOWER USING A BAYESIAN APPROACH ...................................... 45
Yan-Long Xie, Yi-Chen Zhu, Siu-Kui Au

DATA–DRIVEN IDENTIFICATION, CLASSIFICATION AND UPDATE OF DECISION TREES FOR MONITORING AND DIAGNOSTICS OF WIND TURBINES ...................................................................................... 55
Imad Abdallah, Vasilis Dertimanis, Eleni Chatzi

MODEL UPDATING OF A NONLINEAR EXPERIMENTAL VEHICLE USING SUBSTRUCTURING AND UNSCENTED KALMAN FILTERING .................................................................................................................. 66
Maria Tsotalou, Dimitrios Giagopoulos, Vasilis Dertimanis, Eleni Chatzi

RELIABILITY PREDICTION OF FATIGUE DAMAGE ACCUMULATION ON WIND TURBINES SUPPORT STRUCTURES .. 76
Konstantinos Tatsis, Eleni Chatzi, Eliz-Mari Lourens

MS 2: ENGINEERING ANALYSES WITH VAGUE AND IMPRECISE INFORMATION

EFFICIENT RELIABILITY AND UNCERTAINTY ASSESSMENT ON LIFELINE NETWORKS USING THE SURVIVAL SIGNATURE ................................................................................................................................. 90
Geng Feng, Sean Reed, Edoardo Patelli, Michael Beer, Frank P.A. Coolen

RANDOM VIBRATION ANALYSIS OF A SINGLE DEGREE OF FREEDOM SYSTEM UNDER IMPRECISE PROBABILITY ASSIGNMENTS ............................................................................................................. 100
Alice Cicirello, Robin S. Langley
MS 3: MULTISCALE ANALYSIS AND DESIGN OF RANDOM HETEROGENEOUS MEDIA

PROBABILISTIC REDUCED-ORDER MODELING FOR STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS .............. 111
Constantin Grigo, Phaedon-Stelios Koutsourelakis

MS 4: ADVANCED SIMULATION-BASED APPROACHES TO UNCERTAINTY QUANTIFICATION AND RELIABILITY ANALYSIS

PERFORMANCE PREDICTIONS OF COASTAL DEFENCES USING STOCHASTIC DETERIORATION MODELLING .......... 130
Mehrdad Bahari Mehrabani, Hua-Peng Chen

FLOOD RISK ASSESSMENT OF MASONRY ARCH BRIDGES ........................................................................ 140
Enrico Tubaldi, Lorenzo Macorini, Bassam Izzuddin

SAMPLING SCHEMES FOR HISTORY MATCHING USING SUBSET SIMULATION ........................................ 154
Zitong Gong, Francisco Alejandro Díaz De la O, Michael Beer

MARKOV CHAIN MONTE CARLO METHODS FOR UNCERTAINTY PROPAGATION AND RELIABILITY ANALYSIS ...... 165
Carsten Proppe

ZERO-VARIANCE SIMULATED ANNEALING FOR BAYESIAN SYSTEM IDENTIFICATION ................................. 176
Peter L. Green

MS 5: SURROGATE MODELS FOR UNCERTAINTY QUANTIFICATION, RELIABILITY ANALYSIS AND ROBUST DESIGN

AN EFFICIENT METHODOLOGY FOR THE ANALYSIS AND MODELING OF COMPUTER EXPERIMENTS WITH LARGE NUMBER OF INPUTS .................................................................................. 187
Bertrand Iooss, Amandine Marrel

ADAPTIVE EMULATION-BASED RELIABILITY ANALYSIS ............................................................................ 198
P.O. Hristov, F.A. DiazDelaO, K.J. Kubiak, U. Farooq

MS 6: SOFTWARE FOR UNCERTAINTY QUANTIFICATION

COSSAN SOFTWARE: A MULTIDISCIPLINARY AND COLLABORATIVE SOFTWARE FOR UNCERTAINTY QUANTIFICATION ........................................................................................................... 212
Edoardo Patelli, Matteo Broggi, Silvia Tolo, Jonathan Sadeghi

GAUSSIAN PROCESS RESPONSE SURFACE MODELING AND GLOBAL SENSITIVITY ANALYSIS USING NESSUS .... 225
John M. McFarland, John A. Dimeo, Barron J. Bichon
THE GRAPHICAL USER INTERFACE OF OPENTURNS, A UQ SOFTWARE IN SIMULATION ........................................ 238
Michaël Baudin, Anne Dutfoy, Anthony Geay, Anne-Laure Popelin, Aurélié Ladier, Julien Schueller, Thierry Yalamas

AERODYNAMIC SHAPE OPTIMIZATION UNDER FLOW UNCERTAINTIES USING NON-INTRUSIVE POLYNOMIAL
CHAOS AND EVOLUTIONARY ALGORITHMS ........................................................................................................... 258
Athanasios G. Liatsikouras, Varvara G. Asouti, Kyriakos C. Giannakoglou, Guillaume Pierrot, Mustafa Megahed

MS 7: NON-PROBABILISTIC APPROACHES FOR UNCERTAINTY REPRESENTATION AND ANALYSIS IN ENGINEERING

IDENTIFICATION AND QUANTIFICATION OF MULTIVARIATE POSSIBILISTIC MANUFACTURING UNCERTAINTY
IN DYNAMIC NUMERICAL MODELS .......................................................................................................................... 269
Matthias Faes, Dirk Vandepitte, David Moens

AN APPROACH TO ROBUST DESIGN OF A CRUMPLE-ZONE STRUCTURE USING FUZZY ARITHMETIC .............. 279
Markus Mäck, Michael Hanss

INTERVAL DYNAMIC RESPONSE OF DISCRETIZED STRUCTURES WITH UNCERTAIN-BUT-BOUNDED
PARAMETERS ......................................................................................................................................................... 289
Filippo Giunta, Giuseppe Muscolino, Alba Sofi

RESPONSE OF BEAMS WITH CRACK OF UNCERTAIN-BUT BOUNDED DEPTH SUBJECTED TO DETERMINISTIC
OR STOCHASTIC LOADS ........................................................................................................................................... 300
Giuseppe Muscolino, Roberta Santoro

MODELLING OF MATERIAL PARAMETER UNCERTAINTY OF RESONANCE WOOD USED IN VIOLINS USING
THE INTERVAL FIELD .............................................................................................................................................. 314
Maurice Imholz, Dirk Vandepitte, David Moens

REAL-TIME FUZZY ANALYSIS OF MACHINE DRIVEN TUNNELING ..................................................................... 329
Ba Trung Cao, Steffen Freitag, Günther Meschke

MS 8: BAYESIAN ANALYSIS OF NUMERICAL MODELS

BAYESIAN UPDATING FOR PROBABILISTIC CLASSIFICATION USING RELIABILITY METHODS ...................... 339
Paul Byrnes, Francisco Alejandro DiazDelaO

A BAYESIAN APPROACH TO MODEL CALIBRATION AND PARAMETER ESTIMATION IN POTENTIAL DROP
MEASURING .............................................................................................................................................................. 350
Thomas Berg, Sven von Ende, Rolf Lammering

PARAMETER IDENTIFIABILITY THROUGH INFORMATION THEORY ..................................................................... 372
Giovanni Capellari, Eleni Chatzi, Stefano Mariani
MS 10: CURRENT TOPICS IN UNCERTAINTY CHARACTERIZATION

INERENCE AND SENSITIVITY IN STOCHASTIC WIND POWER FORECAST MODELS. .......................................................... 381
Soumaya Elkantassi, Evangelia Kalligiannaki, Raul Tempone

FROM ATOMISTIC TO SYSTEMATIC COARSE-GRAINED MODELS FOR MOLECULAR SYSTEMS .............................. 394
Vagelis Harmandaris, Evangelia Kalligiannaki, Markos Katsoulakis, Petr Plechac

MS 11: UNCERTAINTY QUANTIFICATION ACROSS MULTIPLE SCALES FOR SOLID MECHANICS

MULTISCALE SURROGATE MODELING AND UNCERTAINTY QUANTIFICATION FOR PERIODIC COMPOSITE STRUCTURES .................................................................................................................. 406
Charilaos Mylonas, Bemetz Valentin, Eleni Chatzi

MS 12: COMPUTATIONAL STOCHASTIC MULTISCALE MODELLING

CONSTRUCTION OF STATISTICALLY SIMILAR RVES FOR THE QUANTIFICATION OF UNCERTAINTY ASSOCIATED WITH THE MATERIAL’S MICROSTRUCTURE MORPHOLOGY .......................................................................................................................... 419
Daniel Balzani, Niklas Miska, Stefan Prüger

Regular Sessions

RS 6: EFFICIENT STOCHASTIC SIMULATION METHODS

STOCHASTIC GALERKIN PROJECTION AND NUMERICAL INTEGRATION FOR STOCHASTIC SYSTEMS OF EQUATIONS .......................................................................................................................... 426
Markus Wahlsten, Jan Nordström

APPROACH TO PROVE THE EFFICIENCY OF THE MONTE CARLO METHOD COMBINED WITH THE ELEMENTARY EFFECT METHOD TO QUANTIFY UNCERTAINTY OF A BEAM STRUCTURE WITH PIEZO-ELASTIC SUPPORTS .......... 441
Sushan Li, Benedict Goetz, Maximilian Schaeffner, Roland Platz

A CONTRIBUTION TO THE EVALUATION OF IMPRECISE AVAILABILITY OF COMPLEX SYSTEMS USING MARKOV MODELS .......................................................................................................................... 456
Joanna Akrouche, Mohamed Sallak, Eric Châtelet, Fahed Abdallah, Hiba Haj Chehade

FORMULATION OF POTENTIAL FOR DYNAMICAL PARTICLE SYSTEM APPLIED TO MONTE CARLO SAMPLING ........ 467
Jan Mašek, Miroslav Vořechovský
VORONOI WEIGHTING OF SAMPLES IN MONTE CARLO INTEGRATION .......................................................... 478
Miroslav Vořechovský, Václav Sadílek, Jan Eliáš

BAYESIAN FORMULATIONS FOR FORCE RECONSTRUCTION PROBLEMS .................................................. 492
Mathieu Aucejo, Olivier De Smet

MLS BASED SEQUENTIAL SRSM IN SPARSE GRID FRAMEWORK FOR EFFICIENT UNCERTAINTY
QUANTIFICATION ................................................................................................................................. 507
Amit Kumar Rathi, Arunasis Chakraborty

MULTILEVEL PRECONDITIONING OF POLYNOMIAL CHAOS METHOD FOR QUANTIFYING UNCERTAINTIES IN
A BLOOD PUMP .................................................................................................................................. 517
Chen Song, Vincent Heuveline

REDUCED ORDER MODEL-BASED UNCERTAINTY MODELING OF STRUCTURES WITH LOCALIZED RESPONSE ...... 531
Pengchao Song, Marc Mignolet

RS 12: UNCERTAINTY QUANTIFICATION

DIFFERENT VIEWS ON ADDITIONAL RANDOM PARAMETERS IN EXPERIMENT DESIGN FOR THERMOPHYSICAL
PARAMETERS ESTIMATION .................................................................................................................. 543
Daniela Jaruskova, A. Kucerova

MULTIDIMENSIONAL STOCHASTIC MATERIAL MODELING AT LARGE DEFORMATIONS CONSIDERING
PARAMETER CORRELATIONS .............................................................................................................. 554
Eduard Penner, Ismail Caylak, Rolf Mahnken

A POSSIBILISTIC APPROACH FOR LINEAR ISOTROPIC ELASTICITY USING THE FUZZY FINITE ELEMENT METHOD .. 565
Alex Dridger, Ismail Caylak, Rolf Mahnken

A COUPLED NUMERICAL AND LASER OPTICAL METHOD FOR ON-SITE CALIBRATION OF FLOW METERS .......... 576
Andreas Weissenbrunner, André Fiebach, M. Juling, P.U. Thamsen

ON THE DEVELOPMENT OF THE 3D EULER EQUATIONS USING INTRUSIVE PCE FOR UNCERTAINTY
QUANTIFICATION .................................................................................................................................. 588
Kyriakos Dimitrios Kantarakias, Michalis Chatzimanolakis, Varvara G. Asouti, Kyriakos C. Giannakoglou

THE NON-PARAMETRIC APPROACH TO THE QUANTIFICATION OF THE UNCERTAINTY IN THE DESIGN OF
EXPERIMENTS MODELLING .................................................................................................................. 598
Jacek Pietraszek, Renata Dwornicka, Mariusz Krawczyk, Maciej Kołomycki
RS 16: STOCHASTIC FRACTURE AND DAMAGE

RANDOM AND GRADIENT BASED FIELDS IN DISCRETE PARTICLE MODELS OF HETEROGENEOUS MATERIALS .... 605
Jan Podroužek, Jan Vorel, Roman Wan-Wendner

FATIGUE DAMAGE MODEL OF WIND TURBINE COMPOSITE BLADES UNDER UNCERTAIN WIND SPEED ........... 616
Chi Zhang, Hua-Peng Chen

RS 17: SYSTEM RELIABILITY ANALYSIS, DESIGN AND RISK ASSESSMENT

RELIABILITY OF PRESTRESSED CONCRETE BRIDGE GIRDERS USING FIELD INFORMATION AND THE COMBINED APPROACH ................................................................. 627
Adrián David García-Soto, Alejandro Hernández-Martínez, Jesús Gerardo Valdés-Vázquez

ASSESSING HOW THE DEPENDENCE STRUCTURE AFFECTS THE RELIABILITY PARAMETER OF THE STRENGTH-STRESS MODEL ................................................................. 640
Alessandro Barbiero

ROBUST ARTIFICIAL NEURAL NETWORK FOR RELIABILITY ANALYSIS ........................................... 651
Uchenna Oparaji, Rong-Jiun Sheu, Edoardo Patelli

A NEW NESTED SPACE-FILLING DESIGNS ON EXPERIMENTS WITH DIFFERENT RANGES OF FACTORS ............. 663
Jin Xu, Xiaojun Duan, Zhengming Wang, Liang Yan

SEISMIC SAFETY ASSESSMENT BY AN ENHANCED MONTE CARLO METHOD ........................................... 671
Christian Gasser, Christian Bucher

VULNERABILITY OF HYDROPOWER INSTALLATIONS TO CLIMATE CHANGE: PRELIMINARY STUDY .................. 680
Silvia Tolo, Edoardo Patelli, Diyi Chen

RS 21: STOCHASTIC OPTIMAZATION METHODS

STOCHASTIC MULTI-OBJECTIVE OPTIMISATION OF COMPOSITES MANUFACTURING PROCESS ...................... 690
Konstantinos I. Tifkitsis, Alexandros A. Skordos
PROBABILISTIC LEARNING ON MANIFOLD FOR OPTIMIZATION UNDER UNCERTAINTIES

C. Soize\textsuperscript{1} and R. Ghanem\textsuperscript{2}

\textsuperscript{1}Université Paris-Est
Laboratoire MSME, UMR 8208 CNRS, 5 bd Descartes, 77454 Marne-la-Vallée, France
e-mail: christian.soize@univ-paris-est.fr

\textsuperscript{2}University of Southern California
210 KAP Hall, Los Angeles, CA 90089, United States
e-mail: ghanem@usc.edu

Keywords: Optimization under uncertainty, Probabilistic optimization, Nonconvex constrained optimization, Probability distribution on manifolds, Concentration of probability, Random sampling generator, MCMC generator, Diffusion maps, Statistics on manifolds.

Abstract. This paper presents a challenging problem devoted to the probabilistic learning on manifold for the optimization under uncertainties and a novel idea for solving it. The methodology belongs to the class of the statistical learning methods and allows for solving the probabilistic nonconvex constrained optimization with a fixed number of expensive function evaluations. It is assumed that the expensive function evaluator generates samples (defining a given dataset) that randomly fluctuate around a ”manifold”. The objective is to develop an algorithm that uses a number of expensive function evaluations at a level essentially equal to that of the deterministic problem. The methodology proposed consists in using an algorithm to generate additional samples in the neighborhood of this manifold from the joint probability distribution of the design parameters and of the random quantities that defined the objective and the constraint functions. This is achieved by using the probabilistic learning on manifold from the given dataset generated by the optimizer without performing additional expensive function evaluations. A statistical smoothing technique is developed for estimating the mathematical expectations in the computation of the objective and constraint functions at any point of the admissible set by using only the additional samples. Several numerical illustrations are presented for validating the proposed approach.
1 INTRODUCTION

The present paper addresses a novel approach for solving the probabilistic nonconvex constrained optimization by using only a fixed number of the expensive function evaluations. It is assumed that the expensive function evaluator generates samples that fluctuate around a manifold. An algorithm is then introduced to sample the neighborhood of this manifold from the joint probability distribution of the random parameters and the design variables of the stochastic computational model. The underlying manifold is learned from a diffusion process on the dataset that is generated by the optimizer. This paper extends recent work by the authors [1], where the above sampling on manifolds was first introduced, to the case where the joint probability distribution of multiple vectors, is constructed and used to evaluate the conditional expectations that define objective functions and constraints in optimization under uncertainties. The paper is organized as follows. In Section 2, the problem and novel methodology proposed for solving it is presented. Section 3 deals with the probabilistic learning on manifold. Section 4 is devoted to the probabilistic nonconvex constrained optimization to be solved with a fixed number of function evaluations. A numerical illustration is presented for validating the method proposed.

Notations

A lower case letter such as \( x \), \( \eta \), or \( u \), is a real deterministic variable.
A boldface lower case letter such as \( \mathbf{x} \), \( \mathbf{\eta} \), or \( \mathbf{u} \) is a real deterministic vector.
An upper case letter such as \( X \), \( H \), or \( U \), is a real random variable.
A boldface upper case letter, \( \mathbf{X} \), \( \mathbf{H} \), or \( \mathbf{U} \), is a real random vector.
A lower case letter between brackets such as \( [x] \), \( [\eta] \), or \( [u] \), is a real deterministic matrix.
A boldface upper case letter between brackets such as \( \mathbf{[X]} \), \( \mathbf{[H]} \), or \( \mathbf{[U]} \), is a real random matrix.

\( \mathbb{N} = \{0, 1, 2, \ldots\} \): set of all the null and positive integers.
\( \mathbb{R} \): set of all the real numbers.
\( \mathbb{R}^n \): Euclidean vector space on \( \mathbb{R} \) of dimension \( n \).
\( \|x\|_F \): usual Euclidean norm in \( \mathbb{R}^n \).
\( \mathcal{M}_{n,N} \): set of all the \( (n \times N) \) real matrices.
\( \mathcal{M}_\nu \): set of all the square \( (\nu \times \nu) \) real matrices.
\( [x]_{kj} \): entry of matrix \( [x] \).
\( [x]^T \): transpose of matrix \( [x] \).
\( \|[x]\|_F \): Frobenius norm of matrix \( [x] \) such that \( \|x\|_F^2 = \text{tr}\{[x]^T [x]\} \).
\( [I_\nu] \): identity matrix in \( \mathcal{M}_\nu \).
\( \delta_{kk'} \): Kronecker’s symbol such that \( \delta_{kk'} = 0 \) if \( k \neq k' \) and \( = 1 \) if \( k = k' \).
\( \mathbb{1}_A(a) \) is the indicator function of set \( A \): \( \mathbb{1}_A(a) = 1 \) if \( a \in A \) and \( = 0 \) if \( a \notin A \).
\( E \): Mathematical expectation.
pdf: probability density function.
ISDE: Itô Stochastic Differential Equation.
MCMC: Markov Chain Monte Carlo.

2 PRESENTATION OF THE PROBLEM AND NOVEL METHODOLOGY PROPOSED FOR SOLVING IT

Increasingly, the design of engineered systems that either involve complex interacting processes or new composite materials, must rely on computational models that resolve the under-
lying physics with great detail. The computational burden associated with such a design, in particular in the presence of model uncertainty or parametric uncertainty, quickly becomes prohibitive as it entails iterating over an already expensive function evaluation. Novel perspectives, methodologies, and algorithms must be developed to fulfill the promise of model-assisted design for such complex systems. The probabilistic learning on manifold that is proposed in this paper is one possible way for solving design optimization problems under uncertainties.

2.1 What is the problem that is considered?

In order to properly define the problem that is analyzed in the framework of this paper, we first detail what we mean by the “probabilistic learning on manifold from a dataset” and what we mean by the “optimization under uncertainties”.

Meaning of “probabilistic learning on manifold from a dataset”

In the framework of this paper, “probabilistic learning on manifold from a dataset” is related to the novel methodology that is proposed in [1, 2]

\( \eta^\ell = (\eta^\ell_1, \ldots, \eta^\ell_\nu) \) in \( \mathbb{R}^\nu \) with \( \ell = 1, \ldots, N \) of a \( \mathbb{R}^\nu \)-valued random variable \( \mathbf{H} = (H_1, \ldots, H_\nu) \), its non-Gaussian probability distribution that is unknown and that is concentrated on an unknown subset \( S_\nu \) of \( \mathbb{R}^\nu \).

- for identifying, from a database made up of \( N \) samples \( \eta^\ell = (\eta^\ell_1, \ldots, \eta^\ell_\nu) \) in \( \mathbb{R}^\nu \) with \( \ell = 1, \ldots, N \) of a \( \mathbb{R}^\nu \)-valued random variable \( \mathbf{H} = (H_1, \ldots, H_\nu) \), its non-Gaussian probability distribution that is unknown and that is concentrated on an unknown subset \( S_\nu \) of \( \mathbb{R}^\nu \).

- for generating additional samples that follow the unknown probability distribution in preserving the concentration on \( S_\nu \) and consequently, by avoiding the scattering of the generated samples.

(i) What is a dataset generated by a probability distribution that is concentrated on a subset \( S_\nu \) of \( \mathbb{R}^\nu \)? In Figure 1, the three figures are related to a subset \( S_\nu \) of \( \mathbb{R}^\nu \) with \( \nu = 3 \) and \( N \) samples \( \eta^\ell = (\eta^\ell_1, \eta^\ell_2, \eta^\ell_3) \). In the left figure, the statistical mean line of the dataset can easily be identified as a helical. In the central figure, the statistical mean surface of the dataset is concentrated around a surface with a complex geometry as shown in the right figure. For instance, for such examples, the general method proposed in this paper will allow for estimating the statistics of the real-valued random variable \( H_3(\eta_1, \eta_2) \) by using only dataset \( \{\eta^\ell\}_{\ell=1,\ldots,N} \). It will be not assumed that \( N'_\nu \gg 1 \) points \( \eta^\ell_{3,\nu'} \) are available for \( \ell' = 1, \ldots, N'_\nu \), and consequently, the classical empirical estimation

\[
E\{H_3(\eta_1, \eta_2)\} \simeq \frac{1}{N'_\nu} \sum_{\ell'=1}^{N'_\nu} \eta^\ell_{3,\nu'},
\]

cannot be used. In addition, we want to estimate the statistics of \( H_3(\eta_1, \eta_2) \) at any point \( (\eta_1, \eta_2) \) and not only at the points \( \{((\eta^\ell_1, \eta^\ell_2), \ell = 1, \ldots, N\} \) of the dataset.

(ii) What is the scattering of the generated samples if a classical generator is used? Using the nonparametric statistics for estimating the probability distribution of random vector \( \mathbf{H} \) with the dataset made up of \( N = 400 \) samples that are plotted in Figure 2 (left figure) and if a MCMC generator is used for generating 8,000 additional samples plotted in Figure 2 (right figure), then it can be seen that a scattering of the generated samples is obtained and the concentration around the statistical mean helical line is lost.

Meaning of “optimization under uncertainties” in the framework of this paper

The terminology “Optimization Under Uncertainties” (OUU) refers to as optimization algorithms with underlying stochastic operators and stochastic constraints. An efficient exploration
Figure 1: Data set with $N = 400$ for which the statistical mean line of this dataset can easily be identified as a helical (left figure). Data set with $N = 900$ (central figure) for which the statistical mean surface of this dataset has a complex geometry (right figure).

Figure 2: Data set for which the $N = 400$ samples are concentrated around a statistical mean helical line (left figure). Generation with a MCMC generator of 8,000 additional samples that are scattered (right figure) and consequently, that are not concentrated around the statistical mean helical line.

of the admissible set of the design parameters is crucial to the optimization of a problem with expensive functions (nonconvex objective function and nonlinear constraint function). The development of mathematical and algorithmic constructs that promote learning with successive optimization steps continues to be a key challenge in that regard. In the framework of this paper, "optimization under uncertainties" is related to the novel methodology that is proposed in [3]

- for solving a probabilistic nonconvex constrained optimization (an OUU),
- by using a dataset made up of a small number of points generated by the optimizer for which only a small number of expensive function evaluations is carried out.

A few words about optimization under uncertainties

For solving optimization problems under uncertainties, the methods have progressed along many directions, including gradient-based learning, adapted to convex problems [4, 5], and global search algorithms including stochastic, genetic, and evolutionary algorithms [6, 7]. Statistical learning methods, whereby a deterministic problem is construed as the representative from a class of stochastic problems have also been developed with the benefit of enabling statistical learning [8]. The learning process is typically manifested in the form of a surrogate model from which approximations of the expensive function can be readily evaluated [9]. The resulting error and its repercussions on the attained optimal solution distinguish the various algorithms.
The global character of the surrogate is typically achieved either through a deterministic interpolation process, or a stochastic model whereby biases induced by complex dependencies between model outputs and design parameters are captured through statistical correlations over parameter space. Although Gaussian process models are most commonly used in this context [10, 11], more robust alternatives based on Bayesian optimization [8, 12] have also proven useful. Recent research in the field of uncertainty quantification [13, 14, 15, 16, 17] has underscored the need for optimization algorithms with underlying stochastic operators and constraints. In these situations, that we have previously referred to as OUU, the challenge is magnified since for each design point along the optimization path, a sufficiently large statistical sample of function outputs must be computed to evaluate the required expectations. In essence, the function output must be characterized as a stochastic process over the set of design variables in order to facilitate such evaluations. For expensive function evaluations exhibiting uncertainty, computational challenges remain currently significant enough to require simplifying assumptions in the form of surrogate models for the stochastic function itself or approximations to relevant probabilities [18, 19, 20].

2.2 What is the novel methodology proposed for solving the problem that is considered?

Class of the methodology, fundamental hypothesis, and objective

The methodology belongs to the class of the statistical learning methods. It allows for solving a probabilistic nonconvex constrained optimization with a fixed number of expensive function evaluations. It is assumed that the expensive function evaluator generates samples (the given dataset) that randomly fluctuate around a “manifold”. The objective is to develop an algorithm that uses a number of expensive function evaluations at a level essentially equal to that of the deterministic problem.

Principle of the methodology proposed

The methodology proposed [3] consists

- in using an algorithm to generate additional samples in the neighborhood of this manifold from the joint probability distribution of the design parameters and of the random quantities that defined the objective and the constraint functions. This is achieved by using the probabilistic learning on manifold without performing additional expensive function evaluations.

- in developing a statistical smoothing technique for estimating the mathematical expectations in the computation of the objective and constraint functions at any point of the admissible set, by using only the given dataset and the additional samples.

3 PROBABILISTIC LEARNING ON MANIFOLD

3.1 Short summary of the methodology and algorithm for a concentrated probability distribution

In this Section, we summarize the methodology and the algorithm for generating additional samples from a given dataset whose points are the samples of a random vector that follows an unknown concentrated probability distribution, which allows for avoiding the scattering of the generated samples. The details of this approach can be found in [1, 2].
Defining the random matrix \([X]\) and the dataset \([x_d]\) as its given sample

Let \(X = (X_1, \ldots, X_n)\) be a \(\mathbb{R}^n\)-valued random variable defined on a probability space \((\Theta, \mathcal{T}, \mathcal{P})\). Let \(p_X\) be the pdf of \(X\), which is unknown and concentrated on an unknown subset \(S_n\) of \(\mathbb{R}^n\). The dataset is defined by \(N\) given points that are the vectors \(x_{d,1}, \ldots, x_{d,N}\) in \(\mathbb{R}^n\), which correspond to \(N\) independent samples of random vector \(X\), and which is represented by the \((n \times N)\) real matrix \([x_d]\) such that

\[
[x_d] = \begin{bmatrix} x_{d,1} \\ \vdots \\ x_{d,N} \end{bmatrix} \in \mathbb{M}_{n,N}.
\]

We define the random matrix \([X]\) on \((\Theta, \mathcal{T}, \mathcal{P})\) with values in \(\mathbb{M}_{n,N}\) such that

\[
[X] = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix},
\]

in which the columns are \(N\) independent copies \(X_1, \ldots, X_N\) of random vector \(X\). Consequently, the dataset represented by matrix \([x_d]\) is a sample of \([X]\).

Reduced normalized random matrix \([H]\) and its sample \([\eta_d]\) constructed by the principal component analysis

For \(\nu \leq n\), the normalized random matrix \([H] = [H_1, \ldots, H_N]\) with values in \(\mathbb{M}_{\nu,N}\) for which the columns are \(N\) independent copies of a random vector \(H = (H_1, \ldots, H_\nu)\) defined on \((\Theta, \mathcal{T}, \mathcal{P})\) with values in \(\mathbb{R}^\nu\), is defined by the following equation that corresponds to a principal component analysis,

\[
[X] = [x] + [\phi] [\lambda]^{1/2} [H],
\]

in which \([\lambda]\) is the \((\nu \times \nu)\) diagonal matrix of the \(\nu\) positive eigenvalues of the empirical estimate \([\text{cov}]\in \mathbb{M}_n\) of the covariance matrix of \(X\) (computed with the dataset), where \([\phi]\) is the \((n \times \nu)\) matrix of the associated eigenvectors such \([\phi]^T [\phi] = [I_\nu]\), and where \([x]\) is the matrix in \(\mathbb{M}_{n,N}\) with identical columns, each equal to the empirical estimate \(\bar{x} \in \mathbb{R}^n\) of the mean value of random vector \(X\) (computed with the dataset). The sample

\[
[\eta_d] = \begin{bmatrix} \eta_{d,1} \\ \vdots \\ \eta_{d,N} \end{bmatrix} \in \mathbb{M}_{\nu,N}
\]

of \([H]\) (associated with the sample \([x_d]\) of \([X]\)) is computed by

\[
[\eta_d] = [\lambda]^{-1/2} [\phi]^T ([x_d] - [x]).
\]

The empirical estimates of the mean value and of the covariance matrix of random vector \(H\) are therefore \(0_\nu\) and \([I_\nu]\).

Methodology of the mathematical formulation

The methodology of the proposed mathematical formulation that is detailed in [1] can be summarized by the five following steps.

**Step 1.** A multidimensional kernel-density estimation [21] of the pdf \([\eta] \mapsto p_H([\eta])\) of random matrix \([H]\) is constructed by using the normalized dataset represented by matrix \([\eta_d]\).

**Step 2.** A Markov chain Monte Carlo (MCMC) generator for random matrix \([H]\) is constructed by using [22], which belongs to the class of Hamiltonian Monte Carlo methods [22, 23, 24]. The samples are obtained by solving an Itô stochastic differential equation (ISDE) corresponding to
a stochastic nonlinear dissipative Hamiltonian dynamical system, for which \( p_\eta(\eta) \, d\eta \) is the unique invariant measure.

**Step 3.** A diffusion-map approach \([25]\) is used to discover and to characterize the local geometry structure of the normalized dataset concentrated in the neighborhood of the unknown subset \( \mathcal{S}_\nu \) of \( \mathbb{R}^\nu \). The method consists in introducing the transition matrix \([P]\) in \( \mathbb{M}_N \) such that

\[
[P] = [b]^{-1} [K] , \quad [b]_{ij} = \delta_{ij} \sum_{j'=1}^N [K]_{ij'} , \quad [K]_{ij'} = \exp(-\frac{1}{4\varepsilon} ||\eta^{d,i} - \eta^{d,j'}||^2),
\]

in which \( \varepsilon > 0 \) is a real smoothing parameter. Let \( \psi^1, \ldots, \psi^m \) be the right eigenvectors associated with the \( m \) positive eigenvalues \( 1 = \Lambda_1 > \ldots \geq \Lambda_m \) of the eigenvalue problem \([P] \psi^\alpha = \Lambda_\alpha \psi^\alpha \). The eigenvectors are normalized such that \( [\psi]^T [b] [\psi] = [I_m] \). A reduced order diffusion-maps basis of \( \mathbb{R}^N \), of order \( m < N \), is defined by

\[
[g] = [g^1 \ldots g^m] \in \mathbb{M}_{N,m}.
\]

in which \( g^1, \ldots, g^m \) are the vectors in \( \mathbb{R}^r \) that are associated with the first \( m \) eigenvalues of transition matrix \([P]\) relative to the local geometric structure of the given normalized dataset, and that are written as

\[
g^\alpha = \Lambda_\alpha^{\frac{\zeta}{2}} \psi^\alpha \in \mathbb{R}^N , \quad \alpha = 1, \ldots, m ,
\]

in which \( \zeta \) is a given positive integer. For \( m = N \), \( \{g^1, \ldots, g^N\} \) is an algebraic basis of \( \mathbb{R}^N \).

**Step 4.** The following reduced-order representation of random matrix \( \mathbf{H} \),

\[
[\mathbf{H}] = [\mathbf{Z}] [g]^T,
\]

is constructed on the manifold in which \([\mathbf{Z}]\) is a random matrix with values in \( \mathbb{M}_{r,m} \). The value of \( m \) is chosen as explained in \([11]\). As \( m < N \), this equation defines a statistical reduction of random matrix \([\mathbf{H}]\) with respect to data dimension \( N \), which allows for keeping the concentration in \( \mathcal{S}_\nu \subset \mathbb{R}^\nu \) and consequently, for avoiding the scattering of the generated samples.

**Step 5.** A reduced-order ISDE is constructed for generating additional samples concentrated in subset \( \mathcal{S}_\nu \) without scattering of the generated samples. This MCMC generator on the manifold is obtained by projecting the ISDE introduced in Step 2 onto the diffusion manifold by using the reduced-order diffusion-maps basis represented by matrix \([g]^T\). The constructed reduced-order ISDE is then used for generating \( n_{\text{MC}} \) additional samples,

\[
[z^1_{\text{ar}}, \ldots, z^{n_{\text{MC}}}_{\text{ar}}] \in \mathbb{M}_{r,m}
\]

of random matrix \([\mathbf{Z}]\), and therefore, for deducing the \( n_{\text{MC}} \) additional samples

\[
[\eta^1_{\text{ar}}, \ldots, \eta^{n_{\text{MC}}}_{\text{ar}}] \in \mathbb{M}_{r,N}
\]

of random matrix \([\mathbf{H}]\), such that \( [\eta^\ell_{\text{ar}}] = [z^\ell_{\text{ar}}] [g]^T \) for \( \ell = 1, \ldots, n_{\text{MC}} \). Let \( \{([\mathbf{Z}(r)], [\mathbf{Y}(r)]), r \in \mathbb{R}^+ \} \) be the unique asymptotic (for \( r \to +\infty \)) stationary and ergodic diffusion stochastic process with values in \( \mathbb{M}_{r,m} \times \mathbb{M}_{r,m} \), of the following reduced-order ISDE (stochastic nonlinear second-order dissipative Hamiltonian dynamical system \([26][22]\)), for \( r > 0 \),

\[
d[\mathbf{Z}(r)] = [\mathbf{Y}(r)] \, dr ,
\]

\[
d[\mathbf{Y}(r)] = [\mathcal{L}([\mathbf{Z}(r)])] \, dr - \frac{1}{2} f_0 [\mathbf{Y}(r)] \, dr + \sqrt{f_0} [d\mathbf{W}(r)] ,
\]

with appropriate initial conditions for \( r = 0 \), and where
• \([L([Z(r)])] = [L([Z(r)] [g]^T)] [a]\) in which \([η] \mapsto [L(η)]\) is a nonlinear function from \(\mathbb{M}_\nu,\mathbb{N}\) into \(\mathbb{M}_\nu,\mathbb{N}\), which is expressed as a function of pdf \(p_H\),

• \([d\mathbf{W}(r)] = [d\mathbf{W}(r)] [a]\) where \([d\mathbf{W}(r)]\) is the normalized Wiener process with values in \(\mathbb{M}_\nu,\mathbb{N}\),

• the matrix \([a]\) belongs to \(\mathbb{M}_{\nu,m}\) and is such that \([a] = [g] ([g]^T [g])^{-1}\),

• the free parameter \(f_0 > 0\) allows the dissipation term of the nonlinear second-order dynamical system (dissipative Hamiltonian system) to be controlled in order to kill the transient part induced by the initial conditions.

We then have

\[ [Z] = \lim_{r \to +\infty} [Z(r)] \quad \text{in probability distribution}, \]

which allows for generating the additional samples, \([z_1], \ldots, [z_{\text{MCMC}}]\).

Remark on the methodology proposed

The stochastic germ of the reduced-order ISDE that is used for generating the samples of \([Z]\), lives on the ”manifold” that is identified by the diffusion maps, that is to say, lives on a subset of the set \(\mathbb{M}_{\nu,m}\), which has a small dimension because \(m \ll N\). The samples of \([Z]\) are directly generated by the reduced-order ISDE on the ”manifold”, which is the subset of \(\mathbb{M}_{\nu,m}\) with a small dimension.

3.2 Numerical illustration

As the numerical illustration, we present the simple one introduced in Section 2 for which \(n = \nu = 3\) with \(N = 400\) given points in the dataset. Another one, corresponding to a petrophysics data base of experiments for which \(n = 35, \nu = 32,\) and \(N = 13,056\) given points in the dataset, can be found in [1]. Figure 3 displays the dataset that is made up of 400 given points \(\{\eta^\ell\}\ell\) concentrated around a statistical mean helical line (left figure), and the eigenvalues \(\Lambda_{\alpha\alpha}=1,\ldots,20\). The convergence analysis leads us a reduction order \(m = 4\). Figure 4 (left) shows the 400 given points of the dataset and the 8,000 additional realizations generated by using the reduced-order ISDE with \(m = 4\). It can be seen that the concentration of the additional samples is kept. Figure 4 (right) displays 8,000 additional realizations generated by the MCMC generator without using the reduced diffusion-maps basis. In such a case the samples are scattered and the concentration is lost.
4 PROBABILISTIC NONCONVEX CONSTRAINED OPTIMIZATION TO BE SOLVED WITH FIXED NUMBER OF FUNCTION EVALUATIONS

4.1 Definition of a probabilistic nonconvex constrained optimization problem

Algebraic form of the optimization problem

Let \( w \in C_w \subset \mathbb{R}^{m_w} \) be the admissible set of the vector of the design parameters. The nonconvex objective function is defined by the function \( w \mapsto f(w) \) from \( C_w \) into \( \mathbb{R} \). The nonlinear constraints are represented by the function \( w \mapsto c(w) \) from \( C_w \) into \( \mathbb{R}^{m_c} \). The Probabilistic Nonconvex Optimization Problem with nonlinear constraints (PNOP) is written as

\[
\mathbf{w}^{opt} = \arg \min_{w \in C_w, c(w) < 0} f(w).
\]

Probabilistic aspects of the optimization problem

The objective function and the constraint function are assumed to be written as

\[
f(w) = E\{Q(w)\}, \quad c(w) = E\{B(w)\}.
\]

The stochastic process \( \{Q(w), w \in C_w\} \) and \( \{B(w) = (B_1(w), \ldots, B_{m_c}(w)), w \in C_w\} \) are defined on a probability space \((\Theta, T, \mathcal{P})\), are indexed by \( C_w \), are with values in \( \mathbb{R} \) and \( \mathbb{R}^{m_c} \) respectively, are statistically dependent, and are second-order stochastic processes. Consequently, for all \( w \) fixed in \( C_w \), the real-valued random variable \( Q(w) : \theta \mapsto Q(w; \theta) \) and the \( \mathbb{R}^{m_c} \)-valued random variable \( B(w) : \theta \mapsto B(w; \theta) \) are such that

\[
E\{Q(w)^2\} = \int_{\Theta} Q(w; \theta)^2 d\mathcal{P}(\theta) < +\infty,
\]

\[
E\{\|B(w)\|^2\} = \int_{\Theta} \|B(w; \theta)\|^2 d\mathcal{P}(\theta) < +\infty.
\]

4.2 Framework and objective

Framework

For \( w \) given in \( C_w \), \( f(w) \) and \( c(w) \) are calculated by using the Stochastic Computational Model (SCM) in which a probabilistic model of uncertainties is implemented. It is assumed that the PNOP defined before has a unique solution \( w^{opt} \) in \( C_w \).
Objective

The objective is the development of a formulation that permits to solve the PNOP by using a small number of numerical evaluations of $f(w)$ and $c(w)$ in order to limit the calls to the expensive SCM.

4.3 Methodology

Ingredients

The first ingredient is the probabilistic learning on manifold that has been presented in Section 3, which allows for generating additional samples concentrated on the manifold that has been identified by using the dataset $[1, 2]$, without performing additional function evaluations by the use of the SCM. The second ingredient is a smoothing technique that allows for estimating the mathematical expectations in the computation of $f(w^0)$ and $c(w^0)$ at any point $w^0$ in $C_w$, by using only the given dataset and the additional samples $[3]$.

What would be the consequences of the use of the classical procedure?

Let us assume that the PNOP requires $N$ evaluations $f(w_\ell)$ and $c(w_\ell)$ at points $w_\ell$ for $\ell = 1, \ldots, N$. For a given $w^\ell$, the use of the classical estimation would lead us to compute the samples $Q(w^\ell; \theta_\ell')$ and $B(w^\ell; \theta_\ell')$ for $\ell' = 1, \ldots, N'$ with the SCM. For $N'$ sufficiently large, the following empirical estimations would be performed,

$$ f(w_\ell) \approx \frac{1}{N'} \sum_{\ell' = 1}^{N'} Q(w^\ell; \theta_\ell'), \quad c(w_\ell) \approx \frac{1}{N'} \sum_{\ell' = 1}^{N'} B(w^\ell; \theta_\ell'). $$

With such a classical approach, the SCM would be called $N' \times N$ times, which would be prohibitive for expensive function calls.

Method proposed for avoiding $N' \times N$ evaluations with the SCM and based on the use of only $N$ evaluations

Step 1: Construction of the dataset by using only a fixed number $N$ of evaluations. For $\ell = 1, \ldots, N$ (with $N$ fixed), let $w_\ell$ be the $N$ values of $w$, which correspond either to a training procedure applied to $w$ or are some values of $w$ generated by an optimizer as it explores the admissible domain. Let $q_\ell = Q(w_\ell; \theta_\ell)$ and $b_\ell = B(w_\ell; \theta_\ell)$ be the $N$ corresponding samples that are computed by using the SCM (therefore, there are only $N$ evaluations). The dataset is made up of the $N$ data points $x^1, \ldots, x^N$ in $\mathbb{R}^n$ such that,

$$ x^\ell = (w_\ell, q_\ell, b_\ell), \quad \ell = 1, \ldots, N, $$

with $n = m_w + 1 + m_c$.

Step 2: Construction of the diffusion-maps basis and generating additional samples with the reduced ISDE. We introduce the random variable $X = (W, Q, B)$ with values in $\mathbb{R}^n$ such that $x^\ell = (w^\ell, q^\ell, b^\ell)$ are $N$ independent samples. The diffusion-maps basis is constructed by using $\{x^\ell\}_{\ell=1,\ldots,N}$. We can then generate $\nu_{\text{sim}} \gg N$ additional samples such that,

$$ x_{\text{ar}}^\ell = (w_{\text{ar}}^\ell, q_{\text{ar}}^\ell, b_{\text{ar}}^\ell), \quad \ell = 1, \ldots, \nu_{\text{sim}}, $$

without performing additional function evaluations.
Step 3: Introduction of a smoothing technique by using a nonparametric statistical estimation and solving the PNOP. A smoothing technique has been developed and is written as

\[ f(w^0) \simeq E\{Q \mid W = w^0\} \quad , \quad c(w^0) \simeq E\{B \mid W = w^0\}, \]

in which the conditional mathematical expectations are estimated by using the kernel density estimation method with the additional samples \( \{x^e_{ar} = (w^e_{ar}, q^e_{ar}, b^e_{ar})\} \). It should be noted that an explicit numerical formula exists for \( f(w^0) \) and for \( c(w^0) \) for any value of dimensions \( m_w \) and \( m_c \). The probabilistic nonconvex optimization problem with nonlinear constraints is therefore solved,

\[ w^{opt} = \arg \min_{w^0 \in C_w, c(w^0) < 0} f(w^0). \]

4.4 Numerical illustration

Description

The design parameter \( w = (w_1, w_2) \) is with values in \( \mathbb{R}^2 \) \( (m_w = 2) \). The nonconvex objective function \( w \mapsto f(w) = E\{Q(w)\} \) with values in \( \mathbb{R} \) is defined on a subset \( C_w = [0, 1.1] \times [0, 1.1] \) of \( \mathbb{R}^2 \). The constraint function is an affine function \( w \mapsto c(w) = E\{B(w)\} \) defined on \( C_w \) with values in \( \mathbb{R}^3 \) \( (m_c = 4) \). Consequently, we have \( n = m_w + 1 + m_c = 7 \).

Solution of reference

The solution of reference is computed by using the classical procedure with a Cartesian grid of 3,600 points uniformly distributed in \( C_w \) and for which \( N^\prime = 10,000 \) samples. The optimal solution is \( w^{opt}_{r,1} = 0.74, w^{opt}_{r,2} = 0.49, f(w^{opt}) = -0.123 \) for which the 4 constraints are active. In Fig. 6, the six figures show the reference solution. Figure 5-(a) displays the graph of objective function \( w \mapsto f(w) \) while Figure 5-(b) corresponds to its contour plot in which the white lozenge marks the location of the optimal solution. Figures 5-(c) to 5-(f) correspond to the graphs of the four components of constraint function \( w \mapsto c(w) \).

![Figure 5: Reference solution: Graph of \( w \mapsto f(w) \) (up left figure), contour plot of \( w \mapsto f(w) \) (b-figure), graphs of \( w \mapsto c_1(w) \) (c-figure), \( c_2(w) \) (d-figure), \( c_3(w) \) (e-figure), and \( c_4(w) \) (f-figure). In b-figure, the white lozenge marks the location of the optimal solution.](image_url)
**Dataset generated by the optimizer**

The number of numerical evaluations of the objective function $f$ and of the constraint function $c$ by using the SCM is $N = 900$. We thus have computed $x^\ell = (w^\ell, q^\ell, b^\ell)$ for $\ell = 1, \ldots, N$ with $q^\ell = Q(w^\ell; \theta_c)$ and $b^\ell = B(w^\ell; \theta_c)$. Figure 6 (left) shows the $N = 900$ given/computed data points $q^\ell = Q(w^\ell; \theta_c)$ generated by the optimizer for estimating the objective function. Figure 6 (right) gives a view of the statistical fluctuations of these 900 given/computed data points around the surface corresponding to the objective function, $w \mapsto f(w) = E\{Q(w)\}$.

![Dataset for random vector $H = (H_{\ell b}, H_{\ell c}, H_{\ell d})$](image1)

Figure 6: Left figure: $N = 900$ given/computed data points $q^\ell = Q(w^\ell; \theta_c)$ generated by the optimizer for estimating the objective function. Right figure: View of the statistical fluctuations of these 900 given/computed data points around the surface corresponding to the objective function, $w \mapsto f(w) = E\{Q(w)\}$.

**Solution given by the proposed probabilistic learning and data smoothing**

The same grid of 3,600 points is used for the computation. Figure 7 displays the contour plot of the graph of objective function $w \mapsto f(w)$ constructed with the $N = 900$ given/computed data points (Figure 7-(left)), with $\nu_{\text{sim}} = 9,000$ additional samples (Figure 7-(central)), and with $\nu_{\text{sim}} = 90,000$ additional samples (Figure 7-(right)). In these figures, the white lozenge marks the location of the reference solution while the white circle marks the location of the optimal solution computed with $N = 900$ given/computed data points, $\nu_{\text{sim}} = 9,000$ additional samples, and $\nu_{\text{sim}} = 90,000$ additional samples. For $\nu_{\text{sim}} = 9,000$ additional samples \{$x^\ell_{\text{ar}} = (w^\ell_{\text{ar}}, q^\ell_{\text{ar}}, b^\ell_{\text{ar}})\}_{\ell=1,\ldots,\nu_{\text{sim}}}$ (Figure 7-(central)), we obtain a good approximation $w_1^{\text{opt}} = 0.70$, $w_2^{\text{opt}} = 0.49$, $f(w^{\text{opt}}) = -0.112$ of the reference solution, which is confirmed for $\nu_{\text{sim}} = 90,000$. It should be noted that the image of the contour plot of the objective function is really well represented for $\nu_{\text{sim}} = 9,000$ and for $\nu_{\text{sim}} = 90,000$ when comparing these two figures to Figure 5-(b).

![Contour plot of the objective function computed with $N = 900$ given/computed data points](image2)

Figure 7: Contour plot of the objective function computed with $N = 900$ given/computed data points (left figure), with $\nu_{\text{sim}} = 9,000$ additional samples (central figure), and with $\nu_{\text{sim}} = 90,000$ additional samples (right figure). The white lozenge marks the location of the reference solution while the white circle marks the location of the optimal solution computed in each case.
5 CONCLUSIONS

A novel methodology has been presented for generating samples of an $\mathbb{R}^n$-valued random vector from a dataset of length $N$, for which the probability distribution is unknown and is concentrated on an unknown subset $S_n$ of $\mathbb{R}^n$. The method is robust and efficient for high dimension $n$ and for a big dataset of length $N$. A new perspective on optimization under uncertainty is given for a nonconvex objective function with a nonlinear constraints function that are viewed as the average of data concentrated around a manifold. For such a case, standard procedures typically require a very large number of function evaluations at each design point. With the method proposed, only a limited number of expensive function evaluations is used.

Acknowledgments. Part of the research devoted to Data-driven probability concentration and sampling on manifold was supported by the U.S. Department of Energy Office of Advanced Scientific Computing Research. Part of the research devoted to Probabilistic nonconvex constrained optimization with fixed number of function evaluations was supported by the ScramJet-UQ project funded under U.S. DARPA’s EQUIPS Program.

REFERENCES


OBTAINING HIGH ACCURACY VIBRATION MEASUREMENTS WITH LOW-COST SENSORS USING BAYESIAN VIRTUAL SENSING

Jyrki Kullaa

Department of Automotive and Mechanical Engineering
Metropolia University of Applied Sciences
P.O. Box 4021, 00079 Metropolia, Finland
e-mail: jyrki.kullaa@metropolia.fi

Keywords: Virtual Sensing, Bayes’ Rule, Optimal Sensor Placement, Structural Dynamics, Response Estimation, Sensor Network.

Abstract. Vibration measurements are utilized in many applications, e.g. structural health monitoring, vibration control, fatigue assessment, system identification, and model updating. Requirements for data accuracy are often high, which can be obtained by using high-quality sensors. However, wireless sensor networks often consist of a large number of low-cost sensor nodes, the accuracy of which may not fulfill the strict requirements of the applications.

The possibility to replace high-quality vibration sensors with a larger number of low-cost sensors is studied. In order to achieve the required accuracy with low-cost sensors, empirical virtual sensing is introduced, which uses hardware redundancy for estimation. In empirical virtual sensing, the signal of a single sensor can be estimated from the data acquired by the whole sensor network. A Bayesian virtual sensor is derived, resulting in a posterior mean that is more accurate than the actual measurement provided the sensor noise is known. Also optimal sensor placement is studied to minimize the number of sensors.

Numerical simulations are performed for a structure subject to unknown random excitation. Noisy response is measured and the accuracy of virtual sensors is evaluated. Given the original reference sensor network with a small number of high-quality sensors, it is possible to determine the number of low-cost sensors needed to achieve the same accuracy. From this result, the cost-effectiveness can be assessed.
1 INTRODUCTION

Vibration measurements used in many applications, such as structural health monitoring, system identification, model updating, and vibration control, often require a sensor network with simultaneous sampling in order to capture the mode shape information at several locations of the structure. Traditionally, expensive high-quality accelerometers are used to acquire measurements with a high signal-to-noise ratio (SNR). A typical value of SNR in a measurement system is 30 dB. Excitation is often unknown, and response is only measured. Measured variables are usually accelerations, strains, displacements, or velocities.

With new technologies in intelligent structures and systems, digitalization, Internet of Things (IoT), MEMS sensors, and wireless sensor networks, an increasing number of sensors can be installed anywhere. Deployment of sensors is easy, because no cables are used, and the configuration of the sensor network is automatic. The sensor nodes have to be low-cost due to their large number in a single application. Therefore, the quality of the sensors may need to be compromised, which can result in a higher measurement error than with traditional high-quality sensors.

In the linear vibration theory, the structural response can be assumed to consist of the sum of modal contributions, in which only a few natural modes are active. With this assumption, a finite number of sensors is sufficient to make the sensor network redundant. The redundancy can be utilized to decrease the measurement error using virtual sensing techniques.

Virtual sensing (VS) can be either model-based (analytical) or data-driven (empirical) [1]. In analytical virtual sensing, in addition to measurement data, a finite element model is needed to estimate the unmeasured degrees of freedom. For example, full-field dynamic stress/strain field can be estimated using a limited number of sensors [2, 3].

Empirical virtual sensing is based on data from a redundant sensor network. It can be used, for example, to replace a temporarily installed or failed sensor [4]. Empirical virtual sensing has also been used for damage or sensor fault detection in structural health monitoring [5]. Also, combined empirical and analytical VS has been introduced for more accurate full-field response estimation than what can be obtained with analytical VS alone [6, 7].

In this study, empirical virtual sensing is applied to a large sensor network with low-cost sensors. The objective is to design a sensor network resulting in the same accuracy as a small number of high-quality sensors at the reference locations. It would then be possible to replace the original sensor network with a higher number of low-cost sensors. The quality of data is preserved and the decision between the two systems can be made with other criteria, e.g. hardware, installation, or maintenance costs.

The accuracy of the virtual sensors depends on the sensor locations. In order to minimize the required number of sensors, the sensors must be placed in optimal positions. Some review papers and comparisons of different optimal sensor placement (OSP) algorithms exist [8–11]. They present the most commonly applied algorithms and criteria. The sensor placement is a discrete optimization problem, for which genetic algorithms have been proposed [12–14]. Alternatively, a computationally efficient and widely used algorithm is to start with a large set of candidate sensor locations and removing one sensor in each round based on the selected criterion until the selected number of sensors remains. This backward sequential sensor placement (BSSP) algorithm has been used in many studies [15–17]. Another iterative method is to start with a small number of sensors and add one sensor in each round to the sensor network until the required criterion is fulfilled. The algorithm is called forward sequential sensor placement (FSSP) algorithm [13, 17, 18].

Iterative OSP algorithms are studied in this paper. First, a reference sensor configuration is designed using a widely used EFI method [15]. Three different OSP algorithms are studied
for virtual sensing. The cost function in the optimization is the error in the virtual sensors, which must be minimized under constrains concerning the number of sensors, types of sensors, possible locations of the sensors, etc. Bayesian analysis is applied to assess the estimation error. The sensor network must include the reference degrees of freedom (DOFs). The result is an optimal sensor network with a minimum number of sensors for the required accuracy.

It is assumed that measurements or simulations are available at all possible sensor locations and that measurement errors are Gaussian and known. The required input parameters include: (1) the candidate DOFs for sensors; (2) the virtual sensor DOFs (reference sensor network); (3) sensor noise information; (4) the criterion (cost function) for sensor network assessment; and (5) the stopping criterion, for example the desired accuracy or the number of sensors.

The paper is organized as follows. Empirical virtual sensing using Bayesian estimation is derived in Section 2. Optimal sensor placement for virtual sensing is discussed in Section 3. In Section 4, the method is validated by numerical simulations of ambient vibration measurements. Two types of physical and virtual sensors are studied. Concluding remarks are given in Section 5.

2 EMPIRICAL BAYESIAN VIRTUAL SENSING

Virtual sensing (VS), or soft sensing, is used to provide an alternative to physical measurement instrument. The quantity of interest is estimated using the available measurements and the system model. Virtual sensing can be classified into empirical and analytical techniques. Empirical VS is only studied in this paper.

Empirical virtual sensing is based on available current or historical measurements. Consider a sensor network measuring \( p \) simultaneously sampled variables \( y = y(t) \) at time instant \( t \). Each measurement \( y \) includes measurement error \( w = w(t) \):

\[
y = x_m + w
\]

where \( x_m = x_m(t) \) are the exact values of the measured (\( m \)) degrees of freedom. All vectors are divided into predicted DOFs \( u \) and the remaining DOFs \( v \):

\[
y = \begin{bmatrix} y_u \\ y_v \end{bmatrix}, \quad x = \begin{bmatrix} x_{m,u} \\ x_{m,v} \end{bmatrix}, \quad w = \begin{bmatrix} w_u \\ w_v \end{bmatrix}
\]

For simplicity but without loss of generality, assume zero-mean variables \( y \). The partitioned data covariance matrix \( \Sigma_y \) is

\[
\Sigma_y = E(yy^T) = \begin{bmatrix} \Sigma_{y,uu} & \Sigma_{y,uv} \\ \Sigma_{y,vu} & \Sigma_{y,vv} \end{bmatrix} = \begin{bmatrix} \Gamma_{y,uu} & \Gamma_{y,uv} \\ \Gamma_{y,vu} & \Gamma_{y,vv} \end{bmatrix}^{-1} = \Gamma_y^{-1}
\]

where the precision matrix \( \Gamma_y \) is defined as the inverse of the covariance matrix \( \Sigma_y \) and is also written in partitioned form. \( E(\cdot) \) denotes the expectation operator.

A linear minimum mean square error (MMSE) estimate for \( y_u | y_v \) (\( y_u \) given \( y_v \)) is obtained by minimizing the mean-square error (MSE) and can be computed either using the covariance or precision matrix [5, 19]. The expected value, or the conditional mean, of the predicted variable is:

\[
\hat{y}_u = E(y_u | y_v) = -\Gamma_y^{-1} \Gamma_{y,uv} y_v
\]
The error covariance MSE is
\[ \text{cov}(\mathbf{y}_u | \mathbf{y}_v) = \Gamma^{-1}_{y,au} \] (5)

Although Equation 4 may give an accurate estimate, an even better estimate for \( x_{m,u} \) can be derived using Bayes’ rule:

\[ p(x_{m,u} | \mathbf{y}) = \frac{p(x_{m,u}, \mathbf{y}_v) p(\mathbf{y}_u | \mathbf{y}_v)}{p(\mathbf{y}_u | \mathbf{y}_v)} \] (6)

The three terms in Equation 6, the likelihood function, the prior distribution, and the evidence, are derived in the following. Measurement error \( \mathbf{w} \) is assumed to be zero mean Gaussian, independent of \( x_m \), with a (known) covariance matrix \( \Sigma_w = E(\mathbf{w}\mathbf{w}^T) \) (7)

The likelihood in (6) is, according to (1) and (7):

\[ p(\mathbf{y}_u | x_{m,u}, \mathbf{y}_v) = p(\mathbf{y}_u | x_{m,u}) = N(x_{m,u}, \Sigma_{w,au}) \] (8)

Using Equation 1 and the assumed noise model, the conditional means of \( y \) and \( x_m \) are equal:

\[ E(x_{m,u} | \mathbf{y}_v) = E(\mathbf{y}_u | \mathbf{y}_v) \] (9)

Using Equations 1 and 7, the MMSE error covariance contains both the estimation error and noise:

\[ \text{cov}(\mathbf{y}_u | \mathbf{y}_v) = \text{cov}(x_{m,u} | \mathbf{y}_v) + \Sigma_{w,au} \] (10)

The prior distribution is derived using (9) and (4):

\[ p(x_{m,u} | \mathbf{y}_v) = N(K\mathbf{y}_v, \Sigma_{\text{prior},u}) \] (11)

where \( K = -\Gamma^{-1}_{y,au} \Gamma_{y,au} \), and the prior covariance is obtained from (5) and (10):

\[ \Sigma_{\text{prior},u} = \text{cov}(x_{m,u} | \mathbf{y}_v) = \Gamma_{y,au}^{-1} - \Sigma_{w,au} \] (12)

The denominator \( p(\mathbf{y}_u | \mathbf{y}_v) \) in (6) is the normalizing factor, which does not depend on \( x_{m,u} \). It is Gaussian with mean (4) and covariance (5), and could be easily evaluated. However, it is not necessary, because it is merely a scaling factor.

The posterior distribution (6) is derived by some manipulation, resulting also in a Gaussian distribution:

\[
p(x_{m,u} | \mathbf{y}) = c_1 p(y_u | x_{m,u}) p(x_{m,u} | \mathbf{y}_v)
= c_2 \exp \left[ -\frac{1}{2} (\mathbf{y}_u - x_{m,u})^T \Sigma_{w,au}^{-1} (\mathbf{y}_u - x_{m,u}) - \frac{1}{2} (x_{m,u} - K\mathbf{y}_v)^T \Sigma_{\text{prior},u}^{-1} (x_{m,u} - K\mathbf{y}_v) \right]
= c_2 \exp \left[ -\frac{1}{2} (\mathbf{x}_{m,u} - \hat{x}_{m,u})^T \Sigma_{\text{post},u}^{-1} (\mathbf{x}_{m,u} - \hat{x}_{m,u}) \right]
\] (13)
\[
\Sigma_{\text{post}, u} = \text{cov}(x_{m,u} \mid y) = (\Sigma_{w,u,u}^{-1} + \Sigma_{\text{prior}, u}^{-1})^{-1}
\]

and the posterior mean is
\[
\hat{x}_{m,u} = E(x_{m,u} \mid y) = \Sigma_{\text{post}, u}(\Sigma_{w,u,u}^{-1}y_u + \Sigma_{\text{prior}, u}^{-1}K_y)
\]

Notice that the posterior mean (15) is a weighted sum of the noisy measurement \(y_u\) and the MMSE estimate \(Ky_v\).

Equation 15 can also be written in the following matrix form.
\[
\hat{x}_{m,u} = E(x_{m,u} \mid y) = \left[\Sigma_{\text{post}, u} \Sigma_{w,u,u}^{-1} \Sigma_{\text{post}, u}^{-1} \Sigma_{\text{prior}, u}^{-1} K\right] \begin{bmatrix} y_u \\ y_v \end{bmatrix} = a^T \ y
\]

where \(a^T_u\) is
\[
a_u^T = \left[\Sigma_{\text{post}, u} \Sigma_{w,u,u}^{-1} \Sigma_{\text{post}, u}^{-1} \Sigma_{\text{prior}, u}^{-1} K\right]
\]

Generally, the predicted DOFs \(u\) may include several variables. In the sequel, \(u\) is one-dimensional including one sensor only. For each sensor \(u\), a corresponding vector \(a^T_u\) is computed. All these vectors can be assembled in a coefficient matrix \(A\) to compute all estimates simultaneously:
\[
\hat{x}_m = Ay
\]

where each row \(u\) of matrix \(A\) represents the corresponding sensor. The rows corresponding to the reference sensor DOFs are only needed in this study.

3 OPTIMAL SENSOR PLACEMENT

In order to achieve the required accuracy of virtual sensors with a minimum number of physical sensors, the sensor locations must be optimized. The starting point is the reference sensor network with a small number of high-quality sensors. The objective is to replace this network with low-cost sensors using empirical virtual sensing.

Three algorithms are studied for optimal sensor placement: effective independence (EFI), backward sequential sensor placement (BSSP), and forward sequential sensor placement (FSSP).

EFI is a widely used algorithm, because no measurement data are needed but the mode shapes only. It has also shown to result in good results in many applications. However, it is not directly related to the accuracy of Bayesian virtual sensors. The effective independence vector is defined as (using a slightly modified expression from [16]):
\[
E_D = \text{diag}\left[\Phi_m(\Phi_m^T W \Phi_m)^{-1} \Phi_m^T W\right]
\]

where \(\Phi_m\) is the truncated modal matrix including the selected modes and the candidate measurement DOFs only. \(W\) is a weighting matrix defined as the inverse of the noise covariance matrix \(\Sigma_u\). The values of \(E_D\) represent the contributions of the corresponding sensor locations to the linear independence of the mode shape vectors in \(\Phi_m\). The sensor corresponding to the smallest value of \(E_D\) is removed and the \(E_D\) coefficients are then updated using the reduced modal matrix. The process is repeated iteratively until the number of sensors equals the required value.
The BSSP algorithm is iterative like EFI and it also starts with a large initial sensor network including all candidate DOFs. A single sensor is removed and the posterior variances are computed for the virtual sensors at the reference locations. The cost function is evaluated for the reduced sensor network. The cost function is evaluated for different sensor networks of the same number of sensors by removing each sensor in turn. Finally, the minimum cost is found, and the corresponding reduced sensor network is selected. The process is repeated until the required minimum accuracy is reached. Contrary to EFI, measurement data are needed but no mode shape information. The evaluated cost function is directly related to the accuracy of the virtual sensors. However, the optimal sensor placement may vary with different data sets, because the accuracy depends on the data covariance matrix \( \Sigma_y \) that is estimated independently for each data set.

The FSSP algorithm is also iterative, starting with a small initial sensor network including only the reference DOFs. A single sensor is added and the posterior variances are computed for all virtual sensors at the reference locations. The cost function is evaluated for the expanded sensor network. The cost function is evaluated for different sensor networks of the same number of sensors by adding one sensor in turn. Finally, the minimum cost is found, and the corresponding expanded sensor network is selected. The process is repeated until the required accuracy is reached. The FSSP algorithm is a potential alternative, because in the present application an initial small reference sensor network is available, and the algorithm does not need to start from the scratch. In addition, FSSP can result in a dramatic reduction in the required computational effort [18].

The cost function used in this study is the error of the worst virtual sensor after removing (EFI, BSSP) or adding (FSSP) one sensor. Increasing the accuracy of the worst virtual sensor would finally result in the required accuracy of all virtual sensors.

4 NUMERICAL SIMULATIONS

4.1 Model and reference sensor network

An experiment was performed with a numerical model of a steel frame (Figure 1) with a height of 4.0 m and a width of 3.0 m. Both columns were fixed at the bottom. The frame was also supported with a horizontal spring at an elevation of 2.75 m with a spring constant of 2.0 MN/m. The frame was modelled with simple beam elements with hollow square cross section of 100 mm \( \times \) 100 mm \( \times \) 5 mm. The FE model consisted of 176 beam elements 62.5 mm in length and a single spring element.

Horizontal random loading was applied to the right column at nodes 113, 129, and 145, corresponding to elevations of 4 m, 3 m, and 2 m, respectively (Figure 1). The loads were mutually independent having standard deviations of 9 kN, 7 kN, and 5 kN, respectively. All load signals were low-pass filtered below 50 Hz.

Seven first modes were used in the simulation together with a static correction procedure [20]. Modal damping was assumed with a damping ratio of 0.01 for modes 1–2, 0.015 for mode 3, and 0.02 for modes 4–7.

The reference sensor network with a minimum number (7) of sensors was designed using the EFI method. The results are shown in Figure 1 both for acceleration and strain measurements. The same positions were also included in the new sensor network for virtual sensing.

Forced vibration was simulated and the response of all DOFs was recorded. One set of data was generated for accelerations and another for strains. Gaussian random noise was added to each sensor. The noise level was equal in all sensors and it was computed as follows. In the low-cost sensors, the average SNR was 21 dB in the network consisting of the reference
DOFs only. For acceleration measurements (case 1), it resulted in noise standard deviation of 9.47 m/s^2. For strain measurements (case 2), the noise standard deviation was 18.2×10^{-6}. The same data were used with all OSP algorithms.

Define the reference sensors and reference DOFs as the sensor network with high-quality sensors at the locations shown in Figure 1. Define the virtual sensors as the Bayesian estimates at the reference DOFs. Notice that the other locations for virtual sensing were ignored in this study.

The objective was to decrease noise in the virtual sensors by adding more sensors to the network. Three different accuracy requirements were studied: an increase of the average SNR by 3 dB, 6 dB, and 9 dB, which corresponded to average SNRs in the virtual sensors of 24 dB, 27 dB, and 30 dB, respectively. In acceleration measurements, they corresponded to the noise standard deviations of 6.71 m/s^2, 4.75 m/s^2, and 3.36 m/s^2, respectively. In strain measurements, the corresponding standard deviations of the noise were 12.9×10^{-6}, 9.10×10^{-6}, and 6.44×10^{-6}. These absolute standard deviations of the noise were used as the criteria for each sensor. Therefore, the resulted average SNRs probably exceeded the aforementioned values.

![Figure 1: Finite element model of the frame structure with loads and reference sensor networks. The horizontal black line is the spring element. Left: accelerometers. Right: strain sensors.](image)

### 4.2 Virtual accelerometers

In the first case, low-cost accelerometers were installed on the structure and the reference DOFs were estimated using empirical Bayesian VS. The average SNR of the sensors in the reference positions was 21 dB, and the objective was to study, how many sensors would be needed to increase the value of the virtual sensors to 24 dB, 27 dB, or 30 dB.

Three OSP algorithms were studied. First, the EFI method resulted in an increasing cost function with a decreasing number of sensors, as shown in Figure 2 left. Standard deviations of the estimation errors in the virtual sensors are shown in Figure 2 right for different number of sensors corresponding to the three required accuracy levels (the horizontal lines in the figure). The corresponding values of the cost function are shown with red circles in the left plot.

Starting from the top, the horizontal lines in Figure 2 right represent the physical sensors and the three criteria: average SNRs of 21 dB, 24 dB, 27 dB, and 30 dB, or equivalently, changes of the average SNR by 0 dB, +3 dB, +6 dB, and +9 dB of the virtual sensors compared to the physical measurements.
It can be seen that 12, 28, and 51 sensors were needed for an increase of the average SNR by 3 dB, 6 dB, and 9 dB, respectively. The sensor networks corresponding to those requirements are shown in Figure 3. It can be seen that all sensors had a tendency to be located at the reference locations. It should be noted that the same DOF was not allowed for multiple sensors.

![Figure 2: Optimal sensor placement using the EFI method. Left: Cost function with different number of sensors. Right: Standard deviations of the estimation errors in the virtual sensors with different number of sensors. The horizontal lines represent changes of the average SNR of the virtual sensors by 0 dB, +3 dB, +6 dB, and +9 dB compared to the physical measurements.](image1)

![Figure 3: Optimal sensor placement using the EFI method. Sensor networks with 12, 28, and 51 accelerometers corresponding to increases of the average SNR of the virtual sensors by 3 dB, 6 dB, and 9 dB compared to the physical measurements.](image2)

In the BSSP algorithm, one sensor was removed in each round by minimizing the cost function, resulting in an increasing cost function with a decreasing number of sensors shown in Figure 4 left.

Standard deviations of the estimation errors in the virtual sensors are shown in Figure 4 right for different sensor networks fulfilling the given criteria. The corresponding values of the cost function are shown with red circles in the left plot. It can be seen that 10, 19, and 37 sensors were needed for an increase of the average SNR by 3 dB, 6 dB, and 9 dB, respectively. The sensor networks corresponding to those requirements are shown in Figure 5.
It can be seen that the sensors were more widely spread from the reference locations than when using the EFI method. Also, the same accuracy could be obtained with a smaller number of sensors than with the EFI method.

Figure 4: Optimal sensor placement using the BSSP algorithm. Left: Cost function with different number of sensors. Right: Standard deviations of the estimation errors in the virtual sensors with different number of sensors. The horizontal lines represent changes of the average SNR of the virtual sensors by 0 dB, +3 dB, +6 dB, and +9 dB compared to the physical measurements.

Figure 5: Optimal sensor placement using the BSSP algorithm. Sensor networks with 10, 19, and 37 accelerometers corresponding to increases of the average SNR of the virtual sensors by 3 dB, 6 dB, and 9 dB compared to the physical measurements.

In the FSSP algorithm, starting with the DOFs of the reference sensor network, one sensor was added in each round by minimizing the cost function, resulting in a decreasing cost function with an increasing number of sensors shown in Figure 6 left. The errors of all virtual sensors were evaluated and plotted in Figure 6 right for different number of sensors corresponding to the achieved criterion. The corresponding values of the cost function are shown with red circles in the left plot.

It can be seen that 12, 24, and 52 sensors were needed for an increase of the average SNR by 3 dB, 6 dB, and 9 dB, respectively. The sensor networks corresponding to those requirements are shown in Figure 7.

Figure 6: Optimal sensor placement using the FSSP algorithm. Left: Cost function with different number of sensors. Right: Standard deviations of the estimation errors in the virtual sensors with different number of sensors. The horizontal lines represent changes of the average SNR of the virtual sensors by 0 dB, +3 dB, +6 dB, and +9 dB compared to the physical measurements.
It can be seen that the sensors were more widely spread from the reference locations than when using the EFI or BSSP algorithms. In addition, the same accuracy could be obtained with a similar number of sensors as with the EFI method.

As a conclusion, the BSSP algorithm resulted in the highest accuracy with the least number of sensors. The difference between EFI and FSSP was not significant. They resulted in a similar number of sensors but placed in different positions.

Noise reduction of virtual sensor 2 in all algorithms was higher than what was required. Therefore, the resulted average SNRs of the virtual sensors were higher than the reported values.

![Figure 6: Optimal sensor placement using the FSSP algorithm. Left: Cost function with different number of sensors. Right: Standard deviations of the estimation errors in the virtual sensors with different number of sensors. The horizontal lines represent changes of the average SNR of the virtual sensors by 0 dB, +3 dB, +6 dB, and +9 dB compared to the physical measurements.](image)

![Figure 7: Optimal sensor placement using FSSP. Sensor networks with 12, 24, and 52 accelerometers corresponding to increases of the average SNR of the virtual sensors by 3 dB, 6 dB, and 9 dB compared to the physical measurements.](image)

4.3 Virtual strain sensors

In the second case, the accuracy of virtual strain sensors at the reference DOFs was increased by adding more strain sensors in the network. The average SNR of the sensors in the
reference positions was 21 dB. The reference positions are shown in Figure 1 right. These DOFs were also included in the new sensor networks. The cost function was evaluated only at these reference locations.

First, the EFI method resulted in an increasing cost function with a decreasing number of sensors, as shown in Figure 8 left. Standard deviations of the estimation errors in the virtual sensors are shown in Figure 8 right for different number of sensors corresponding to the three required accuracy levels (the horizontal lines in the figure). Starting from the top, they represent the average SNRs of 21 dB, 24 dB, 27 dB, and 30 dB, or equivalently, an increase of the average SNR of 0 dB, +3 dB, +6 dB, and +9 dB compared to the physical measurements. The values of the cost function corresponding to the optimized sensor networks are shown with red circles in the left plot.

It can be seen that 13, 49, and 144 sensors were needed for an increase of the average SNR by 3 dB, 6 dB, and 9 dB, respectively. The sensor networks corresponding to those requirements are shown in Figure 9. It can be seen that the sensors were located in clusters. The same DOF was not allowed for multiple sensors. A very large number of sensors was needed for an increase of 9 dB. Sensors 1 and 7 seemed to be critical for the required accuracy (Figure 8 right).

Figure 8: Optimal sensor placement using EFI. Left: Cost function with different number of sensors. Right: Standard deviations of the estimation errors in the virtual sensors with different number of sensors. The horizontal lines represent changes of the average SNR of the virtual sensors by 0 dB, +3 dB, +6 dB, and +9 dB compared to the physical measurements.
The BSSP algorithm was started with all candidate sensor positions, and by removing a single sensor in each round resulted in increasing values of the cost function shown in Figure 10 left. Standard deviations of the estimation errors of all virtual sensors at the reference locations were computed and plotted in Figure 10 right for different sensor networks satisfying the given criteria. The corresponding values of the cost function are shown with red circles in the left plot.

It can be seen that 12, 27, and 78 sensors were needed for an increase of the average SNR by 3 dB, 6 dB, and 9 dB, respectively. The sensor networks corresponding to those requirements are shown in Figure 11.

It can be seen that the sensors were more widely spread from the reference locations than when using the EFI method. Also, the same accuracy could be obtained with a smaller number of sensors than with the EFI method. However, for the same increase in accuracy, the required number of strain sensors was much larger than that of accelerometers.
Figure 11: Optimal sensor placement using the BSSP algorithm. Sensor networks with 12, 27, and 78 strain sensors corresponding to increases of the average SNR of the virtual sensors by 3 dB, 6 dB, and 9 dB compared to the physical measurements.

The FSSP algorithm started with 7 strain sensors at the reference positions and added one sensor in each round so that the cost function was minimized. The decreasing cost function with an increasing number of sensors is shown in Figure 12 left.

Standard deviations of the virtual sensors at the reference DOFs are shown in Figure 12 right for different sensor networks fulfilling the given criteria. The corresponding values of the cost function are shown with red circles in the left plot.

It can be seen that 18, 39, and 83 sensors were needed for an increase of the average SNR by 3 dB, 6 dB, and 9 dB, respectively. The sensor networks corresponding to those requirements are shown in Figure 13.

It can be seen that the sensors were more widely spread from the reference locations than when using the EFI method. Also, the same accuracy could be obtained with a larger number of sensors than with BSSP but with a smaller number than with EFI.

As a conclusion, the BSSP algorithm resulted in the highest accuracy with the least number of sensors. EFI resulted in a much larger network than the other two algorithms. All methods yielded higher noise reduction in virtual sensors 2 and 4 than what was required.

Figure 12: Optimal sensor placement using the FSSP algorithm. Left: Cost function with different number of sensors. Right: Standard deviations of the estimation errors in the virtual sensors with different number of sen-
The horizontal lines represent changes of the average SNR of the virtual sensors by 0 dB, +3 dB, +6 dB, and +9 dB compared to the physical measurements.

Figure 13: Optimal sensor placement using the FSSP algorithm. Sensor networks with 18, 39, and 83 strain sensors corresponding to increases of the average SNR of the virtual sensors by 3 dB, 6 dB, and 9 dB compared to the physical measurements.

As a summary, the number of sensors in the two cases using different OSP algorithms is shown in Table 1. It can be seen that increasing the accuracy by 3 dB could be done with just a few additional sensors, whereas the increase of 9 dB required very dense sensor networks. Another important result was that for the same accuracy increase, a smaller number of accelerometers was needed than strain sensors. In both cases, the BSSP algorithm showed the best performance.

<table>
<thead>
<tr>
<th>Absolute Change</th>
<th>Acceleration measurements</th>
<th>Strain measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>dB</td>
<td>dB</td>
<td>EFI</td>
</tr>
<tr>
<td>24</td>
<td>+3</td>
<td>12</td>
</tr>
<tr>
<td>27</td>
<td>+6</td>
<td>28</td>
</tr>
<tr>
<td>30</td>
<td>+9</td>
<td>51</td>
</tr>
</tbody>
</table>

Table 1: The number of acceleration or strain sensors needed for the required increase of the average SNR of virtual sensors using different OSP algorithms.

5 CONCLUSIONS

Empirical virtual sensing was studied in an application where low-cost sensors replace traditional high-quality sensors. The research question was that how many additional sensors would be needed to achieve the same accuracy as with a smaller number of high-quality sensors. It was shown that only a few additional sensors were needed to increase the SNR by 3 dB. However, in order to increase SNR by 9 dB, a dense sensor network would be needed.

BSSP was shown to be the best OSP algorithm in all cases. With EFI or FSSP, more sensors were needed for the same accuracy. For the same increase of the average SNR, the required number of accelerometers was smaller than that of strain sensors. The measurement error was assumed to be equal in all sensors. However, the resulting accuracy of the virtual sensors was not uniform. The cost function in the OSP algorithm was the maximum error of the worst sensor. Also other cost functions could be studied, e.g. the average error in the virtual sensors.
Also virtual sensors at other DOFs than the reference locations would be available. However, they were not used in this study, because the objective was to achieve the required accuracy only at the reference DOFs. This additional information could be useful in many applications.

The virtual sensors were estimated using measurement data. Because the excitation was random, different realizations would generally result in slightly different results. Therefore, it would be important to have different data sets to investigate the effect of excitation variability. Also the errors in the finite element model parameters were ignored. Empirical VS was based on the assumption that the noise was Gaussian and known. If noise is unknown, it is possible to use the MMSE estimate (4) instead, which yields accurate virtual sensors if the number of sensors is large enough.

REFERENCES


FINITE ELEMENT MODEL DEVELOPED AND MODAL ANALYSIS OF LARGE SCALE STEAM TURBINE ROTOR: QUANTIFICATION OF UNCERTAINTIES AND MODEL UPDATING

Dimitrios Giagopoulos¹, Alexandros Arailopoulos¹, Ilias Zacharakis¹, Eleni Pipili¹

¹Department of Mechanical Engineering, University of Western Macedonia
Kozani 50100, Greece
e-mail: dgiagopoulos@uowm.gr, aarailopoulos@uowm.gr, liakos_z@hotmail.com, elenipipili92@gmail.com

Keywords: Structural Dynamics, Large scale structures, Modal identification, Model updating, Substructuring, Rotordynamics.

Abstract.

In this work, the effectiveness of a computational framework to handle large scale linear and nonlinear models is presented by calibrating a high-fidelity FE model of a steam turbine rotor with several millions of degrees of freedom, using experimentally identified modal parameters. An extensible framework for Bayesian Uncertainty Quantification and Propagation of complex and computationally demanding physical models, was connected in an efficient way with a numerical code leading to an automated determination of dynamic response of linear and nonlinear mechanical systems. The effect of correlation in the prediction error models postulated in the Bayesian model selection and parameter estimation technique is investigated. First, using an integrated reverse engineering strategy, the digital shape of the three sections of a steam turbine rotor was developed and the final parametric CAD model was created. The finite element model of the turbine were created using tetrahedral solid elements. Due to complex geometry of the structure, the developed model consists of about fifty-five million DOFs. The identification of modal characteristics of the frame is based on acceleration time histories, which are obtained through an experimental investigation of its dynamic response in a support-free state by imposing impulsive loading. The developed computational framework with appropriate substructuring methods, are used for estimating the parameters (material properties) of the finite element model, based on minimizing the deviations between the experimental and analytical modal characteristics (modal frequencies and mode shapes). Direct comparison of the numerical and experimental data verified the reliability and accuracy of the methodology applied. The identified finite element model is representative of the initial structural condition of the turbine and is used to develop a simplified finite element model, which then used for the turbine rotordynamic analysis. Also, this model can be further used for structural health monitoring purposes of the rotor.
1 INTRODUCTION

Reverse engineering is a modern field of engineering, which finds application in many areas of industry. Current industrial design requirements, tend to have the need of improving, modifying and developing new and optimized versions of various mechanical parts or even entire structures. For many of these structures, there is no available information not only about their geometric and designing details, but also about their material properties and mechanical treatment and procedures carried out during the construction process. To address this issue, an integrated reverse engineering strategy is necessary to be applied [1-5]. In this process, many issues are taken into account, related to the development of the geometry and finite element model, with the experimental modal analysis procedures and the application of robust and effective computational model updating techniques. The main objective of the present work is to demonstrate the advantages of a developed computational framework to handle large scale linear and nonlinear models, by calibrating a high-fidelity FE model of a steam turbine rotor with several millions of degrees of freedom, using experimentally identified modal parameters.

Applying classical finite element techniques, the equations of motion of mechanical systems with complex geometry are first set up. As the order of these models increases, the existing numerical and experimental methodologies for a systematic determination of their dynamic response become inefficient to apply. Therefore, there is a need for the development, improvement and application of new suitable methodologies for investigating dynamics of large scale mechanical models in a systematic and efficient way. Traditionally, in the area of structural dynamics this is done by first employing methodologies that reduce the dimensions of the original system. In this paper examined a time domain reduction method [1-5]. In order to improve the FE model of the structure, structural model updating techniques [16], have been proposed in order to reconcile the numerical (FE) model, with experimental data. Structural model parameter estimation based on measured modal data (e.g. [6-12]) are often formulated as weighted least-squares estimation problems in which metrics, measuring the residuals between measured and model predicted modal characteristics, are build up into a single weighted residuals metric formed as a weighted average of the multiple individual metrics using weighting factors. Standard gradient-based optimization techniques are then used to find the optimal values of the structural parameters that minimize the single weighted residuals metric representing an overall measure of fit between measured and model predicted modal characteristics.

The accuracy of the predictions will be uncertain, on the one hand, because of the uncertainty of all future structural excitations and on the other hand, because the structural model will always involve approximations of the real dynamic behavior that affects in an uncertain manner the predicted responses. Uncertainties related to model-structure errors, arise from the assumptions made to parameterize and describe the behaviour of the physical structure, because of simplifications and erroneous assumptions, inexact modeling of the material constitutive behaviour, inexact modeling of boundary conditions, errors because of the spatial discretization of the distributed structural system, unmodeled features such as neglected nonstructural components, as well as errors introduced by numerical methods. Such uncertainties, in this work, are treated meticulously and are minimized to the most plausible extent according to each problem, as cannot be regarded and tuned by model updating methodologies. Model updating basically regards, erroneous assumptions of model parameters such as material parameters (Young’s modulus and mass density), cross section properties (moments of inertia), shell or plate thickness, spring stiffnesses and non-structural mass.

The organization of this paper is as follows. First, in the following section, is presented an overview of the formulation for finite element model updating based on modal data. In the third
section, the structure examined (steam turbine) is introduced. More specifically, first presented the procedure followed in order to develop the digital shape of the steam turbine (rotor and blades), using a 3D Laser Scanner. The development of the detailed FE model of the structure analyzed next and finally is given a brief review of the experimental modal analysis results. Finally, the parametric studies on updating finite element model of the system are presented in the fourth section. Conclusions are summarized in the fifth section.

2 FINITE ELEMENT MODEL UPDATING METHOD

Let \( D = \{ \omega_r, \hat{\phi}_r \in \mathbb{R}^{N_o}, r = 1, \ldots, m \} \) be the measured modal data from a structure, consisting of modal frequencies \( \omega_r \) and mode shape components \( \hat{\phi}_r \) at \( N_o \) measured DOFs, where \( m \) is the number of observed modes. Consider a parameterized class of linear structural models used to model the dynamic behavior of the structure and let \( \Theta \in \mathbb{R}^{N_0} \) be the set of free structural model parameters to be identified using the measured modal data. The objective in a modal-based structural identification methodology is to estimate the values of the parameter set \( \Theta \) so that the modal data \( \{ \omega_r(\Theta), \hat{\phi}_r(\Theta) \in \mathbb{R}^{N_o}, r = 1, \ldots, m \} \) predicted by the linear class of models at the corresponding \( N_o \) measured DOFs best matches the experimentally obtained modal data in \( D \).

For this, let

\[ e_{\omega_r}(\Theta) = \frac{\omega_r^2(\Theta) - \hat{\omega}_r^2}{\hat{\omega}_r^2} \quad \text{and} \quad e_{\hat{\phi}_r}(\Theta) = \left\| \frac{\beta_r(\Theta)\hat{\phi}_r(\Theta) - \hat{\phi}_r}{\hat{\phi}_r} \right\|^2 \]

be the measures of fit or residuals [16] between the measured modal data and the model predicted modal data for the \( r \)-th modal frequency and mode shape components, respectively, where \( \|z\|^2 = z^T z \) is the usual Euclidean norm, and \( \beta_r(\Theta) = \frac{\hat{\omega}_r}{\hat{\omega}_r^2} \) is a normalization constant that guaranties that the measured mode shape \( \hat{\phi}_r \) at the measured DOFs is closest to the model mode shape \( \beta_r(\Theta)\hat{\phi}_r(\Theta) \) predicted by the particular value of \( \Theta \). To proceed with the model updating formulation, the measured modal properties are grouped into two groups. The first group contains the modal frequencies while the second group includes the mode shape components for all modes. For each group, a norm is introduced to measure the residuals of the difference between the measured values of the modal properties involved in the group and the corresponding modal values predicted from the model class for a particular value of the parameter set \( \Theta \). For the first group, the measure of fit \( J_1(\Theta) \) is selected to represent the difference between the measured and the model predicted frequencies for all modes. For the second group, the measure of fit \( J_2(\Theta) \) is selected to represent the difference between the measured and the model predicted mode shape components for all modes. Specifically, the two measures of fit are given by

\[ J_1(\Theta) = \sum_{r=1}^{m} e_{\omega_r}^2(\Theta) \quad \text{and} \quad J_2(\Theta) = \sum_{r=1}^{m} e_{\hat{\phi}_r}^2(\Theta) = \sum_{r=1}^{m} \left[ 1 - MAC^2_r(\Theta) \right] \]

where \( MAC_r(\Theta) = \frac{\omega_r^2(\Theta)\hat{\phi}_r^2(\Theta)}{\|\hat{\phi}_r\|^2} \) is the Modal Assurance Criterion [13] between experimentally identified and estimated mode shapes for the \( r \)-th mode. Alternative measures of fit can easily be used and found in literature [14-17].
Derived from the MAC for any measured frequency point, $\omega_k$ a global correlation coefficient may be used [18, 19]:

$$x_s(\omega_k) = \frac{\left| \{H_X(\omega_k)\}^H \{H_A(\omega_k)\} \right|^2}{\left( \{H_X(\omega_k)\}^H \{H_X(\omega_k)\} \right) \left( \{H_A(\omega_k)\}^H \{H_A(\omega_k)\} \right)}$$  \hspace{1cm} (3)

where $\{H_X(\omega_k)\}$ and $\{H_A(\omega_k)\}$ are the experimental (measured) and the analytical (predicted) response vectors at matching excitation - response locations. As the MAC value, $x_s(\omega_k)$ assumes a value between zero and unity and indicates perfect correlation with $x_s(\omega_k) = 1$. For $x_s(\omega_k) = 0$, no correlation exists. Similar to the MAC, $x_s(\omega_k)$ is unable to detect scaling errors and is only sensitive to discrepancies in the overall deflection shape of the structure. To emphasize this characteristic, $x_s(\omega_k)$ is accordingly called the shape correlation coefficient [19].

The lack of sensitivity to scaling of the shape correlation coefficient does not allow the identification of identical FRFs. This insufficiency becomes even more dramatic if just one measurement and its corresponding prediction are correlated. In this case, the column vectors reduce to scalars and $\{H_A(\omega_k)\} = k \{H_X(\omega_k)\}$ is always satisfied (constant $k$ may be complex), therefore leading to $x_s = 1$ across the full frequency spectrum for uncorrelated FRFs.

As a result, a supplementary correlation coefficient $x_a(\omega_k)$ is proposed by targeting the discrepancies in amplitude. The amplitude correlation coefficient is defined as:

$$x_a(\omega_k) = \frac{2\left| \{H_X(\omega_k)\}^H \{H_A(\omega_k)\} \right|}{\left( \{H_X(\omega_k)\}^H \{H_X(\omega_k)\} \right) + \left( \{H_A(\omega_k)\}^H \{H_A(\omega_k)\} \right)}$$  \hspace{1cm} (4)

where the response vectors are identical to those used for $x_s(\omega_k)$. As for the shape correlation coefficient, $x_a(\omega_k)$ is defined to lie between zero and unity. This time, however, the correlation measure is more stringent and only becomes unity if $\{H_A(\omega_k)\} = \{H_X(\omega_k)\}$. All elements of the response vectors must be identical in both phase and amplitude even if only one measurement is considered. Similarly, to modal residuals, two measures of fit are proposed using $x_a(\hat{\omega}_r)$ and $x_a(\hat{\omega}_s)$ which correspond to the identified resonant frequencies of the system:

$$J_3(\hat{\theta}) = \sum_{r=1}^{m} \left[ 1 - x_a(\hat{\omega}_r, \hat{\theta})^2 \right] \quad \text{and} \quad J_4(\hat{\theta}) = \sum_{r=1}^{m} \left[ 1 - x_a(\hat{\omega}_r, \hat{\theta})^2 \right]$$  \hspace{1cm} (5)

Minimizing at global minimum the following single objective, traditionally solves the parameter estimation problem:

$$J(\hat{\theta}; \hat{\omega}) = w_1 J_1(\hat{\theta}) + w_2 J_2(\hat{\theta}) + w_3 J_3(\hat{\theta}) + w_4 J_4(\hat{\theta})$$  \hspace{1cm} (6)

formed by the four objectives $J_i(\hat{\theta})$, using the weighting factors $w_i \geq 0$, $i = 1, 2, 3, 4$, with $w_1 + w_2 + w_3 + w_4 = 1$. The objective function $J(\hat{\theta}; \hat{\omega})$ represents an overall measure of fit between the measured and the model predicted characteristics. The relative importance of the residual errors in the selection of the optimal model is reflected in the choice of the weights. The results of the identification depend on the weight values used. The optimal solutions for the parameter set $\hat{\theta}$ for given $\hat{\omega}$ are denoted by $\hat{\theta}(\hat{\omega})$ [6, 20, 23].
3 EXPERIMENTAL APPLICATION

In this section, the proposed computational framework is applied to update the FE model of the three sections of a real steam turbine, shown in Figure 1. More specifically, the turbine type is LMZ K-300-170 with total power 310MW, which is placed in the unit IV of the thermal power plant in the area of Agios Dimitrios. The turbine consists of three sections, (Low pressure LP, Intermediate Pressure IP, High Pressure HP), which are connected in line. The total mass of the turbine is approximately 76 tones and consisted of 37 stages (10 stages in Low pressure section, 17 stages in Intermediate pressure section and 10 stages in High pressure section) and of 27 different types of blades (total number of blades 4126). The low pressure section consisted from the same blade types used in the intermediate section. The 27 different blade types are presented in Figure 2.

Figure 1: The three sections of a real steam turbine, (a) low pressure turbine, (b) intermediate pressure turbine and, (c) high pressure turbine.

Figure 2: The 27 different blade types of the three sections of a real steam turbine
3.1 Digitization and CAD Model of the Steam Turbine

Due to large size of the results of the complete steam rotor, is selected to present the basic idea of the applied method only for the intermediate pressure rotor. First, exploiting a 3D Laser Scanner, the digital shape of each blade type which used in the three turbine sections was developed by using the DSR (Digital Shape Reconstruction) method. In this process, four basic steps are being followed in order to collect, process and design the final CAD model. First, the geometrical data of each blade were captured, exploiting the 3D scanner’s functionalities, as well as its software tools in order to produce a primary stereo-lithography (STL) file. As a second step, compatible utilities were used to pre-process the initial raw model in order to create the final STL file of the digitized geometry, before designing the CAD surfaces [6, 24-26]. Next, in order to produce the initial CAD model, a segmentation of the triangulated STL model and NURBS (Non-uniform Rational B-Splines) surface fitting, were applied. The whole procedure is depicted in Figure 3, for one blade of the section IP15 of the intermediate pressure section.

Figure 3: Digitization and final 3D CAD model of one blade of the section IP15 of the intermediate pressure section.

The geometry of the rotor shaft is also unknown. The 3D CAD model of the rotor shaft was developed using the digital 2D drawings (Figure 4).

Figure 4: Digitization and final 3D CAD model of the rotor shaft of the intermediate pressure section.
The complete 3D CAD assembly of the intermediate pressure section, including the rotor shaft and the blades of all the 17 sections is presented in Figure 5.

Following the same procedure was carried out for the other two sections (low and high pressure) and, the final 3D CAD model of the complete steam turbine was developed (Figure 6).

3.2 Finite Element Model

The geometry of the turbine sections discretized by solid elements (tetrahedral). Due to complex geometry of the structure, the total number of degrees of freedom of the resulting complete rotor model is about fifty-five million degrees of freedom (55,000,000). The detailed FE Model of the frame presented in Figure 7. For the development and solution of the finite element model some appropriate software was used [27, 28]. Typical eigenmodes of the three sections and of the complete model presented in Figure 8.
Figure 7: Final finite element model of the complete steam turbine.

Figure 8: Typical eigenmodes of the three sections and of the complete model.
3.3 Experimental Modal Analysis

Next, in order to quantify the dynamic characteristics of the turbine, an experimental modal analysis of the three turbine sections was performed. Again, due to large size of the results, is selected to present the results only for the intermediate pressure rotor. The rotor, was hung up with the help of a crane and straps, to approximate free-free boundary conditions for the test.

First, all the necessary elements of the FRF matrix required for determining the response of the rotor were determined by imposing impulsive loading [6-10, 12]. The measured frequency range was 0-2048 Hz, which includes the analytical frequency range of interest, 0-700 Hz. A schematic illustration of the measurement geometry for the modal analysis of the intermediate pressure rotor, with the real experimental set up of this test is presented Figure 9. For instance, Figure 10 shows the magnitude of typical elements of the FRF matrix.

![Figure 9: Schematic illustration of the measurement geometry and real experimental set-up for the modal analysis of the intermediate pressure rotor.](image)

![Figure 10: Typical elements of the experimental FRF matrix for the intermediate pressure rotor.](image)
Based on the measured FR functions, the natural frequencies and the damping ratios of the rotor were estimated by applying the “Rational Fraction Polynomial Method” (RFPM). This method has certain attractive merits, especially for systems with high modal density, like the system under consideration \[11, 22, 29\]. The identified mode shapes have also been recorded so that they can be used for updating the finite element models. As an outcome of the above procedure, the first column of Table 1 presents the values of the lowest 11 natural frequencies \( \omega_{\text{RE}} \) of the frame, while the corresponding damping ratios are included in the fourth column. In the same table, the second column presents the values of the natural frequencies obtained from the analysis of the nominal finite element model \( \omega_{\text{RN}} \) and the third column compares these frequencies with the corresponding frequencies obtained by the experimental data.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Identified Modal Frequency</th>
<th>Nominal FE Predicted Modal Frequency</th>
<th>Difference between Identified and FE Predicted Modal Frequencies</th>
<th>Identified Modal Damping Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \omega_{\text{RE}} ) [Hz]</td>
<td>( \omega_{\text{RN}} ) [Hz]</td>
<td>( \frac{\omega_{\text{RE}} - \omega_{\text{RN}}}{\omega_{\text{RN}}} \times 100% )</td>
<td>( \zeta_{\text{RE}} ) (%)</td>
</tr>
<tr>
<td>1</td>
<td>58.70</td>
<td>63.04</td>
<td>6.88</td>
<td>0.21</td>
</tr>
<tr>
<td>2</td>
<td>101.30</td>
<td>89.52</td>
<td>13.16</td>
<td>0.19</td>
</tr>
<tr>
<td>3</td>
<td>110.60</td>
<td>127.22</td>
<td>13.06</td>
<td>0.12</td>
</tr>
<tr>
<td>4</td>
<td>126.10</td>
<td>135.62</td>
<td>7.02</td>
<td>0.08</td>
</tr>
<tr>
<td>5</td>
<td>134.80</td>
<td>143.45</td>
<td>6.03</td>
<td>0.13</td>
</tr>
<tr>
<td>6</td>
<td>135.03</td>
<td>145.41</td>
<td>7.15</td>
<td>0.13</td>
</tr>
<tr>
<td>7</td>
<td>142.32</td>
<td>156.12</td>
<td>8.84</td>
<td>0.12</td>
</tr>
<tr>
<td>8</td>
<td>147.40</td>
<td>158.23</td>
<td>8.84</td>
<td>0.16</td>
</tr>
<tr>
<td>9</td>
<td>156.77</td>
<td>171.22</td>
<td>8.44</td>
<td>0.16</td>
</tr>
<tr>
<td>10</td>
<td>161.30</td>
<td>181.36</td>
<td>11.06</td>
<td>0.35</td>
</tr>
<tr>
<td>11</td>
<td>167.70</td>
<td>183.42</td>
<td>8.57</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Table 1: Modal frequencies and modal damping ratios for the intermediate pressure rotor.

4 \hspace{1em} \textbf{FINITE ELEMENT MODEL UPDATING}

4.1 \hspace{1em} \textbf{FE model parameterization}

The parameterization of the finite element model of the intermediate section is introduced in order to demonstrate the applicability of the proposed finite element model updating method. This model consists of about twenty-three million degrees of freedom. The parameterized model consisting of forty-four parts which is shown in Figure 11. At each of these parts are used as design variables the Young’s modulus and the density. Thus, the final number of the design parameters are eighty-eight (88) variables.

The finite element model is updated using the lowest eleven identified modal frequencies and mode shapes shown in Table 1. The identified mode shapes include components at all 17 sensor locations. Additionally, we define as design response and total weight of the model, in order to be taken into consideration during the optimization process. The results from the FE model updating method are shown in Table 2. In this table presented a comparison between identified \( \omega_{\text{RE}} \) and optimal FE predicted modal frequencies \( \omega_{\text{RO}} \).
Figure 11: Parts of the parameterized FE model of the intermediate pressure rotor.

Table 2: Comparison between identified and optimal FE predicted modal frequencies for the intermediate pressure rotor.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Identified Modal Frequency</th>
<th>Optimal FE Predicted Modal Frequency</th>
<th>Difference between Identified and FE Predicted Modal Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \omega_{iz}[Hz] )</td>
<td>( \omega_{opt}[Hz] )</td>
<td>( \frac{\omega_{iz} - \omega_{opt}}{\omega_{opt}} \times 100% )</td>
</tr>
<tr>
<td>1</td>
<td>58.70</td>
<td>59.45</td>
<td>1.26</td>
</tr>
<tr>
<td>2</td>
<td>101.30</td>
<td>99.62</td>
<td>1.69</td>
</tr>
<tr>
<td>3</td>
<td>110.60</td>
<td>112.22</td>
<td>1.44</td>
</tr>
<tr>
<td>4</td>
<td>126.10</td>
<td>128.62</td>
<td>1.96</td>
</tr>
<tr>
<td>5</td>
<td>134.80</td>
<td>134.85</td>
<td>0.04</td>
</tr>
<tr>
<td>6</td>
<td>135.01</td>
<td>135.92</td>
<td>0.67</td>
</tr>
<tr>
<td>7</td>
<td>142.32</td>
<td>144.24</td>
<td>1.33</td>
</tr>
<tr>
<td>8</td>
<td>147.40</td>
<td>148.64</td>
<td>0.83</td>
</tr>
<tr>
<td>9</td>
<td>156.77</td>
<td>154.22</td>
<td>1.65</td>
</tr>
<tr>
<td>10</td>
<td>161.30</td>
<td>160.04</td>
<td>0.79</td>
</tr>
<tr>
<td>11</td>
<td>167.70</td>
<td>164.54</td>
<td>1.92</td>
</tr>
</tbody>
</table>

5 SUMMARY

The applicability and effectiveness of a computational framework namely the model reduction method, model updating method and experimental modal analysis, is explored by calibrating a high-fidelity FE model of a steam turbine rotor with fifty-five million degrees of freedom. First, using a reverse engineering method, the digital shape of the three sections of a steam turbine rotor was developed and the final parametric CAD model was created. The finite
element model of the turbine were created using tetrahedral solid elements. Modal analysis techniques were applied in order to identify the modal parameters. Direct comparison of the numerical and experimental data verified the reliability and accuracy of the methodology applied.

6 REFERENCES


[18] Grafe, H., Model updating of large structural dynamics models using measured response function, in Department of Mechanical Engineering. 1999, Imperial College: London.


OPERATIONAL MODAL ANALYSIS OF BRODIE TOWER USING A BAYESIAN APPROACH

Yan-Long Xie¹, Yi-Chen Zhu², and Siu-Kui Au³

¹ Institute for Risk and Uncertainty and Centre for Engineering Dynamics
University of Liverpool, United Kingdom
e-mail: yanlong.xie@liverpool.ac.uk

² Institute for Risk and Uncertainty and Centre for Engineering Dynamics
University of Liverpool, United Kingdom
e-mail: sgyzhu7@liverpool.ac.uk

³ Institute for Risk and Uncertainty and Centre for Engineering Dynamics
University of Liverpool, United Kingdom
e-mail: siukuiau@liverpool.ac.uk

Keywords: Ambient Vibration Test, Modal Identification, Bayesian, Multiple Setups, Operational Modal Analysis, Uncertainty, Global Mode Shape.

Abstract. This paper presents a full-scale ambient vibration test and modal identification of Brodie Tower, an eight-storey office building at the University of Liverpool. Five triaxial force-balance accelerometers were deployed. The measurement scheme comprised seven setups covering four locations on each floor, acquiring twenty minutes of ambient data for each setup. The modal properties of the building are identified using a Bayesian fast Fourier transform modal identification method incorporating multiple setup data, which accounts for the variability and the quality of data among different setups. Besides providing the estimates of modal parameters through the posterior (i.e., given data) most probable values, this method can also quantitatively assess their identification precision through the posterior covariance. The identified modal properties and their accuracy in different setups are compared and discussed. The identified global mode shapes are also compared with those obtained by global least square method. The data quality in each setup and how they affect the identification results are investigated.
1 INTRODUCTION

Operational Modal Analysis (OMA) allows one to identify the modal properties (e.g., natural frequencies, damping ratios and mode shapes) of a structure from ambient ‘output-only’ data. Modal properties are often among the first few quantities to be identified in a structural vibration project. They are often examined to provide information for structural model updating, structural modification, damage detection, or more generally, structural health monitoring [1–4].

In field vibration tests, one common situation is to measure a large number of DOFs (degrees of freedom) with a limited number of sensors. In order to identify the ‘global’ mode shape comprising a potentially large number of DOFs, a multiple-setup scheme is required where sensors are ‘roved’ to different DOFs in different setups to cover all DOFs of interest [5–7]. Conventionally, the modal parameters are identified individually using the data from each setup. The ‘representative’ natural frequencies and damping ratios may be taken as the averaged values from different setups. The global mode shapes need to be assembled (or ‘glued’) from the partial ones that only cover the DOFs measured in individual setups. Techniques such as the post separate estimation rescaling method [8] and the global least square (LS) method [9] can be used for assembling the global mode shapes. For good quality data in all setups, reported cases show that different methods yield practically the same results. In general, the quality of the global mode shape depends critically on whether the identified mode shapes of different setups match at the reference DOFs, which should not be taken for granted. Recently, identification methods capable of incorporating data directly from multiple setups have been developed based on stochastic subspace identification method [6] and Bayesian fast Fourier transform (FFT) approach [10]. Reported cases show that incorporating data from multiple setups allows one to have reasonable identification results even when the data quality in some setups is low, for which methods based on analysing data in individual setups give poor or questionable results.

Field OMA data has a variety of complications that are difficult to replicate numerically by synthetic data, especially for multiple setup data where environmental (e.g., excitation) and human (e.g., sensor alignment) factors can change over different setups. Investigation with field data is therefore indispensable for OMA research with multiple setups. As a contribution along this line, this paper presents an ambient vibration test of the Brodie Tower at the University of Liverpool. The modal properties of the building are identified using a recently developed fast Bayesian FFT method incorporating multiple setup data [10]. The identified modal properties and their uncertainties in different setups are compared and discussed. The data quality in each setup and how it affects the identification results are investigated. The identified global mode shapes are also investigated by comparing the results with those assembled by the global LS method.

2 AMBEINT VIBRATION TEST OF BRODIE TOWER

The Brodie Tower is an eight-storey concrete building located in the main campus of the University of Liverpool, UK, see Figure 1. It has a ‘T’ shape floor plan measuring 25 m (length) × 28 m (width) and approximately 25 m high. It hosts the civil engineering department, with a heavy structure laboratory in the basement.
An ambient vibration test with multiple setups was performed to determine the modal properties of the Brodie Tower, aiming at a mode shape that covers the T-shape of the floor plan in different storeys. Figure 2 shows a set of equipment used at each measurement location. Five triaxial force-balance accelerometers were deployed. The sensors have a noise level of approximately 0.1μg/√Hz for frequencies above 1 Hz. To obtain synchronised data among all the sensors, high precision clocks were used during the test. These clocks can provide accurate timing for the sensors so that their data can be logged locally without interconnection but still in a synchronous manner. Before the test, each clock was first synchronised using a GPS receiver. After that, the GPS receiver was removed and the clock was able to continuously provide practically synchronised time stamping for the sensor within the frequency range of interest for one or two days.

Figure 3 shows the sensor location and setup plan. Four locations forming a T-shape were measured on each floor. With seven floors this gives 7×4=28 locations. Sensor 1 was used as a reference and its location remained the same in all setups. The remaining four sensors were roved from the top floor to the bottom, covering all locations in seven setups. The data in each setup was recorded for 20 minutes at a sampling rate of 50 Hz. The transition from one setup to another took about 5 minutes. Synchronising clocks in advance took approximately 20 minutes. The whole test lasted for approximately 3.25 hours from 1:30 pm to 4:45 pm.
The data obtained from the field test in the last section is analysed using a recently developed Bayesian modal identification method that is capable of incorporating data from different setups directly [10]. The theory is summarised in this section. Consider ambient vibration data from setups, each covering a possibly different set of DOFs. Let \( \theta \) denote a set of modal parameters of a well-separated mode, which includes natural frequencies \( \{f_i\} \), damping ratios \( \{\zeta_i\} \), the spectral density of modal excitations \( \{S_i\} \), the spectral density of prediction errors \( \{S_e\} \) and partial mode shapes \( \{\phi_k \in \mathbb{R}^{n \times 1}\} \) (i.e., covering the DOFs in a particular setup only), where \( n \) is the number of measured DOFs in setup \( i \). Here, the subscript \( i \) denotes that the quantity refers to the one in the \( i \)-th setup. Let \( L_i \in \mathbb{R}^{n \times n} \) be the selection matrix, where \( n \) is the total number of measured DOFs. The \( (j,k) \) entry of \( L_i \) equals to 1 for DOF \( k \) measured at the \( j \)-th channel in setup \( i \), and zero for the remaining entries. The partial mode shape \( \phi_i \) is related to the global mode shape \( \Phi \in \mathbb{R}^{n \times 1} \) by:

\[
\phi_i = L_i \Phi \quad (i = 1, \ldots, n)
\]

The measured acceleration data is modelled as

\[
\hat{x}_j = \bar{x}_j + \varepsilon_j \quad (j = 1, \ldots, N)
\]

where \( \bar{x}_j \) denotes the theoretical modal acceleration that depends on \( \theta \); \( \varepsilon_j \) is the prediction error; \( N \) is the number of samples per channel. The scaled FFT of the data is defined as

\[
F_k = \sqrt{(2\Delta t/N)} \sum_{j=1}^{N} \hat{x}_j \exp\{-2\pi i((k-1)(j-1)/N)\} \quad (k = 1, \ldots, N)
\]

where \( \Delta t \) is the time interval and \( i^2 = -1 \). Let \( \{F_k\} \) denote the collection of \( F_k \) over a selected band around the mode of interest. For a small \( \Delta t \) and large \( N \), it can be shown that \( \{F_k\} \) are asymptotically independent and jointly ‘circularly complex Gaussian’ [11]. Applying Bayes’ theorem and assuming a non-informative prior distribution for \( \theta \), the posterior probability density function is proportional to the likelihood function, i.e.,

\[
p(\theta|Z_k) \propto p(\{Z_k\}|\theta)
\]

The most probable value (MPV) of the modal parameters can be determined by maximising \( p(\{Z_k\}|\theta) \), or equivalently, minimising its negative logarithm of the likelihood function (NLLF)
with respect to $\theta$. As the NLLF depends on $\theta$ in a general nonlinear manner, analytical solution for the MPV has not been sought. Rather, it can be shown that the MPV of the global mode shape can be determined analytically in terms of other parameters. This allows the MPV of the full setup of modal parameters to be determined by an iterative algorithm, where parameters of different groups are updated until convergence.

4 DATA ANALYSIS

Data analysis and modal identification are presented in this section. In section 4.1, singular value (SV) spectrum is first examined to locate potential modes. Modal identification results are presented in section 4.2. Section 4.3 investigates the identified global mode shapes by comparing the results with those assembled by the global LS method.

4.1 The SV spectrum

The SV spectrum is often used to discover potential modes in frequency domain. It is a plot of the eigenvalues of power spectra density matrix of data with frequency. Lines that describe the shape of dynamic amplification in the SV spectrum indicate potential modes. Figure 4 shows the root SV spectrum of the data in setup 2. Six potential modes below 10 Hz are highlighted. The bar below each mode indicates the band whose FFT will be used for identifying the mode; the circle indicates the initial guess of natural frequency.

![Figure 4: The root singular value spectrum of the data in setup 2.](image)

4.2 The identified modal properties

Figure 5 shows the identified natural frequencies and damping ratios in different setups. The blue circles represent the MPVs of the modal parameters and the error bars cover +/- 2 posterior standard deviation. The sample mean and sample coefficient of variation (c.o.v. = sample standard deviation / sample mean) of the modal parameters among different setups are shown in Table 1. The results show that the identified natural frequencies and damping ratios vary among the setups, but the variability of the natural frequencies is relatively small compared to the damping ratios.
Table 1: The sample mean and sample c.o.v. of the natural frequencies and damping ratios among the setups.

<table>
<thead>
<tr>
<th>Mode</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Natural Frequency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean (Hz)</td>
<td>2.422</td>
<td>2.707</td>
<td>3.750</td>
<td>7.405</td>
<td>7.995</td>
<td>9.782</td>
</tr>
<tr>
<td>c.o.v. (%)</td>
<td>0.16</td>
<td>0.2</td>
<td>0.25</td>
<td>0.22</td>
<td>0.42</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>Damping Ratio</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean (%)</td>
<td>1.031</td>
<td>0.905</td>
<td>0.754</td>
<td>2.525</td>
<td>2.624</td>
<td>6.439</td>
</tr>
<tr>
<td>c.o.v. (%)</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>15</td>
<td>14</td>
</tr>
</tbody>
</table>

Figure 5: The identified natural frequencies and damping ratios among the setups, the error bars cover +/- 2 standard deviation.
Based on the identification results, the modal signal-to-noise (s/n) ratio can be calculated, which reflects how well-excited of a mode relative to the prediction error level. It is given by the ratio of the spectral density of modal excitation to the spectral density of prediction error at the resonance peak [12], i.e.,

$$\gamma_i = S_i / 4S_e s_i^2$$

where $\gamma_i$ is the modal s/n ratio of the setup $i$.

Figure 6 shows the modal s/n ratio of mode 1 to 6 among different setups, where different colour of ‘*’ represents the modal s/n ratios in different setups. It can be seen that the modal s/n ratio of the first three modes are very high (over 500). The modal s/n ratio of mode 4 and 5 are moderate (30~100), while mode 6 has the lowest modal s/n ratio, in the order of 10. The level of modal s/n ratio also can be predicted from the SV spectrum. The first three peaks in the SV spectrum (Figure 4) suggest that the modal s/n ratios are relatively high, while the same is not true for the mode 4 to 6.

![Figure 6: The modal s/n ratio among different setups of the mode 1 to 6.](image)

**4.3 The global mode shapes**

Figure 7 and Figure 8 show the global mode shapes obtained by the Bayesian method incorporating multiple setups and the global LS method, respectively. The sample mean of the natural frequencies and damping ratios are also shown in these figures. The modal assurance criterion (MAC) values between the global mode shapes obtained by the two methods are listed in Table 2.

![Figure 7: The global mode shapes obtained by the Bayesian method incorporating multiple setups.](image)
Figure 8: The global mode shapes assembled by the global LS method.

<table>
<thead>
<tr>
<th>Mode</th>
<th>MAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9999</td>
</tr>
<tr>
<td>2</td>
<td>1.0000</td>
</tr>
<tr>
<td>3</td>
<td>0.9999</td>
</tr>
<tr>
<td>4</td>
<td>0.9999</td>
</tr>
<tr>
<td>5</td>
<td>0.9999</td>
</tr>
<tr>
<td>6</td>
<td>0.8244</td>
</tr>
</tbody>
</table>

Table 2: The MAC values between the global mode shapes obtained by the Bayesian method incorporating multiple setups and the global LS method.

It can be seen that the first five global mode shapes obtained by these methods are very close to each other and their MAC values are almost equal to 1. The mode 1 and 2 are translational modes in x and y direction, respectively. The mode 3 is a torsional mode with its torsional centre at the left side of the T-shape connection. The mode 4 and 5 are second harmonics of the translational modes in the x and y direction, respectively.

Figure 9: Multiple views of the global mode shape of mode 6 obtained by the Bayesian method incorporating multiple setups (left) and the global LS method (right).

For the mode 6, the global mode shape obtained by the Bayesian method incorporating multiple setups is different from the one assembled by the global LS method. The MAC value between the two methods is 0.8244. Figure 9 shows a detailed plot of the mode shape obtained by these methods. The one obtained by the Bayesian method incorporating multiple setups is a combination of a vertical mode and a second translational mode in the y direction. The one
assembled by the global LS method is not physically plausible. The result obtained by the Bayesian method incorporating multiple setups is more reasonable, which demonstrates the potential benefit of incorporating information directly from multiple setup data.

5 CONCLUSIONS

This paper investigates the quality of identification results based on field test data using the Bayesian modal identification method incorporating multiple setup data. A full-scale ambient vibration test of an eight-storey building has been presented. The identified modal parameters and their c.o.v.s in different setups are compared and discussed. The identified global mode shapes are investigated by comparing the results with those assembled by the global LS method.

Based on the identification results, it can be seen that when the modal s/n ratio is high or moderate (e.g., > 100), the global mode shapes obtained by the Bayesian method incorporating multiple setups and the global LS method are very close to each other. When the modal s/n ratio is low (e.g., ~10), the global LS method may give questionable results, while the Bayesian method incorporating multiple setups still gives reasonable results.

ACKNOWLEDGEMENTS

This work is funded by the UK Engineering & Physical Sciences Research Council (EP/N017897/1). The financial supports from the China Scholarship Council (first author) and the Tung’s scholarship (second author) are gratefully acknowledged.

REFERENCES


DATA–DRIVEN IDENTIFICATION, CLASSIFICATION AND UPDATE OF DECISION TREES FOR MONITORING AND DIAGNOSTICS OF WIND TURBINES

Imad Abdallah, Vasilis K. Dertimanis, and Eleni N. Chatzi

Institute of Structural Engineering, ETH Zürich
Stefano-Franscini-Platz 5, 8093 Zürich, Switzerland
e-mail: abdallah,v.derti,chatzi@ibk.baug.ethz.ch

Keywords: Wind turbines, data–driven updating, probabilistic decision trees, quantitative risk analysis, O&M, object–oriented design

Abstract. This paper describes a conceptual framework for the online monitoring of wind turbines (WTs) relying on an object–oriented (OO), event–and–decision tree–driven platform for information processing and propagation. To this end, a WT is viewed as a multilayered system of objects (e.g. structure, controller, actuator, etc.) that are defined on the basis of abstract superclasses, attributed with specific properties and methods. The former generally provides insight about the current state of the respective object, while the latter communicates state information and determines the interaction among objects and events. The term state refers herein to a set of mutually exclusive “positions”, which a specific object may reach (e.g. safe, critical, fail, etc.), while an unknown state is also included in order to take into account possible combinations of events that have not been registered during the design phase and would eventually be identified in the decision tree through the real–time telemetry. The envisioned platform is purely probabilistic, e.g. a set of probabilities is initially assigned to all events and updated accordingly, based on actual information extracted from the WT. This information may either be acquired using sensors (through corresponding sensor objects), or may be estimated using appropriate algorithms (through corresponding methods, such as Kalman-based filters). A paradigm of the proposed conceptual framework focuses on the tower substructure of the WT and indicates the potential of the proposed approach, especially in respect to the design of specialized software for monitoring and diagnostics of both new and existing WT installations.
1 INTRODUCTION

Despite the fact that wind turbines (WTs) constitute one of the most important, continuously evolving and expensive structures in modern engineering [1, 2], their maintenance and inspection is still succeeded using conventional methods [3], such as visual inspection (planned and unplanned), non-destructive evaluation (i.e., crack detection using ultrasound technologies) and post–processing of SCADA data. Specialized methods are only focused on specific components, as for example the condition monitoring (CM) of elements of the drive train (i.e., gearbox and main bearing) [4], while structural health monitoring (SHM) systems are deployed mostly for research purposes, or only during the certification stage, and are far from forming part of the actual engineering practice [5]. As a result, there exist currently no systematic, quantitative and automated tools for monitoring, detection and diagnostics of WTs, for operation, maintenance (O&M) and decision making within their life-cycle.

In identifying possible reasons for this discrepancy, one may argue that state–of–the–art CM and SHM methods are still not straightforward to adopt, when O&M engineers are in fact seeking robust and reliable decision making tools that compress data in user–friendly output formats. They are also by default “local”, in the sense that they are configured for monitoring only a specific WT component (e.g. the blade, the gearbox, the tower etc.), and they generally lack cross communication and interaction capabilities. As a result, in this way, interconnecting links or the monitoring of critical events among the components of a WT is a non-trivial task.

A remedy solution to this problem may be provided via implementation of event and decision trees. Event trees are traditionally used in quantitative risk analysis of engineering systems, laying a path from an initiating event to an end state of a system. An initiating event may correspond to a sudden increase in wind speed (gust) or an earthquake tremor, and the corresponding end states may pertain to critical interference of a wind turbine blade with the tower surface or a crack in the concrete foundation, respectively. For a given initiating event, multiple end states are possible. Intermediate chronological events make up the branches leading from the initiating event to the end state, and each event is associated with a probability of occurrence.

Under this perspective, previous studies have presented general purpose solutions that are based on event and decision trees [6, 7, 8], as well as on Bayesian networks [9]. However, their common feature is that they only correspond to the design phase of the system and, once formulated, they remain static. In this work, we propose real-time updating of decision trees as a decision-making tool for O&M that integrates live data (or estimates) from the individual components of WTs. Decision trees are preferred over Bayesian Networks, as the former can be easily updated from data, they are visually more appealing and simpler to interpret, while it is easier to follow and track an event path, a way that follows the sequence and chronology of how events are interlinked.

On the basis of our proposed supervisory platform, (i) events are classified and event probabilities are updated via real time telemetry from the WT; and (ii) new initiating events and end states are identified. In the first step, a set of condition and/or structural data samples can be trained using decision trees, which are initially provided by engineers. Then, by running new data through the trees, classification and prediction may be carried out. Data that does not fit the existing decision tree structure is used to identify new initiating events and end states. We stipulate that such a framework can be used for real-time monitoring and diagnostics of structures, root cause analysis of future failures and quantitative risk analysis in the context of operation and maintenance scheduling of components. The use of decision trees is motivated by the fact that they tend to be easier to interpret than other quantitative data-driven methods such
Figure 1: Schematic of a wind turbine and its interaction with the environment.

as Bayesian networks (Bayesian networks may link more variables in complex, direct and indirect ways, rendering interpretation oftentimes problematic) [10]. Furthermore, the learning and updating process of decision trees from real–time data is a far more mature field as compared to the training of Bayesian Networks [11].

Our envisioned approach follows an object–oriented (OO) architecture. This implies that all the aforementioned features are developed as corresponding properties and methods of abstract classes and interfaces, instances of which are used to interpret a multi–layered diagram of WT objects and the interactions among them. Decision trees are embedded locally in a specific object (i.e., the tower, the controllers, the blade, etc), which is essentially a finite state machine. Under this layout, possible CM and SHM tools are also locally registered and they are used to determine the current state of an object and/or notify neighbouring objects for the occurrence of an event or a change in state.

2 PROBLEM FORMULATION

2.1 The wind turbine

Figure 1 sketches a representative wind turbine (WT) facility and its interaction with the environment. The WT is exposed to wind and wave loads, which act primarily on the structural components, i.e., the foundation, the tower and the blades. While the wave motion is a typically uncontrolled (and often unmeasurable) form of structural excitation, the wind speed and direction are essentially responsible for the production of energy. The latter is achieved via the effective cooperation of the pitch and yaw control systems that optimally adjust the angles of the blades and the nacelle.
Component Function
Rotor blades Convert aerodynamic forces into rotational motion
Drive train Converts the rotational motion of the rotor blades into electric power, and consists of main bearing, main shaft, gearbox, brake, coupling and generator
Nacelle and main frame Houses and supports the drive train and transfers the loads from the rotor to the tower, respectively
Tower Erects the rotor up in the air, and insures that the loads are transferred from the rotor down to the foundations
Foundation Ensures load bearing capabilities and support of the wind turbine (resist overturn)
Controllers Ensure that the turbine operates within the structural design limits, while maximizing power extraction from the wind
Sensors Measure quantities critical for operation and monitoring (e.g. wind speed & direction, pitch and yaw angles, bearing’s acceleration, etc.)
Actuators Transfer the control commands to the structural components (e.g. nacelle and blade hub)
Safety systems Override control commands to ensure safety and structural integrity of the wind turbine

Table 1: The main components of a WT and their functionality.

A modern WT is composed of several interrelated subsystems that fulfill specific tasks. The most important of these are listed in Table 1 along with their main functionality. It must be noted that each subsystem consists of a number of subcomponents itself, rendering the WT a complex multi–level engineering system. Indicatively, the drivetrain consists of a frame construction that supports the main bearing, main shaft, the gearbox, a brake and a high–speed coupling, as well as the generator that transforms the mechanical motion into electrical energy. It is thus apparent that the structural and operational integrity of the facility are strongly dependent on the constitutive sub–components of each sub–system.

2.2 The monitoring problem

Under this setting, the problem considered herein pertains to the establishment of a robust, data–driven supervision platform for the monitoring of WTs over their whole life-cycle, which combines three main drivers:

[D1] An effective tool attributed with the ability of tracking, identifying and classifying indi–
individual events based on real-time telemetry from components and propagating their implications to the global state.

[D2] The integration of a probabilistic framework, aiming at projecting the current state of all individual components to the future, and providing both short and long-term predictions of the local and global safety margins.

[D3] Flexibility of the tool to integrate new components (e.g. new sensors, new controller, etc.) in the context of evolving structures.

Due to the inherent complexity of the monitored system, the envisioned supervision platform should be further attributed with adaptive tools that may create sequences of events that were not predicted during the design stage. The conceptual framework of such a platform is outlined in what follows.

3 THE SUPERVISORY MONITORING PLATFORM

3.1 Description

The proposed monitoring framework is based on the OBEST paradigm [6, 7] and its predecessor [8], in which engineering systems and their constitutive components are represented as objects (e.g. instances of predefined classes) that are interlinked and exchange information with respect to their current state. OBEST is composed of four different views that describe the functionality of a system in various levels. These are the system structure diagram, the interaction diagram, the state transition diagram and the data flow diagram. In our implementation we currently focus on the first three, as the latter is strongly dependent on the availability of measured information from the WT, rather than to an explicit scheme that is defined during the design stage of the WT.

The system structure and interaction diagrams describe the constitutive objects of the system and their interaction, respectively. For a WT, a merged view of both is displayed in Fig. 2, where the boxes correspond to the individual WT components and subsystems and the arrows to the interactions among them. In general, a piece of information received by an object results in altering its internal state or a subset of its fundamental characteristics (attributes). This accordingly implies the transmission of another piece of information that communicates its new condition to further objects. Thus, the behaviour or state of each component in the system is entirely encapsulated within the confines of an object and the performance of the entire WT is represented by combining and connecting object models for individual components or subsystems.

Each component/subsystem has an internal representation that is modelled using discrete, mutually exclusive states and corresponding state transition events (refer to Sec. 4 for a visual interpretation of such a diagram). The initial state transition trees are herein provided by the component engineer/designer. However, unlike in traditional quantitative risk assessment and system reliability, the initial (a-priori) event-state transition trees are further updated by actual measurements of the WT system; we propose to perform the updating using a decision tree learning approach (ID3, CART, C4.5/5.0, Random forests, Naive Bayes, etc.) [12].

3.2 Data-driven probabilistic updating

A novelty of our framework is the integration of Structural Health Monitoring (SHM) data, such as strains and accelerations, complementary to Condition Monitoring (CM) ones such as
control signals, status logs, temperature, etc. when monitoring a structure and diagnosing any faults or failure. This is observed in Fig. 2, where the Sensors object (see below) receives both wind speed and strain measurements as inputs from the blade and subsequently forwards this information to the pitch control algorithm object. Both condition and structural signals are then propagated through the pitch control object internal event tree until an end state is reached and an output message is accordingly transmitted to the pitch actuators object.

The updating aspect of the framework consists in identifying new branches in the event tree and classifying initially unknown end states. Such end states could be, for example, operating modes such as normal, critical, abnormal, fault, failure, and unclassified. As the continuous stream of telemetry is propagated through the event trees of each component, classifications and predictions are continually accomplished. Decision trees are admittedly not the the most competitive means of classifying telemetric data or for assessing conditional probabilities in complex systems, but they offer the benefit of straightforward interpretation by non-expert operators via visualization of the chains of events following an accidental end state. Another advantage of decision trees is related to diagnosis, which is the task of locating the source of a system fault once it is detected by tracing the sequence of events in the tree that lead to fault or failure [13].

Such an object oriented framework together with updating through decision tree learning is appealing in the context of a system that is continuously evolving/ageing in time, which is particularly true for a wind turbine system. For instance, ageing and degradation of the blade composites or the gearbox bearings introduce new and previously unaccounted for states, which could be captured via decision tree learning. Furthermore, wind turbines are operating under significantly diverse environmental conditions, hence the resulting ageing process and the faults appearing may not be identical across similar systems. The decision tree learning offers the possibility to infer new and distinct states from each of operating turbine. Finally, a common scenario involves renewal and replacement of components on wind turbines; for instance ten years into the operation of a wind turbine, the operator may decide to introduce a new pitch control algorithm/features for load reduction, which then introduce new operating modes (events) and states. In the context of the proposed object oriented framework for monitoring and diagnostics, this can be readily achieved and seamlessly integrated with the already existing components (objects).

The framework we propose is probabilistic in nature. The first possibility here is to use the decision tree, where the conditional probability of an end state $\text{EndState}_i$, depends on the
conditional probabilities of the preceeding events $e_i$ on a path such that:

$$P(EndState_i) = P(e_1) \cdot P(e_2 | e_1) \cdot P(e_3 | e_2, e_1) \ldots$$  \hspace{1cm} (1)$$

The limitation appears when the component displays a behaviour with feedback (i.e., after a repair or update in the system) or for evolving systems (e.g. when new sensors are integrated or aging of the system), which implies a need to establish several decision trees based on the possible ordering of the events or based on new initiating events. One way around this is to convert decision tree learners to undirected Markov networks [14] or make use of hidden Markov decision trees [15, 16]. An alternative is to map the decision trees learners into Bayesian Networks for further assessment of the conditional probabilities [17]. The conditional probabilities of the end states are a means to predicting the likelihood of potential fault or failure occurring. Such a probabilistic approach would inherently include uncertainties stemming from measurements.

For systems comprising several elements, the number of different combinations of inspection and maintenance could be large [18]. The possibility to update the conditional probabilities of end states and intermediate events based on real-time telemetry facilitates pre-posterior type of analysis (i.e., assessment of the value of data before they become available), and allows for optimal planning of inspection and maintenance on the structure.

### 3.3 Objects

An extensive outlining of the individual classes that define the objects of the WT subsystems falls outside the aim of the current study. However, our framework relies on a number of abstract class and interface definitions, which may be easily inherited to other pending subclasses. These are illustrated in Fig. 3 and follow the conventional object oriented design pattern. In the current stage of our conceptual interpretation, these are separated into two broad categories: the first corresponds to classes directly related to the WT and its constituent subsystems (e.g.
windTurbine class, structure class, controller class, etc.), while the second category defines the aspects of the supervisory platform that correspond to interactions, events and information flow (e.g. associated interfaces and eventTree, state and observer classes).

The specified properties and methods of the definitions of Fig. 3 are indicative and may include any individual method that is integrated into the monitoring platform. For example, an adopted SHM method that may be used for tracking the remaining fatigue life of a structural element can be placed within the setState() method of the structure class. Similarly, a Kalman filter that may be established for estimating unmeasured states required for the controller class can be registered within the readEstimators() method of the observer class.

4 EXAMPLE: THE TOWER SUBSTRUCTURE

As an indicative example of our framework, we focus on the tower component of the WT. A tower object is an instance of the tower class, which is a child of the structure superclass and inherits its properties and methods. Figure 4 illustrates the adopted interaction diagram. It is assumed that both waves and wind do not contribute in a significant way to the excitation of the tower. Thus, its structural response is only affected by the interaction forces that are developed with its neighbouring components, the foundation and the rotor-nacelle assembly. The latter are also instances of the corresponding foundation and nacelle classes, again children of the structure superclass.

Fig. 5 displays an indicative event-state transition tree for the tower. The properties of the state class of Fig. 3 are clearly indicated, along with events that have been specifically defined for this object. Notice the presence of an Unclassified state and Unknown events in Fig. 5. The decision tree would, based on the real-time telemetry from the tower, learn those unknown events and the corresponding unclassified states. In a fully probabilistic framework, all the events among the different states must be attributed with corresponding conditional probabilities. The latter are initially assigned during the design stage and updated accordingly with respect to the actual structural behaviour based on the real-time telemetry. The latter could be achieved by combining both sensors, which provide a direct indication of the structural response at the measurement points, and estimators, which can be utilized in order to estimate the structural response in unmeasured spots along the tower [19]. The term estimator could refer to, but is not restricted to, Kalman filters, surrogate models, and reduced time marching simulators which could be run in near real-time.
5 CONCLUSIONS

We provided a conceptual description of a monitoring and diagnostics framework of wind turbines. The framework combines three main drivers:

[D1] tracking, identifying and classifying individual components’ decision trees based on real–
time telemetry and propagating their implications to the global state

[D2] real-time updating of conditional probabilities of events and end states of each component
aiming at providing both short and long–term predictions of the safety margins

[D3] flexibility to integrate new system components (e.g. new sensors, new controller, etc.) in
the context of evolving structures

The above drivers are rendered possible by the OO architecture of the framework, where each
component (object) is defined as an extension of an abstract superclass. Two broad families
of superclasses are defined, the first pertaining to the wind turbine and the second relating to
interaction, events and information flow. We illustrate the framework with an example case-
study involving the tower object.

ACKNOWLEDGEMENTS

The authors would like to gratefully acknowledge the support of the European Research
Council via the ERC Starting Grant WINDMIL (ERC-2015-StG #679843) on the topic of Smart
Monitoring, Inspection and Life-Cycle Assessment of Wind Turbines.
REFERENCES


MODEL UPDATING OF A NONLINEAR EXPERIMENTAL VEHICLE USING SUBSTRUCTURING AND UNSCENTED KALMAN FILTERING
Maria Tsotalou¹, Dimitrios Giagopoulos¹*, Vasilis Dertimanis², Eleni Chatzi²

¹Department of Mechanical Engineering, University of Western Macedonia
Kozani, Greece
mtsotalou@uowm.com, dgiagopoulos@uowm.gr

²ETH Zurich, Inst. of Structural Engineering, Dpt. Of Civil, Environmental and Geomatic Engineering
Zurich, Switzerland
v.derti@ibk.baug.ethz.ch, chatzi@ibk.baug.ethz.ch

Keywords: Nonlinear Dynamics, Finite Elements, Model Updating, Unscented Kalman Filter, Substructuring

Abstract. This study establishes a computational framework for nonlinear finite element (FE) model updating of large-scale structures, through the exploration of two case studies pertaining to a spring-mass-damper chain model and to a laboratory vehicle with nonlinear suspensions. The proposed approach combines a substructuring model reduction approach with a near real-time system identification scheme, namely the Unscented Kalman Filter (UKF). The former aims at isolating and locally updating individual structural subsystems of a large-scale structure, while the latter, in contrast to other alternatives (e.g. the Extended Kalman Filter), offers a number of advantages in treating nonlinear systems, such as a derivative free calculation and a capacity for handling higher order nonlinearities. To this end, after formulating a detailed large-scale FE model for the vehicle frame substructure, a lumped model is adopted for the description of the nonlinear suspensions. Accordingly, a joint state and parameter estimation (JS&PE) problem is formulated on the basis of the lumped model. The proposed framework uses acceleration measurements from a limited number of sensors attached on the structure and a UKF observer for fusing these with a nonlinear FE substructure model, resulting into a JS&PE problem. The results indicate the validity of the proposed framework and motivate further implementation in large-scale structural systems with nonlinear components.
1 INTRODUCTION

A structural system may be often formulated as an assembly of distinct linear and nonlinear substructures [1]. For instance, a vehicle structure consists of the body frame, which is designed to behave linearly, and the four suspension-wheel substructures, which typically exhibit nonlinear behavior and are subjected to external excitation. In order to simulate the dynamic behavior of such systems it is important to develop a high fidelity finite element (FE) model, reducing the discrepancies between analytical predictions and the real structure. To this end, a model updating procedure need be put in place in order to fine-tune and adjust the parameters of numerical models on the basis of experimental measurements. A main challenge involved in the updating procedures pertains to the oftentimes large number of degrees of freedom (DOFs) involved. To tackle with this problem, appropriate substructuring methods in either the time or the frequency domain have been developed and are commonly employed [2-4].

Furthermore, large-scale nonlinear FE models are further characterized by uncertainties, which are usually represented as unknown model parameters (e.g. individual flexibility, material properties, etc.) [5]. The main objective of this study is to establish a computational framework for identifying and adjusting these parameters, while estimating the structural states, in a problem that is referred to as joint state and parameter estimation (JS&PE) [6-7]. To achieve this, the Unscented Kalman Filter (UKF) is utilized [8], due to its efficient performance of in real-time nonlinear system identification problems [9-13].

Under this setting, this study proposes a multi-stage optimization process for the nonlinear model updating of large-scale uncertain structures, which is based on the integration of substructuring and the UKF. To this end, a nonlinear and uncertain structural system is decomposed into linear and nonlinear uncertain substructures and the kinematic and dynamic constraints along the boundaries are established. The UKF is then applied to the uncertain and/or nonlinear substructure(s), resulting in the estimation of both the unknown structural states and parameters. The acceleration time histories of each local substructure are imported as base excitations in the linear substructures, in order to calculate the dynamic response of the latter. Finally, the calculated acceleration time histories at the measurement locations of the linear substructure are compared with the experimental measurements, in order to quantify the discrepancy between analytical and experimental models and define a response residual. These residuals are minimized to acquire a best fit model between the estimated quantities and those identified from the experiments.

The organization of this paper is as follows. In the following section, a brief review of UKF formulation for state and parameter estimation is presented. Then, the effectiveness and accuracy of the developed computational framework is demonstrated by presenting numerical results obtained for two selected examples. The study concludes by presenting a summary of the results.

2 REVIEW OF THE UNSCENTED KALMAN FILTER FORMULATION FOR STATE AND PARAMETER ESTIMATION

A broad range of mechanical systems may be composed into individual structural components that are linearly deformable and interconnected by local elements of purely nonlinear behavior. In this case, the corresponding equation of motion is

$$\mathbf{M}\ddot{\mathbf{z}}(t) + \mathbf{C}\dot{\mathbf{z}}(t) + \mathbf{K}\mathbf{z}(t) + \mathbf{h}(\mathbf{z}, \dot{\mathbf{z}}) = \mathbf{F}(t)$$

(1)

where the vector $\mathbf{z}(t)$ corresponds to the displacement of the system, $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$ are the mass, damping and stiffness matrices, respectively, $\mathbf{h}(\mathbf{z}, \dot{\mathbf{z}})$ includes the nonlinear forces imposed by
the interconnecting elements, herein assumed to be “smooth”, and $\mathbf{F}(t)$ is the vector of externally applied forces. The prediction of the response of systems represented by Eq.(1) is a difficult task, since in most practical cases the number of the equations of motion is quite large and the nonlinearities are dominant. As a result, such systems can only be studied by applying special numerical methodologies. In many cases, the underlying computations are facilitated after applying appropriate model reduction methodologies that lead to a significant “compression” of the original deformable coordinates, without affecting considerably the accuracy of the results.

The KF is a Bayesian approximation technique for state estimation using noise-corrupted measurements from a subset of the structural degrees of freedom [14]. The KF assumes availability of system matrices, as well as the presence of process and measurement noise, and in its discrete-time version is represented by a state-space model of the form

$$
\begin{align*}
x_{k+1} &= A x_k + B u_k + w_k \\
y_k &= C x_k + D u_k + v_k
\end{align*}
$$

where $A \in \mathbb{R}^{n \times n}$ is the state matrix, $B \in \mathbb{R}^{n \times l}$ is the input matrix, $C \in \mathbb{R}^{m \times n}$ is the output matrix, $D \in \mathbb{R}^{m \times l}$ is the feedforward matrix, $u \in \mathbb{R}^{l \times 1}$ is the control vector, $x \in \mathbb{R}^{n \times 1}$ is the state vector, comprising the system displacements and velocities, $x = [z, \dot{z}]^T$, $y \in \mathbb{R}^{m \times 1}$ is the observation vector and $k = 1, 2, \ldots$ are the discrete-time steps. The stochastic nature of the KF is introduced by the process and measurement noise vectors $w \in \mathbb{R}^{n \times 1}$ and $v \in \mathbb{R}^{m \times 1}$, respectively, assumed to belong to the $N(0, Q_k)$ and $N(0, R_k)$ Gaussian distributions, respectively.

The effective performance of the KF depends on (a) how accurately the process model can track the actual system; (b) the assumption of additive, independent white Gaussian noise; and (c) the accuracy of the process and measurement noise processes. For large modeling uncertainties and/or unknown structural parameters, the KF is generally amenable to inconsistencies, or even fails at providing state estimates. To this end, the use of an alternate Bayesian filter is proposed herein, namely, the Unscented Kalman Filter (UKF), which is capable of joint parameter and state identification.

In handling linear systems of the type of Eq.(2), where the system properties, i.e., the elements of matrix $A$, are considered as unknowns, a joint state and parameter formulation is adopted. This demands an augmentation of the regular state vector $x$, in order to include those properties of the system that are considered as unknowns and which can be gathered in a parameter vector $\theta$. The augmented state vector is defined as $\bar{x} = [x, \theta]^T$. The resulting system is of nonlinear nature since it comprises bilinear products of the components $x$ & $\theta$ of $\bar{x}$. The UKF is chosen herein in place of the widely used nonlinear filter alternative, the extended KF (EKF) [15, 16], since it is able to overcome some significant shortcomings of the latter when dealing with higher order nonlinearities and noise contamination. The UKF models the state as a Gaussian random variable whose distribution can be approximated by a carefully chosen set of deterministic points, namely the sigma points. These points capture the prior mean and covariance of the state and when propagated through the nonlinear function, provide an improved posterior estimate of the transformed state. This process is known as the unscented transformation (UT) [8]. The process and observation equations are in this case reformulated in the general case of a nonlinear system as,

$$
\begin{align*}
\bar{x}_{k+1} &= f(x_k, u_k) + \bar{w}_k \\
y_k &= h(x_k, u_k) + v_k
\end{align*}
$$

(3)
where \( f(\cdot) \) and \( h(\cdot) \) are generally nonlinear functions and \( \bar{w}_k \) is defined as previously. The main steps of the UKF are summarized in Tab.1.

### Initialize:

1. Set initial values for the augmented state vector mean and covariance: \( \bar{x}_0 \) and \( \bar{P}_0 \)
2. Set the parameters of the UKF (\( n_x \) the size of the augmented state vector)
   \( a = 1, b = 2, \kappa = 0 \)
   \( \lambda = \alpha^2(n_x + \kappa) - n_x, c = \alpha^2(n_x + \kappa) \)
   \( W_m^0 = \frac{\lambda}{(n_x + \lambda)}, W_m^i = \frac{\lambda}{2(n_x + \lambda)}, i = 1, \ldots, 2n_x \)
   \( W_c^0 = \frac{\lambda}{(n_x + \lambda)} + (1 - \alpha^2 + b), W_c^i = W_m^i, i = 1, \ldots, 2n_x \)
   \( \mu_x = [W_m^0 \ldots W_m^{2n_x}]^T \)
   \( M_x = (I - [\mu_x \ldots \mu_x]) \times diag(W_c^0 \ldots W_c^{2n_x})(I - [\mu_x \ldots \mu_x])^T \)

### Update:

1. Calculate Sigma points: \( X_k^- = [\bar{x}_k^- \ldots \bar{x}_k^-] + \sqrt{c} [0 \sqrt{P_k^-} \ldots - \sqrt{P_k^-}] \)
2. Propagate sigma points through the output equation: \( Y_k^- = h(X_k^-, k) \)
3. Calculate output mean and covariance: \( \hat{y}_k = Y_k^- \mu_x, P_{xy}^k = Y_k^- M_x Y_k^-^T + R \)
4. Calculate cross covariance between state and output: \( P_{xy}^k = Z_k^- M_x D_k^-^T \)
5. Calculate state gain: \( P_{xy}^k K_k = P_{xy}^k \)
6. Update state mean and covariance:
   \( \bar{x}_k = \bar{x}_k^- + K_k (y_k - \hat{y}_k) \)

### Predict:

1. Calculate Sigma points: \( X_k = [x_k \ldots x_k] + \sqrt{c} [0 \sqrt{P_k} \ldots - \sqrt{P_k}] \)
2. Propagate sigma points through the state equation: \( \tilde{X}_k = f(X_k, k) \)
3. Predict state mean and covariance for \( k + 1 \):
   \( \tilde{x}_{k+1}^- = \tilde{X}_k \mu_x, P_{k+1}^- = \tilde{X}_k M_x \tilde{X}_k^T + Q \)

Table 1: The general scheme of the UKF algorithm for joint state and parameter estimation
3 NUMERICAL APPLICATION TO VEHICLE-LIKE CASE STUDIES

The UKF outlined in the previous section is now validated and assessed in two numerical case studies. The first pertains to a spring-mass-damper system, while the second corresponds to a complex laboratory model of a vehicle.

3.1 A spring-mass-damper model

Figure 1a displays a spring-mass chain-like mechanical model with 11-DOFs, which comprises two substructures. The lower one contains the first DOF, of mass $m_w$, which is connected to the ground through a linear spring of stiffness $k_w$ and to the first mass ($m_1$) of the upper substructure via a nonlinear spring-damper, in which the associated restoring and damping forces are given by (see also Fig.1b) respectively, with $x = x_1 - x_w$.

![Figure 1a: 11-DOF spring-mass-damper model](image)

![Figure 1b: The nonlinear restoring and damping forces.](image)

Figure 1: First case study.
\[ f_k = k_s x + \mu k_1 \frac{x}{\mu k_2 + |x|} \quad \text{and} \quad f_c = c_s \dot{x} + \mu c_1 \frac{x}{\mu c_2 + |x|} \]  

(4)

The upper substructure includes the remaining 10 DOFs and it is assumed linear and proportionally damped, with 1% modal damping ratio in every mode. Due to the presence of the interconnecting nonlinear spring-damper in the lower substructure, the equations of motion for this example are strongly nonlinear, since the values of the nonlinear terms in Eq.(4) are comparable to the ones of the linear terms. To this end, the simulation of the structure is succeeded using a variant of the Newmark’s method [2]. A random displacement base excitation \( r(t) \) is used as an input to the system and the calculated acceleration responses of a subset of the DOFs (see below), noise-corrupted with 15% measurement error, are considered as the available output measurements. In addition, it is assumed that the values \( \mu k_1 \) and \( \mu c_1 \) of the nonlinear restoring and damping forces of the lower substructure are unknown, along with the structural states, e.g. the DOF displacements and velocities.

Under this setting, the applied computational framework consists of the following steps:
1. The UKF is applied to the lower (nonlinear) substructure, in order to estimate the acceleration at the boundary between the two substructures.
2. The UKF-estimated acceleration time history is subsequently used as a base excitation to the upper substructure, which is simulated separately, in order to obtain the dynamic response of the trailing DOFs.
3. The calculated acceleration time histories of steps 1-2 are compared to their counterparts from the simulation of the total structure.

In order to obtain an insight about the optimal number and location of the “sensors”, several trials are executed revealing that four (4) sensors are capable of rendering the UKF effective. It is nevertheless important that the corresponding locations are distributed along the structure, instead of being concentrated around a specific area. Indicatively, the locations «3-5-7-11» were found to be a good choice.

Figures 2-3 display indicative results of the applied method. Both unknown parameters converge to their actual values (Fig.2), with the one of the restoring force reaching quite rapidly. This efficiency is also reflected to the estimation quality of the states, as for example the displacement of the 4\(^{th}\) DOF (Fig.3), verifying the efficiency of the devised JS&PE scheme.

![Figure 2: Convergence of the nonlinear parameters (a) \( \mu k_1 \) , (b) \( \mu c_1 \).](image)
3.2 Small scale vehicle-like frame structure

The proposed computational framework is now applied to the laboratory structure of Fig.4, that has been designed for small-scale simulations of vehicle frames [3,4]. The frame structure is characterized by predominantly linear response and high modal density and it is supported to four “suspension” systems with strongly nonlinear behavior. They consist of a lower set of linear discrete spring-damper units, connected to concentrated masses that simulate the wheelsets, as well as of an upper set of a nonlinear discrete spring-damper units connected to the frame that simulate the action of the suspension. The measurement points, indicated by 1-4, correspond to connection points between the frame and its supporting structures, while points 5-7 correspond to available measurements along the frame.

Figure 3: 4th DOF’s displacement (a) total timeframe, (b) after 2s.

Figure 4: Small scale vehicle model with frame and nonlinear supports.
Based on the geometric details and the material properties of the structure, a detailed FE model of the vehicle frame is created, consisting of 45,564 DOFs. The nonlinear restoring and damping forces in the suspensions assume the earlier form, and the same computational steps are applied herein as well. That is, a nonlinear transient response analysis of the full vehicle model (frame and supports) is performed first, by applying four different transient displacement base excitation histories to the wheel subsystems in the vertical direction. The model is then solved by using the same direct integration scheme as in the previous example and the acceleration histories at the selected locations 1-7 are calculated. Again, the values $\mu_{k1}$ and $\mu_{c1}$ of the nonlinear restoring and damping forces of the suspension subsystem which connected to the boundary location 2 are considered as unknown, along with the displacements and velocities.

The results are expanded over Figs.5-6 and confirm the efficiency of the proposed UKF scheme. Both unknown nonlinear parameters converge rapidly to their true values, while the estimated displacements at the measurement locations 5, 6, 7 and 8 result almost identical to their counterparts, which have been calculated from the simulation of the full structure.

![Figure 5: Convergence of the nonlinear parameters of suspension subsystem 2 (a) $\mu_{k1}$, (b) $\mu_{c1}$.](image)

![Figure 6: UKF state estimation of displacement time histories in measurement locations 5, 6, 7 and 8.](image)
4 SUMMARY

This work outlines a computational framework for the effective finite element (FE) model updating of large-scale nonlinear uncertain structures. The proposed approach combines a substructuring model reduction stage with a near real-time system identification scheme, the UKF. The nonlinear observer is employed herein for the JS&PE of the nonlinear substructure, which is successfully handled using acceleration measurements in a limited number of DOFs. The application of the proposed scheme to the selected numerical case studies verifies its efficacy, encouraging the further investigation of its performance in more complicated structures.

5 REFERENCES


RELIABILITY PREDICTION OF FATIGUE DAMAGE ACCUMULATION ON WIND TURBINE SUPPORT STRUCTURES

K. Tatsis\(^1\), E. Chatzi\(^1\) and E. Lourens\(^2\)

\(^1\)ETH Zürich, Institute of Structural Engineering
Stefano-Franscini-Platz 5, 8093 Zürich, Switzerland
e-mail: tatsis@ibk.baug.ethz.ch, chatzi@ibk.baug.ethz.ch

\(^2\)Delft University of Technology, Faculty of Civil Engineering and Geosciences
Stevinweg 1, 2628 CN Delft, The Netherlands
e-mail: e.lourens@tudelft.nl

Keywords: Wind turbine, Structural monitoring, Input-state estimation, Response identification, Fatigue damage, Reliability.

Abstract. The evaluation of fatigue damage accumulation on wind turbine support structures under operational conditions is heavily influenced by a number of uncertainties. These uncertainties may, firstly, be attributed to the highly variable and complex environmental loads, and secondly, to the unavoidable modelling errors which mainly originate from the inherent randomness in both material properties and fatigue resistance of structural components. It is therefore essential that assessment of fatigue life is carried out within a probabilistic framework; one that accounts for the stochastic nature of the phenomenon. The present study proposes a strategy for real-time reliability prediction of accumulated fatigue damage on wind turbine support structures by taking into account the above-mentioned uncertainties. To this end, the availability of structural monitoring information for the identification of the global response on wind-turbine support structures is exploited in order to address the discrepancies between actual and predicted damage accumulation. This is carried through utilization of an augmented version of the Kalman filter, which is capable of jointly estimating the response and the unknown inputs of the structure while relying on a limited number of noisy observations and a presumably uncertain model of the real system. A fixed-lag smoother is further deployed for the attenuation of the estimation error in an on-line mode and the smoothed stochastic estimates of the response are propagated over the model at the level of stresses. The accumulated damage along with the corresponding reliability level is finally predicted using a stochastic nonstationary fatigue damage model. The proposed scheme is demonstrated via implementation on the NREL 5.0 MW wind turbine under different operational conditions, on the basis of dummy vibration data generated via the FAST software.
1 INTRODUCTION

Fatigue is regarded as a critical and highly-uncertain factor for wind turbine structures, where it is essential to ensure a certain life span under irregular and constantly varying operational and environmental conditions. Conventionally, fatigue life predictions are conducted on the basis of numerical simulations in conjunction with the information provided from historical metocean data. In the wake of recent advancements in Structural Health Monitoring technologies and methodologies, significant attention has been redirected to vibration-based approaches for fatigue estimation [1], particularly to what pertains to response prediction under unknown inputs.

Although fatigue has been vastly and exhaustively investigated under different perspectives [2], not many studies have been relied on a probabilistic framework. In recent years, an attempt has been made to approach damage accumulation due to fatigue as a stochastic process [3]. To account for the randomness in both the loading process and fatigue resistance of materials, Shen et al. [4] established a probabilistic model of fatigue damage based on the distribution of stress amplitudes. Despite the efficiency in obtaining the distribution of fatigue damage, the approach in [4] is not appropriate for real-time applications. On the contrary, Rathod et al. [5] proposed a more universal methodology for the stochastic modelling of damage accumulation under multilevel loading, which may be easily tailored to an online framework. In a more recent work and in the field of wind turbine structures, Thöns et al. [6] conducted a sensitivity study for fatigue limit state on the basis on the Spearman rank coefficient. The outcome of the study in [6] provides a useful insight into the most influential sources of uncertainty in fatigue damage. Within the same context, Thöns et al. [7] conducted a quantitative study for the value of structural health monitoring towards the integrity management of fatigue deteriorating structural systems.

In what concerns the real-time state estimation of systems with unknown inputs, a variety of methods is currently available. The primary step was taken by Kitanidis [8] through the development of a linear state estimator in the presence of unknown inputs. Under this perspective, Gillijns et al. [9] proposed an unbiased minimum-variance filter for the joint input-state estimation of linear time invariant systems. Although optimal in terms of second order statistics, the estimator in [9] was proven to be susceptible to numerical issues when the number of measurements is larger than the order of the system. This was recently alleviated by Lourens et al. [10] with an extension of the previously mentioned algorithm. An alternative for the dual state and input estimation of structural systems was implemented also by Lourens et al. [11], using an augmented version of the standard Kalman filter. To improve the poor performance of the latter when acceleration-only measurements are employed, Naets et al. [12] proposed a stabilized version of the augmented Kalman filter by using dummy displacement measurements. Finally, in more recent years Azam et al. [13] suggested a dual implementation of the Kalman filter in order to resolve the numerical issues that arise in the augmented formulation of input-state estimation problem.

This study presents a vibration-based probabilistic framework for fatigue assessment on wind turbine structures accounting for uncertainties in the level of the structural model, the estimated stresses time histories and the employed fatigue model. The paper is organized as follows: in Section 2 the mathematical formulation of the response estimation problem for systems with unknown inputs is established. This is carried out using the augmented form of the Kalman filter which is further enhanced with a fixed-lag smoother for the attenuation of the underlying model uncertainty and the improvement of state estimates. In Section 3, the stochastic framework for fatigue assessment is presented. This comprises a non-stationary fatigue model capable of...
capturing the expected value and variance of fatigue damage as well as their evolution in time as a function of the loading history. Finally, the implementation of the proposed approach is illustrated in Section 4 by means of an application to a wind turbine support structure on the basis of simulated measurement data.

2 RESPONSE ESTIMATION

The starting point for the response estimation on wind turbine structures is the continuous-time system of equations of motion

\[ M \ddot{u}(t) + C \dot{u}(t) + Ku(t) = S_p f(t) \]  

where \( u(t) \in \mathbb{R}^n \) denotes the displacement vector, \( M, C, K \in \mathbb{R}^{n \times n} \) are the mass, damping and stiffness matrices, \( f(t) \in \mathbb{R}^{n_p} \) is the force vector, with \( n_p \) designating the number of input forces, and \( S_p \in \mathbb{R}^{n \times n_p} \) is the corresponding selection matrix.

Upon introduction of the state vector \( x(t) = \text{vec}([\dot{u}(t), u(t)]) \in \mathbb{R}^{2n} \), Eq. (1) may be transformed into the state equation and additionally fused with a measurement process in order to form the deterministic state-space model in the continuous-time domain. The latter can be further transferred, through temporal discretization, to the discrete-time domain, which yields the following stochastic state and observation equations

\[ x_{k+1} = Ax_k + Bp_k + w_k \]  
\[ y_k = Gx_k + Jp_k + v_k \]

where it should be noted that Eqs. (2) and (3) are additionally supplemented with the zero-mean white noise processes \( w_k \in \mathbb{R}^{2n} \) and \( v_k \in \mathbb{R}^{n_y} \) that represent the system and measurement noise of covariance matrices \( Q \in \mathbb{R}^{2n \times 2n} \) and \( R \in \mathbb{R}^{n_y \times n_y} \), respectively.

In the absence of knowledge with respect to the driving forces of the system, the state-space model described in Eqs. (2) and (3) may be written in the augmented form. This is accomplished by supplementing the initial state vector with the input force vector in order to construct an augmented state vector as follows

\[ z_k = \begin{bmatrix} x_k \\ p_k \end{bmatrix} \]

To derive the system matrices of the augmented state-space, it is postulated that the evolution of the input dynamics may be captured by a random-walk process

\[ p_{k+1} = p_k + \eta_k \]

upon proper tuning of the random variable \( \eta_k \in \mathbb{R}^{n_p} \) which represents a zero-mean white Gaussian process with covariance matrix \( S \in \mathbb{R}^{n_p \times n_p} \). The augmented state and measurement equations may then be formulated as follows

\[ z_{k+1} = A_az_k + \zeta_k \]  
\[ y_k = G_az_k + v_k \]
where the subscript $a$ denotes the augmented state-space matrices,

$$A_a = \begin{bmatrix} A & B \\ 0 & I \end{bmatrix}, \quad G_a = \begin{bmatrix} G \\ J \end{bmatrix},$$

and $\zeta_k = \text{vec}(\begin{bmatrix} w_k \\ \eta_k \end{bmatrix}) \in \mathbb{R}^{2n+n_p}$ is the augmented noise vector with covariance matrix $Q_a \in \mathbb{R}^{2n+n_p \times 2n+n_p}$. Thereafter, both input and state estimation may be accomplished recursively through the standard Kalman filter operating on the augmented state-space formulation in two steps

**Time update**

$$\hat{z}_{k+1|k} = A_a \hat{z}_{k|k}$$

$$P_{k+1|k} = A_a P_{k|k} A_a^T + Q_a$$

**Measurement update**

$$K_k = P_{k|k-1} G_a^T (G_a P_{k|k-1} G_a^T + R)^{-1}$$

$$\hat{z}_{k|k} = \hat{z}_{k|k-1} + K_k (y_k - G_a \hat{z}_{k|k-1})$$

$$P_{k|k} = P_{k|k-1} - K_k G_a P_{k|k-1}$$

2.1 Smoothing and backward sampling

In the previous section, it was demonstrated how optimal *a priori* and *a posteriori* state and input estimates may be obtained on the basis of a limited number of response measurements. Due to the probabilistic origin of the Kalman filter, these state estimates are delivered in terms of the first two moments of their probability distribution, which essentially reflects the uncertainty associated with both the system dynamics and the measurement quality. An effective way of attenuating this uncertainty and improving the performance of the estimator is to further condition the forward results on posterior measurements via the so-called smoothing process.

Among the three basic classes of smoothers, that is fixed-point, fixed interval and fixed-lag smoothers, the latter is considered as the optimal estimator since, apart from allowing for processing delay, it is also tailored for on-line operation [14]. The aim of such a smoother is to provide the optimal state estimate at time $k - N$ conditioned on measurements up to and including time instant $k$. It is therefore required that $N$ future measurements are available for the smoothed state estimation at time instant $k - N$.

Upon obtaining the state estimates at time $k$ using the augmented Kalman filter, the fixed interval smoother is running backwards from $k$ up to $k - N$ in order to provide the smoothed estimate of each state with delays between 0 and $N$. This process is summarized in four steps performed for $i = 2, \ldots, N + 1$ at each time instant $k$ as follows

**Fixed-lag smoother**

$$L_{k,i} = P_{k}^{0,i-1} G^T (G P_{k}^{0,0} G^T + R_k)^{-1}$$

$$\tilde{z}_{k+1-i|k} = \tilde{z}_{k+2-i|k} + L_{k,i} (y_k - G \tilde{z}_{k|k-1})$$

$$P_{k+1}^{0,i} = P_{k}^{0,i-1} (A - L_{k,0} G)^T$$

$$P_{k+1}^{i,i} = P_{k+1}^{i-1,i-1} - P_{k}^{0,i-1} G^T L_{k,i} A^T$$

79
It should be noted that for \( i = 1 \) the matrix \( L_{k,i} \) is equal to the Kalman gain and the above process represents the measurement update of the standard Kalman filter. Accordingly, the covariance matrix \( P_{k,0} \) designates the \textit{a priori} estimate error covariance while in the general case the covariance matrices \( P_{i,j}^k \) are defined as

\[
P_{i,j}^k = E \left[ (z_{k-j} - \hat{z}_{k-j,k-1}) (z_{k-i} - \hat{z}_{k-i,k-1})^T \right] \tag{11}
\]

Once the smoothed state and input estimates at time instant \( k - N \) are obtained, it may easily be deduced from Eqs. (10) that the final state estimate is delivered as a Gaussian distribution with mean and variance

\[
E [z_{k-N}|y_1, \ldots, y_k] = \hat{z}_{k-N+1|k} + L_{k,N+1} (y_k - G \hat{z}_{k|k-1}) \tag{12}
\]

\[
E \left[ (z_{k-N} - \hat{z}_{k-N,k}) (z_{k-N} - \hat{z}_{k-N,k})^T \right] = P_{k,k}^{N,N} - P_{k}^{0,N} G^T L_{k,N}^T A^T \tag{13}
\]

where every quantity on the right-hand side is obtained through the last step of smoothing process. Within the context of fatigue assessment, this uncertainty in state estimates may be propagated over the model at the level of stresses and subsequently quantified through the backward sampling of the stress time histories in order to account for the modelling and measurement uncertainties.

### 3 FATIGUE DAMAGE

Fatigue damage accumulation is a stochastic process characterized by several uncertainties. These uncertainties are associated with a number of sources, such as material properties, modeling errors and loading conditions among others. Under this perspective, it is essential that damage accumulation is treated within a probabilistic framework, one that is able to account for the stochastic nature of the phenomenon. In so doing, the present study is based on the premise of non-stationary and normally distributed damage accumulation process

\[
D(t) \sim N (\mu_D(t), \sigma_D^2(t)) \tag{14}
\]

with \( \mu_D(t) \) and \( \sigma_D^2(t) \) denoting the time-varying mean and variance respectively. The elaboration of this assumption is presented in three steps through the following sections.

#### 3.1 Model of damage accumulation

A common practice for the evaluation of fatigue damage in steel structures is the linear accumulation rule, also known as Palmgen-Miner rule [15, 16], whereby damage at a given stress level is defined as the ratio of operational cycles to the number of failure cycles. For varying stress level, this is formulated as follows

\[
D = \sum_{j=1}^{k} D_j = \sum_{j=1}^{k} \frac{n(\Delta \sigma_j)}{N_f(\Delta \sigma_j)} \tag{15}
\]

where \( n(\Delta \sigma_j) \) is the number of cycles with stress amplitude \( \Delta \sigma_j \), \( N_f(\Delta \sigma_j) \) denotes the number of cycles to failure at stress level \( \Delta \sigma_j \) and \( k \) is the number of stress ranges contained in the examined time history. In the case of directly measured or estimated strain/stress time histories,
the number of cycles at stress amplitude $\Delta \sigma_j$ may be determined using counting techniques. Among others, the rainflow counting algorithm, which is thoroughly described in [17] and employed in this study as well, constitutes the most accurate and commonly used method in fatigue analysis. Finally, the relationship between fatigue life in terms of cycles $N_f$ and stress range $\Delta \sigma$ is obtained by the well-known $S - N$ curve [18] which is expressed by

$$N_f \Delta \sigma^m = A$$

(16)

where $A$ represents a fatigue strength constant and $m$ denotes the slope of the curve, with both variables being material dependent. Combining Eq. (15) and the $S - N$ curve model described by Eq. (16), the expression of damage accumulation for multi-stress levels may be written as

$$D = \sum_{j=1}^{k} C (\Delta \sigma_j)^m n (\Delta \sigma_j)$$

(17)

with $C$ denoting the reciprocal of fatigue strength constant $A$.

### 3.2 Distribution of fatigue damage

In establishing a probabilistic representation of damage accumulation, fatigue life is considered to be described by a probabilistic $S - N$ curve, as illustrated in Fig. 1. In practice, this is accomplished by treating the failure cycles $N_f$ of each stress level as a random variable that follows a certain distribution. Since fatigue life of components under constant or random amplitude stress conditions may be adequately represented via normal or log-normal distributions, as underlined by Wu et al. [19], it is assumed that the number of cycles to failure at a certain stress level is normally distributed

$$f_n (N_f) = \frac{1}{\sigma_{N_f} \sqrt{2\pi}} \exp \left( -\frac{1}{2} \left( \frac{N_f - \mu_{N_f}}{\sigma_{N_f}} \right)^2 \right)$$

(18)

with mean $\mu_{N_f}$ and standard deviation $\sigma_{N_f}$. The corresponding distribution of damage accumulation at the same stress level may then be derived through the one-to-one transformation of Eq. (18). The latter requires the functional relationship $N_f = h(D)$ between failure life $N_f$ and damage accumulation $D$, which may be obtained through evaluation of Eq. (17) at failure.
life for a single stress level. Thereafter, the probability density function of cumulative damage $D$ may be calculated through the differentiation of its cumulative distribution as shown below

\[ f_D(D) = \frac{d}{dD} F_D(D) \]

\[ = \frac{D}{dD} F_n \left( h^{-1}(D) \right) \]

\[ = \frac{D}{dD} \left( \int_{-\infty}^{h^{-1}(D)} f_n(N_f) dN_f \right) \]

\[ = \frac{dh^{-1}(D)}{dD} f_n \left( h^{-1}(D) \right) \]

\[ = \frac{dN_f}{dD} f_n \left( N_f \right) \quad (19) \]

where $dN_f/dD = 1/s$ with $s = C (\Delta \sigma)^m$ denoting the slope of the $S - N$ curve at the stress level $\Delta \sigma$. Substituting in Eq. (19) and making use of Eq. (17) for a single stress level and Eq. (18), the probability density function of $D$ may be written as follows

\[ f_D(D) = \frac{1}{s \sigma_{N_f} \sqrt{2\pi}} \exp \left( -\frac{1}{2} \left( \frac{D - s \mu_{N_f}}{s \sigma_{N_f}} \right)^2 \right) \quad (20) \]

where it may be observed that fatigue damage follows a normal distribution as well

\[ D \left( N_f \right) \sim N \left( s \mu_{N_f}, s \sigma_{N_f} \right) \quad (21) \]

with mean $s \mu_{N_f}$ and standard deviation $s \sigma_{N_f}$.

### 3.3 Evolution of variance

As illustrated in Fig. 1, fatigue life follows an increasing trend in variability as stress level decreases, thus resulting in low variability of failure cycles at high stress conditions and higher variability at low stress levels. Additionally, as demonstrated by Wang et al. [20], damage accumulation at constant stress level exhibits a monotonically increasing variability as usage cycles increase. This second source of variability is graphically depicted in Fig. 2, where it

![Figure 2: Evolution of variance for different stresses](image-url)
may be inferred that the initial stage is represented by zero usage cycles and zero variability for all stress levels. Subsequently, as the number of usage cycles at a certain stress level increases, the variability is accordingly increasing until it reaches a certain value $\sigma_{N_f}$ at failure life. This trend may be geometrically interpreted, as elaborated by Rathod et al. [5], and captured through the rate of change of the standard deviation $r_\sigma$, given by

$$r_\sigma = \frac{\sigma_{N_f}}{N_f}$$  \hspace{1cm} (22)

Subsequently, the standard deviation of the loading cycles $n$ may be derived from the product of the rate of change and the number of cycles $n$, given by the following expression

$$\sigma_n = r_\sigma n = \left(\frac{\sigma_{N_f}}{N_f}\right) n$$  \hspace{1cm} (23)

Considering finally that the damage index is related to the number of cycles via the slope $s = C(\Delta \sigma)^m$ of the $S - N$ curve, the standard deviation of cumulative damage may be written as follows

$$\sigma_D = \left(\frac{\sigma_{N_f}}{N_f}\right) ns = C(\Delta \sigma)^m n \left(\frac{\sigma_{N_f}}{N_f}\right)$$  \hspace{1cm} (24)

Although the above formula represents the variability in damage due to single stress-level condition, it can be readily extended to account for the variability in damage accumulation under multi-level stress conditions, as graphically presented in Fig. 3. In so doing, it is assumed that damage accumulation at each stress level bears an independent stochastic contribution to the total variability of damage accumulation which is given by

$$\sigma_D = \sqrt{\sum_{j=1}^{k} C(\Delta \sigma_j)^m n_j \left(\frac{\sigma_{N_f_j}}{N_f_j}\right)^2}$$  \hspace{1cm} (25)

where $N_{f_j}$ denotes the cycles to failure at stress level $\Delta \sigma_j$ and $\sigma_{N_{f_j}}$ the corresponding standard deviation.

![Figure 3: Damage accumulation for multi-stress loading](image-url)
3.4 Reliability prediction

Based on the above probabilistic model for fatigue damage accumulation, which relies on the assumption of normally distributed usage and failure cycles, the reliability of a structural component with respect to fatigue damage may be calculated through the limit state function

\[ Z(n) = D_c - D(n) \]  

(26)

where \( D_c \) is the critical damage with \( \mathbb{E}[D_c] = 1 \) and \( Z(n) = 0 \) describes the limit state that separates the safe domain, for which \( Z(n) > 0 \), from the failure region where \( Z(n) < 0 \). Thereafter, the reliability is calculated on the basis of the limit state function as follows

\[
R = P(Z(n) > 0)
= 1 - P(Z(n) \leq 0)
= 1 - \Phi \left( - \frac{\mu_{D_c} - \mu_D}{\sqrt{\sigma_{D_c}^2 + \sigma_D^2}} \right)
\]  

(27)

with \( \mu_D \) and \( \sigma_D \) denoting the mean and the standard deviation of damage accumulation while \( \mu_{D_c}, \sigma_{D_c} \) designate the mean the standard deviation of threshold damage.

Thereafter, substitution of Eqs. (17) and (25) into Eq. (27) yields the reliability index of a structural component under multi-level stress conditions as given below

\[
R = 1 - \Phi \left( - \frac{\mu_{D_c} - \sum_{j=1}^{k} C(\Delta \sigma_j)^m n(\Delta \sigma_j)}{\sqrt{\sigma_{D_c}^2 + \sum_{j=1}^{k} C(\Delta \sigma_j)^m n(\Delta \sigma_j) \left( \sigma_{N_{f_j}}/N_{f_j} \right)^2}} \right)
\]  

(28)

where it should be noted that the variability of threshold damage \( \sigma_{D_c} \) at failure life, i.e. when \( n = N_{f} \), is equal to the variability of damage accumulation \( \sigma_D \), given by Eq. (25). This is also implied from Fig. 2 where it is shown that the variability of damage follows a monotonically increasing trend until it reaches the value of threshold damage at failure level.

4 RESULTS

The demonstration of the proposed scheme for reliability prediction of fatigue damage accumulation is carried out through an application to the NREL 5.0 MW land-based wind turbine, whose features are in detail presented by Jonkman et al. [21]. To investigate the fatigue limit state (FLS), the considered structure is modelled using the FAST v8 software platform for the generation of artificial measurements from aero-servo-elastic simulations in operational conditions. A total of two hundred simulations, each with a duration of ten minutes, is used for this purpose. Each simulation is performed at a mean wind speed sampled from a Weibull distribution with mean equal to 10 m/s and a turbulence intensity derived from a log-normal distribution conditioned on the previously sampled mean wind speed.

To enable the implementation of the proposed approach in a long-term period, the structural response is assumed to be measured with sensors appropriate for continuous and permanent
monitoring. In this study, availability of acceleration and inclination measurements is assumed, whose adopted configuration is depicted in Figs. 4(a) and 4(b). Their number and location on the structure are determined by the stability and observability conditions, in accordance with the number of inputs to be identified and their corresponding locations. These inputs are considered to be the interface forces between the tower top and the nacelle, as illustrated in Fig. 4(c).

To additionally account for modelling errors, a stochastic Finite Element (FE) model of the wind turbine structure is employed for the solution of the inverse problem. Namely, while the forward process for the data generation is carried out with a perfectly known deterministic model, the one implemented in FAST, the identification part is performed with a perturbed FE model. This is a refined shell-element model, reduced with a component mode synthesis technique, that involves a randomness in the material properties, introduced at the level of the constitutive matrix $C$ as follows

$$C = C_0(\alpha + f(x))$$  \hspace{1cm} (29)$$

where $\alpha$, equal to 0.95, is a perturbation factor for the mean value of $C$ which is denoted by $C_0$ and $f(x)$ is a zero-mean stochastic process which is herein assumed to be log-normally distributed with standard deviation $2 \cdot 10^{10}$ N/m². Within this context, the stochastic stiffness matrix of an element $(e)$ may be expressed on the basis of the principle of virtual work as

$$k^{(e)} = k_0^{(e)} + \Delta k^{(e)} = \int_{V^{(e)}} B^{(e)} C_0 B^{(e)} dV^{(e)} + \int_{V^{(e)}} B^{(e)} C_0 f^{(e)}(x) B^{(e)} dV^{(e)}$$  \hspace{1cm} (30)$$

where $k_0^{(e)}$ is the perturbed deterministic part and $\Delta k^{(e)}$ denotes the fluctuating part. For the sake of completeness, it should be noted that $V^{(e)}$ is the volume of element $(e)$ and $B^{(e)}$ represents the deformation matrix.

The simulation measurements obtained from FAST are polluted with 3% white Gaussian noise and subsequently fed to the augmented Kalman filter which is operating along with the fixed-lag smoother for the dual input and state estimation of the perturbed model. The algorithm is initialized with a zero state vector, the measurement error covariance is obtained as a function of the noise level and the system error covariance is calculated on the basis of the L-curve, as shown in Fig. 5. The smoothed state estimates, which are delivered as Gaussian random
variables with mean and variance given by Eqs. (12) and (13), are finally propagated over the model in order to obtain the corresponding distributions of stress estimates at critical locations.

The expected value of stress time histories for locations 1 and 2 are illustrated in Fig. 6 and 7 respectively, along with their 95% confidence intervals for two different time frames of 100s and 5s. Although an erroneous model is used for the response identification, it is seen that the stress time histories are identified with a sufficiently high degree of accuracy. This may be primarily attributed to the capability of the filter to provide robust response estimates, when an erroneous model is deployed, at the expense of the input predictions and secondly to the smoothing effect which yields a substantial improvement on the peak estimates.

To quantify the uncertainty of the estimated stress time histories, which implicitly comprises all kinds of uncertainty associated with the FE model and the response measurements, the time histories are backwards sampled in a Monte Carlo framework and subsequently counted with the rainflow algorithm, yielding thus a stochastic representation of the stress cycle distribution. The latter is then processed with the fatigue model of Section 3 and produces the time-dependent distribution of damage accumulation. It should be noted that in the context of this study, the fatigue strength parameters are chosen such that a substantial part of the fatigue life of the structure is wasted throughout the simulation time.

The expected value of accumulated fatigue damage with respect to the number of usage cycles for the whole simulation time of two thousand minutes at locations 1 and 2 is depicted.
in Fig. 8. As observed, the evolution of damage is identified with a high degree of accuracy, yet there is a small drift on the estimates which mainly stems from spurious cycles due to noise on measurements. The total damage accumulation for all five locations is presented in the bar-chart of Fig. 9, showing the sufficiently good agreement between actual and estimated values, with a 7% maximum error.

Finally, the time-dependent reliability index of a structural component under the estimated stress conditions is evaluated on the basis of Eq. (28). Figure 10 shows the evolution of reliability with respect to the number of usage cycles for locations 1 and 2, where it is evident that reliability follows a decreasing trend as the number of usage cycles increases. As underlined in Section 3, the rate of change of the reliability index is dependent on the stress conditions, resulting to higher rate of loss for high-low cycles and accordingly to lower one for low stress levels. An additional information revealed from Fig. 10 pertains to the evolution of fatigue mechanisms. Namely, it may be deduced that the initial and highly reliable period represents the phase of crack initiation while the crack propagation period is indicated by loss of reliability. This implies that high stress-level conditions result in small crack initiation periods and faster loss of reliability during the crack propagation phase, which in turn implies faster damage accumulation.

![Figure 8: Expected value of fatigue damage accumulation on locations 1 and 2](image)

![Figure 9: Expected value of accumulated fatigue damage on locations 1-5](image)

![Figure 10: Evolution of reliability index at locations 1 and 2](image)
5 CONCLUSIONS

This study presents a probabilistic framework for real-time reliability estimation of fatigue damage accumulation on wind turbine support structures. The approach is based on output-only measurements from wind turbine support structures and a combination of the augmented version of the Kalman filter with a fixed-lag smoother. The smoothed uncertainty of the estimated states is transferred through the FE model of the wind turbine substructure at the level of stresses and subsequently propagated over a stochastic non-stationary fatigue model for the identification of damage accumulation. The latter is finally obtained at unmeasured critical locations along with a time-dependent reliability index. Despite the fact that a strongly perturbed numerical model is implemented for the inverse problem, the structural quantities of interest are estimated with sufficiently high accuracy. This is mostly due to the capability of the algorithm to provide robust response predictions at the expense of the input estimates and partially due to the smoothing effect. However, it is seen that even a small amount of noise can give rise to spurious stress cycles and lead to an error accumulation on fatigue damage estimates.

ACKNOWLEDGEMENTS

The authors would like to gratefully acknowledge the support of the European Research Council via the ERC Starting Grant WINDMIL (ERC-2015-StG #679843) on the topic of Smart Monitoring, Inspection and Life-Cycle Assessment of Wind Turbines.

REFERENCES


EFFICIENT RELIABILITY AND UNCERTAINTY ASSESSMENT ON LIFELINE NETWORKS USING THE SURVIVAL SIGNATURE

Geng Feng\textsuperscript{1}, Sean Reed\textsuperscript{2}, Edoardo Patelli\textsuperscript{1}, Michael Beer\textsuperscript{3} and Frank P.A. Coolen\textsuperscript{4}

\textsuperscript{1}Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK  
  e-mail: \{fenggeng, epatelli\}@liverpool.ac.uk

\textsuperscript{2}Resilience Engineering Research Group, University of Nottingham, Nottingham, UK  
  e-mail: sean.reed@nottingham.ac.uk

\textsuperscript{3}Institute for Risk and Reliability, Leibniz University Hannover, Hannover, Germany  
  Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK  
  School of Civil Engineering & Shanghai Institute of Disaster Prevention and Relief, Tongji University, Shanghai, China  
  e-mail: beer@irz.uni-hannover.de

\textsuperscript{4}Department of Mathematical Sciences, Durham University, Durham, UK  
  e-mail: frank.coolen@durham.ac.uk

\textbf{Keywords:} Reliability Analysis, Uncertainty Assessment, Imprecision, Lifeline Networks, Survival Signature, Binary Decision Diagrams.

\textbf{Abstract.} Lifeline networks, such as water distribution and transportation networks, are the backbone of our societies, and the study of their reliability of them is required. In this paper, a survival signature-based reliability analysis method is proposed to analyse the complex networks. It allows to consider all the characters of the network instead of just analysing the most critical path. What is more, the survival signature separates the system structure from its failure distributions, and it only needs to be calculated once, which makes it efficient to analyse complex networks. However, due to lack of data, there often exists imprecision within the network failure time distribution parameters and hence the survival signature. An efficient algorithm which bases on the reduced ordered binary decision diagrams (BDD) data structure for the computation of survival signatures is presented. Numerical example shows the applicability of the approaches.
1 INTRODUCTION

Nowadays reliability engineering is used in a wide range of applications on complex lifeline networks, which are a series of components interconnected by communication paths. The analysis of these networks becomes more and more important as they are the backbone of our societies. Examples include the Internet, social networks of individuals or businesses, transportation network, power plant system, aircraft and space flights, metabolic networks, and many others. Since the breakdown of lifeline networks might have catastrophic effects, it is essential to assess the reliability and availability of such networks.

System signatures [1] have been recognized as an important tool to quantify the reliability of systems, however, the use of the system signature is associated with the assumption that all components in the system are of the same type. Survival signature which does not rely any more on the restriction to one component type has been proposed by Coolen and Coolen-Maturi in [2]. Recent developments have opened up a pathway to perform a survival analysis using the concept of survival signature even for complex lifeline networks. Aslett [3] developed a Reliability Theory R package which was used to calculate the survival signature. Feng et al. [4] considered imprecise system reliability and component importance measures. Patelli and Feng [5] proposed efficient simulation approaches based on survival signature for reliability analysis of large systems. Coolen and Coolen-Maturi [6] linked the (imprecise) probabilistic structure function to the survival signature. An algorithm for exact computation of system and survival signatures is proposed by Reed [7].

Most existing models assume that there are precise parameter values available, so the quantification of uncertainty is mostly done by the use of precise probabilities [8]. However, due to lack of perfect knowledge, imprecision often exists within component failure times or their distribution parameters. Hence, the reliability analysis for the lifeline network is affected by the imprecision and uncertainty.

Augustin et al. gave a detailed introduction of imprecise probability in [9]. In order to deal with the uncertainty, Beer et al. [10] introduced fuzzy set theory in engineering analyses. An integrated framework to deal with scarce data, aleatory and epistemic uncertainties is presented by Patelli et al. [11], and OpenCossan is an efficient tool to perform uncertainty management of large finite element models [12]. Also, the use of probability box in risk analysis offers many significant advantages over a traditional probabilistic approaches because it provides convenient and comprehensive ways to handle several of the most practical serious problems face by analysts [13].

In this paper, survival signature is used perform reliability and uncertainty assessment on complex networks. The remainder of the paper is organized as follows. Section 2 gives a brief overview of survival signature-based reliability analysis on lifeline networks with imprecision. Then, an efficient algorithm which bases on the reduced ordered binary decision diagrams (BDD) data structure for calculating the survival signature is proposed in Section 3. In Section 4 a numerical example is analysed to show the performance and applicability of the proposed methods. Finally, the paper is concluded in Section 5 with some discussions.

2 RELIABILITY ANALYSIS ON LIFELINE NETWORKS WITH IMPRECISION

2.1 Reliability Assessment on Lifeline Networks

Lifeline network may be represented with $m$ components which belong to $K \geq 2$ component types, with $m_k$ components of type $k \in \{1, 2, \ldots, K\}$ and $\sum_{k=1}^{K} m_k = m$. Assume that the random failure times of components of the same type are exchangeable, while full independence
is assumed for components belong to different types (iid), the survival signature which can be denoted by $\Phi(l_1, l_2, ..., l_K)$, with $l_k = 0, 1, ..., m_k$ for $k = 1, 2, ..., K$. It defines the probability that the system functions given that $l_k$ of its $m_k$ components of type $k$ work, for each $k \in \{1, 2, ..., K\}$. There are $\binom{m_k}{l_k}$ state vectors $x^k$ with $\sum_{i=1}^{m_k} x^k_i = l_k$ ($k = 1, 2, ..., K$), where $x^k = (x^k_1, x^k_2, ..., x^k_{m_k})$. Let $S_{l_1, l_2, ..., l_K}$ denote the set of all state vectors for the whole system, and it can be known that all the state vectors $x^k \in S_{l_k}^k$ are equally likely to occur. Therefore, the survival signature can be expressed as:

$$
\Phi(l_1, ..., l_K) = \left[ K \prod_{k=1}^{K} \binom{m_k}{l_k}^{-1} \right] \times \sum_{\vec{x} \in S_{l_1, ..., l_K}} \phi(\vec{x})
$$

where $\phi = \phi(\vec{x}) : \{0, 1\}^m \rightarrow \{0, 1\}$ is the system structure function, i.e., the system status based on all possible state vectors $\vec{x}$. $\phi$ is 1 if the system functions for state vector $\vec{x}$ and 0 if not.

Let $C_k(t) \in \{0, 1, ..., m_k\}$ denote the number of $k$ components working at time $t$, the survival function of the system with $K$ types of components becomes:

$$
P(T_s > t) = \sum_{l_1=0}^{m_1} \sum_{l_2=0}^{m_2} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, ..., l_K) P(\bigcap_{k=1}^{K} \{C_k(t) = l_k\})
$$

If the components of type $k$ have a known cumulative distribution function $F_k(t)$, let make the assumptions of independence of failure times for components of different types and of iid given a distribution function for components of the same type, then:

$$
P(\bigcap_{k=1}^{K} \{C_k(t) = l_k\}) = \prod_{k=1}^{K} P(C_k(t) = l_k) = \prod_{k=1}^{K} \binom{m_k}{l_k} [F_k(t)]^{m_k-l_k} [1 - F_k(t)]^{l_k}
$$

Equation 2 shows that the structure of the system is separated from the its components failure times, which is the typical advantage of the survival signature. The survival signature is a summary of structure functions and only needs to be calculated once for the same system. As a result, it is an efficient method to perform system reliability analysis on complex systems with multiple component types.

2.2 Uncertainty Assessment on Lifeline Networks

Uncertainty is an unavoidable component affecting the behaviour of systems and more so with respect to their limits of operation. In spite of how much effort is dedicated into improving the understanding of systems, components and processes through the collection of representative data, the appropriate characterization, representation, propagation and interpretation of uncertainty will remain a fundamental element of the reliability analysis of any complex systems 14. If only few data points are available it might be difficult to identify the parameters of the components precisely. However, it is essential to take the uncertainty into account when analyse the lifeline network reliability.

Considering the imprecision in the component parameters will lead to bounds of the Lifeline network survival function. The lower bound of the survival function can be got through Equation 4.
\[
P(T_S > t) = \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} \bar{D}(C_k(t) = l_k)
\]

{\text{(4)}}

\[
= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} (\bar{P}(C_k(t) \leq l_k) - \bar{P}(C_k(t) \leq l_k - 1))
\]

While the corresponding upper bound of the survival function can be calculated as:

\[
\bar{P}(T_S > t) = \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} D(C_k(t) = l_k)
\]

{\text{(5)}}

\[
= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} (P(C_k(t) \leq l_k) - P(C_k(t) \leq l_k - 1))
\]

3 PROPOSED ALGORITHM FOR CALCULATING SURVIVAL SIGNATURE OF LIFE-LINE NETWORKS

The state vector count or survival signature values for a system, such as a lifeline network, can be represented by a multidimensional array. The values stored at index \((l_1, \ldots, l_K)\) of the array store the value corresponding to \(l_1, \ldots, l_K\) components of types 1 to \(K\) surviving. However, computing these arrays using enumerative methods becomes quickly infeasible since the number of state vectors to consider is equal to \(2^m\) and therefore the computational complexity grows exponentially with the number of components in the network. An efficient algorithm for computing the multidimensional array representation of the survival signature for a system, based on the use of the reduced ordered binary decision diagrams (BDD) data structure, is proposed.

A BDD \[15\] is a data structure in the form of a rooted directed acyclic graph which can be used to compactly represent and efficiently manipulate a Boolean function. They are based upon Shannon decomposition theory \[16\]. The Shannon decomposition of a Boolean function \(f\) on Boolean variable \(x_i\) is defined as \(f = x_i \land f_{x_i=1} + \bar{x_i} \land f_{x_i=0}\) where \(f_{x_i=v}\) is \(f\) evaluated with \(x_i = v\). Each BDD contains two terminal nodes that represent the Boolean constant values 1 and 0, whilst each non-terminal node represents a subfunction \(g\), is labelled with a Boolean variable \(v\) and has two outgoing edges. By applying a total ordering on the \(m\) Boolean variables for function \(f\) by mapping them to the integers \(x_0, x_m\}\), and applying the Shannon decomposition recursively to \(f\), it can be represented as a binary tree with \(m + 1\) levels. Note that the chosen ordering can have a significant influence on the size of the BDD \[17\]. Each intermediate node, referred to as an if-then-else (ite) node, at level \(l \in \{0, \ldots, m - 1\}\) (where the root node is at level 0 and the nodes at level \(m - 1\) are adjacent to the terminal nodes) represents a Boolean function \(g\) on variables \(x_l, x_{l+1}, \ldots, x_{m-1}\). It is labelled with variable \(x_l\) and has two out edges called 1-edge and 0-edge linking to nodes labelled with variables higher in the ordering. 1-edge corresponds to \(x_l = 1\) and links to the node representing \(g_{x_l=1}\), whilst 0-edge corresponds to \(x_l = 0\) and links to the node representing \(g_{x_l=0}\). In addition, the following two reduction rules are applied. Firstly, the isomorphic subgraphs are merged; and secondly, any node whose two children are isomorphic is eliminated.

Complement edges \[18\] are an extension to standard BDDs that reduce memory size and the computation time. A complement edge is an ordinary edge that is marked to indicate that the connected child node (at a higher level) has to be interpreted as the complement of its Boolean function. The use of complement edges is limited to the 0-edges to ensure canonicity.
The BDD representing the system structure function for a network can be computed in various ways, e.g. from its cut-sets or network decomposition based methods [19]. In order to show the implementation of the approach, a simple network with 4 nodes and 4 edges is considered and shown in Figure 1.

The corresponding BDD representing the structure function of this network, where the dashed edges represent 0-edges (marked with -1 if complemented) and solid edges represent 1-edges, is shown in Figure 2. The survival signature from a BDD representation of the system structure function for a network can then be calculated through the iterative algorithm described by Figure 3.

The number of operations performed during the execution of the algorithm grows approximately linearly with the number of nodes in the BDD. In general, the BDD representation of the structure function for a network has far fewer nodes than $2^m$ nodes. It is therefore far more computationally efficient than using enumerative algorithms.

4 NUMERICAL EXAMPLE

Figure 4 shows a lifeline network of 17 nodes and 32 edges. The source is the node $s$ and the sink is the node $t$. All the nodes are assumed to be perfectly reliable in the network.

Three cases are considered. The first case is used to compare results between the former improved recursive decomposition method and the presented survival signature-based method. And the proposed approach is extended to analyse complex network with multiple component types in the second case. For the last case, imprecision is taken into consideration.

4.1 Network with Single Type of Components

Reliability analysis on the network show in Figure 4 considering only one type of components, which has been studied by Liu and Li [20]. In this Case, there has an assumption that the...
edges of the network are independent and identical distributed. All edges are undirected edges (which means all edges are connected by nodes), let all edges reliability be 0.9 (Case I), 0.8 (Case II), 0.2 (Case III) and 0.1 (Case IV).

The network reliability calculated by the improved recursive decomposition algorithm in [20] is 0.999930 (Case I), 0.996522 (Case II), 0.017194 (Case III) and 0.000777 (Case IV), respectively.

By using the efficient algorithm which proposed in this paper, the survival signature of the complex network can be calculated in 28.07 seconds, and the results can be seen in Table 1.

<table>
<thead>
<tr>
<th>l</th>
<th>Φ(l)</th>
<th>l</th>
<th>Φ(l)</th>
<th>l</th>
<th>Φ(l)</th>
<th>l</th>
<th>Φ(l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.00014</td>
<td>5</td>
<td>0.00081</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.00285</td>
<td>7</td>
<td>0.0077</td>
<td>8</td>
<td>0.01765</td>
<td>9</td>
<td>0.03597</td>
</tr>
<tr>
<td>12</td>
<td>0.18409</td>
<td>13</td>
<td>0.27635</td>
<td>14</td>
<td>0.38916</td>
<td>15</td>
<td>0.51445</td>
</tr>
<tr>
<td>18</td>
<td>0.18409</td>
<td>19</td>
<td>0.90414</td>
<td>20</td>
<td>0.94679</td>
<td>21</td>
<td>0.97271</td>
</tr>
<tr>
<td>24</td>
<td>0.99799</td>
<td>25</td>
<td>0.99938</td>
<td>26</td>
<td>0.99985</td>
<td>27</td>
<td>0.99998</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>31</td>
<td>1</td>
<td>32</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Survival signature of the network in Figure 4.

In all four Cases, the network reliabilities calculated through the survival signature-based reliability method given by Equation 2 are identical to those calculated using the method from Liu and Li [20]. However, the survival signature-based method only needs to calculate the survival signature of the network once and store the results, so it is efficient to calculate the network reliability for more cases. Furthermore, the proposed method is powerful at dealing with the complex networks with multiple component types and components with time varying distributions.

4.2 Network with Multiple Types of Components

Assume that the edges within the network are belonging to three types instead of one single type. To be specific, edges 1, 2, 3, 4, 5, 28, 29, 30, 31 and 32 are in type one with reliability is 0.9; edges 6, 7, 8, 9, 15, 16, 17, 18, 24, 25, 26 and 27 are in type two with reliability is 0.8; edges 10, 11, 12, 13, 14, 19, 20, 21, 22 and 23 are in type three with reliability is 0.2.

In order to estimate the network reliability, the survival signature of this network can be calculated by the proposed algorithm in 23.78 seconds, and then the reliability of the network is 0.3746931 by using Equation 2.

It can be seen from the above examples that the network reliability is time independent, because we assume the edge reliability values are stable as time goes. In the real engineering world, however, the failure times of edges are according to different distribution types (i.e., Exponential, Weibull, Normal and Lognormal distribution) sometimes. All of these distribution are time dependent, and will lead to the network reliability values are time varying.

Now let assume the failure times of edges type one are according to Exponential distribution with parameter \( \lambda = 0.12 \). Similarly, type two \( \sim \) Weibull(0.4,4.2) and type three \( \sim \) Normal(0.7,0.02).

The survival signature remains the same as the network does not change its configuration. The survival function of the network is shown in Figure 5. It can be seen that the survival function is time varying, thus, it is easy to know the network reliability at each time.
4.3 Imprecise Network Reliability

Due to lack of data or limited knowledge, there are not always precise data for edges failure time distributions. For instance, Table 2 shows the failure types and imprecise distribution parameters of edges in the network.

<table>
<thead>
<tr>
<th>Edge Type</th>
<th>Distribution Type</th>
<th>Imprecise Parameters $\lambda$ or $(\alpha, \beta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exponential</td>
<td>[0.08, 0.18]</td>
</tr>
<tr>
<td>2</td>
<td>Weibull</td>
<td>([0.3, 0.6], [3.8, 4.6])</td>
</tr>
<tr>
<td>3</td>
<td>Normal</td>
<td>([0.5, 0.8], [0.01, 0.03])</td>
</tr>
</tbody>
</table>

Table 2: Failure types and imprecise distribution parameters of edges in the network of Figure 4.

According to Equations 4 and 5, the lower and upper bounds of survival function of the network can be estimated by means of a double loop approach. The double loop sampling involves two layers of sampling: the outer loop, called the parameter loop, samples values from the set of distribution parameters; while the inner loop computes the survival function stating for the network knowing the precise probability distribution functions. Figure 5 shows the interval of the survival function. It can conclude that imprecision either within the components failure time distribution parameters can lead to survival function intervals of the complex network.

5 CONCLUSIONS

Survival signature opens a new pathway for analysing complex network with multiple component types, and it just needs to be calculated once for a specific network, which represents a significant computational advantage. An efficient algorithm has been proposed for calculating the survival signature of large and complex networks, and then used for reliability and uncertainty assessment on lifeline networks based on the survival signature.

The proposed approach allows to include imprecision and vagueness both within the components failure time distribution parameters and lifeline network configuration. Either analytical approaches or simulation methods can be applied to deal with the uncertainty efficiently. The case study presented in this paper indicates that the proposed approaches can be used to evaluate the reliability and uncertainty of complex networks efficiently.

REFERENCES


Figure 3: Algorithm for computing signature from the BDD representation of a system structure function.
Figure 4: A lifeline network with 17 nodes and 32 edges [20].

Figure 5: Time varying precise survival function alongside with lower and upper bounds of survival function of the network in Figure 4 (imprecise distribution parameters).
RANDOM VIBRATION ANALYSIS OF A SINGLE DEGREE OF FREEDOM UNDER IMPRECISE PROBABILITY ASSIGNMENTS

Alice Cicirello¹, Robin S. Langley²

¹ Department of Engineering Science
University of Oxford
Parks Road, Oxford OX1 3PJ
alice.cicirello@eng.ox.ac.uk

² Department of Engineering
University of Cambridge
Trumpington Street, Cambridge CB2 1PZ
rsl21@eng.cam.ac.uk

Keywords: Imprecise probability, random vibration, Miles’ equation, bounds on failure probability, design of aircraft structural components.

Abstract. The random vibration analysis of aircraft structural components is often performed with simplified techniques, such as Miles’ Equation, which yields the root mean square of the acceleration response. The failure probability of the structural component is then established as the probability that the displacement exceeds a given threshold with a velocity in a certain period of time. Nonetheless, the dominant frequency of the system may be uncertain because of manufacturing variability, and because of variation in joints and connections of structural components. In this paper the conventional approach is extended in two ways: (i) firstly, the system parameters are taken to be uncertain rather than known, (ii) secondly, the uncertainty is modelled with an imprecise probability description, rather than with conventional pdfs (which usually are not known, due to a lack of empirical information). When the system parameters are modelled with a probability density function, the proposed approach yields: (a) the unconditional failure probability across an ensemble of Single Degree of Freedom (SDoF) systems; (b) the probability that the failure probability exceeds a specified target level across the ensemble of SDoFs. By introducing uncertainty in the probabilistic assignment of the system parameters, the bounds on these quantities are readily obtained. The feasibility of the proposed approaches are shown through their application to a structural panel of a spacecraft.
1 INTRODUCTION

In the design of aircraft structural components is often assumed that only the randomness in the excitation (of mechanical and acoustic nature) is of concern, while the parameters of the structural components are assumed to be deterministic [1]. The random vibration analysis of aircraft structural components is often performed with simplified techniques, such as Miles’ Equation, which assumes that a structural component, or a secondary structure, has a dominant natural frequency with respect to the structural response [1]. This equation yields the root mean square of the acceleration, as a function of the natural frequency of an equivalent Single Degree of Freedom (SDoF), of the quality factor and of the value of the random external force (expressed as a Power Spectral Density) at the natural frequency of the SDoF. The failure probability of the structural component is then established as the probability that the displacement exceeds a given threshold with a velocity in a certain period of time, by using well established techniques [1-3]. Nonetheless, the dominant frequency of the system may be uncertain because of manufacturing variability, and because of variation in joints and connections of structural components.

In this paper the conventional approach is extended in two ways: (i) firstly, the system parameters are taken to be uncertain rather than known, (ii) secondly, the uncertainty is modelled with an imprecise probability description [4], rather than with conventional pdfs (which usually are not known, due to a lack of empirical information). When the system parameters are modelled with a probability density function, the proposed approach yields: (a) the unconditional failure probability across an ensemble of SDoFs; (b) the probability that the failure probability exceeds a specified target level across the ensemble of SDoFs. By introducing uncertainty in the probabilistic assignment of the system parameters, the bounds on these quantities are readily obtained.

The standard random vibration analysis of aircraft structural components is reviewed in section 2. In section 3, the system parameters are taken to be uncertain rather than known, and described using a specified probability density function. Two methods are proposed: (i) Method A addresses the problem of estimating the probability that a deterministic limit value of the SDoF displacement response is exceeded by the response variable across the ensemble of SDoFs; (ii) Method B yields the probability that the failure probability itself exceeds a critical value. In section 4, an imprecise probability model based on a generalization of the Maximum Entropy distribution under uncertain statistical information is reviewed, and then Method A and B are extended to include this type of description to yield bounds on the two probability estimates. The random vibration analysis of a Single Degree of Freedom system with uncertainty in the system parameters, and in the probabilistic assignment of the system parameters is addressed in section 5.

2 REVIEW OF RANDOM VIBRATION ANALYSIS OF AIRCRAFT STRUCTURAL COMPONENTS

Miles’ Equation [1] is a well-established technique used to assess the performance of aircraft structural components under random vibration. In this context it is assumed that a structural component, or a secondary structure, has a dominant natural frequency with respect to the structural response. The random vibration analysis of aircraft structural components reduces to the study of a Single Degree of Freedom (SDOF), consisting of a mass $m$, a spring
k and a damper c, subject to a random external force \( F(t) \) which produces a displacement of the mass \( y(t) \), as shown in Figure 1.

The governing equation of the problem can be written as:

\[
\ddot{y}(t) + 2\zeta \omega_n \dot{y}(t) + \omega_n^2 y(t) = \frac{F(t)}{m}
\]  

(1)

being \( \zeta = c/(2\sqrt{k \, m}) \) the damping ratio, and \( \omega_n = \sqrt{k / m} = 2\pi f_n \) the SDOF natural frequency corresponding to the dominant natural frequency of the structural component.

If the input is a broadband random loading with a constant spectral density over the frequency range of interest, \( S_{FF}(\bar{f}) = S_0 \) (with units of \( (N/kg)^2/Hz \)), being \( \bar{f} \) the frequency expressed in Hz, the solution of the SDoF governing equation yields the Miles’ Equation [1]:

\[
\sigma_y = y_{rms} = \sqrt{\frac{Q S_0 (\bar{f}_n)}{32\pi^3 f_n^3}}
\]  

(2)

This equation expresses the root mean square (rms) of the mass displacement, \( \sigma_y \) as a function of the natural frequency of the SDoF, \( \bar{f}_n \), of the quality factor \( Q = 1/(2\zeta) = (\sqrt{k \, m})/c \) and of the value of the excitation force spectral density at the natural frequency \( S_0(\bar{f}_n) \).

The rms of the mass displacement is then used to verify that the structure is be able to withstand a certain stress levels, such as the yield stress. This is achieved by defining the failure probability of the structural component, as the probability that the displacement \( y(t) \) exceeds a given level \( b \) with a velocity \( \dot{y}(t) \) in a certain period of time \( \tau \). It is well known that by
assuming that each crossing of the barrier is independent from each other and randomly distributed along the time axis, the failure probability $P_f$ can be expressed as [1-3]:

$$P_f = 1 - \exp\left[ -\nu_b^+ \tau \right]$$

(3)

Being $\nu_b^+$ the average number of positive crossing per unit of time of a barrier $b$, whose distribution is given by [1-3]:

$$\nu_b^+ = \frac{\sigma_y}{2\pi\sigma_y} \exp\left[ -\frac{1}{2} \left( \frac{b}{\sigma_y} \right)^2 \right]$$

(4)

Being $\sigma_y$ the mean square of the velocity which is given by [1-3]:

$$\sigma_y = \sqrt{\frac{S_0}{8\xi \left( 2\pi \bar{f}_n \right)}}$$

(5)

The main aim of the analysis is therefore to establish the failure probability obtained with Equation (3) and compare it to a failure probability target. If the failure probability obtained is higher than the target one, a different design of the aircraft structural component should be explored. Nonetheless, the dominant frequency of the system may be uncertain, leading to uncertainty in the failure probability estimate. This conventional approach is extended in the next sections to account for the effect of the uncertainties in the system parameters.

3 RANDOM VIBRATION ANALYSIS OF AIRCRAFT STRUCTURAL COMPONENTS WITH UNCERTAIN SYSTEM PARAMETERS

Let us include uncertainty in the system parameters and model this uncertainty with a probability density function. The problem concern with an ensemble of SDoFs, where each element of the ensemble has a specified value of the natural frequency. When subject to the random loading, each element of the ensemble is characterized by a random displacement response. If the natural frequency of the SDOF $\bar{f}_n$ is described by a probability distribution $p(\bar{f}_n)$, two different approaches can be considered to yield two probability estimates which can be very useful from the engineering point of view.

3.1 Method A – unconditional failure probability

Method A enables the evaluation of the unconditional failure probability across an ensemble of SDoFs. This is expressed as:

$$\bar{P}_f = \int P_f(\bar{f}_n) p(\bar{f}_n) d\bar{f}_n$$

(6)

Being $P_f(\bar{f}_n)$ the failure probability of a member $n$ of the ensemble (for fixed $\bar{f}_n$) which is computed using Equation (3). This type of integral can be evaluated numerically by considering a grid of integration points (direct integration) or by employing Monte Carlo Simulations [5].
3.2 Method B – probability of failure probability exceeding a target level

Method B yields the probability that the failure probability exceeds a specified target level across the ensemble of SDoFs. This is expressed as:

\[ P \left[ P_f > \theta \right] = \int_{P_f > \theta} p(\bar{f}_n) d\bar{f}_n \]  

(7)

In most cases the analytical solution of this integral is not known and its approximate solution can be obtained by using numerical integration methods or Monte Carlo Simulations [5].

While Method A addresses the problem of estimating the probability that a deterministic limit value of the SDoF displacement response is exceeded by the response variable across the ensemble of SDoFs; Method B addresses a different type of problem: by specifying a critical value of the failure probability, it yields the probability that the failure probability itself exceeds this critical value. Therefore, these approaches yield two probability estimates which can be very useful from the engineering point of view.

4 INTRODUCING UNCERTAINTY IN THE PROBABILISTIC ASSIGNMENT OF THE SYSTEM PARAMETERS

4.1 Generalization of the Maximum Entropy distribution under uncertain statistical information

One of the main difficulties in performing a probabilistic uncertainty analysis is to specify an appropriate pdf of the uncertain system parameters. This might be caused by insufficient data required to empirically determine the pdf, because of cost or time constraints, or because the structure does not yet exist. Considering a wrong pdf might lead to overestimating or underestimating the system performance, therefore it is extremely important to represent the current state of knowledge using only the available information.

Often the only available information on the probabilistic assignment of some system parameters is that some statistical moments are known to lie in a certain domain (e.g. the mean can take values between an upper and lower bounds, and/or there is a certain probability of finding the uncertain variable within certain bounds). For example, this domain can be defined as a set of inequality constraints on the statistical expectation:

\[ v_{j,\min} \leq v_j = \mathbb{E}[f_j(x)] = \int f_j(x) p(x) dx \leq v_{j,\max}, \quad j = 2, 3, ..., n \]  

(8)

where \( v_{j,\min} \) and \( v_{j,\max} \) are the lower and upper bound on the \( j \)th statistical expectation \( v_j \) and \( f_j(x) \) is a specified function of \( x \). For example, if \( f_j(x) = x \) then the constraints are specified on the mean value, alternatively if \( f_j(x) = x^2 \) they are specified on the second moment.

Many distributions belonging to the family of polynomial distributions, exponential distributions, maximum entropy distributions, and others can satisfy these inequality constraints. The principle of Maximum Entropy [6] allows the estimation of a subjective pdf \( p(x) \) of an un-
certain variable $x$ which best represent the current state of knowledge by maximizing the relative entropy subject to constraints representing the available information [6]. However, if Maximum Entropy is used to find the distribution with the largest entropy whose statistical moments lie within the domain, the information at only a single point in the domain would be used and other information would be discarded. Recently a generalisation of this approach has been proposed [4], that first constructs a family of maximum entropy distributions consistent with the statistical inequality constraints, then propagates the family of pdfs through the equations describing the problem on hand in order to yield the pdf which maximizes (or equivalently minimizes) a specified engineering metric [4].

With this approach [4], the pdf of a vaguely known variable $x$ is expressed as the exponential of a linear combination of specified functions of the random variable $f_j(x)$ and some bounded variables $a$ (referred to here as basic variables), so that:

$$p(x | a \in R) = t(x) \exp \left[ \sum_{j=1}^{n} a_j f_j(x) \right].$$

The basic variables are contained in the vector $a$, which has entries $a_j$ with $j = 2,3,...,n$ that lie within an admissible region $R$ (which can be an interval, a convex region, etc.). These basic variables substitute the Lagrange multipliers of a standard Maximum Entropy distribution and are such that they can have any possible pdf within certain bounds, including the extreme case of a delta function at any point between the bounds. Therefore, Eq. (9) represents a family of maximum entropy distributions defined over the set of basic variables $a$. If a parameter is not “basic”, then its pdf can be expressed in terms of the basic parameters, and thus only this type of parameter is considered in what follows. The term $f_1(x) = 1$, the coefficient $a_i$ is dependent on the basic variables $a_j$ (with $j = 2,3,...,n$) and it is chosen to satisfy the normalisation condition. The function $t(x)$ is the reference pdf introduced to allow the pdf to be frame invariant. In the forthcoming discussions, without loss of generality, it will be assumed that $t(x) = 1$.

The construction of the family of Maximum Entropy distributions consists of the following steps:

1. Express the pdf of a vaguely known variable in the form of Eq. (9).
2. Convert the statistical moments domain ($m$-domain) into a basic variable domain (so-called $a$-domain) [4].
3. Subdivide the $a$-domain into a sufficient number of grid of points.
4. Compute the coefficient $a_i$ for each point of the grid (by using the normalisation condition) to derive the corresponding pdf.

The mapping of the $m$-domain to the $a$-domain (Step 2) can be very computationally demanding even for a simple domain, since it requires the solution of a set of non-linear equations for each point that need to be mapped, and convergence problems may occur.

In reference [4] an approach for computing an approximate mapping has been presented. This approach is based on expressing the variation of the pdf over the $m$-domain as a Taylor series, centered at $a^*$ which is a point of the $a$-domain obtained from the mapping of the mid-point
of the \(m\)-domain (denoted by \(\nu^*\)), so that:

\[
v_s(a) = c_s^1(a^*) - \sum_{j=1}^{n} (a_j - a^*_j) c_s^2(a^*) - \frac{1}{2} \sum_{j=1}^{n} \sum_{k=2}^{n} (a_j - a^*_j)(a_k - a^*_k) c_s^3(a^*)
\]

\[s = 1, 2, \ldots, n\]

where \(c_s^1(a^*), c_s^2(a^*)\) and \(c_s^3(a^*)\) are the first, second and third order cumulants, respectively, calculated at the \(a^*\) point. The term \(c_s^1(a^*)\) corresponds to the statistical expectation calculated at \(a^*\), therefore it is equivalent to the term \(v_s^*\) which can be obtained directly from the \(m\)-domain, while \(c_s^2(a^*)\) and \(c_s^3(a^*)\) are calculated by numerical integration of:

\[
c_s^2(a^*) = \int (f_s(x) - v_s^*) (f_j(x) - v_j^*) p(x|a^*) \, dx,
\]

\[
c_s^3(a^*) = \int (f_s(x) - v_s^*) (f_j(x) - v_j^*) (f_k(x) - v_k^*) p(x|a^*) \, dx.
\]

Each point of the \(a\)-domain \(a\) corresponding to a point \(v(a)\) of the \(m\)-domain can be then computed by solving a set of \(n\) (where \(n\) corresponds to the domain dimension) equations (in the form of Eq. (10)) where a linear or quadratic approximation might be considered. High order terms may also be included [4]. Therefore, if \(r\) points of a 3-dimensional \(m\)-domain are mapped into the \(a\)-domain, this would require solving \(r\) sets of 3 equations.

By adopting this approach, the edges of the \(m\)-domain can be mapped to the \(a\)-domain in an efficient way, thus allowing a permissible region of the \(a\)-domain to be determined.

### 4.2 Extension of Method A and Method B

The uncertainty in the probabilistic assignment of the system parameters of the SDOF is now expressed as in Equation (9), so that the pdf of the natural frequency of the system is given by \(p(f|\bar{\nu})\). For fixed basic variables \(P_f(\bar{\nu}|a)\) is then obtained using Eq. (3), then Equation (6) can be generalized to yield the unconditional failure probability for fixed basic variables:

\[
P_f(a) = \int P_f(\bar{\nu}|a) p(\nu|a) \, d\nu
\]

As a result the upper and lower bounds on the failure probability are given by:

\[
\min_{a \in R} \left( P_f(a) \right) \leq \bar{P}_f \leq \max_{a \in R} \left( P_f(a) \right)
\]

Similarly, Method B can be generalized to yield the probability that the failure probability exceed a certain threshold \(\theta\) for fixed basic variables \(a\):

\[
P\left( \left( P_f > \theta \right) | a \right) = \int_{P_f > \theta} p(\nu|a) \, d\nu
\]

To yield the following bounds:
\[
\min_{a \in R} \{ P \left( \left[ P_f > \theta \right] \bigg| a \right) \} \leq P \left( \left[ P_f > \theta \right] \right) \leq \max_{a \in R} \{ P \left( \left[ P_f > \theta \right] \bigg| a \right) \} \tag{16}
\]

5 NUMERICAL APPLICATION

Let us consider a structural panel of a spacecraft with fundamental frequency of \( f_n = 250 \) Hz, damping ratio \( \zeta = 0.05 \) (quality factor \( Q = 10 \)), subject to a white noise Power Spectral Density function \( S_0 = 100 \) (N/kg \( )^2 / \)Hz. A displacement threshold level \( b = 7 \times \sigma_y = 0.0018 \) mm (\( \sigma_y \) being computed with Eq. (2)) during a time interval \( \tau = 3 \) hours is specified.

Using the standard approach (Eq. (3)), the resulting failure probability is: \( P_f = 6.1 \times 10^{-5} \).

Now, let us consider that the fundamental frequency of the panel is uncertain since the spring stiffness is uncertain. Specifically, let us consider the case that the following information are available on the spring stiffness: (i) the variable is positive (ii) the vertices of a convex region of the statistical expectations \( E[x] \) and \( E[\ln(x)] \) are known, and as shown in Figure 2.

![Figure 2: Statistical expectation domain (m-domain). The circles indicate points along the m-domain, while the “x” indicates the middle point of the m-domain](image)

The vertices of the domain have the following coordinates: point 1 \((18.0\times10^3,14.273)\); point 5 \((22.0\times10^3,14.518)\); point 9 \((22.0\times10^3,14.498)\); and point 13 \((18.0\times10^3,14.243)\).

Following the procedure described in subsection 4.1, the pdf of \( x \) is written in the same form of Equation (9), that is a Gamma distribution:

\[
p(x|a) = \exp \left[ a_1 - a_2 x - a_3 \ln(x) \right], \tag{17}
\]

where \( a_2 \) and \( a_3 \) are the two unknown basic variables, while \( a_1 \) is obtained by applying the normalisation condition (for \( \Re(a_2) > 0; \Re(a_3) < 1 \)) as:

\[
a_1 = -\ln \left( a_2^{(a_3-1)} \Gamma(1-a_3) \right), \tag{18}
\]

107
where $\Gamma(\bullet)$ is the gamma function.

Although the $m$-domain is characterized by linear surfaces, the corresponding $a$-domain is not; 16 points along the $m$-domain are mapped onto the $a$-domain by using the strategy summarized in section 4.1. The approximate mapping strategy would consider a Taylor series expansion over the whole domain, taking the mid-point of the domain as the reference point for the expansion. The mid-point has coordinates $E[\chi^z] = 20.0 \times 10^9$ N/m and $E[\ln(\chi)] = 14.383$, and it is indicated with an “x” in Figure 2, and its mapping to the $a$-domain corresponds to $a^* = (20.696 \times 10^3, -3.139)$. Each point in the $a$-domain obtained with the approximate procedure has been verified by solving a set of non-linear equations showing a good agreement. The $a$-domain so obtained is shown in Figure 3.

A set of 430 pdfs was generated by considering the 414 point inside the $a$-domain (considering a grid of $50 \times 50$ equally-spaced points overlaying the $a$-domain) and the 16 points along the $a$-domain.

The Maximum Entropy distribution, which is the gamma pdf with the largest entropy that satisfies the statistical moment conditions, can be computed by calculating the entropy (measured in nats) associated with each gamma distribution:

$$h_e = \ln \left( \frac{\Gamma(1-a_3)}{a_2} \right) + a_3 \Psi(1-a_3) + 1 - a_3$$

(19)

and then searching in the $a$-domain, the combination of basic variables which maximise Eq. (19). For the case under investigation, the pdf with the largest entropy is the one with basic variables corresponding to point 10 of the $a$-domain.

By performing a transformation of variables, the pdf of the natural frequency conditional on the basic variables, $p(\mathcal{F}_n | a)$, is obtained as:

$$p(\mathcal{F}_n | a) = 2\gamma\mathcal{F}_n \exp \left[ a_1 - a_2 \gamma \mathcal{F}_n^2 - a_3 \ln \left( \gamma \mathcal{F}_n^2 \right) \right],$$

(20)
where $\gamma = 4\pi^2m$ with $m=0.81$ kg, and $a_1$ is obtained by applying the normalisation condition (for $\text{Re}[a_2] > 0; \text{Re}[a_3] < 0$) as:

$$a_1 = \ln \left( \frac{a_2\gamma^{a_3}(a_2\gamma)^{-a_3}}{\Gamma(1-a_3)} \right).$$

The maximum entropy distribution (obtained at point 10 of the domain) would yield a failure probability of: $P_f|_{\text{point 10}} = 0.2935$ (solving Eq. (6) using direct integration), that is much larger than the result obtained using a deterministic value. This results can be explained by considering the sharp variation of the $P_f$ close to 250 Hz, as shown in Figure 4.

![Figure 4: $P_f$ as a function of frequency](image)

If we now set a failure probability target $\theta = 0.001$, the probability of exceedance of this target (obtained solving Eq. (7) using numerical integration) for the MaxEnt distribution is $P\left[\left(P_f > \theta\right)\right]|_{\text{point 10}} = 0.4605$.

By letting varying the basic variable within the $a$-domain (using the 430 distributions), it has been found that the unconditional failure probability can take values between $0.2126 \leq P_f \leq 0.4303$ (obtained at points 7 and 13, respectively); therefore the MaxEnt pdf would underestimate the upper bound.

Similarly, the probability that the failure probability target is exceeded can vary from $0.3876 \leq P\left[\left(P_f > \theta = 0.001\right)\right] \leq 0.5918$ (obtained at points 7 and 13, respectively); also in this case the MaxEnt pdf would underestimate the response upper bound.

It can be concluded that employing the proposed uncertainty model enable an enhanced description of structural panel performance yielding the maximum and minimum values that the engineering metric of interest (eg the failure probability) might take based on the available information, rather than yielding a single value which might significantly underestimate or overestimate the engineering metric of interest.
6 CONCLUSIONS

Two approaches for the random vibration analysis of a single degree of freedom under uncertainty in the system parameters have been presented. These two approaches yield: (i) the probability that a deterministic limit value of the SDoF displacement response is exceeded by the response variable across the ensemble of SDoFs; and the (ii) probability that the failure probability itself exceeds a critical value. These two probability estimates can be very useful from the engineering point of view, and in particular for the design of aircraft structural components which are usually performed assuming deterministic system parameters.

These two approaches have been also extended to account for the uncertainty in the probabilistic assignment of the system parameters, yielding bounds on the two probability estimates. Through a numerical application, it has been shown that including uncertainty in the probabilistic assignment lead to an enhanced description of the system response that support the development of a design solution which can be more robust with respect to the design constraints under uncertainties in the models parameters.

REFERENCES

PROBABILISTIC REDUCED-ORDER MODELING FOR STOCHASTIC
PARTIAL DIFFERENTIAL EQUATIONS

Constantin Grigo\textsuperscript{1} and Phaedon-Stelios Koutsourelakis\textsuperscript{1}

\textsuperscript{1}Continuum Mechanics Group,
Department of Mechanical Engineering,
Technical University of Munich,
D-85748 Garching b. München

Keywords: Reduced-order modeling, generative Bayesian model, SPDE, effective material properties

Abstract. We discuss a Bayesian formulation to coarse-graining (CG) of PDEs where the coefficients (e.g. material parameters) exhibit random, fine scale variability. The direct solution to such problems requires grids that are small enough to resolve this fine scale variability which unavoidably requires the repeated solution of very large systems of algebraic equations.

We establish a physically inspired, data-driven coarse-grained model which learns a low-dimensional set of microstructural features that are predictive of the fine-grained model (FG) response. Once learned, those features provide a sharp distribution over the coarse scale effective coefficients of the PDE that are most suitable for prediction of the fine scale model output.

This ultimately allows to replace the computationally expensive FG by a generative probabilistic model based on evaluating the much cheaper CG several times. Sparsity enforcing priors further increase predictive efficiency and reveal microstructural features that are important in predicting the FG response. Moreover, the model yields probabilistic rather than single-point predictions, which enables the quantification of the unavoidable epistemic uncertainty that is present due to the information loss that occurs during the coarse-graining process.
1 INTRODUCTION

Many engineering design problems such as flow in porous media and mechanical properties of composite materials require simulations that are capable of resolving the microstructure of the underlying medium. If the material components under consideration exhibit fine-scale heterogeneity, popular discretization schemes (e.g. finite elements) yield very large systems of algebraic equations. Pertinent solution strategies at best (e.g. multigrid methods) scale linearly with the dimension of the unknown state vector. Despite the ongoing improvements in computer hardware, repeated solutions of such problems, as is required in the context of uncertainty quantification (UQ), poses insurmountable difficulties.

It is obvious that viable strategies for such problems, as well as a host of other deterministic problems where repeated evaluations are needed such as inverse, control/design problems etc, should focus on methods that exhibit sublinear complexity with respect to the dimension of the original problem. In the context of UQ a popular and general such strategy involves the use of surrogate models or emulators which attempt to learn the input-output map implied by the fine-grained model. Such models, e.g. Gaussian Processes [26], polynomial chaos expansions [8], neural nets [2] and many more, are trained on a finite set of fine-grained model runs. Nevertheless, their performance is seriously impeded by the curse of dimensionality, i.e. they usually become inaccurate for input dimensions larger than a few tens or hundreds, or equivalently, the number of FG runs required to achieve an acceptable level of accuracy grows exponentially fast with the input dimension.

Alternative strategies for high-dimensional problems make use of multi-fidelity models [10, 16] as inexpensive predictors of the FG output. As shown in [12], lower-fidelity models whose output deviates significantly from that of the FG can still yield accurate estimates with significant computational savings, as long as the outputs of the models exhibit statistical dependence.

In the case of PDEs where finite elements are employed as the FG, multi-fidelity solvers can be simply obtained by using coarser discretizations in space or time. While linear and nonlinear dimensionality reduction techniques are suitable for dealing with high-dimensional inputs [14], it is known which of the microstructural features are actually predictive of FG outputs [22].

The model proposed in the present paper attempts to address this question. By using a two-component Bayesian network, we are able to predict fine-grained model outputs based on only a finite number of training data runs and a repeated solution of a much coarser model. Uncertainties can be easily quantified as our model leads to probabilistic rather than point-like predictions.

2 THE FINE-GRAINED MODEL

Let \((\Omega, \mathcal{F}, P)\) be a probability space. Let \(\mathcal{H}\) be the Hilbert space of functions defined over the domain \(\mathcal{D}\) over which the physical problem is defined. We consider problems in the context of heterogeneous media which exhibit properties given by a random process \(\lambda(x, \xi)\) defined over the product space \(\mathcal{D} \times \Omega\). The corresponding stochastic PDE may be written as

$$\mathcal{L}(x, \lambda(x, \xi)) u(x, \lambda(x, \xi)) = f(x), \quad \text{B.C.} \quad (1)$$

where \(\mathcal{L}\) is a stochastic differential operator and \(x \in \mathcal{D}, \xi \in \Omega\) are elements of the physical domain and the sample space, respectively. Discretization of the random process \(\lambda(x, \xi) \xrightarrow{\text{discretize}} \lambda_f \in \mathbb{R}^{n_f}\) as well as the governing equation leads to a system of \(n_f\) (potentially
nonlinear) algebraic equations, which can be written in residual form as

$$r_f(U_f; \lambda_f) = 0,$$

(2)

where $U_f(\lambda_f) \in \mathbb{R}^{n_f}$ is the $n_f$-dimensional discretized solution vector for a given $\lambda_f$ and $r_f : \mathbb{R}^{n_f} \times \mathbb{R}^{n\lambda_f} \rightarrow \mathbb{R}^{n_f}$ the discretized residual vector. It is the model described by equation (2) which is denoted as the fine-grained model (FG) henceforth.

3 A GENERATIVE BAYESIAN SURROGATE MODEL

Let

$$p(\lambda_f) = \int \delta(\lambda_f - \lambda_f(\xi)) p(\xi) d\xi$$

(3)

be the density of $\lambda_f$. The density of the fine-scale response $U_f$ is then given by

$$p(U_f) = \int p(U_f|\lambda_f)p(\lambda_f)d\lambda_f,$$

(4)

where the conditional density $p(U_f|\lambda_f)$ degenerates to a $\delta(U_f - U_f(\lambda_f))$ when the only uncertainties in the problem are due to $\lambda_f$.

The objective of this paper is to approximate this input-output map implied by $U_f(\lambda_f)$, or equivalently in terms of probability densities, the conditional distribution $p(U_f|\lambda_f)$. The latter case can also account for problems where additional sources of uncertainty are present and the input-output map is stochastic. To that end, we introduce a coarse-grained model (CG) leading to an approximate distribution $\tilde{p}(U_f|\lambda_f)$ which will be trained on a limited number of FG solutions $D_N = \left\{ \lambda_f^{(i)}, U_f^{(i)}(\lambda_f^{(i)}) \right\}_{i=1}^N$.

Our approximate model $\tilde{p}(U_f|\lambda_f)$ employs a set of latent [3], reduced, collective variables which we denote by $\lambda_c \in \mathbb{R}^{n\lambda_c}$ for reasons that will be apparent in the sequel, such that

$$\tilde{p}(U_f|\lambda_f) = \int \tilde{p}(U_f, \lambda_c|\lambda_f) d\lambda_c$$

$$= \int \frac{\tilde{p}(U_f|\lambda_c) \tilde{p}(\lambda_c|\lambda_f)}{\text{decoder}} d\lambda_c. $$

(5)

As it can be understood from the equation above, the latent variables $\lambda_c$ act as a probabilistic filter (encoder) on the FG input $\lambda_f$, retaining the features necessary for predicting the FG output $U_f$. In order for $\tilde{p}(U_f|\lambda_f)$ to approximate well $p(U_f|\lambda_f)$, the latent variables $\lambda_c$ should not simply be the outcome of a dimensionality reduction on $\lambda_f$. Even if $\lambda_f$ is amenable to such a dimensionality reduction, it is not necessary that the $\lambda_c$ found would be predictive of $U_f$. Posed differently, it is not important that $\lambda_c$ provides a high-fidelity encoding of $\lambda_f$ but it suffices that it is capable of providing a good prediction of the corresponding $U_f(\lambda_f)$.

The aforementioned desiderata do not unambiguously define the form of the encoding/decoding densities in (5) nor the type/dimension of the latent variables $\lambda_c$. In order to retain some of the physical and mathematical structure of the FG, we propose employing a coarsened, discretized version of the original continuous equation (1). In residual form this can again be written as

$$r_c(U_c; \lambda_c) = 0,$$

(6)
where $U_c \in \mathbb{R}^{n_c}$ is the $n_c$-dimensional ($n_c \ll n_f$) discretized solution and $r_c : \mathbb{R}^{n_f} \times \mathbb{R}^{n_c} \rightarrow \mathbb{R}^{n_c}$ is the discretized residual vector. Due to the significant discrepancy in dimensions $n_c \ll n_f$, the cost of solving the CG in (6) is negligible compared to the FG in (2).

It is clear that $\lambda_c$ plays the role of effective/equivalent properties but it is not obvious (except for some limiting cases where homogenization results can be invoked [23]) how these should depend on the fine-scale input $\lambda_f$ nor how the solution $U_c(\lambda_c)$ of the CG should relate to $U_f(\lambda_f)$ in (2). Furthermore, it is important to recognize a priori that the use of the reduced variables $\lambda_c$ in combination with the coarse model in (6) would in general imply some information loss. The latter should introduce an additional source of uncertainty in the predictions of the fine-scale output $U_f$ [25], even in the limit of infinite training data. For that purpose and in agreement with (5), we propose a generative probabilistic model composed of the following two densities:

- A probabilistic mapping from $\lambda_f$ to $\lambda_c$, which determines the effective properties $\lambda_c$ given $\lambda_f$. We write this as $p_c(\lambda_c|\lambda_f, \theta_c)$ where $\theta_c$ denotes a set of model parameters,

- A coarse-to-fine map $p_{cf}(U_f|U_c, \theta_{cf})$, which is the PDF of the FG output $U_f$ given the output $U_c$ of the CG. It is parametrized by $\theta_{cf}$.

This model is illustrated graphically in Figure 1. The density $p_c$ encodes $\lambda_f$ into $\lambda_c$ and the coarse-to-fine map $p_{cf}$ plays the role of a decoder, i.e., given the CG output $U_c$, it predicts $U_f$.

Using the abbreviated notation $\theta = [\theta_c, \theta_{cf}]$, from (5) we obtain

$$
\bar{p}(U_f|\lambda_f, \theta) = \int p_{cf}(U_f|U_c(\lambda_c), \theta_{cf})p_c(\lambda_c|\lambda_f, \theta_c)d\lambda_c,
$$

(7)

where $U_c(\lambda_c)$ is the solution vector to equation (6). The previous discussion suggests the following generative process for drawing samples from $\bar{p}(U_f|\lambda_f, \theta)$ i.e. predicting the FG output $U_f$ given a FG input $\lambda_f$,

- draw a sample $\lambda_c \sim p_c(\lambda_c|\lambda_f, \theta_c),$

- solve the CG to obtain $U_c(\lambda_c),$

- draw a sample $U_f \sim p_{cf}(U_f|U_c(\lambda_c), \theta_{cf}).$

3.1 Model training

In order to train the model described above, it is a reasonable strategy to minimize the Kullback-Leibler divergence [5] between the target density $p(U_f, \lambda_f) = \delta(U_f - U_f(\lambda_f))$
and \( \tilde{p}(U_f|\lambda_f, \theta) \). As these are conditional distributions, the KL divergence would depend on \( \lambda_f \). In order to calibrate the model for the \( \lambda_f \) values that are of significance, we operate on the augmented densities \( p(U_f, \lambda_f) = p(U_f|\lambda_f)p(\lambda_f) \) and \( \tilde{p}(U_f, \lambda_f|\theta) = \tilde{p}(U_f|\lambda_f, \theta)p(\lambda_f) \), where \( p(\lambda_f) \) is defined by (3). In particular, we propose minimizing with respect to \( \theta \)

\[
\text{KL}(p(U_f, \lambda_f)||\tilde{p}(U_f, \lambda_f)) = \text{KL}(p(U_f, \lambda_f)||\tilde{p}(U_f, \lambda_f, \theta)p(\lambda_f))
= \int p(U_f, \lambda_f) \log \left( \frac{p(U_f, \lambda_f)}{\tilde{p}(U_f, \lambda_f, \theta)p(\lambda_f)} \right) dU_f d\lambda_f
= - \int p(U_f, \lambda_f) \log \tilde{p}(U_f, \lambda_f|\theta)dU_f d\lambda_f + H(p(U_f, \lambda_f))
\approx - \frac{1}{N} \sum_{i=1}^{N} \log \tilde{p}(U_f|\lambda_f(i), \theta) + H(p(U_f, \lambda_f)),
\]

(8)

where \( N \) is the number of training samples drawn from \( p(U_f, \lambda_f) \), i.e.

\[
\lambda_f(i) \sim p(\lambda_f), \quad U_f(i) = U_f(\lambda_f(i)),
\]

(9)

and \( H(p(U_f, \lambda_f)) \) is the entropy of \( p(U_f, \lambda_f) \) that is nevertheless independent of the model parameters \( \theta \). It is obvious from the final expression in (8) that \( \sum_{i=1}^{N} \log \tilde{p}(U_f|\lambda_f(i), \theta_c, \theta_c)f \) is a log-likelihood function of the data \( D_N \) which we denote by \( L(D_N|\theta_c, \theta_c)f \). In a fully Bayesian setting, this can be complemented with a prior \( p(\theta_c, \theta_c)f | D_N \) leading to the posterior

\[
p(\theta_c, \theta_c)f | D_N \propto e^{L(D_N|\theta_c, \theta_c)f}p(\theta_c, \theta_c)f.
\]

(10)

It is up to the analyst if predictions using equation (7) are carried out using point estimates of \( \theta \) (e.g. maximum likelihood (MLE) or maximum a posteriori (MAP)) or if a fully Bayesian approach is followed by averaging over the posterior \( p(\theta_c, \theta_c)f | D_N \). The latter has the added advantage of quantifying the uncertainty introduced due the finite training data \( N \). We pursue the former in the following as it is computationally more efficient.

3.1.1 Maximizing the posterior

Our objective is to find \( \theta^* = [\theta_c^*, \theta_c^*f] \) which maximizes the posterior given in equation (10), i.e.

\[
[\theta_c^*, \theta_c^*f] = \arg \max_{\theta_c, \theta_c f} e^{L(D_N|\theta_c, \theta_c f)}p(\theta_c, \theta_c f)
= \arg \max_{\theta_c, \theta_c f} \left( L(D_N|\theta_c, \theta_c f) + \log p(\theta_c, \theta_c f) \right)
= \arg \max_{\theta_c, \theta_c f} \left( \sum_{i=1}^{N} \log \tilde{p}(U_f(i)|\lambda_f(i), \theta_c, \theta_c f) + \log p(\theta_c, \theta_c f) \right)
\]

(11)

The main difficulty in this optimization problem arises from the log-likelihood term which involves the log of an analytically intractable integral with respect to \( \lambda_c \) since

\[
L_i(\theta_c, \theta_c f) = \log \tilde{p}(U_f(i)|\lambda_f(i), \theta_c, \theta_c f)
= \log \int p_c(U_c(i)|\lambda_c(i), \theta_c f)p_c(\lambda_c(i)|\lambda_f(i), \theta_c) d\lambda_c(i),
\]

(12)
where we note that an independent copy of $\lambda_c^{(i)}$ pertains to each data point $i$. Due to this integration, typical deterministic optimization algorithms are not applicable.

However, as $\lambda_c$ appears as a latent variable, we may resort to the well-known Expectation-Maximization algorithm [6]. Using Jensen’s inequality, we establish a lower bound on every single term $L_i$ of the sum in the log-likelihood $L(D_N|\theta_c, \theta_{cf})$ by employing an arbitrary density $q_i(\lambda_c^{(i)})$ (different for each sample $i$) as

$$
L_i(\theta_c, \theta_{cf}) = \log \tilde{p}(U_f^{(i)}|\lambda_c^{(i)}, \theta_c, \theta_{cf})
$$

$$
= \log \int p_{cf}(U_f^{(i)}|U_c(\lambda_c^{(i)}), \theta_{cf})p_c(\lambda_c^{(i)}|\lambda_f^{(i)}, \theta_c)d\lambda_c^{(i)}
$$

$$
= \log \int q_i(\lambda_c^{(i)}) \frac{p_{cf}(U_f^{(i)}|U_c(\lambda_c^{(i)}), \theta_{cf})p_c(\lambda_c^{(i)}|\lambda_f^{(i)}, \theta_c)}{q_i(\lambda_c^{(i)})} d\lambda_c^{(i)}
$$

$$
\geq \int q_i(\lambda_c^{(i)}) \log \left( \frac{p_{cf}(U_f^{(i)}|U_c(\lambda_c^{(i)}), \theta_{cf})p_c(\lambda_c^{(i)}|\lambda_f^{(i)}, \theta_c)}{q_i(\lambda_c^{(i)})} \right) d\lambda_c^{(i)} \text{(Jensen)}
$$

$$
= \mathcal{F}_i(q_i; \theta_c, \theta_{cf}),
$$

(13)

Hence, the log-posterior in (10) can be lower bounded by

$$
\log p(\theta_c, \theta_{cf}|D_N) = \log L(D_N|\theta_c, \theta_{cf}) + \log p(\theta_c, \theta_{cf})
$$

$$
= \sum_{i=1}^N \log L_i(\theta_c, \theta_{cf}) + \log p(\theta_c, \theta_{cf})
$$

$$
\geq \sum_{i=1}^N \mathcal{F}_i(q_i; \theta_c, \theta_{cf}) + \log p(\theta_c, \theta_{cf})
$$

$$
= \mathcal{F} \left( \{q_i\}_{i=1}^N, \theta_c, \theta_{cf} \right) + \log p(\theta_c, \theta_{cf}).
$$

(14)

The introduction of the auxiliary densities $q_i$ suggests the following recursive procedure [4] for maximizing the log-posterior:

**E-step:** Given some $\theta_c^{(t)}, \theta_{cf}^{(t)}$ in iteration $t$, find the auxiliary densities $q_i^{(t+1)}$ that maximize $\mathcal{F}$,

**M-step:** Given $q_i^{(t+1)}$, find the parameters $\theta_c^{(t+1)}, \theta_{cf}^{(t+1)}$ that maximize $\mathcal{F}$.

It can be readily shown that the optimal $q_i$ is given by

$$
q_i(\lambda_c^{(i)}) \propto p_{cf}(U_f^{(i)}|U_c(\lambda_c^{(i)}), \theta_{cf})p_c(\lambda_c^{(i)}|\lambda_f^{(i)}, \theta_c)
$$

(15)

with which the inequality in (13) becomes an equality. In fact, both E- and M-steps can be relaxed to find suboptimal $q_i^{(t)}, \theta_c^{(t)}, \theta_{cf}^{(t)}$, which enables the application of approximate schemes such as e.g. Variational Inference (VI) [24, 9]. For the M-step, we may resort to any (stochastic) optimization algorithm [17] or, on occasion, closed-form updates might also be feasible.
Random conductivity field $\lambda_f$ for $\phi_{hi} = 0.35$, $l = 0.0976$ and $c = \frac{\lambda_{hi}}{\lambda_{lo}} = 100$

![Random microstructure samples and corresponding fine-grained model output](image)

Full-order model output $U_f$

![Full-order model output](image)

Figure 2: Random microstructure samples and corresponding fine-grained model outputs.

4 SAMPLE PROBLEM: 2D STATIONARY HEAT EQUATION

As a sample problem, we consider the 2D stationary heat equation on the unit square $[0, 1]^2$

$$\nabla_x (-\lambda(x, \xi(x)) \nabla_x U(x, \xi(x))) = 0 \tag{16}$$

where $U(x, \xi(x))$ represents the temperature field. For the boundary conditions, we fix the temperature $U$ in the upper left corner (see Figure 2) to $-50$ and prescribe the heat flux $Q(x) = (150 - 30y)$ on the remaining domain boundary $\partial D$.

We consider a binary random medium whose conductivity $\lambda(x)$ can take the values $\lambda_{hi}$ and $\lambda_{lo}$. To define such a field we consider transformations of a zero-mean Gaussian process $\xi(x)$ of the form $[18, 13]$

$$\lambda(x, \xi(x)) = \begin{cases} \lambda_{hi}, & \text{if } \xi(x) > c, \\ \lambda_{lo}, & \text{otherwise} \end{cases} \tag{17}$$

where the thresholding constant $c$ is selected so as to achieve the target volume fraction $\phi_{hi}$ (or equivalently $\phi_{lo} = 1 - \phi_{hi}$) of the material with conductivity $\lambda_{hi}$ (or equivalently of $\lambda_{lo}$). In the following we use an isotropic squared-exponential covariance function for $\xi(x)$ of the form

$$\text{cov}(x_i, x_j) = \exp \left\{ -\frac{|x_i - x_j|^2}{l^2} \right\}. \tag{18}$$

In the following studies, we used values $l \approx 0.01$.

Due to the small correlation length $l$, we discretize the SPDE in Equation (16) using $256 \times 256$ standard quadrilateral finite elements, leading to a linear system of $N_{eq} = \dim(U_f) - 1 =$
257 \times 257 - 1 = 66048 \text{ algebraic equations.} \text{ We choose the discretization mesh of the random process } \lambda(x, \xi(x)) \text{ to coincide with the finite element discretization mesh of the SPDE, i.e. dim(} \lambda_f) = 256 \times 256 = 65536.

Samples } D_N = \{ \lambda_f^{(i)}, U_f^{(i)} \}_{i=1}^N \text{ are readily obtained by simulating realizations of the discretized Gaussian field, transforming them according to (17) and solving the discretized SPDE. Three such samples can be seen in Figure 2.

4.1 \textbf{Model specifications}

We define a coarse model employing } n_{\lambda_c} \text{ quadrilateral elements, the conductivities of which are given by the vector } \lambda_c. \text{ Since these need to be strictly positive, we operate instead on } z_c \text{ defined as}

\[ z_c = \log \lambda_c. \]

(19)

For each element } k = 1, \ldots, n_{\lambda_c} \text{ of the coarse model/mesh, we assume that}

\[ z_{c,k} = \sum_{j=1}^{N_{\text{features}}} \theta_{c,j} \chi_j(\lambda_{f,k}) + \sigma_k Z_k, \quad Z_k \sim N(0, 1), \]

(20)

where } \lambda_{f,k} \text{ is the subset of the vector } \lambda_f \text{ that belongs to coarse element } k \text{ and } \chi_j \text{ some feature functions of the underlying } \lambda_{f,k} \text{ that we specify below. In a more compact form, we can write}

\[ p_c(z_c|\lambda_f, \theta_c, \sigma) = N(z_c|\Phi(\lambda_f)\theta_c, \text{diag}(\sigma^2)), \]

(21)

where } \Phi(\lambda_f) \text{ is an } n_{\lambda_c} \times N_{\text{features}} \text{ design matrix with } \Phi_{kj}(\lambda_f) = \chi_j(\lambda_{f,k}) \text{ and } \text{diag}(\sigma^2) \text{ is a diagonal covariance matrix with components } \sigma_k^2.

For the coarse-to-fine map } p_{cf}, \text{ we postulate the relation}

\[ p_{cf}(U_f|U_c(z_c), \theta_{cf}) = N(U_f|W U_c(z_c), S), \]

(22)

where } \theta_{cf} = (W, S) \text{ are parameters to be learned from the data. The matrix } W \text{ is of dimension } n_f \times n_c \text{ and } S \text{ is a positive definite covariance matrix of size } n_f \times n_f. \text{ To reduce the large amount of free parameters in the model, we fix } W \text{ to express coarse model’s shape functions. Furthermore, we assume that } S = \text{diag}(s) \text{ where } s \text{ is the } n_f \text{-dimensional vector of diagonal entries of the diagonal matrix } S. \text{ The aforementioned expression implies, on average, a linear relation between the fine and coarse model outputs, which is supplemented by the residual Gaussian noise implied by } S. \text{ The latter expresses the uncertainty in predicting the FG output when only the CG solution is available. We note that the relation in Equation (20) (i.e. } \theta_c \text{ and } \sigma^2 \text{) will be adjusted during training so that the model implied in Equation (22) represents the data as good as possible.

From equation (15), we have that

\[ \log q_i^{(t+1)}(z_c^{(i)}) \propto \frac{1}{2} \sum_{k=1}^{N_{\text{el}}} \log \left( (\sigma_k^{(t)})^{-2} \right) - \frac{1}{2} \sum_{k=1}^{N_{\text{el}}} (\sigma_k^{(t)})^{-2} \left( z_c^{(i)} - \Phi(\lambda_f^{(i)})\theta_c^{(t)} \right)^2_k

- \frac{1}{2} \sum_{j=1}^{N_f} \log s_j^{(t)} - \frac{1}{2} \sum_{j=1}^{N_f} s_j^{(t)} \left( T_f^{(t)} - W U_c(z_c^{(i)}) \right)^2_j, \]

(23)
where with $(\cdot)_{i}$ we mean the $i$-th component of the vector in brackets. For use in the M-step, we compute the gradients
\[
\frac{\partial F}{\partial \sigma_k^2} = \frac{N}{2} \sigma_k^2 - \frac{1}{2} \sum_{i=1}^{N} \left( \frac{\mathbf{z}_i^{(i)} - \Phi(\lambda_f^{(i)}) \theta_c^{(i)}}{q_i} \right)^2,
\]
(24)

\[
\nabla_{\theta_c} F = \sum_{i=1}^{N} \left( \Phi^T(z_c^{(i)}) \Sigma^{-1} z_c^{(i)} q_i^{(i)} - \Phi^T(\lambda_f^{(i)}) \Sigma^{-1} \Phi(\lambda_f^{(i)}) \theta_c \right)
\]
(25)

where $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2 \lambda_c)$. We can readily compute the roots to find the optimal $\Sigma, \theta_c$. Furthermore, from $\frac{\partial F}{\partial \sigma_k^2} = 0$ we get
\[
s_j^{(i+1)} = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{T_f^{(i)} - W \Phi(\mathbf{z}_c^{(i)})}{q_i^{(i)}} \right)^2.
\]
(26)

4.2 Feature functions $\chi$

The framework advocated allows for any form and any number of feature functions $\chi_j$ in (20). Naturally such a selection can be guided by physical insight [11]. In practice, the feature functions we are using can be roughly classified into two different groups:

4.2.1 Effective-medium approximations

We consider existing effective-medium approximation formula that can be found in the literature [23] as ingredients for construction of feature functions $\chi_j$ in equation (20). The majority of commonly used such features only retain low-order topological information. Only the following approximations to the effective property $\lambda_{\text{eff}}$ can be used as building blocks for feature functions $\chi$ in the sense that we can transform them nonlinearly such that $\chi(\lambda_f) = f(\lambda_{\text{eff}}(\lambda_f))$. In particular, as we are modeling $\mathbf{z}_c = \log \lambda_c$, we include feature functions of type $\chi(\lambda_f) = \log(\lambda_{\text{eff}}(\lambda_f))$.

Maxwell-Garnett approximation (MGA) The Maxwell-Garnett approximation is assumed to be valid if the microstructure consists of a matrix phase $\lambda_{\text{mat}}$ and spherical inclusions $\lambda_{\text{inc}}$ that are small and dilute enough such that interactions between them can be neglected, another. Moreover, the heat flux far away from any inclusion is assumed to be constant. Under such conditions, the effective conductivity can be approximated in 2D as
\[
\lambda_{\text{eff}} = \lambda_{\text{mat}} + \lambda_{\text{inc}} + \phi_{\text{inc}}(\lambda_{\text{inc}} - \lambda_{\text{mat}}),
\]
where the volume fraction $\phi_i(\lambda_f)$ is given by the fraction of phase $i$ elements in the binary vector $\lambda_f$.

Self-consistent approximation (SCA) The self-consistent approximation (or Bruggeman formula) was originally developed for effective electrical properties of random microstructures. It also considers non-interacting spherical inclusions and follows from the assumption that perturbations of the electric field due to the inclusions average to 0. In 2D, the formula reads
\[
\lambda_{\text{eff}} = \frac{\alpha + \sqrt{\alpha^2 + 4 \lambda_{\text{mat}} \lambda_{\text{inc}}}}{2}, \quad \alpha = \lambda_{\text{mat}} (2 \phi_{\text{mat}} - 1) + \lambda_{\text{inc}} (2 \phi_{\text{inc}} - 1).
\]
(28)
Figure 3: Left: Sample microstructure for $l = 0.078$ and $\phi_{\mathrm{hi}} = 0.7$. The blue line encompasses the convex area of the encircled low conducting phase blob, the two red lines are its maximum extent in $x$- and $y$-direction. The green lines are paths along which we count pixels and compute generalized means in the pixel-cross and straight path mean functions. Right: distance transform (distance to nearest black pixel) of the microstructure.

Note that the SCA exhibits phase inversion symmetry.

**Differential effective medium (DEM)** From a first-order expansion in volume fraction of the effective conductivity in the dilute limit of spherical inclusions, one can deduce the differential equation [23]

\[
(1 - \phi_{\mathrm{inc}}) \frac{d}{d\phi_{\mathrm{inc}}} \lambda_{\mathrm{eff}}(\phi_{\mathrm{inc}}) = 2\lambda_{\mathrm{eff}}(\phi_{\mathrm{inc}}) \frac{\lambda_{\mathrm{inc}} - \lambda_{\mathrm{eff}}(\phi_{\mathrm{inc}})}{\lambda_{\mathrm{inc}} + \lambda_{\mathrm{eff}}(\phi_{\mathrm{inc}})},
\]

which can be integrated to

\[
\left(\frac{\lambda_{\mathrm{inc}} - \lambda_{\mathrm{eff}}}{\lambda_{\mathrm{inc}} - \lambda_{\mathrm{mat}}}\right) \sqrt{\frac{\lambda_{\mathrm{mat}}}{\lambda_{\mathrm{eff}}}} = 1 - \phi_{\mathrm{inc}}.
\]

We solve for $\lambda_{\mathrm{eff}}$ and use it as a feature function $\chi$.

**4.2.2 Morphology-describing features**

Apart from the effective-medium approximations which only take into account the phase conductivities and volume fractions, we wish to have feature functions $\chi_j$ that more thoroughly describe the morphology of the underlying microstructure. Popular members of this class of microstructural features are the two-point correlation, the lineal-path function, the pore-size density or the specific surface to mention only a few of them [23].

We are however free to use any function $\chi_j : (\mathbb{R}_+)^{\dim(x,y)} \to \mathbb{R}$ as a feature, no matter from which field or consideration it may originate. We thus make use of existing image processing features [15] as well as novel topology-describing functions. Some important examples are
Convex area of connected phase blobs  This feature identifies distinct, connected “blobs” of only high or low conducting phase pixels and computes the area of the convex hull to each blob. We then can e.g. use the mean or maximum value thereof as a feature function.

Blob extent  From an identified phase blob, we can compute its maximum extension in x- and y-directions. One can for instance take the mean or maximum of maximum extension among identified blobs as a feature.

Distance transformation functions  The distance transform of a binary image assigns a number to each pixel \( i \) that is the distance from that pixel \( i \) to the nearest nonzero pixel in the binary image, see right part of Figure 3. One can use either phase to correspond to nonzero in the binary image as well as different distance metrics. As a feature, one can e.g. take the mean or maximum of the distance transformed image.

Pixel-cross function  This feature counts the number of high or low conducting pixels one has to cross going on a straight line from boundary to boundary in x- or y-direction. One can again take e.g. mean, maximum or minimum values as the feature function outputs.

Straight path mean function  A further refinement of the latter function is to take generalized means instead of numbers of crossed pixels along straight lines from boundary to boundary. In particular, we use harmonic, geometric and arithmetic means as features.

4.3 Sparsity priors

It is clear from the previous discussion that the number of feature functions is practically limitless. The more such \( \chi_j \) one introduces, the more unknown parameters \( \theta_{n,j} \) must be learned. From the modeling point of view, while ML estimates can always be found, it is desirable to have as clear of a distinction as possible between relevant and irrelevant features that could provide further insight as well being able to do so with the fewest possible training data available. For that purpose we advocate the use of sparsity-enforcing priors in \( \theta_c \) [1, 7, 19]. From a statistical perspective, this is also motivated by the bias-variance-tradeoff. Model prediction accuracy is adversely affected by two factors: One is noise in the training data (variance), the other is due to overly simple model assumptions (bias). Maximum-likelihood estimates of model parameters tend to have low bias but high variance, i.e. they accurately predict the training data but generalize poorly. To address this issue, a common Bayesian approach to control model complexity is the use of priors, which is the equivalent to regularization in frequentist formulations. A particularly appealing family of prior distributions is the Laplacian (or LASSO regression [21]), as it sets redundant or unimportant predictors to exactly 0, thereby simplifying interpretation.

In particular, we use a prior on the coefficients \( \theta_c \) of the form

\[
\log p(\theta_c) = \log \frac{\sqrt{\gamma}}{2} - \sqrt{\gamma} \sum_{i=1}^{N_{\theta_c}} |\theta_{c,i}|, \tag{31}
\]

with a hyper-parameter \( \gamma \) which can be set by either applying some cross-validation scheme or more efficiently by minimization of Stein’s unbiased risk estimate [20, 27]. A straightforward implementation of this prior is described in [7].
Predictive error vs. number of training data

Figure 4: Relative squared prediction error \( \frac{\left\langle (U_{f,\text{true}} - \mathbf{U}_f)^2 \right\rangle}{\text{var}(U_f)} \) versus number training data samples \( N \) and different CG mesh sizes \( N_{el,c} = \dim(\lambda_c) \). We set \( \phi_{hi} = 0.2, l = 0.0781 \) and \( c = 10 \).

5 RESULTS

5.1 Required training data

We use a total number of 306 feature functions \( \chi_f \) and a Laplacian prior (31) with a hyperparameter \( \gamma \) we find by cross-validation. We set the length scale parameter to \( l = 0.0781 \) and the expected volume fraction of the high conducting phase to \( \phi_{hi} = 0.2 \). For the contrast, we take \( c = \frac{\lambda_{hi}}{\lambda_{lo}} = 10 \), where we set \( \lambda_{lo} = 1 \). The full- and reduced-order models are both computed on regular quadrilateral finite element meshes of size \( 256 \times 256 \) for the FG and \( 2 \times 2, 4 \times 4 \) and \( 8 \times 8 \) for the CG.

Our goal is to measure the predictive capabilities of the described model. To that end, we compute the mean squared distance

\[
d^2 = \frac{1}{N_{\text{test}} n_f} \sum_{j=1}^{n_f} \sum_{i=1}^{N_{\text{test}}} \left( U_{f,\text{true},j}^{(i)} - \left\langle U_{f,j}^{(i)} \right\rangle_{\mathcal{P}(U^{(i)}_f | \lambda_c^{(i)}, \sigma^*)} \right)^2
\]

(32)

of the predictive mean \( \left\langle U_{f,j}^{(i)} \right\rangle_{\mathcal{P}(U^{(i)}_f | \lambda_c^{(i)}, \sigma^*)} \) to the true FG output \( U_f^{(i)} \) on a test set of \( N_{\text{test}} = 256 \) samples. Predictions are carried out by drawing 10,000 samples of \( \lambda_c \) from the learned distribution \( p_c(\log \lambda_c | \lambda_f, \theta_c^*, \sigma) = \mathcal{N}(\log(\lambda_c) | \Phi(\lambda_f)\theta_c^*, \text{diag}(\sigma^2)) \), solving the coarse model \( U_c^{(i)} = U_c(\lambda_c^{(i)}) \) and drawing a sample from \( p_{U_f}(U_f^{(i)} | U_c^{(i)}, \theta_c^f) = \mathcal{N}(U_f^{(i)} | \mathbf{W}U_c^{(i)}, \mathbf{S}^*) \) for every test data point. Monte Carlo noise is small enough to be neglected.

As a reference value for the computed error \( d^2 \), we compute the mean variance of the FG output

\[
\text{var}(U_f) = \frac{1}{n_f} \sum_{i=1}^{n_f} \left( \left\langle U_{f,1}^{2} \right\rangle - \left\langle U_{f,1} \right\rangle^2 \right)
\]

(33)
where expectation values $\langle . \rangle$ are estimated on a set of 1024 FG samples such that errors due to Monte Carlo can be neglected. Figure 4 shows the relative squared prediction error $\frac{\hat{d}^2}{\text{var}(\hat{d})}$ in dependence of the number of training data samples for different coarse mesh sizes $\dim(\lambda_c) = 2 \times 2, 4 \times 4$ and $8 \times 8$. We observe that the predictive error converges to a finite value already for relatively few training data. This is due to the inevitable information loss during the coarsening process $\lambda_f \rightarrow \lambda_c$ as well as finite element discretization errors. In accordance with that, we see that the error drops with the dimension of the coarse mesh, $\dim(\lambda_c)$.

5.2 Activated microstructural feature functions

For the same volume fraction and microstructural length scale parameter as above ($\phi_{hl} = 0.2, l = 0.0781$), a varying contrast of $c = \{10, 20, 50\}$, a coarse model dimension of $\dim(\lambda_c) = 4 \times 4$ and a training set of $N = 128$ samples, we find the optimal $\theta^*_c$ shown in Figure 5. Due to the application of a sparsity enforcing prior as described in section 4.3, we observe that most components of $\theta^*_c$ are exactly 0. Comparability between different feature functions can be ensured by standardization or normalization of feature function outputs on the training data. For all three contrast values, we see that the three most important feature functions are given by the maximum convex area of blobs of conductivity $\lambda_{hi}$, the maximum log of the geometric mean of conductivities along a straight line from boundary to boundary in $y$-direction and the log of the self-consistent effective medium approximation as described above. The maximum convex area feature returns the largest convex area of all blobs found within a coarse element.

In Figure 5, it is observed that with increasing contrast, the coefficient $\theta^*_{c,j}$ belonging to the log self-consistent approximation is decreasing in contrast to increasing values of $\theta^*_{c,j}$'s cor-
responding the max. high conducting convex area and the max. log geometric mean along y-direction. We believe that this is because, the higher the contrast $c$ is, the more the exact geometry and connectedness of the microstructure plays a role for predicting the effective properties. As the SCA only considers the volume fraction and the conductivities of both phases, it disregards such information.

### 5.3 Learned effective property $\lambda_c$

Figure 6 shows three test samples of $\phi_{hi} = 0.2, l = 0.0781$ and $c = 10$ (top row) along with the mode $\exp(\Phi(\lambda_f)\theta^*_c)$ of the learned distribution of the effective conductivity $p_c(\lambda_c|\lambda_f, \theta^*_c)$, which is connected to the mean by $\langle \lambda_c \rangle_{p_c} = \exp(\Phi(\lambda_f)\theta^*_c + \frac{1}{2}\sigma^2)$. We emphasize again that the latent variable $\lambda_c$ is not a lower-dimensional compression of $\lambda_f$ with the objective of most accurately reconstructing $\lambda_f$, but of providing good predictions of $U_f(\lambda_f)$. Even though Figure 6 gives the impression of a simple local averaging relation between $\lambda_f$ and $\lambda_c$, this is not always be the case. In particular, $p_c(\lambda_c|\lambda_f, \theta^*_c)$ was found to have non-vanishing probability mass for $\lambda_{c,i} \prec \lambda_{hi}$ or $\lambda_{c,i} \succ \lambda_{hi}$, especially for more general models where the coarse-to-fine mapping (22) is not fixed to be the shape function interpolation $W$. 

![Finescale conductivity field $\lambda_f$](image1.png)  
![Coarse-grained conductivity mode $\exp(\Phi(\theta^*_c)$](image2.png)

**Figure 6:** Three test samples and the corresponding mode $\exp(\Phi(\lambda_f)\theta^*_c)$ of $p_c$. 


5.4 Predictive uncertainty

![Variance parameters $\sigma^2$ and $s^*$ of $p_c$ and $p_{cf}$, respectively.](image)

Figure 7: Variance parameters $\sigma^2$ and $s^*$ of $p_c$ and $p_{cf}$, respectively.

The predictive uncertainty is composed of the uncertainty in having an accurate encoding of $\lambda_f$ in $\lambda_c$ which is described by $\sigma^2_k$ in $p_c$ (21), as well as the uncertainty in the reconstruction process from $U_c$ to $U_f$, which is given by the diagonal covariance $S = \text{diag}(s)$ in $p_{cf}$ (22). Both are shown after training ($\phi_2 = 0.2, l = 0.0781, c = 10$) in Figure 7. For $\sigma^2$, we observe that values in the corner elements always converge to very tight values, whereas some non-corner elements can converge to comparably large values. The exact location of these elements is data dependent. The coarse-to-fine reconstruction variances $s^*$ is depicted on the right column of Figure 7. As expected, we see that the estimated coarse-to-fine reconstruction error is largest in the center of coarse elements i.e. at large distances from the coarse model finite element nodes.
5.5 Predictions

As in section 5.1, for a model with $N_{\text{train}} = 32$ and $\dim(\lambda_c) = 8 \times 8$ and random microstructures with parameter values $\phi_{hi} = 0.2$, $l = 0.0781$, $c = 10$, we consider predictions by sampling from $\ddot{\rho}(U_f | \lambda_f, \theta_{ef}, \theta_c)$ using 10,000 samples. The predictive histogram for the temperature $U_{f,lr}$ in the lower right corner of the domain can be seen in Figure 8. Figure 9 shows a surface plot of the true solution (colored), the predictive mean (blue) $\pm \sigma$ (gray). As can be seen, the true solution $U_f$ is nicely included in $\ddot{\rho}$ everywhere.

Figure 8: Predictive histogram (samples from $\ddot{\rho}(U_{f,lr} | \lambda_f, \theta^*)$) for the temperature $U_{f,lr}$ of the lower right corner of the domain. The true solution $U_{f,lr,\text{true}}$ is nicely captured by the distribution $\ddot{\rho}$.

Figure 9: Predictions over the whole domain on four different test samples. The true solution (colored) lies in between $\pm \sigma$ (grey). The predictive mean (blue) is very close to the true solution.
6 CONCLUSION

In this paper, we described a generative Bayesian model which is capable of giving probabilistic predictions to an expensive fine-grained model (FG) based only on a small finite number of training data and multiple solutions of a fast, but less accurate reduced order model (CG).

In particular, we consider the discretized solution of stochastic PDEs with random coefficients where the FG corresponds to a fine-scale discretization. Naturally, this comes along with a very high-dimensional vector of input uncertainties. The proposed model is capable of extracting the most relevant features of those input uncertainties and gives a mapping to a much lower dimensional space of effective properties (encoding). These lower dimensional effective properties serve as the input to the CG, which solves the PDE on a much coarser scale. The last step consists of a probabilistic reconstruction mapping from the coarse- to the fine-scale solution (decoding).

We demonstrated features and capabilities of the model proposed for a 2D steady-state heat problem, where the fine scale of the conductivity implies (upon discretization) a random input vector of dimension $256 \times 256$ and the solution of a discretized system of equations of comparable size. In combination with a sparsity-enforcing prior, the proposed model identified the most salient features of the fine-scale conductivity field and allowed accurate predictions of the FG response using a CGs of size only $2 \times 2$, $4 \times 4$ and $8 \times 8$. The predictive distribution always included the true FG solution as well as provided uncertainty bounds arising from the information loss taking place during the coarse-graining process.
REFERENCES


PERFORMANCE PREDICTIONS OF COASTAL DEFENCES USING STOCHASTIC DETERIORATION MODELLING

Mehrdad Bahari Mehrabani\(^1\), and Hua-Peng Chen\(^{1*}\)

\(^1\)Department of Engineering Science, University of Greenwich
Chatham Maritime, Kent, ME4 4TB, UK

\(*Corresponding author: h.chen@gre.ac.uk

Keywords: Reliability analysis, Markov model, Grade-based condition assessment, Stochastic deterioration model, Performance prediction.

Abstract. Performance prediction of coastal defence structures becomes more challenging due to uncertainties arising from changing environments and deterioration processes; hence, requiring sophisticated time-variant reliability analysis in order to manage the risk of floods and erosions. This paper proposes a method for assessing failure probability of coastal defences with respect to their present condition grade. A Markov chain model is utilised to predict the transition probabilities between different condition grades of a coastal defence over time with respect to available maintenance plans. The deterioration level is categorised into 5 stages (each stage for an appropriate condition grade) in relation to the initial resistance, and then translated into a probabilistic framework in order to consider in the performance deterioration evaluation. A case study of a sea dyke section, located in Portsmouth, England, is employed to demonstrate the effectiveness of the proposed method. Finally, the time-variant failure probability are illustrated in order to provide a clear understanding of performance evaluation in the future. The results show that the failure probability of the sea dyke associated with overtopping failure mode will increase significantly due to deterioration processes.
1 INTRODUCTION

Coastal defence structures are the first lines to protect low-lying areas and their valuable assets against risks posed by flooding, coastal erosion and landslides. It is necessary to manage the risks effectively to avoid casualties and damages concerning the economic matters. However, coastal defence structures are often located in harsh and changeable environments; therefore, it becomes challenging to assess their future performance level accurately. It is necessary to inspect a coastal flood defence regularly to evaluate its current conditions including damage level. Environment Agency has been recording the inspection data since the 1990s and introduced Condition Assessment Manual in 2006 [1] to assess coastal defence general condition grade in a standard framework.

During last decades, many studies have been carried out on performance predictions of flood defences. Flikweert and Simm [2] have suggested estimating fragility curves for different condition grades defined by Environment Agency [1] as a new approach for performance prediction and risk assessment of flood defences. Simm et al. [3] have developed the mentioned method in 2009 to analyse reliability of flood defences using a condition-grading system. Many studies considered the deteriorating resistance and the uncertainties associated with the operating environment as major challenges to evaluate performance deterioration in the future (e.g. [4, 5]). Gouldby et al. [6] developed a methodology for assessing the flood risk under different deterioration scenarios in the future in a regional-scale model. However, the application of the condition-based system in the most studies is deterministic, and a grade-based stochastic deterioration model is not developed for coastal flood defence structures.

This paper aims to propose a stochastic performance analysis of coastal defence structures using a grade-based Markov model. Crest height degradation due to settlement is also considered for different condition grades. Time-dependent transition probability matrices are estimated by utilising the Markov model with Weibull sojourn times. Finally, the time-dependent failure probability of a sea dyke for its overtopping failure mode is presented for consideration of the future maintenance plans.

2 COASTAL DEFENCE CONDITION GRADING SYSTEM

Efficient management of assets at coastal areas is one of the main objectives for Environment Agency in the UK. Routine inspections of these assets help authorities to have essential information about the likely assets’ performance and condition grade in order to make decision for future. For this purpose, Environment Agency published Condition Assessment Manual in 2006 in order to offer a standard framework for evaluating asset’s condition grade [1]. Table 1 shows the definition of the condition grades defined by Environment Agency.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Very Good</td>
<td>Cosmetic defects that will have no effect on performance.</td>
</tr>
<tr>
<td>2</td>
<td>Good</td>
<td>Minor defects that will not reduce the overall performance of the asset.</td>
</tr>
<tr>
<td>3</td>
<td>Fair</td>
<td>Defects that could reduce performance of the asset.</td>
</tr>
<tr>
<td>4</td>
<td>Poor</td>
<td>Defects that would significantly reduce the performance of the asset. Further investigation needed.</td>
</tr>
<tr>
<td>5</td>
<td>Very poor</td>
<td>Severe defects result in complete performance failure.</td>
</tr>
</tbody>
</table>

Table 1: Condition grades and descriptions adopted by Environment Agency [1].
In 2013, Environment Agency offered a series of asset deterioration curves for different types of coastal defences in order to estimate the residual life of an asset or structure related to with the condition grading system [7]. Deterioration curves predict the condition grade of assets associated with time, relying on historical data and expert judgment. The curves are explicitly focused on condition grades and do not necessarily provide information about structural deteriorations such as settlement. This paper uses the deterioration curves to predict the lifecycle performance of coastal defenses over time.

3 MARKOV MODEL

A Markov model is a memoryless stochastic process that evolves over time or space, and the future of the process is only conditional on the current condition of the system or asset [8]. Let \( X = \{x_1, x_2, ..., x_M\} \) be a set of finite states of a flood defence asset. A Markov process on a flood defence is defined once its transition probability matrix and initial condition state \( X_0(t) \) at time \( t \geq 0 \) are specified, where \( X_0(t) = \{\alpha^t_1, \alpha^t_2, ..., \alpha^t_M\}, \sum \alpha^t_i = 1 \) and \( \alpha^t_i \geq 0 \). A discrete-time Markov property for a stochastic process \( X \) at time \( t > 0 \) is expressed as \[ p_{ij}^{t,t+1} = \Pr(X^{t+1} = x_{j+1} | X^t = x_i) \] (1) where conditional probability \( p_{ij}^{t,t+1} \), with constraints \( p_{ij}^{t,t+1} \geq 0 \) and \( \sum_{j=1}^{M} p_{ij}^{t,t+1} = 1 \), denotes the transition of the asset from condition grade \( i \) to \( j \) during a given period of time. It is assumed that the structure can deteriorate only one state at a time (with consideration of an appropriate transition period).

3.1 Transition probability matrix

As discussed earlier, coastal defence structures are categorised into 5 condition grades, where condition grade 1 denotes good as new and condition grade 5 denotes failure. Hence, if a structure is in condition grade 5 under no-maintenance plan, then cannot transit to a higher (worst) state, and always remains in condition grade 5. The one-step transition probability matrix under do-nothing maintenance plan is expressed here as

\[
P^{t,t+1} = \begin{bmatrix}
1 - p_{12}^{t+1} & p_{12}^{t+1} & 0 & 0 & 0 \\
0 & 1 - p_{23}^{t+1} & p_{23}^{t+1} & 0 & 0 \\
0 & 0 & 1 - p_{34}^{t+1} & p_{34}^{t+1} & 0 \\
0 & 0 & 0 & 1 - p_{45}^{t+1} & p_{45}^{t+1} \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\] (2)

In a Markov process, a state space remains in a particular state for a given length of time, then transits to another state. Let \( \{T_1, T_2, ..., T_M\} \) be random variables denote the sojourn time in states \( \{x_1, x_2, ..., x_M\} \), respectively. Their corresponding probability density function (PDF), cumulative distribution function (CDF), and survival function are denoted \( f_i(t), F_i(t) \) and \( S_i(t) \), respectively. Thus, \( T_{i,i+1} \) is the time from state \( i \) to state \( (i + 1) \). If the deterioration process is in state \( i \) at time \( t \), then the conditional probability that the structure transits to the next state in one time-step is expressed as \[ p_{i,i+1}(t) = \Pr[X^{t+1} = x_{i+1} | X^t = x_i] = \frac{f_{i-1}(t)}{S_{i-1}(t) - S_{i-1}(t-1)} , \quad i = \{1,2,3,4\} \] (3)
where the transition probability \( p_{i,i+1}(t) \) is time-dependent and non-homogenous. Assume that the waiting time \( T_i \) follows Weibull distribution. The CDF \( F(t) \), PDF \( f(t) \), and survival \( S \) function of Weibull distribution are expressed, respectively, as [10]

\[
F_i(t, \theta_i, \beta_i) = 1 - \exp(-\theta_i t)^{\beta_i} , \quad t > 0; \quad (F(t) = 0 \text{ for } t \leq 0)
\]

\[
f_i(t, \theta_i, \beta_i) = \theta_i \beta_i (\theta_i t)^{\beta_i-1} \exp(-\theta_i t)^{\beta_i} , \quad t > 0
\]

\[
S_i(t, \theta_i, \beta_i) = 1 - F_i(t, \theta_i, \beta_i) = \exp(-\theta_i t)^{\beta_i}
\]

where \( \theta_i \) and \( \beta_i \) are scale and shape parameters, respectively. Hence, under the specification of Weibull parameters, the probability of the transition from state \( i \) to state \( (i+1) \) can be estimated. For example the transition probability from state 4 to state 5 at time \( t \) is expressed as

\[
p_{4,5}(t) = \frac{f_{1\rightarrow4}(t)}{S_{1\rightarrow4}(t) - S_{1\rightarrow3}(t)}
\]  

### 3.2 Parameter estimation

The scale \( \theta_i \) and shape \( \beta_i \) parameters are estimated based on historical observations and condition assessments, which are available via Environment Agency publications as discussed in Section 2 (e.g. [7]). Deterioration curves are developed by experts to predict the condition grade of assets associated with time according to their environment, material, deterioration rate and maintenance regime. For example, a deterioration curve for a specific asset describes that the structure remains in condition grade \( i \) for \( T \) years. After selecting an appropriate deterioration curve for the case study, a two-state Markov chain is utilised to estimate the transition probability for condition grade \( i \). The one-step transition matrix is expressed here as

\[
TP_{ii} = \begin{bmatrix}
1 - p_{ii} & p_{ii} \\
0 & 1
\end{bmatrix}
\]  

where \( TP_{ii} \) is a discrete-time and homogenous Markov transition matrix; and \( p_{ii} \) is the probability of being in condition grade \( i \) at time-step (year) one. Transition probability that the coastal defence goes from condition grade \( i \) to next condition grade in \( n \) steps is called \( n \)-step transition probability, and is calculated from the \( n \)th power of the transition matrix \( TP_{ii} \). Let \( \tau \) be the time-step (year) that the structure remains in the same condition grade with probability of \( p_a = 50\% \). From Equation 4(c), gives

\[
\exp(-\theta_i \tau)^{\beta_i} = p_a = p_{ii}^\tau
\]

Also, it was assumed that the survival function at condition grade \( i \) follows a Weibull survival function \( S_i \), as expressed in Equation 4. A non-linear least squares method is suggested to find the scale \( \theta_i \) and shape \( \beta_i \) parameters by minimising the difference between two functions. The scale \( \theta_i \) and shape \( \beta_i \) parameters can be estimated by solving the minimization, together with Equation 5, as

\[
\text{minimise} \sum_i (S_i(t) - p_{ii}^t)^2 ; \quad i = \{1,2,3,4\}
\]  

where \( S_i(t) = \exp(-\theta_i t)^{\beta_i} \) is the Weibull survival probability at time (year) \( t \).
4 SEA DYKE PERFORMANCE ASSESSMENT

The reliability of coastal flood defence structures depends on two main factors, the hydraulic conditions and the structural resistance. The resistance of the structures decreases over time due to the deteriorating processes such as settlement. If \( R(t) \) denotes as the resistance over time, and \( S(t) \) denotes as the loading, then the limit state function \( Z(t) = R(t) - S(t) \) describes the reliability of the system over time. Therefore, the structure fails if the loading is equal or more than the resistance [3]. The probability of failure \( p_f \) can be expressed as

\[
p_f(t) = \Pr[Z(t) \leq 0]
\]

In this paper, reliability analysis is carried out for one major failure mode; the overtopping over the dyke crest. The limit state function for overtopping \( Z_q \) over time \( t \) is given as [4]

\[
Z_q(t) = q_{cr} - \chi q(t)
\]

where \( q_{cr} \) is a predefined mean overtopping discharge; \( \chi q \) is model uncertainty coefficient associated with overtopping parameters; \( q(t) \) is average overtopping discharge over time and can be calculated by using Eurotop manual [11], given here as

\[
\frac{q(t)}{\sqrt{g \cdot H_{m0}^3}} = \frac{B_1}{\sqrt{\tan \alpha}} \cdot \gamma_f \cdot \varepsilon_{m-1,0} \cdot \exp\left(-B_2 \cdot \varepsilon_{m-1,0} \cdot \gamma_b \cdot \gamma_f \cdot \gamma_{\beta} \cdot H_{m0}\right)
\]

with a maximum of:

\[
\frac{q(t)}{\sqrt{g \cdot H_{m0}^3}} = B_3 \cdot \exp(-B_4 \cdot \frac{R_{ci}(t)}{\gamma_f \gamma_{\beta} H_{m0}})
\]

where \( B_1 = 0.067, B_2 = 4.30, B_3 = 0.20, \) and \( B_4 = 2.30 \) are empirical coefficients for deterministic uses, and normally distributed stochastic parameters with means of \( B_2 = 4.75 \) and \( B_4 = 2.60 \) and standard deviations of 0.50 and 0.35, respectively, are taken for probabilistic calculations; \( H_{m0} \) is significant wave height at the toe of structure; \( g \) is the acceleration due to the gravity; \( \varepsilon_{m-1,0} \) is breaker parameter; \( \gamma_b, \gamma_f \) and \( \gamma_{\beta} \) are correction factors for berm, roughness and oblique wave attack, respectively; \( \alpha \) is the angle between overall structure slope and horizontal line; and \( R_{ci}(t) \) is crest freeboard (over still water level) of the structure at time \( t \) given is in condition grade \( i \in \{1,2,3,4,5\} \). The future freeboard \( R_{ci}(t) \) at time \( t \) is described here as

\[
R_{ci}(t) = R_c(0) - \Delta L_{Z,i}(t)
\]

where \( R_c(0) \) is initial freeboard at initial time; and \( \Delta L_{Z,i}(t) \) is the deterioration of crest level (Z direction) associated with certain condition grade at time \( t \). In order to characterise the deterioration level \( \Delta L_{Z,i} \) regarding the initial freeboard, the table below is suggested (Table 2). Table 2 suggests the variation of the lifecycle phases according to condition grade scheme based on the crest level elevation loss.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Crest level loss, ( \Delta L_{Z,i} ), (m)</th>
<th>Distribution of deterioration intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00 ≤ ( \Delta L_{Z,i} ) &lt; 0.05</td>
<td>Lognormal</td>
</tr>
<tr>
<td>2</td>
<td>0.05 ≤ ( \Delta L_{Z,i} ) &lt; 0.10</td>
<td>Normal</td>
</tr>
<tr>
<td>3</td>
<td>0.10 ≤ ( \Delta L_{Z,i} ) &lt; 0.20</td>
<td>Normal</td>
</tr>
<tr>
<td>4</td>
<td>0.20 ≤ ( \Delta L_{Z,i} ) &lt; 0.40</td>
<td>Normal</td>
</tr>
<tr>
<td>5</td>
<td>0.40 ≤ ( \Delta L_{Z,i} )</td>
<td>Lognormal</td>
</tr>
</tbody>
</table>

Table 2: Suggested crest level loss (Z direction) deterioration for a sea dyke related with condition grade system.
The crest level loss is defined based on the erosion at a specific condition grade, and it is available in Quantitative Assessment Methods [12]. To translate the condition grades into probabilistic parameters, it is assumed that erosion over time is linear and the deterioration intensity is normally distributed. The distributions are normal except the first and last condition grades, which are assumed to be lognormal, as the dyke will not improve due to deterioration (condition grade 1), and more deterioration considers as the functional failure (condition grade 5).

5 NUMERICAL EXAMPLE

A case study for a simplified model of earth sea dyke (see Figure 1) at Portsmouth, England described by Taylor et al. [13] is employed to demonstrate the applicability of the proposed method for reliability assessment of coastal defence structures using condition-grading system. The structure has a crest height of 3.60 mOD, a seaside slope of 1:8, and landside slope of 1:5. The structure has a bermed slope and it is protected by pitched stones. The earth dyke rests on a layer of impermeable clay soils, and below the clay layer is 5.00 m of water conductive sand layer overlaying impervious bedrock.

The extreme water level height during storm surges and high tides is at a level of 1.50 mOD. Significant wave height in deep water is 2.30 m with a return period of 1000 years. The critical
design value of the wave overtopping discharge is considered 2 l/s/m. It is assumed that the structure is in good condition with the initial condition vector of $X_0(0) = (1,0,0,0,0)$. Figure 2 shows appropriate deterioration curves for this case study under medium deterioration rate, and subjected to three different maintenance regimes. Medium maintenance regime means regular maintenance including minor repair to sea dyke for surface cracking, rutting and erosion. High maintenance regime is similar to medium maintenance regime but with increased frequency and more stringent criteria for repair [7].

<table>
<thead>
<tr>
<th>Grade</th>
<th>$\tau$</th>
<th>$\beta_i$</th>
<th>$\theta_i$</th>
<th>SSE</th>
<th>$\tau$</th>
<th>$\beta_i$</th>
<th>$\theta_i$</th>
<th>SSE</th>
<th>$\tau$</th>
<th>$\beta_i$</th>
<th>$\theta_i$</th>
<th>SSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1→2</td>
<td>4</td>
<td>3.061</td>
<td>0.221</td>
<td>2.1e-7</td>
<td>7</td>
<td>3.061</td>
<td>0.221</td>
<td>2.1e-7</td>
<td>7</td>
<td>3.061</td>
<td>0.221</td>
<td>2.1e-7</td>
</tr>
<tr>
<td>2→3</td>
<td>6</td>
<td>4.814</td>
<td>0.154</td>
<td>2.4e-7</td>
<td>12</td>
<td>7.916</td>
<td>0.086</td>
<td>1.2e-7</td>
<td>12</td>
<td>7.916</td>
<td>0.086</td>
<td>1.2e-7</td>
</tr>
<tr>
<td>3→4</td>
<td>7</td>
<td>4.734</td>
<td>0.132</td>
<td>8.8e-8</td>
<td>11</td>
<td>4.152</td>
<td>0.083</td>
<td>1.8e-7</td>
<td>11</td>
<td>4.152</td>
<td>0.083</td>
<td>1.8e-7</td>
</tr>
<tr>
<td>4→5</td>
<td>4</td>
<td>3.061</td>
<td>0.221</td>
<td>2.1e-7</td>
<td>7</td>
<td>7.032</td>
<td>0.135</td>
<td>1.5e-7</td>
<td>13</td>
<td>5.774</td>
<td>0.072</td>
<td>1.8e-7</td>
</tr>
</tbody>
</table>

Table 3: Parameter estimation for three different scenarios. $\tau$: years after initial date that the sea dyke is in the same condition grade with 50% probability. SSE: sum of squares due to error (Goodness-of-Fit test).

The parameters of the Weibull function are estimated using Equations 7 and 8. Trust-Region algorithm is utilised to solve the minimisation function to estimate the parameters for different condition grades, and results are presented in Table 3. Figure 3 shows the cumulative survival distributions of the transition process and their changes over time under no maintenance plan. For example, the sea dyke at age of 15 is about 6% to be in condition grade 2, about 56% to be in condition grade 3, about 24% to be in condition grade 4, about 13% to be in condition grade 5, and less than 1% to be in condition grade 1. The same procedure is considered to estimate cumulative survival functions for sea dyke under medium and high maintenance regimes. Then, the time-dependent transition probabilities are generated using Equation 3.

![Figure 3](image-url)  
Figure 3: Survival functions of cumulative waiting times in different condition grades (CG) subjected to no maintenance regime.

Figure 4 shows the probability density distribution of crest level loss associated with different condition grades according to Table 2. The distributions of the loss are lognormal for
condition grade 1 and 5, and are normal distributions for condition grades 2, 3 and 4. The mean and standard deviation of the distributions are calculated using the proposed method by Ang and Tang [14]. It is assumed that the accuracy of distributions are 90% of all the inspection results, for example, the mean value for the probability distribution of condition grade 3 is 15.0 cm with the standard deviation of 3.20, assuming 10% of all observed values belong to another condition grades.

![Figure 4: Probability density distribution of the sea dyke crest level loss associated with condition grade (CG).](image)

A reliability analysis carried out for overtopping failure mode for the structure of the sea dyke section shown in Figure 1. Reduction factors for berm $\gamma_b$ and slope roughness $\gamma_f$ are considered 0.85 and 0.90, respectively. Figure 5 shows the failure probability caused by wave overtopping over 80 years after the initial construction date under three maintenance regimes. It is demonstrated that the probability of failure increases over time due to deterioration process and crest level settlement. The failure risk of overtopping varies with the type of maintenance regime that is undertaken over time, as expected. For example, the risk of failure at age 30, is high (about 67%) for the sea dyke under no maintenance regime, while it is about 30% and 9% for the sea dyke under moderate and high maintenance regime, respectively.

![Figure 5: Overall probability of failure caused by wave overtopping with different maintenance scenarios.](image)
6 CONCLUSION

This paper proposes a method to assess failure probability for coastal defence structures by utilising a stochastic time-dependent Markov model with respect to condition grading scheme. The probability of condition grades for the sea dyke during the lifetime is determined by using transition probability matrices estimated from Markov process with Weibull model waiting time at each condition grade. Reliability analysis for overtopping failure mode is carried out, and the structural resistance deterioration due to crest level settlement is considered in the limit state equation for each condition grade. The crest level settlement is categorised into 5 lifecycle stages (each lifecycle for an appropriate condition grade) in relation to the initial resistance, and then translated into a probabilistic framework to consider in the performance deterioration evaluation. A case study is utilised to show the applicability of the proposed method, and the following conclusions are noted.

The Markov model is useful to assess the transition of condition grades over time with respect to the maintenance plans in order to predict the future performance evaluation. The Markov model can estimate distinct transition matrix at each time-step with consideration of undertaken maintenance regime. The deterioration of sea dyke (e.g. crest level settlement) can be defined into condition grading system, and then is translated into a probabilistic framework. Estimating deterioration level with respect to condition grading system through the time-dependent transition probabilities gives a more accurate evaluation of the future performance prediction. Overtopping failure probabilities largely depend on sea dyke’s resistance level; hence, failure risks are expected to be increased over time due to deterioration processes. The dyke crest level settlement increases significantly the risk of overtopping failure of the sea dyke in the future.

7 ACKNOWLEDGMENT

The authors are grateful for the financial support from the Institution of Civil Engineers through the R and D Enabling Fund (Project No. 1305). The findings and opinions expressed in this study are those of the authors alone and are unnecessarily the views of the sponsors.

REFERENCES


FLOOD RISK ASSESSMENT OF MASONRY ARCH BRIDGES

Enrico Tubaldi\textsuperscript{1}, Lorenzo Macorini\textsuperscript{1}, and Bassam Izzuddin\textsuperscript{1}

\textsuperscript{1}Imperial College London
Department of Civil Engineering, South Kensington Campus, London, SW7 2AZ.
e-mail: \{e.tubaldi,l.macorini,b.izzuddin\}@imperial.ac.uk

Keywords: Flood risk assessment, Arch Bridges, Scour, Hydrodynamic forces, Buoyancy

Abstract. Floods are one of the most common natural disasters in Europe, responsible for the damage and collapse of many masonry arch bridges built over rivers and canals. The accurate prediction of the safety of these bridges against flood-induced loading is a task of paramount importance for their preservation. This paper describes the framework developed by the authors for the flood risk assessment of masonry arch bridges, accounting for the specific characteristics of the analysed structures, the most critical types of loading associated with floods, and the various sources of uncertainty relevant to the problem. The proposed framework combines the results of flood hazard analysis and of structural vulnerability analysis to obtain the flood risk estimate. A case study consisting of a three-span bridge under scour is considered to illustrate the application of the proposed framework and to show the capabilities of the advanced modelling technique developed for evaluating the effects of flood actions on masonry arch bridges.
1 INTRODUCTION

Masonry arch bridges still constitute a significant portion of the bridge stock in Europe [1]-[3]. These structures have proven to be very durable, with a service life extending well beyond the design life of modern bridges. However, the combined effects of ageing/deterioration, scarce maintenance as well as the increase of traffic loading and natural/man-made hazards have resulted in a significant growth of their risk of failure. In particular, many masonry arch bridges are built on over rivers and canals, which make them very vulnerable to flood effects. These include [4]: a) hydrodynamic pressure on the submerged surfaces exerted by the water and floating debris, b) buoyant forces on the submerged components, and c) scour at the footings of piers and abutments, which is the most common cause of collapse due to the high vulnerability of arches to foundation settlements. Despite the high rate of failure of masonry arch bridges due to floods, studies on this topic are rather scarce, and cover only specific aspects of the problem [5]-[9]. To account for the joint effect of flood-induced actions on bridges, a probabilistic approach is required given the randomness in the occurrence and magnitude of floods, and the uncertainty in the models employed to predict flood effects.

This paper illustrates the framework developed for the flood risk assessment of masonry arch bridges. This framework accounts for the specific characteristics of the analyzed structures, the most critical types of loading associated with floods, and the various sources of uncertainty relevant to the problem. It aims to overcome current approaches for flood risk assessment which consider only a design flood event with a given return period to estimate the bridge safety [10].

In the first part of the paper, the various steps of the framework and the inherent assumptions are explained, together with the extension to account for the temporal accumulation of scour. The application of the framework to masonry arch bridges also involves the development of accurate and efficient computational modelling strategies to evaluate the effect of flooding. In this regard, an accurate three-dimensional (3D) mesoscale representation [11] of masonry arch bridges has been developed at Imperial College [12]-[16]. The proposed modelling approach allows for a realistic description of the non-homogeneous components of masonry arch bridges and the three-dimensional behaviour induced by the flood-induced actions. In the final part of the paper, a realistic case study is considered to show the application of the proposed framework to the risk assessment of masonry arch bridges under scour, which is the most critical action due to floods.

2 FRAMEWORK FOR FLOOD RISK ASSESSMENT

2.1 Flood-induced actions

Water flow exerts different actions on arch bridges. First of all, it induces a significant hydrodynamic pressure on the submerged surfaces, resulting in horizontal forces which may become very high when the deck is also submerged [7],[8]. Foundation capacities decrease when the water level increases due to the increased eccentricity of the vertical loads. Buoyant forces reduce the effective unit weights of submerged components, thus decreasing the load-carrying capacity of the bridge which strongly depends on the compressive forces within the arch due to the self-weight [5]. Large floating lumps/debris can impact the bridge causing local damage which can jeopardize the bridge integrity. Water flow results also in scour at the footings of piers and abutments which is the most common cause of collapse due to the high vulnerability of arches to foundation settlements [17][18]. The factors causing scour to develop are complex and differ according to the type of structure. Scour may occur as a result of
natural changes of flow in the channel, as part of longer-term morphological changes to the river, or as a result of human activities, such as the building of structures in the channel or the dredging of material from the bed. Undermining the foundation can induce several negative effects which may jeopardize the integrity of arch bridge components or even global stability. These include (i) rupture of the foundation plinth due to the loss of support [19], (ii) failure of the foundation-soil system [20], (iii) cracking and mechanism formation due to angular rotation, (iv) subsidence, and/or (v) shift of the bridge pier’s [21]. Finally, it is worth noting that the low clearance offered by arch bridges makes them very susceptible to debris accumulation, and this may increase both scour development and hydrodynamic forces [22].

The evaluation of the performance of bridges against floods must account for the random nature of flood-induced actions. This entails the development of a probabilistic hazard model capable of describing the frequency of exceedance of the intensity of the single actions as well as their correlation. The model should also account for the fact that flood-induced actions on bridges may often concurrently interact with each other, e.g. the accumulation of debris against bridges might significantly affect the bridge hydraulics, the hydrodynamic forces and the scour at the bridge foundations, as well as the development of a scour hole may increase the hydrodynamic pressure on the bridge.

2.2 Framework description

The proposed framework is very similar in concept to other frameworks that have been recently formalised for evaluating the risk of structures exposed to different natural and man-made scenarios including blast, fire, tsunami and wind scenarios [23]-[25]. The main aim of these methodologies is to provide a general procedure for the evaluation of the performance of structural systems in terms of decision variables (DVs) such as risk of collapse, fatalities, repair costs, and loss of function, while accounting for all the possible sources of uncertainty that characterise the problem at hand. In this way, the structural risk can be efficiently defined in terms of variables of interest for stakeholders, decision makers, and the society.

The flood risk assessment framework disaggregates the performance assessment procedure for bridge structures subject to the flood hazard into elementary tasks that are carried out in sequence. In particular, the risk assessment is disaggregated into: (1) hydrologic analysis, (2) hydraulic analysis, (3) interaction analysis, (4) structural analysis, (5) damage analysis, and (6) loss analysis (Fig.1). The following expression is used to integrate the results of these tasks:

\[
\lambda_{DV} = \int \int \int G_{DV|IM} \left( dq \right) \cdot f_{DM|EDP} \left( dq \right) \cdot f_{EDP|IP} \left( dq \right) \cdot f_{IP|H,Q} \left( dq \right)
\]

where \( G(\cdot) = \) complementary cumulative distribution function, and \( G(\cdot|\cdot) = \) conditional complementary cumulative distribution function; \( f(\cdot) = \) probability density function, and \( f(\cdot|\cdot) = \) conditional probability density function; \( IM = \) vector of intensity measures (i.e., parameters characterizing the environmental hazard); \( SP = \) vector of structural parameters (i.e., parameters describing the relevant properties of the structural system and non-environmental actions); \( IP = \) vector of interaction parameters (i.e., parameters describing the interaction phenomena between the environment and the structure); \( EDP = \) vector of engineering demand parameters (i.e., parameters describing the structural response for the performance evaluation); \( DM = \) vector of damage measures (i.e., parameters describing the physical damage to the structure), \( DV = \) vector of decision variables.
It is worth noting that this decomposition, which is a statement of the Total Probability Theorem, is made possible through the fundamental Markovian assumption that the result of each analysis (e.g. $DV$), conditional on the result of the previous step (e.g. $DM$), is independent from the other preceding steps of the analysis (i.e., $EDP$, $IP$, $H$, $IM$). Another assumption that needs to be introduced is that of stationarity, i.e. the conditional probabilities are the same and the relevant conditional distributions shown in these formulas are independent and identically distributed for successive flood events. This implicitly assumes that the system does not deteriorate/evolve, and that it is instantaneously restored to its original state after each flooding event. In this regard, it should be observed that while the hydrodynamic and buoyancy actions associated with the flood event can be assumed to renew at each flood occurrence by following the same conditional probability distribution, the scouring can be cumulative over the long term, and it may result by the succession of events of flooding of different intensity occurring over time.

The combination of the first steps (1)-(3) of the framework allows the characterisation of the flood hazard in terms of the mean annual frequency (MAF) of exceedance for the IPs that are used as input for the structural analysis such as the hydrodynamic pressure, the buoyancy and the extent of the scour. Under the aforementioned assumption, it can be shown that also the IPs follow a Poisson distribution with a mean rate given by the combination of the results of steps (1)-(3). The combination of the results of steps (4)-(5) provides information on the structural vulnerability. This can be expressed in terms of fragility curves, which yield the probability of exceeding given damage or limit states vs. the values assumed by the IPs.

### 2.3 Incorporation of memory effects in scour risk assessment

As already pointed out in the previous section, scouring can accumulate during successive events, and it may take many events before the maximum scour depth is fully developed. The temporal evolution of scour depends on whether it occurs under so called clear water or live bed conditions. The former case occurs when the velocity of the flow upstream of the bridge is lower than the critical velocity necessary to erode the river bed, whereas the latter when it is higher [26]. Thus, in clear water conditions, erosion processes occur only in proximity of bridge foundations whereas in live-bed conditions sediment transport occurs in the undisturbed channel upstream of the bridge. In live-bed conditions the scour hole is continually supplied with sediment by the approach flow. In the clear-water scenario, the equilibrium scour depth is approached asymptotically when the flow is no longer capable of removing bed...
sediment from the hole. In the live-bed scenario, an equilibrium is attained when, over a period of time, the average amount of sediment transported into the scour hole by the approach flow is equal to the average amount of sediment removed from the scour hole due to erosion processes. Generally, scour holes develop more slowly under clear water conditions, for which the equilibrium scour depth is reached asymptotically. Under live-bed conditions, equilibrium is reached much faster. Contrary to the clear-water case, in live-bed conditions the scour hole tends to be refilled with sediment as the flood recedes. More research is however needed to characterize this phenomenon.

In order to illustrate these concepts, Fig. 2 shows a sample of a flood hydrograph of a channel and the corresponding evolution of the scour depth at the foundation of a bridge pier. The flow in the channel is such that live bed conditions are attained only during the time interval $[t_3,t_4]$, while clear water conditions are observed during many time intervals. For example, during $[t_0,t_1]$, the scour depth increases under clear water conditions, at a rate which is higher during the ascending part of the flood hydrograph and very low during the receding limb. In the time interval $[t_1,t_2]$, the scour depth does not increase because the discharge is below the level $q_0$ necessary to induce scour, in $[t_2,t_3]$ it increases again under clear water conditions. Finally, in $[t_3,t_4]$ the scour evolves under live-bed conditions, and the scour hole is partially refilled with sediment from upstream the bridge.

![Flow chart illustrating the probabilistic framework for scour risk assessment.](image)

While the process described above is only qualitative, it suggests a different analytical treatment of the problem for clear water and live-bed scour. For floods of high intensity, presumably inducing live-bed scouring conditions, it is plausible to assume that the equilibrium scour depth is attained almost instantaneously, followed by a recession stage with refilling. Thus, the probability of exceeding the critical scour depth depends only on the flow discharge (and the relevant values of the hydraulic parameters) of that particular flood event rather than on the history of scour evolution. For floods of lower intensity and hence presumably inducing clear water scouring conditions, a single flood event is unlikely to induce a critical scour depth, but it may only deepen the scour hole. The repeated occurrence of multiple floods may however still result in the attainment of the critical scour depth, with an associated probability which increases by increasing the number of floods, their intensity and duration [27]. Thus, in this case it is very important to account for memory effects.

One way to account for memory effects in evaluating the probabilistic distribution of the scour depth is to assume that the susceptibility to increase (or eventually decrease) scour during one event, given the features of the flood, depends on the scour level accumulated until the occurrence of the event itself, rather than on the entire flood and scour history. This leads to a Markov-process description [28] of the scour evolution process.
Hereinafter, the methodology developed by the authors [29] to account for memory effects is illustrated synthetically. Let \( P^i_n(t) \) denote the number of floods occurring in [0,\( t \)] of intensity higher than a minimum significant threshold, and \( P^0_i \) the probability mass function of the scour depth at year 0. A set of discrete scour states of increasing scour depth is assumed, and the probability of the scour depth \( S \) being equal to \( s_j \) at time \( t \) can be expressed as:

\[
P_j(t) = \sum_{n=0}^{\infty} \Phi^n_i P^0_i \cdot P[N = n]
\]

where \( \Phi = [\phi_{ij}] \) is the state transition matrix, expressing the probability of passing from scour state \( i \) to scour state \( j \) given the occurrence of an event of a given significant magnitude, and \( P[N = n] \) is the probability of having a number \( n \) of events of intensity higher than the minimum threshold. Obviously, \( \Phi \) must be derived by accounting for the randomness in the flood intensity and the actual duration of each flood. Given an hazard model, Monte Carlo simulation can be employed to estimate \( \Phi \). Once \( \Phi \) and \( P^0_i \) are known, the first passage probability of exceedance for the scour depth can be evaluated.

3 MESCOSCALE PARTITIONED MODELLING OF MASONRY ARCH BRIDGES

This section describes the advanced modelling strategy developed for the analysis of masonry arch bridges based on the mesoscale partitioned modelling approach proposed in [12], [13]. This approach has been developed specifically for computation of large-scale unreinforced masonry structures, and it has been already employed and validated for evaluation of arches and single-span masonry arch bridges subjected to vertical loadings and settlements [14],[15].

Masonry is a heterogeneous and strongly nonlinear material whose behaviour depends on the orientation of the loading direction with respect to the masonry bond, where mortar joints represent preferential fracture planes [30]. In this respect, a detailed mechanical model for the masonry arch and piers should take into account not only the mechanical characteristics of units and mortar but also the actual 3D masonry texture. Unlike continuous approaches which assumes masonry a homogeneous material, a discrete modelling strategy [12] is employed to represent the actual masonry bond and model the development of cracks in real brick/stone-masonry arches and in the masonry piers. This numerical strategy allows for an accurate representation of the 3D domain of any masonry arch/pier, as the actual 3D masonry bond is represented using two or more elastic solid elements for each brick and 2D nonlinear interface element for mortar joints. In particular, 20-noded elastic solid elements formulated according to standard finite element (FE) procedures are used together with specific 2D zero-thickness nonlinear interface elements with 16 nodes accounting for material and geometric nonlinearity. In this way, the typical fracture lines which characterise the nonlinear response up to collapse of masonry arches can be represented. These correspond to radial cracks, circumferential cracks leading to ring separation in multi-ring arches and longitudinal cracks caused by transverse bending. While the first two types of crack generally take place in the mortar joints, longitudinal cracks may pass also through the masonry units. Thus nonlinear interface elements are placed also in the middle of each brick to capture the potential development of cracks. This renders the FE mesh for brick-masonry arches relatively simple, as it is made up of identical solid elements connected to each other by nonlinear interface elements as shown in Fig. 3. In the constitutive model for the adopted interface element, material nonlinearity is
taken into account by employing a cohesive model, which enables an effective representation of damage, cracks and plastic separations. More detailed information can be found in [13].

![Image](image1.png)

Fig. 3. (a) 3D mesoscale model for masonry arches, (b) Solid elements connected by nonlinear interfaces.

A realistic representation of the fill behaviour and its interaction with the arch barrel is critical for an accurate response prediction of masonry arch bridges. In the proposed modelling strategy, similarly to other studies on masonry arch bridges [18], the backfill is described as a continuous elasto-plastic material, where a Drucker-Prager plastic criterion with tension cut-off [31] is adopted to describe the development of plastic deformations. In particular, 15-noded elasto-plastic tetrahedral elements are utilised to model the fill domain.

With reference to the foundations and the soil, masonry arch bridges are often built on shallow foundations [32],[33], making them very vulnerable to scour. In order to simulate accurately the scour effects, the bridge model should also include the foundation and the surrounding soil domain. The soil medium has a very complex mechanical behaviour, which is generally nonlinear, stress-dependent, anisotropic and heterogeneous in nature. The difficulty to simulate these features may outweigh the advantages of using complex modelling approaches. For this reason, in order to investigate the effects of scour, a Winkler subgrade modelling approach is adopted in this study, where the three-dimensional soil domain is replaced by a set of interfaces surrounding the foundation. The soil interfaces can resist only normal forces and are characterized by a linear elastic behaviour with a coefficient of subgrade reaction $k_s$, relating deflection (settlement) and soil pressure [34]. Different formula can be employed for estimating the values of the subgrade parameters. In this study, references is made to the expressions developed by Gazetas [35], reported in [36] in function of the shear modulus of the soil $G_s$ and of the foundation geometry. Fig. 4 shows a model of a two span masonry arch bridge, including the backfill, the spandrel walls, and the foundation with the surrounding interfaces.

![Image](image2.png)

Fig. 4. Bridge-foundation-soil model with soil modelled via Winkler-type interfaces.
The hierarchical partitioned modelling method [37],[38], available in the FE code ADAPTIC [39], is employed to minimize the computational cost of the analyses. This approach consists in dividing the structure into super-elements, with each super-element corresponding to one partition. The connection between the different hierarchical levels is achieved by ensuring that the number of element nodes connected to the parent structure is equal to the number of nodes at the partition boundary in the corresponding partition. Each partition is then modelled using the mesoscale approach described earlier and the parent structure corresponds to the partitioned boundary. Dual super-elements allow for partitions to be modelled as separate processes, where communication between each parent/child superelement pair ensures that the analysis for all partitions is run in parallel. A pre-processor has been developed in Matlab to automate the generation and partitioning of the bridge model for ADAPTIC, by accounting for the actual texture and arrangement of the masonry bricks.

4 CASE STUDY

This section presents a case study, where the proposed flood risk assessment framework is applied to evaluate the scour risk for a realistic two-span bridge. The focus here is on pier scour, which is one of the most critical flooring actions, whose modelling has not received adequate attention to date.

4.1 Bridge and numerical model description

The analysed structure is a two-span arch bridge with a length of 20.30m (left to right abutment) and a width of 3.80m. The two arches are segmental in shape, with a radius of 9.26m, a rise of 3.45m, and a thickness of 0.50m. The pier is cylindrical with diameter $D = 2m$. Fig. 5 reports a schematic illustration of the bridge.

![Fig. 5. Schematic illustration of the bridge (dimensions in mm).](image)

The river in proximity of the bridge is assumed to have a trapezoidal section, with base $B = 20.32m$, a riverbank slope (vertical over horizontal) of $1/3$, a channel slope $S_0=0.0006$. The river bed is composed of sand with relative density $\gamma = 16.5 \text{ kN/m}^3$ and uniform diameter $D_{50} = 2 \text{ mm}$, corresponding to a Manning coefficient $n=0.035 \text{ s/m}^{1/3}$. The friction angle of the bed material is taken equal to $30^\circ$.

The flood occurrence is described by a homogenous Poisson process, with a mean annual frequency of occurrence of flood events of peak discharge $Q > q_0=10 \text{ m}^3/\text{s}$ equal to $\lambda=2 \text{ yrs}^{-1}$. The peak flood discharge $Q$ for a given event is assumed to follow an exponential distribution, with scale parameter $\beta =25 \text{ m}^3/\text{s}$ and $q_0= 10 \text{ m}^3/\text{s}$. For simplicity, a triangular shape is assumed for the flood hydrograph, with a duration in hours described by the relation...
\( t_{\text{flood}}(q) = 0.63(q - q_0) \) [40]. According to this relationship, the flood duration increases linearly with its peak discharge and it assumes a value equal to 5 days for \( q=200 \text{ m}^3/\text{s} \). The hydraulic properties of the flow for the different discharge values are evaluated in a simplified way through the Manning formula.

\[
P(\text{P}>q) = -\log(1-P)\]

\( P(\text{P}>q_0) \) for different magnitudes of the flood event. The hydraulic properties of the flow for the different discharge values are evaluated in a simplified way through the Manning formula.

Fig. 6 a) Probability mass function of the number \( N \) of occurrences of flood events with \( Q>q_0 \) for different design life times, b) Empirical cumulative distribution function of the flow discharge for a given flood event of magnitude \( Q>q_0 \).

4.2 Scour risk assessment

This section illustrates some preliminary results of the analyses carried out to evaluate the probability of bridge collapse under scour. The vulnerability analysis of the bridge against scour is carried out for the collapse limit state, which is analytically defined as \( G=R-S \), where \( R \) denotes the value of the scour depth that induces collapse of the bridge under the permanent loads, and \( S \) is the scour depth.

The value of \( S \) is evaluated by following the procedure outlined in Section 2.3, assuming that the initial scour depth is known deterministically and equal to 0m. Fig. 7a,b show respectively the empirical probability mass function of the scour depth and the risk of exceedance of the scour depth at different times of observation between 0 and 100 yrs. The probabilistic distribution of the scour depth evolves in time by exhibiting an increase of the mean value of \( S \), and a decrease of standard deviation. In fact, the scour depth \( S \) slowly tends to a stationary value which is attained ideally for an infinite time of observation.

Fig. 7 Probability mass function and risk of exceedance of the scour depth at different times (between 1 yr and 100 yrs after the initial one.)
It is noteworthy that the value of $S$ represents here the maximum depth of the hole created by scouring. The geometric domain of the scour hole is defined by an inverted pyramid. The upstream surface has a slope corresponding to an angle equal to the soil friction angle ($\phi=30^\circ$), the downstream one has a slope corresponding to $\phi/2$, whereas the lateral surfaces have a slope corresponding to $3/4\phi$. The maximum scour depth is assumed to be located along the vertical plane containing the upstream pier surface.

The value of $R$ is evaluated by performing the structural analysis on the FE bridge model under progressively increasing scour depths. The mechanical properties of the model are representative of typical bridge constructions \[14,42\]. In particular, a tensile strength $\sigma_{t0} = 0.06\text{MPa}$, a cohesion $c_0 = 0.1\text{MPa}$ and a friction angle with tangent $\tan(\phi) = 0.75$ were considered for the mortar joints, while a Young’s Modulus $E_b = 3918\text{MPa}$ and a Poisson’s ratio $\nu_b = 0.15$ were utilised for the brick units. The compressive strength of masonry is assumed equal to $\sigma_{c0} = 8\text{MPa}$. The backfill material model has a friction angle of $\phi_{bf} = 60^\circ$, a dilation angle of $\phi_{bdf} = 30^\circ$, and a cohesion $c_{bf} = 0.01\text{MPa}$. The elements describing the spandrel walls are assigned a value of the Young modulus $E_w = 3100\text{MPa}$, a cohesion $c_w = 0.1\text{MPa}$, a friction angle $\phi_w = 43^\circ$, a dilation angle $\phi_{wd} = 10^\circ$, and a tensile cap on the first invariant of stress tensor $I_{wt} = 0.03\text{ Mpa}$. The compressive strength is assumed equal to $\sigma_{c0} = 8\text{MPa}$. Finally, the riverbed sand is assumed to have a Young’s Modulus $E_r = 500\text{MPa}$ and a Poisson’s ratio $\nu_r = 0.15$, leading to a normal stiffness of the interfaces below the foundation equal to $k_r = 0.184\text{N/mm}^3$.

The scour action is simulated via time-history analysis by progressively degrading the stiffness of the interfaces representing the soil and located within the scour hole. Fig. 8 shows the evolution of the scour hole and the progressive removal of the interfaces located within the scour hole.

![Fig. 8. Assumed scour hole shape and scour progression modelling.](image-url)

Fig. 9 shows the deformed shape of the bridge at different increasing scour levels. The scour action induces a rotational mechanism at the base of the pier, with non-uniform vertical displacements along the pier base, and a formation of a crack between the foundation and the pier. This mechanism has been responsible of the collapse of Copley Bridge, a two-span bridge located near Halifax, West Yorkshire, whose geometrical properties are very similar to those of the model considered as case study. Fig. 10 shows the cracks in the arch barrel for a
maximum scour depth \( s = 2.6 \text{m} \), and Fig. 11 shows the damage in tension of the mortar interfaces between the bricks. It can be observed that scour induces the formation of both transversal and diagonal cracks in the arch barrel. The diagonal cracks are located in proximity of the pier and of the abutments, whereas the transversal cracks are localized at midspan of each span. When the maximum scour depth attains the value \( s = 2.6 \text{m} \), the extent of cracking is very significant and the bridge is at collapse. The corresponding risk of failure, based on Fig. 7a, is established as 2% in 100 years.

Fig. 9. Deformed shape of the bridge model for increasing levels of maximum scour depth.

Fig. 10. Cracking pattern in the arch barrel for maximum scour depth \( s=2.6 \text{m} \).

Fig. 11. Damage parameter in tension in the arch and the pier for maximum scour depth \( s=2.6 \text{m} \).
5 CONCLUSIONS

In this paper, a probabilistic framework for the flood risk assessment of masonry arch bridges is illustrated. The framework integrates information on the most critical actions associated with the flood hazard analysis with those of structural vulnerability analysis. The computational strategy developed by the authors for the numerical analysis of the structural response of masonry arch bridges subjected to flooding actions is also described.

The proposed framework and modelling strategy are applied to the scour risk assessment of a realistic case study. It is shown that an accurate 3D model, including the foundation and the surrounding soil, can be used to describe the rotational-translational mechanism induced by the scour action at the base of the pier, allowing the prediction of cracks forming in the pier and arches. Future studies will address the impact of other critical actions induced by floods on masonry arch bridges, the influence of the uncertainty of the model parameters on the risk estimates, and also the development of simplified and efficient modelling strategies leading to a reduction of the computational cost.

AKNOWLEDGEMENTS

The financial support of the European Commission through the Marie Skłodowska-Curie Individual fellowship IF ("FRAMAB", Grant Agreement 657007) for the first author is greatly acknowledged. The authors also acknowledge the High Performance Computing (HPC) Services at Imperial College for providing and supporting the required computing facilities.

REFERENCES


SAMPLING SCHEMES FOR HISTORY MATCHING USING SUBSET SIMULATION

Z. Gong\(^1\), F. A. DiazDelaO\(^1\), and M. Beer\(^2\)

\(^1\)Institute for Risk and Uncertainty, School of Engineering
University of Liverpool, Liverpool, United Kingdom
e-mail: \{gong7, f.a.diazdelao\}@liverpool.ac.uk

\(^2\)Institute for Computer Science in Civil Engineering
Leibniz University Hannover, Hannover, Germany
e-mail: beer@bauinf.uni-hannover.de

Keywords: Subset Simulation, Rare Event Simulation, History Matching, Non-implausibility, Bayesian Emulation.

Abstract. History Matching (HM) is a form of model calibration suitable for high-dimensional and computationally expensive numerical models. It sequentially cuts down the input space to find the non-implausible domain that provides a reasonable match between the output and experimental data. The non-implausible domain can be orders of magnitude smaller than the original input space and it can have a complex topology. This leads to one of the most challenging open problems in implementing HM, namely, the efficient generation of samples in the non-implausible set. Previous work has shown that Subset Simulation can be used to solve this problem. Unlike Direct Monte Carlo, Subset Simulation progressively decomposes a rare event (here is the non-implausible set), which has very small failure probabilities, into sequential less rare nested events. The original Subset Simulation uses a Modified Metropolis algorithm to generate the conditional samples that belong to intermediate less rare failure events. Generating samples moving forwards to the target space is the heart for Subset Simulation. This work considers different sampling strategies to generate samples and compares their performance in the context of expensive model calibration. A numerical example is provided to show the potential of HM using different Subset Simulation sampling schemes.
1 INTRODUCTION

Numerical models, also known as simulators, are universally designed and employed to represent and study complex real-world systems. In order for a simulator to be reliable, it requires calibration using experimental data. Unfortunately, high-dimensional input and computational cost often hinder the calibration process. This results in simple goodness-of-fit methods such as distance-based methods or likelihood functions to become impracticable [1].

History matching (HM) is a form of model calibration that is suitable for high-dimensional and expensive simulators. Anterion [2] first applied it to the oil industry to diminish computation time in inverse problems for reservoir simulation. Since then, more applications using HM to get reservoir characteristics have emerged [3, 4, 5]. HM sequentially cuts down the initial input space using an implausibility threshold that includes various of uncertainty sources. At each iteration, HM rules out samples in the implausible input domain, and generates samples that eventually define the non-implausible domain providing a reasonable match between the model output and experimental data. Evaluating the model output at each sample point is typically expensive, therefore HM also involves Bayesian emulation [6], which reduces the running cost of complex simulators and quantifies the variance of the predictive posterior output.

Recently, HM has been successfully applied to highly nonlinear geophysical simulators [7], to galaxy formation models [8], and to large climate systems modeling [9] amongst other applications. In the above literature, the non-implausible domain can reduce to orders of magnitude smaller than the initial input space, and might present a complex topology or be disconnected. Therefore, generating samples from such non-implausible domain has remained an open and challenging problem.

Traditionally, there have been at least three classes of methods to generate samples in the non-implausible set: an acceptance-rejection strategy [10], an implausibility driven evolutionary Monte Carlo algorithm [11], and a perturbation approach [11]. However, their efficiency depends on several requirements. Recent work [12] has shown that Subset Simulation can be employed as an efficient sampling scheme.

Unlike Direct Monte Carlo, Subset Simulation progressively decomposes a rare event, which has a small failure probability, into sequential less rare nested events. Samples providing better non-implausibilities in intermediate events remain as seeds to generate more samples towards the final target space. Eventually, rare samples of interests can be targeted down given the occurrence of less rare events.

The original Subset Simulation algorithm uses Modified Metropolis algorithm (MMA) [13] to generate conditional samples that belong to intermediate less rare failure events. During the last decade, multiple versions of sampling schemes for Subset Simulation have emerged in different contexts [14, 15]. Santoso et al. [16] repeat the sample generation process until the samples are accepted at the first step. Miao and Ghosn [17] proposed a delayed rejection approach. Recently, Papaioannou et al [18] proposed a conditional sampling from the current sample, avoiding the first step in the traditional Markov Chain Monte Carlo (MCMC) algorithm. Au and Patelli [19] interpreted this sampling method from a different perspective and validated its powerful efficiency in reducing correlation between samples. This work considers different MCMC strategies and compares their performance in the context of model calibration.

The paper is organized as follows. Section 2 presents important components of HM: initial design for simulation, Bayesian emulation, uncertainty quantification, implausibility threshold, and sampling on the non-implausible domain. Section 3 presents different sampling algorithms for Subset Simulation. In section 4, an example combining Subset Simulation and HM is pre-
presented. Finally, section 5 provides some conclusions.

2 HISTORY MATCHING

HM aims to find the input domain (representing value ranges of parameters) that provides a match between computer model outputs and observations for a complex physical system. HM sequentially cuts down the non-implausible input domain, which leads to a higher concentration of non-implausible samples in each subsequent iteration (also known as wave), using an implausibility threshold that takes into account diverse sources of uncertainty. To reduce the computation time, HM adopts emulation to make inference about the simulator’s output. If a large number of non-implausible samples are found in a dramatically reduced input space, HM can be stopped. non-implausible in a wave. If the stopping criterion is not met, HM refocuses on the non-implausible domain via sampling on it and conduct another wave of model running and implausibility testing. A typical HM workflow is summarized in Figure 1.

Figure 1: Typical workflow of HM. This work focuses on the sampling from the non-implausible domain efficiently.

An initial design to run the simulator is the first step for a typical HM workflow. Let \( z \) be an observation of a physical system \( y \) and \( g(\cdot) \) be a simulator. Given \( z \), wide possible ranges of the input domain are considered at the beginning. As a rule of thumb, Loepky et al. [20] suggest \( n = 10d \) as the number of sample points, where \( d \) is the input dimension. We resort to Latin hypercube sampling (LHS) to efficiently explore the input space. Once an LHS design is specified, \( g(\cdot) \) is run at each input point \( x_i \in \mathbb{R}^d \). Hence, the initial design as training data for the emulation is \( \{(x_1, g(x_1)), \ldots, (x_n, g(x_n))\} \).

Based on the simulation data, a Bayesian emulator [21, 22, 23] interpolates the model output to reduce the computational cost. A Bayesian emulator (also called Gaussian process emulator or Kriging) is a stochastic approximation to the output of an expensive computer model, which is widely used as a surrogate for complex simulators. Bayesian emulators have been found effective in multiple disciplines: modeling for structural dynamic analyses [24, 25], stochastic mechanical responses [26] and reliability assessments [27], amongst any others.

In this paper, we use a Bayesian emulator of the form:

\[
g(x) = \sum_{i=1}^{d} h_i(x)^T \beta_i + Z(x) \quad (1)
\]
where \( g(x) \) is the emulator’s inference based on the training data; each \( \beta_i \) is a regression coefficient; the function \( h(x)^T \beta \) models the global trend of the output, whereas a Gaussian process \( Z(x) \) models local variations. Note that, the emulator output interpolates the simulator output, for which \( g(\cdot) \) is used to denote both.

The emulator’s output \( g(x) \) follows the student \( t \) distribution: 
\[
t_{n-p}(E^*(g(x)), \sqrt{V_c(x)}),
\]
where \( n - p \) is the degree of freedom, \( E^*(g(x)) \) is the predictive mean and \( V_c((x)) \) is the predictive variance.

HM takes into account three sources of uncertainty: \( V_o \) is the observation uncertainty (OU), which stands for the experimental error, such as the finite accuracy of measurement instruments or human operations; \( V_c \) is the code uncertainty (CU), which stands for the error generated by the emulation (implemented as computer codes) being an approximation; \( V_m \) is the model discrepancy (MD), which stems from the inability to model a physical system \( y \) perfectly. Summarizing, the total uncertainty considered is as follows:
\[
V(x) = V_o + V_c(x) + V_m
\]

Let \( I(x) \) denote the implausibility that an output matches the specified observation, quantified via the difference between them with uncertainty margins [1]:
\[
I(x) = \frac{|z - E^*(g(x))|}{\sqrt{V(x)}}
\]

If at a sample point \( x_i \), the implausibility measure \( I(x_i) \) returns a small value, it is very likely that this input is an acceptable match between the model output and the experimental data. A criterion of the implausibility threshold \( I(x) \leq 3 \) is applied [1] following the Pukelsheim three sigma rule [28] (around 99%, or to say nearly all values in a distribution lie within a three-fold standard variance band on either hand of the mean). Sample points that fail the criterion are considered implausible.

Once the non-implausible samples at the current wave are defined by the implausibility measure, if the stopping criterion of HM is not met, the new wave begins by sampling from the non-implausible domain. As previously mentioned, the initial input design can be generated easily by an LHS plan. However after the first wave, sampling from a non-implausible domain with a complex topology can be challenging. It has been recently proposed [12] to use Subset Simulation to sample new emulation points since the non-implausibility domain can very naturally be interpreted as a failure set, both requiring the definition of a threshold.

The conceptual core of Subset Simulation is: generate a sequence of subsets \( F = \{x : I(x) < 3\} = F_d \subset F_{d-1} \subset \cdots \subset F_1 \), so that the probability for the event of interest \( F \) (here is the non-implausible set) can be calculated as \( P(F_1) \times P(F_2|F_1) \times \cdots \times P(F_d|F_{d-1}) \). This decomposes the event \( F \) of small probability into conditional events that are more likely to happen and easier to sample from. Every intermediate event corresponds to a level in Subset Simulation. An MCMC algorithm is applied to populate samples for intermediate events and eventually for the target event.

There are two important parameters for Subset Simulation: the level probability \( p_L \) and the number of samples in each level \( n_L \). Both of them are determined by the user. The level probability is usually chosen as 0.1 in the literature [29]. The number of samples for each level \( n_L \) should be decided by a balance considering between the dimensions and the computation speed. It must also ensure integer values for both the number of chains in the MCMC algorithm \( n_c = n_L \times p_L \) and the number of new samples in an event \( n_s = (1 - p_L)/p_L \).
At the initial 0\textsuperscript{th} level, Subset Simulation selects the first \(n_c = n_L \times p_L\) samples \(x_1, ..., x_{n_c}\) that have the smallest implausibility, according to Equation 3. Then, for the implausibility measure \(I\), the intermediate threshold defining event \(F_1 = \{x : I(x) < Y_{SS_1}\}\) is:

\[
Y_{SS_1} = \frac{I^{(n_c)}_0 + I^{(n_c+1)}_0}{2}
\]

The goal then becomes the generation of samples from \(F_1\), on which the subsequent levels are conditioned. The Modified Metropolis algorithm (MMA) \cite{13} algorithm applied to sample in \(F_1\) is as follows:

1. Propose a random standard Normal move \(\alpha \sim N(x_k, 1)\) in each dimension from each seed \(x_1, ..., x_k, ..., x_{n_c}\), or alternatively adaptive MMA: a Normal move using the variance of the current sampling seeds \(x_1, ..., x_{n_c}\) \cite{29};

2. Accept the candidate movement \(\xi = \alpha\) with probability \(\min\{1, \frac{\phi(\alpha)}{\phi(x_k)}\}\), otherwise stay at \(x_k, \xi = x_k\);

3. Remain \(x_k\) if the movement \(\xi \notin F_1\).

The process is repeated until:

\[
n_F = \frac{\sum_{k=1}^{n_L} I_{I(x_k) < 3}}{n_L} > p_L
\]

where \(I(\cdot)\) is an indicator function that counts the number of samples in the relevant set. Figure 2 summarises the workflow for Subset Simulation combined with HM found in \cite{12}.

![Figure 2: The workflow for Subset Simulation sampling embedded in HM. This work focuses on the sampling generation step, which is implemented and compared with different schemes in following sections.](image-url)
3 Sampling schemes for Subset Simulation

The original MMA and adaptive MMA are not the only sampling schemes available in the literature [18]. The aim of this paper is to compare the performance of different sampling schemes for Subset Simulation within the HM framework. To this end, some alternative sampling schemes are outlined.

3.1 Delayed rejection

To improve the acceptance rate, Miao and Ghosn [17] repeat the sample generation process by a second proposal PDF if the initial candidate is rejected, thus delaying rejection. The approach can be described as follows:

1. Propose a random standard Normal move $\alpha \sim N(x_k, 1)$ in each dimension from each seed $x_1, \ldots, x_k, \ldots, x_n$;
2. Accept the candidate movement $\xi = \alpha$ with probability $\min\{1, \frac{\phi(\alpha)}{\phi(x_k)}\}$, and go straight to step 5, otherwise go to step 3;
3. Propose another move e.g. $\alpha' \sim U(x_k, 1)$ in each dimension from each seed $x_1, \ldots, x_k, \ldots, x_n$;
4. Accept the candidate movement $\xi = \alpha'$ with probability $\min\{1, \frac{\phi(\alpha')}{\phi(x_k)}\}$, otherwise stay at $x_k$;
5. Remain at $x_k$ if the movement $\xi \notin F_1$.

3.2 Adaptive MCMC with optimal scaling

Roberts et al. [30, 31] found that the optimal efficiency is achieved when the final sample acceptance rate is around 0.44. Based on this, Papaioannou et al. [18] scale standard deviation of the proposal PDF adaptively (using a scaling parameter $\lambda$) to stabilize the acceptance probability. At each level, all $n_c$ seeds are randomly divided into $R$ groups, each with $N_a$ seeds. Given an initial standard deviation for the proposal PDF, the initial $N_a$ seeds are used to generate sample candidates conditioned on them. The value of the scaling parameter $\lambda$ is determined by the difference between the acceptance rate of the first $N_a$ seeds and the optimal value 0.44. The adapted standard deviation is then plugged into the next group of $N_a$ seeds. The adaptive standard deviation algorithm for conditional sampling is as follows:

1. Randomly divide $n_c$ seeds into $R$ groups of $N_a$ samples;
2. For $r^{th}$ group ($r = 1, \ldots, R$), set the initial standard deviation of the proposal distribution to $s_i = 1, i = 1, \ldots, d$, and the initial scaling parameter $\lambda$ to 0.6 [18];
3. Compute the coefficient: $a_i = \sqrt{1 - (\lambda s_i)^2}$, where $a = [a_1, \ldots, a_d]$ and $s = [s_1, \ldots, s_d]$;
4. Generate the sample candidates: $x_k' \sim N(ax_k, \lambda s_i), k = 1, \ldots, N_a$;
5. Accept $x_k'$ if $x_k' \notin F_1$, otherwise the chain remains at $x_k$.
6. Compute the average acceptance rate from $N_a \times N_s$ chains: $A = \frac{I_A(\cdot)}{N_a \times N_s}$, where $I_A(\cdot)$ is an indicator function which counts the number of acceptable samples;
7. Upgrade the adaptive scaling parameter: $\lambda = 10^{(\log_{10}(\lambda) + \zeta(A-0.44))}$, with $\zeta = r^{-1/2}$;
8. Repeat steps 2 to 9 until finishing all $R$ groups of $n_c$ seeds.
### 3.3 Subset-infinity

Au and Patelli [19] extended the conditional sampling to decomposing the sample $x_i$ of standard Gaussian distribution by an arbitrary number ($1 \leq N \leq \infty$) of i.i.d. standard Gaussian variables. Let

$$x_i = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} Z_j$$

where $Z_j \sim N(0, 1)$. When $N \to \infty$, sample candidates $x_i'$ conditioned on $x_i$ can be generated by a proposal PDF ($i = 1, \ldots, d$):

$$p(x_i'|x_i) = \frac{1}{\sqrt{2\pi s_i}} \exp \left[ -\frac{1}{2s_i^2}(x_i' - ax_i)^2 \right]$$

where

$$a_i = 1 - 2\kappa_i$$

$$s_i^2 = 4\kappa_i - 4\kappa_i^2$$

$$\kappa_i = \int_0^\infty w^2 \Phi(-w/2)p_i^*(w)dw$$

It can be shown that $a_i^2 + s_i^2 = 1$, and $0 \leq \kappa_i \leq 1$, hence $a_i \in [-1, 1]$ and $s_i \in [0, 1]$. To sample in $F_1$, the conditional sampling algorithm for Subset-infinity is:

1. Assign values for parameters of the proposal PDF: e.g. $s_i = 0.5$ (suggested by [32]),
   $$a_i = \sqrt{1 - s_i^2}, \quad i = 1, \ldots, d;$$
2. Generate the sample candidate dimension-wise: $x_k' \sim N(a_i x_k, s_i)$, $k = 1, \ldots, n_c$;
3. Remain at $x_k$ if the candidate $x_k' \notin F_1$.

### 4 Numerical example

The different sampling schemes described in the previous section are now compared under the HM framework using an 8-dimensional robot arm model [33].

$$f(x) = (u^2 + v^2)^{0.5}$$

where

$$u = \sum_{i=1}^{4} L_i \cos \left( \sum_{j=1}^{i} \theta_j \right), \quad v = \sum_{i=1}^{4} L_i \sin \left( \sum_{j=1}^{i} \theta_j \right)$$

The response $f(x)$ is the distance from the end of the robot arm to the origin, on the $(u, v)$-plane. The input variables and their ranges are shown in Table 1.

In order to implement HM, an observation of 4 units is assumed. The observational uncertainty and model discrepancy were both given fixed values of 0.01. For Subset Simulation, the parameters $p_L = 0.1$ and $n_L = 8000$ were chosen. Table 2 compares the results of the different sampling schemes. It is worth noting that all sampling schemes required 4 levels of Subset Simulation. Also, since by construction MCMC rejects samples from proposal, this yields repetitive sample points. Our aim is to determine which scheme produces more unique non-implausible

---

**Table 1**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$</td>
<td></td>
</tr>
<tr>
<td>$v$</td>
<td></td>
</tr>
<tr>
<td>$L_i$</td>
<td></td>
</tr>
<tr>
<td>$\theta_j$</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2**

<table>
<thead>
<tr>
<th>Sampling Scheme</th>
<th>Unique Sample Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subset Simulation</td>
<td>Yes</td>
</tr>
<tr>
<td>MCMC</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 1: The input variables and their ranges for the robot arm function

<table>
<thead>
<tr>
<th>Input variables’ range</th>
<th>Input variables’ meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_1 \in [0, 2\pi] )</td>
<td>angle of the first arm segment</td>
</tr>
<tr>
<td>( \theta_2 \in [0, 2\pi] )</td>
<td>angle of the second arm segment</td>
</tr>
<tr>
<td>( \theta_3 \in [0, 2\pi] )</td>
<td>angle of the third arm segment</td>
</tr>
<tr>
<td>( \theta_4 \in [0, 2\pi] )</td>
<td>angle of the fourth arm segment</td>
</tr>
<tr>
<td>( L_1 \in [0, 1] )</td>
<td>length of the first arm segment</td>
</tr>
<tr>
<td>( L_2 \in [0, 1] )</td>
<td>length of the second arm segment</td>
</tr>
<tr>
<td>( L_3 \in [0, 1] )</td>
<td>length of the third arm segment</td>
</tr>
<tr>
<td>( L_4 \in [0, 1] )</td>
<td>length of the fourth arm segment</td>
</tr>
</tbody>
</table>

Table 2: The performance difference among different MCMC sampling approaches, \( p_L = 0.1 \) and \( n_L = 8000 \). samples. The MMA with adaptive variance and the adaptive MCMC with optimal scaling produced the greatest number of such samples.

The resulting non-implausible input domains given by different sampling schemes are shown in the upper triangle panels of Figure 3. The lower triangle panels of Figure 3 show optical depth plots: on each plane of inputs’ pairs, a 20×20 grid is created, with the color denoting the proportion of the non-implausible samples over each grid, which gives plausibility information perpendicular to projective input planes.

5 CONCLUSIONS

The combination of Subset Simulation and HM is possible due to the natural analogy between the non-implausible space in HM and the failure set in reliability analysis. In this paper, different variants of Subset Simulation were presented and illustrated in the numerical example.

In this work, the performance of the sampling scheme was judged by the number of unique non-implausible samples. It was observed that the MMA with adaptive variance and the adaptive MCMC with optimal scaling ranked the highest under this simple criterion.

Future research will concentrate on different metrics of performance, as well as the application of the best-performing sampling schemes to calibrate more complex and realistic models.

REFERENCES


Figure 3: The implausibility plot and optical depth plot of the input after the second wave using different sampling strategies.


MARKOV CHAIN MONTE CARLO METHODS FOR UNCERTAINTY PROPAGATION AND RELIABILITY ANALYSIS

Carsten Proppe

Chair of Engineering Mechanics, Karlsruhe Institute of Technology
Kaiserstr. 10, Bdg. 10.23, 76131 Karlsruhe, Germany
e-mail: proppe@kit.edu

Keywords: Reliability Estimation, Markov Chain Monte Carlo Simulation, Subset Simulation, Moving Particles Algorithm.

Abstract. Two Markov chain Monte Carlo simulation methods for reliability estimation, subset simulation and the moving particles algorithm, are compared. To this end, both low-dimensional and high-dimensional test cases are considered. The investigation discusses differences in the efficiency of both algorithms and sheds light on the parameter settings as well as the avoidance of correlated samples in both algorithms, notably on the necessity of a burn-in period and the influence of a metamodel.
1 INTRODUCTION

In direct Monte Carlo Simulation (MCS), independent and identically distributed samples are drawn in order to obtain an unbiased estimate of a function of random variables. For MCS, the strong law of large numbers yields the $P$-almost sure convergence of the estimator, and from the central limit theorem, one can deduce that the variance of the estimator decreases with $1/N$, where $N$ is the number of samples. In particular, the decrease does not depend on the number of random variables. Unfortunately, direct MCS is not well suited for reliability analysis, where low failure probabilities must be estimated, because a huge number of samples is needed in order to obtain accurate estimates. This leads in turn to a prohibitively large number of calls to the performance function; and if the performance function itself requires a rather large computational time (which is in general the case in structural analysis), the total computational time will be excessive.

Generating samples from a fixed importance sampling density aims to reduce the variance of the MCS estimator while retaining the advantages of independent and identically distributed samples. Thus, the strong law of large numbers and the central limit theorem still apply, yielding the same convergence properties as for direct MCS, but with a reduced variance of the estimator due to a judicious choice of the importance sampling density. For reliability analysis, heuristics to obtain an appropriate importance sampling density have been discussed in the literature, see e.g. [1]. However, as has been shown in [2], sampling with a fixed importance sampling density becomes inefficient for reliability estimation involving a high-dimensional random vector and thus a large number of random variables.

Adaptive importance sampling densities may be constructed by Markov Chain Monte Carlo (MCMC) algorithms. In this case, the samples are not independent (but still identically distributed, if the Markov chain is in stationary state), but the ergodic theorem and the central limit theorem for reversible Markov chains yield the convergence and the asymptotic unbiasedness of the estimator. An MCMC based algorithm that is widely used for reliability estimation is subset simulation [3], which is based on the estimation of conditional probabilities for a nested sequence of sets. In contrast to many other MCMC algorithms, subset simulation does not require a burn-in of the Markov chain, because the seeds of the Markov chains are already distributed according to the target distribution [4]. On the other hand, the classical Metropolis-Hastings (MH) algorithm suffers from a high rejection rate in conjunction with subset simulation, especially for high-dimensional reliability estimation, cf. [2]. Therefore, a componentwise MH algorithm has been introduced in [3] and more recently, direct sampling from a normal transition kernel has been proposed which yields candidates that always differ from the current state [5] and leads to an improved efficiency for high dimensional problems, if the proposal distribution is adapted for each subset.

In subset simulation, the size of the nested sets is usually chosen such that the conditional probabilities are equal to 0.1, i.e. 90% of the samples are discarded and need to be recomputed from the remaining 10% that serve as seeds. This leads in general to less than ten nested sets. Having too few nested sets would lead to a similar inefficiency as for direct MCS. In [3], it is argued that having much more nested sets would lead to an increase in the total number of samples, which again decreases the efficiency of subset simulation. However, it is worthwhile to consider the limit case, where only a single sample is discarded and recomputed. This leads to a maximum number of nested sets. Such an MCMC based algorithm, called moving particles algorithm, has been introduced in [6] and [7].

The objective of this contribution is to compare subset simulation and the moving particles algorithm.
algorithm for reliability estimation on low- as well as on high-dimensional test cases with emphasis on the accuracy, efficiency and the acceptance rate and thus to shed some light on specific features of these two MCMC based algorithms. The paper is organized as follows: in the next section, subset simulation and the moving particles algorithm are introduced. Following this, the test cases and the test methodology are presented. The test results are discussed and explanations for the observed phenomena are given. Finally, a summary of the observed phenomena and recommendations for off-the-shelf application of the simulation algorithms are given.

2 MARKOV CHAIN MONTE CARLO SIMULATION

2.1 Subset simulation

Subset Simulation is based on nested sets, $F_1 \supset F_2 \supset \ldots \supset F_M$, where $F_M = \{ \theta \in \mathbb{R}^n | g(\theta) < 0 \}$ denotes the failure region and $g(\theta)$ is the performance function. If $\theta$ is a vector of random variables, the failure probability is given by

$$P_f = P_{F_1} \prod_{i=1}^{M-1} P(F_{i+1}|F_i). \quad (1)$$

The rather small failure probability is written as the product of larger probabilities that can be estimated with less effort. However, the estimation of the conditional probabilities requires the application of MCMC simulations, because the corresponding conditional probability density function is not known explicitly. In subset simulation, $N_s$ parallel Markov chains are started from seeds that for step $i-1$ lie in $F_i$. Generating the Markov chains with the classical MH algorithm may result in low acceptance rates. Alternatives are the componentwise MH algorithm [3], a single repetition of the candidate sample (delayed rejection) or the direct sampling from a suitable transition kernel such that rejection is limited to the case where the candidate sample is not an element of $F_i$.

The most influential parameters of subset simulation are the transition kernel of the MCMC algorithm and the sets $F_i$, for which sets of equal conditional probability $p_0 = P(F_{i+1}|F_i)$ are preferred. Given $p_0$, the sets are obtained from a percentile estimation for the performance function. The failure probability is then given by

$$P_0^{M-1} \hat{P}_M, \quad (2)$$

where $\hat{P}_M$ is the estimate for $P(F_M|F_{M-1})$.

The coefficient of variation of the estimator for the conditional probability $P_{i+1} = P(F_{i+1}|F_i)$ is given by [3]

$$\sqrt{\frac{1 - \hat{P}_i}{NP_i}} (1 + \gamma_i), \quad (3)$$

where the additional term $\gamma_i$ is

$$\gamma_i = 2 \sum_{k=1}^{N/N_s-1} \left( 1 - \frac{kN_s}{N} \right) \rho_i(k). \quad (4)$$

$N_s = p_0N$ is the number of samples that for step $i-1$ lie in $F_i$ and constitute the seeds of the Markov chains. $\rho_i(k)$ is the correlation coefficient of the series $I_{F_i}(\theta_{jk}^{(i-1)})$, where $\theta_{jk}^{(i-1)}$ is the $k$th sample of the $j$ Markov chain that is generated in step $i-1$. A weak correlation of the
samples produced by the Markov chain is thus necessary for a reduction of the coefficient of variation.

The coefficient of variation for the failure probability estimated with subset simulation can be approximated by

\[ \delta_{\text{sub}} \approx \sqrt{\frac{\log p_F (1 - p_0)(1 + \bar{\gamma})}{N p_0}}, \]

(5)

where \( \bar{\gamma} \) is the average value of \( \gamma_i \) (averaged over the number of subsets) and \( \frac{\log p_F}{\log p_0} \) represents the number of subsets. A typical value for \( \bar{\gamma} \) is \( \frac{1}{1+\bar{\gamma}} = 0.4 \), cf. [5]. The average number of function evaluation is thus approximated by

\[ N_{\text{sub}} = N \left( 1 + (1 - p_0) \frac{\log p_F}{\log p_0} \right), \]

(6)

and is composed by the initial Monte Carlo samples and the Markov chain samples (without burn-in) at each step.

### 2.2 Moving particles algorithm

The moving particles algorithm can be considered as subset simulation with a maximum number of steps. Thus, in each step, only one sample is discarded and resampled by a Markov chain that takes as seed one of the retained samples. As for subset simulation, the algorithm starts with a direct MCS. For each step, the values \( g(\theta_i), i = 1, \ldots, N_m \), of the \( N_m \) samples are ranked. The sample with the maximum value of the performance function is moved: MCMC is carried out starting from one of the remaining samples and the final state of the Markov chain is accepted, if the value of the performance function could be reduced. Otherwise, the sample is simply replaced by the seed of the Markov chain. The Markov chain can be generated either by application of the MH algorithm or by direct sampling from a normal transition kernel. As the classical MH algorithm suffers from low acceptance rates for high-dimensional problems, the componentwise MH algorithm of [3] is applied in the following. Thus, the same Markov chain simulation technique is applied in this study for the moving particles algorithm and subset simulation.

However, instead of computing the probability of failure from eq. (2), each initial sample is moved until it reaches the failure region and the number of moves is counted. As has been shown in [6], the number of moves to get an initial sample into the failure region follows a Poisson distribution with parameter \( \lambda = \log \frac{1}{1-p_F} \). The estimator for the parameter of the Poisson distribution is

\[ \hat{\lambda} = \frac{\sum_{i=1}^{N_m} M_i}{N_m}, \]

(7)

where \( M_i \) denotes the number of moves until seed \( i \) reaches the failure state.

In order to obtain an unbiased estimate, it is mandatory that the trajectories of the Poisson process generated from the initial samples remain independent until the samples finally reach the failure domain. In [7], two means are proposed to maintain the independence:

- **Burn-in**: The Markov chain simulation is carried out with a burn-in period. The burn-in should ensure the independence of the candidate and the seed of the Markov chain.

- **Seed avoidance**: Repeated use of the same seed for the Markov chain should be avoided. Once a sample has been used as seed, the sample and its offspring should not be used as seed again.
The coefficient of variation for the failure probability estimated with the moving particles algorithm is given by

$$\delta_{mp} = \sqrt{\frac{-\log p_F}{N_m}},$$

(8)
cf. [6], and the average number of function evaluations is

$$N_{mp} = N_m(1 - T \log p_F),$$

(9)
where the first term accounts for the initial Monte Carlo simulation and the second term for the Markov chain samples (with burn-in period $T$).

### 2.3 Comparison of the two algorithms

While the moving particles algorithm can be considered as a special case of subset simulation, there are several differences with respect to the original subset simulation algorithm:

- In subset simulation, the number of steps is rather small; however, the number of steps in the moving particles algorithm is maximal.

- In subset simulation, only $p_0$ (usually 10%) of the samples are retained in each step and serve as seed for the Markov chains. In contrast, only one sample is resampled in each step of the moving particles algorithm, and the seed can be selected among the other samples.

- The moving particles algorithm has a clear interpretation from Poisson process theory, but requires that all initial samples finally reach the failure region.

### 3 TEST CASES

Test cases have been considered with standardized parameters for the algorithms in order to investigate the robustness, accuracy and efficiency of both simulation methods. The aim was to find out whether the simulation methods can be applied off-the-shelf without fine tuning to a specific problem. The low-dimensional test cases were those summarized in [9], Table 1. They allowed to carry out parameter studies for the number of initial samples and the Markov chain transition kernel. Moreover, for the moving particles algorithm, the influence of a burn-in period, a seed selection strategy and a low-rank metamodel have been investigated. Following this, both algorithms were applied to high-dimensional examples, a paraboloid with discretely varying principal curvatures, [10], example 4, and a single-degree-of-freedom oscillator with 1500 random variables, [3], example 1, with a threshold value $b = 1.5$.

For each test case and each set of parameters, 100 simulation runs were carried out from which the mean probability of failure, the coefficient of variation, the mean number of performance function evaluations and the mean acceptance rate has been observed.

For the MCMC with MH algorithm, two acceptance rates can be defined:

- Level 1: Acceptance of the pre-candidate by the MH algorithm.

- Level 2: Acceptance of the candidate sample. For subset simulation, the candidate sample at step $i$ must lie in $F_i$, while for the moving particles algorithm, the value of the performance function must decrease.
In this investigation, the level 2 acceptance rate has been monitored only. The simulation platform was Matlab with FERUM [11]. To this end, the moving particles simulation algorithm has been added to FERUM and extensions to the subset simulation algorithm were made.

4 RESULTS

4.1 Number of initial samples

Figure 1: Influence of the number of initial samples. a) Subset simulation. b) Moving particles algorithm.

Figure 1 displays the influence of the number of initial samples on the coefficient of variation of the simulation algorithms for the low-dimensional test cases. It can be seen that the number of initial samples needed to obtain the same level for the coefficient of variation differs by nearly a factor of ten, which is due to the fact that for subset simulation, 90% of the samples are discarded in each simulation step. In fact, considering equations (5) and (8) with $\delta_{sub} = \delta_{mp}$, $\frac{1}{1 + \delta_{sub}} = 0.4$ and $p_0 = 0.1$, one obtains $N_{mp} \approx 0.1N$.

While increasing the number of initial samples increases the accuracy, the efficiency is decreased. The number of initial samples does not influence the mean level 2 acceptance rate very much.

4.2 Influence of the proposal density

As proposal density for the componentwise MH algorithm, a normal probability density function has been chosen. Figure 2 a) and b) displays the influence of the standard deviation on the coefficient of variation of the simulation algorithm. The figure shows the relative coefficient of variation that is computed by taking as reference the coefficient of variation obtained for a normal density with standard deviation of 0.7. It can be seen that increasing the standard deviation decreases the coefficient of variation for subset simulation, but increases the coefficient of variation for the moving particles algorithm. Thus, in comparison to subset simulation, a smaller standard deviation for the proposal density is required for the moving particles algorithm. This can be explained by the fact that the Markov chain simulations in the moving particles algorithm require a burn-in period and thus a repeated application of the proposal density, which is not the case for subset simulation.
Figure 2: Influence of the standard deviation $\sigma$ of the proposal density. a) Subset simulation. b) Moving particles algorithm.

Figure 3: Moving particles algorithm. a) Influence of the burn-in period. b) Influence of seed selection strategies.
4.3 Moving particles algorithm: burn-in period and seed selection strategy

Figure 3 summarizes the influence of the burn-in period on the efficiency and various seed selection strategies on the accuracy of the moving particles algorithm. For the investigation of the burn-in period, simulations without burn-in have been taken as reference and the relative coefficient of variation is shown together with the increase in the number of function calls. The results underline that a burn-in period is necessary for a reduction of the coefficient of variation. However, a burn-in period of 20 samples as proposed in the literature leads to a high number of performance function calls. It was found that a shorter burn-in period (about 5 samples) is a good compromise between accuracy and effort.

Various seed selection strategies have been tested. Figure 3 b) states the results for two seed selection strategies. Simulations without a seed selection strategy were taken as reference. The results indicate that seed avoidance might be beneficial for the accuracy of the simulation algorithm and that blocking the seed of the Markov chain and the generated sample (the offspring) for further use as seed of a Markov chain might be a viable strategy to increase the efficiency. However, the effect is not very pronounced.

Neither the length of the burn-in period nor the seed selection strategy had a significant influence on the level 2 acceptance rate.

4.4 Moving particles algorithm: application of a low rank metamodel

In order to reduce the number of function calls during the burn-in period of the Markov chains, the application of a metamodel has been investigated. A low rank separated representation with at most 50 polynomial terms of order 5 has been calibrated from the initial samples and continuously updated (as described in [7]) during the simulations. Results have been compared to computations without a metamodel. The relative coefficient of variation and the relative number of function calls have been collected for the test cases and are shown in Figure 4.

As can be seen from Figure 4, the application of a metamodel during the burn-in period of the Markov chains for the moving particles algorithm decreases the number of performance function calls, but increases at the same time the coefficient of variation. Moreover, there is a considerable scatter in the increase of the coefficient of variation.
4.5 Comparison of subset simulation and moving particles algorithm

The efficiency of both algorithms can be compared by setting $N_m = 0.1 N$ for the number of initial samples and $p_0 = 0.1$ as before. In this case, the coefficient of variation of both algorithms will be nearly the same. From equations (6) and (9), one obtains the relationship

$$\frac{N_{\text{sub}}}{N_{\text{mp}}} = \frac{10(1 - 0.39 \log p_F)}{1 - T \log p_F} \quad (10)$$

By setting this expression to one, a burn-in period $T$ can be obtained as a function of the failure probability for which both algorithms would require approximately the same amount of function evaluations. This relationship is depicted in Fig. 5 a). It can be seen that the obtained burn-in period is in the range of values that has been found to be sufficient in the previous section and thus the number of function evaluations for both algorithms is of the same order of magnitude. For small failure probabilities, subset simulation becomes slightly more efficient than the moving particles algorithm, while for larger probabilities of failure, the opposite is the case.

The comparison of the two variants of the moving particles algorithm and subset simulation, Figure 5 b), underlines that both algorithms lead to a similar efficiency. However, while the number of function evaluations is approximately the same, the coefficient of variation for the results from the moving particles algorithm is slightly higher than for subset simulation. As a burn-in period of $T = 5$ has been applied for the moving particles algorithm and the failure probabilities are in the range from $10^{-4}$ to $10^{-6}$, this confirms the results of Figure 5 a).

Finally, both algorithms have been applied to high dimensional test cases, a single-degree-of-freedom oscillator with 1500 random variables and a paraboloid with varying curvatures. Both algorithms yielded results of same accuracy and efficiency. The level 2 acceptance rates were similar as for the low-dimensional test cases.

5 CONCLUSIONS

The aim of the simulation study was to investigate the suitability of two MCMC simulation algorithms, subset simulation and the moving particles algorithm, for off-the-shelf reliability computations.
The following conclusions can be drawn: A large initial sample size is beneficial for the accuracy of both algorithms, but decreases the efficiency in terms of performance function evaluations. Comparing the subset simulation algorithm to the moving particles algorithm, it was found that the subset simulation algorithm needs about a factor of ten more initial samples, as most of the initial samples are discarded during the move to the next subset simulation step. For the componentwise MH transition kernel, both algorithms tolerate a wide range of values for the standard deviation of the normal distribution. In order to obtain accurate results, the standard deviation for subset simulation must be higher than for the moving particles algorithm. A burn-in period is necessary in order to obtain a good accuracy of the moving particles algorithm. A seed selection strategy might increase the accuracy of the moving particles algorithm, but has nearly no influence on the level 2 acceptance rate.

For the efficiency, the differences between the moving particles and the subset simulation algorithm were rather small for the low- as well as the high-dimensional test cases. In conclusion, subset simulation and the moving particles algorithm seem to be well suited for off-the-shelf reliability estimations, but with a different setting for the most important parameters ($\sigma$ and the initial sample size).

REFERENCES


ZERO-VARIANCE SIMULATED ANNEALING FOR BAYESIAN SYSTEM IDENTIFICATION

Peter L. Green¹

¹ Institute for Risk and Uncertainty, School of Engineering, University of Liverpool, Liverpool, L69 7ZF, United Kingdom
e-mail: p.l.green@liverpool.ac.uk

Keywords: Markov chain Monte Carlo, Zero Variance, Hamiltonian Monte Carlo, Simulated Annealing

Abstract. Markov chain Monte Carlo (MCMC) algorithms are a set of methods which allow samples to be generated from generic probability distributions. They have been used to aid the simulation of rare events, the Bayesian system identification of systems which are nonlinear and/or are approached using a Bayesian hierarchical structure and the training of a variety of machine learning algorithms (for example). The current paper discusses the ‘Zero-Variance method’, which can be used to greatly reduce the sample variance of quantities that are estimated using Monte Carlo methods. The ability of this approach to increase the efficiency of gradient based MCMC methods is illustrated. Finally, a Zero-Variance version of the well-known simulated annealing algorithm is employed. The algorithm is demonstrated on the Bayesian system identification of a nonlinear dynamical system.
1 INTRODUCTION

Over recent years Bayesian approaches to system identification have been adopted across a wide range of applications within structural dynamics. Markov chain Monte Carlo (MCMC) algorithms, which can be used to generate samples from generic probability distributions, often form a fundamental component of these methods.

Recent works, for example, have focused on the application of hierarchical Bayesian frameworks to aid the identification of structures subject to changes in ambient temperature and excitation amplitude [1], damage detection from noisy and/or incomplete modal data [2] and the multilevel identification of sets of nominally identical systems [3]. These works either make use of MCMC [2] [3] or cite it as an avenue for future work [1]. MCMC can also be applied in various other contexts, such as the efficient simulation of rare events (subset simulation) [4].

For the current paper, it is sufficient to consider the situation where one wishes to generate samples from the posterior distribution

\[ p(\theta|D) \propto p(D|\theta)p(\theta) \]  

where \( \theta \in \mathbb{R}^{N_\theta} \) is a vector of parameters which are to be inferred from a set of observations, \( D \). It is assumed here that closed-form solutions for the posterior are unavailable (such that it is necessary to generate samples from \( p(\theta|D) \)).

2 ZERO-VARIANCE PRINCIPLE

Say \( g(\theta) \) is a quantity whose expected value, with respect to a target distribution \( \pi(\theta) \), is of interest. In the context of this paper \( \theta \) is a vector of a model’s parameters and the expected value of \( g(\theta) \) is typically estimated using

\[ E_{\pi}[g(\theta)] \approx \frac{1}{N} \sum_{i=1}^{N} g(\theta^i) \]  

where \( \{\theta^1, ..., \theta^N\} \) are samples from the posterior which have been realised using MCMC. This estimate is, of course, subject to statistical error which will reduce if larger \( N \) is used (or if correlations between successive MCMC samples is reduced).

The zero-variance (ZV) principle [5] suggests that \( g(\theta) \) should be transformed into a different function, \( \tilde{g}(\theta) \), whose expected value is still equal to \( E_{\pi}[g(\theta)] \) but whose variance is reduced. \( \tilde{g}(\theta) \) can then be used to estimate the quantity of interest with less statistical error, relative to if \( g(\theta) \) had been used.

This method gets its name from the fact that, ideally, the transformation would lead to \( \tilde{g}(\theta) \) having zero variance (thus eliminating the statistical error). In practice, however, this is usually impossible to achieve and approximations to this ideal transformation must be employed. It should be noted that a very general treatment of the zero-variance principle and the nature of the transformations that can be utilised are outlined in [5] but that this is somewhat beyond the current paper, which aims to investigate the applicability of the ZV method to Bayesian system identification problems within structural dynamics. The current work instead focuses on the transformations that are described in [6] [7] – the reader simply needs to be
aware that these are specific examples of the wide variety of transformations that are potentially available.

2.1 First order transformation

At this point it is convenient to denote the negative score of the target distribution as $\mathbf{z}$:

$$\mathbf{z}(\theta) = -\frac{\partial}{\partial \theta} \log \pi$$

Noting that the expected score is zero, the following transformation is defined:

$$\tilde{g}(\theta) = g(\theta) + \mathbf{a}^T \mathbf{z}(\theta)$$

where $\mathbf{a}$ is a vector of parameters which require identification. The values of $\mathbf{a}$ that minimise $\text{Var}[\tilde{g}]$ are

$$\mathbf{a} = -(\text{Var}[\mathbf{z}])^{-1} \text{Cov}(g, \mathbf{z})$$

where

$$\text{Cov}[g, \mathbf{z}] = E[g \mathbf{z}] - E[g]E[\mathbf{z}]$$

(proved in the appendix). Consequently then, samples from the target must be used to realise a Monte Carlo estimate of $\mathbf{a}$ before the transformation can be applied. This may appear to be a somewhat circular argument, as the resulting estimates of $\mathbf{a}$ will still be subject to statistical error. For now, it should simply be observed that, for the examples investigated in the current paper, the results were very insensitive to the statistical variation of $\mathbf{a}$. A more formal investigation of this property is a topic of future work.

With regard to the reduction in variance that can be achieved, it is straightforward to show that

$$\text{Var}[\tilde{g}] = \text{Var}[g] + 2\mathbf{a}^T \text{Cov}[g, \mathbf{z}] + \mathbf{a}^T \text{Var}[\mathbf{z}] \mathbf{a}$$

such that, after substituting in the optimum values of $\mathbf{a}$ and rearranging, the reduction in variance is found to be

$$\text{Var}[\tilde{g}] - \text{Var}[g] = -\text{Cov}[g, \mathbf{z}]^T (\text{Var}[\mathbf{z}])^{-1} \text{Cov}[g, \mathbf{z}]$$

This confirms that $\text{Var}[\tilde{g}] \leq \text{Var}[g]$, as desired (as $(\text{Var}[\mathbf{z}])^{-1}$ is positive definite).

2.2 Second order transformation

A potential second order transformation, defined in [6] [7], is given by

$$\tilde{g}(\theta) = g(\theta) + \mathbf{a}^T \mathbf{z}(\theta) + \theta^T \mathbf{B} \mathbf{z}(\theta) + \mathbf{c}$$

where $\mathbf{B}$ is symmetric and $\mathbf{C}$ must be chosen such that $E_\pi[\tilde{g}(\theta)] = E_\pi[g(\theta)]$. To define $\mathbf{C}$ it is first noted that
\[ \theta^T B z(\theta) \equiv \sum_i B_{ii} \theta_i z_i + \sum_{i \neq j} B_{ij} \theta_i z_j \]  
(10)

(where summations are taken up to \( N_\theta \) and the notation \( z_i \equiv z(\theta_i) \) has been adopted). As

\[ E_\pi[\theta_i z_i] = 1 \]  
(11)

\[ E_\pi[\theta_i z_j] = 0, \quad i \neq j \]  
(12)

then, for the estimator to be unbiased, it follows that \( C = -\text{Tr}(B) \). The second order transformation therefore becomes

\[ \tilde{g}(\theta) = g(\theta) + \alpha^T z(\theta) + \theta^T B z(\theta) - \text{Tr}(B) \]  
(13)

At this point it is convenient to rearrange all of the transformation’s parameters into a single vector, such that equation (5) can be used to estimate their optimum value. This is achieved as follows:

\[ \theta^T B z(\theta) - \text{Tr}(B) = \sum_i B_{ii} \theta_i z_i + 2 \sum_{j > i} B_{ij} \theta_i z_j - b^T 1 \]  
(14)

\[ = b^T u(\theta) + c^T v(\theta) \]  
(15)

where \( b = \text{diag}(B) \),

\[ u(\theta) = \theta \circ z(\theta) - 1 \]  
(16)

(\( \circ \) is the Hadamard product), the \( \frac{1}{2} (j - 1)(j - 2) + i \) th element of \( c \) is equal to \( B_{ij} (j > i) \) and the \( \frac{1}{2} (j - 1)(j - 2) + i \) th element of \( v(\theta) \) is equal to \( 2 \theta_i z_j \). This allows one to write the second order transformation neatly as

\[ \tilde{g}(\theta) = g(\theta) + \alpha^T w(\theta) \]  
(17)

where

\[ \alpha = \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \quad w(\theta) = \begin{pmatrix} z(\theta) \\ u(\theta) \\ v(\theta) \end{pmatrix} \]  
(18)

In a similar manner to the first order case, the reduction in variance can be shown to be

\[ \text{Var}[\tilde{g}] - \text{Var}[g] = -\text{Cov}[g, w]^T (\text{Var}[w])^{-1} \text{Cov}[g, w] \]  
(19)
2.3 Illustrative examples

The method will first be demonstrated on some simple examples, where samples from the target can be generated directly (without having to resort to MCMC). The interested reader may also like to consult the technical report [7] where other relatively simple examples are discussed. Note that the following only explores the first and second order expansions that were outlined in the previous section.

First, Monte Carlo samples are used to estimate the mean of the distribution $\pi(\theta) \propto \exp\left(-\frac{1}{2}\theta^2\right)$ such that $g(\theta) = \theta$ and $z(\theta) = \theta$. For the first order case, it is straightforward to show that $\text{Var}[z] = 1$ and $\text{Cov}[g, z] = 1$. This implies that the optimum choice of transformation parameter is $\alpha = -1$. For the second order case $\nu(\theta) = 0$ (as $\theta \in \mathbb{R}^1$) and so

$$w(\theta) = \begin{pmatrix} z(\theta) \\ \theta z(\theta) - 1 \end{pmatrix} = \begin{pmatrix} \theta \\ \theta^2 - 1 \end{pmatrix} \Rightarrow \text{Var}[w] = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad (20)$$

Also,

$$\text{Cov}[g, w] = E\left(\begin{pmatrix} \theta^2 \\ \theta(\theta^2 - 1) \end{pmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (21)$$

and so, for the second order case, the optimum transformation parameters are

$$\alpha = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \quad (22)$$

This implies that, in the current example, the second order term has no influence and so the second order ZV expansion will not be able to outperform the first order ZV expansion. This is verified by simulation in Figure 1 where 100 estimates of the mean are realised, each using 100 samples from the target. Monte Carlo estimates of $\alpha$ are used throughout. Relative to standard Monte Carlo estimates, the reduction in the statistical error of the estimates is quite remarkable.
The scenario where the aim is to estimate $E[p(\theta)]$ is now considered. For the first order case $\text{Var}[z] = 1$ and $\text{Cov}[g, z] = E[\theta^3] = 0$ and so $\alpha = 0$ (implying that there is no advantage to using the first order expansion). For the second order expansion, $\text{Var}[w]$ is the same as in equation (20) and

$$
\text{Cov}[g, z] = E\left[ g \sigma^2 \left( \theta^2 - 1 \right) \right] = \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix}
$$

(23)

which gives optimum transformation parameters

$$
\alpha = \begin{pmatrix} 0 \\ -1 \end{pmatrix}
$$

(24)

These results are verified by simulation in Figure 2 where it is shown that, to achieve a reduction in the statistical error of the estimated quantity of interest, the second order transformation must be employed. Again, the reduction in statistical error is impressive.
Figure 2. Estimating the variance of \( p(\theta) = N(\theta; 0,1) \)

3 ZERO-VARIANCE BAYESIAN SYSTEM IDENTIFICATION USING MCMC

In the following the aim is to realise a Bayesian estimate of the nonlinear stiffness, \( k_3 \), of a Duffing oscillator:

\[
\ddot{x} + c\dot{x} + kx + k_3x^3 = F(t)
\]

where \( F(t) \), in this case, was a random excitation generated from a zero-mean Gaussian with unit variance. The training data, \( D \), consisted of the excitation time history as well as noisy observations of the system’s displacement response (shown in Figure 3).

Figure 3. Training data for Duffing oscillator example.

The likelihood was created by assuming the following noise model:
\[ y_i = x_i(\theta) + \epsilon_i, \quad \epsilon_i = N(\epsilon_i; 0, \sigma^2) \]  \hspace{1cm} (26)

where \( \theta = k_3 \) in this case, \( y_i \) denotes the \( i \)th noisy observation of the displacement and the noise variance, \( \sigma^2 \), is assumed to be known. The true value of the nonlinear stiffness was \( k_3 = 100 \text{ N/m}^3 \), while the prior was \( U(k_3; 0, 1000) \). The other model parameters were set equal to \( k = 10 \text{ N/m} \) and \( c = 0.1 \text{ Nm/s} \). Samples from the posterior were generated using a simulated annealing MCMC algorithm whose annealing schedule is self-adaptive and is designed to introduce the information contained in the measurements at a constant rate (where information is measured using the Shannon entropy). This is essentially the algorithm outlined in [9] but applied to the situation where the set of measurement data does not grow with time. Throughout, the score of the target distribution was estimated using a finite difference approximation.

The resulting MCMC samples were used to estimate the mean and variance of the posterior (using both standard Monte Carlo and ZV estimation methods). ZV estimates of the variance were realised by defining

\[ g_1(\theta) = \theta, \quad g_2(\theta) = \theta^2 \]  \hspace{1cm} (27)

before computing

\[ E[g_2] - E^2[g_1] = \bar{g}_2 - \bar{g}_1^2 \]  \hspace{1cm} (28)

For the sake of completeness, the following results are also compared with ‘long run’ Monte Carlo estimates (realised with sample size of 10000).

Estimates of the mean and variance are shown in Figure 4 and Figure 5 respectively. The variance of the ZV estimates are far reduced relative to the Monte Carlo estimates. The ZV estimates are also centered on the results realised using the long run Monte Carlo simulations.

![Figure 4. Estimating the posterior mean of the nonlinear stiffness coefficient, \( k_3 \).](image_url)
4 DISCUSSION

The transformations used throughout this paper all make use of the score of the target distribution. The exploitation of the information contained in the score appears to be an important factor in the successful application of the method. With this in mind it seems sensible that, alongside the zero-variance method, a MCMC algorithm which uses the gradient to improve its performance should be employed (Hamiltonian Monte Carlo, for example, is probably the most well-known example of such an algorithm). This is explored in detail in the paper [7]. For future work the author aims to apply the zero-variance method to samples that are generated using an importance-sampling-based Sequential Monte Carlo sampler (recently applied to growing data sets, within the context of mechanical engineering, in [10]). The potential to combine the statistically efficient ZV method with a sampling algorithm that is well suited to parallelisation is particularly attractive.

The current paper demonstrates examples where, despite the statistical errors associated with estimating the parameters $\alpha$, the zero-variance method greatly outperformed standard Monte Carlo. For future work, it would be useful to gain insight into how far the method can be pushed (in other words: how few samples is it possible to ‘get away with’ when estimating $\alpha$). The current results certainly indicate that it is worth investigating the potential of the ZV method further.

With regard to the simulated annealing algorithm used in the current paper, the self-adaptive annealing schedule relies on estimating the variance of the negative log-likelihood [9]. Interestingly, it was found that the ZV methods investigated here were unable to reduce the statistical error of these estimates and so, as a result, were not able to improve the consistency of the simulated annealing algorithm. For the simple Gaussian examples that were considered here, estimating the variance of the negative log-likelihood would be equivalent to estimating
$E_\pi[\theta^4]$, hinting that a higher order ZV expansion may be required to improve the repeatability of the simulated annealing algorithm (another interesting avenue of future work).

REFERENCES


APPENDIX

To derive the optimum values of $\alpha$ it is first convenient to show that:

$$E[\hat{g}^2] = E[g^2] + 2\alpha^T E[gz] + \alpha^T E[zz^T] \alpha$$

(29)

therefore

$$\frac{\partial}{\partial \alpha} E[\hat{g}^2] = 2E[gz] + 2E[zz^T] \alpha$$

(30)

Also:
\[ E[\hat{g}]^2 = E[g]^2 + 2\alpha^T E[g] E[z] + \alpha^T E[z] E[z^T] \alpha \] (31)

therefore

\[ \frac{\partial}{\partial \alpha} E[\hat{g}]^2 = 2E[g] E[z] + 2E[z] E[z^T] \alpha \] (32)

Consequently, setting

\[ \frac{\partial}{\partial \alpha} \text{Var}[\hat{g}] = 0 \] (33)

it can be shown that

\[ (E[zz^T] - E[z] E[z^T]) \alpha + (E[gz] - E[g] E[z]) = 0 \] (34)

\[ \Rightarrow \alpha = -(\text{Var}[z])^{-1} \text{Cov}[g, z] \] (35)

where

\[ \text{Cov}[g, z] = E[gz] - E[g] E[z] \] (36)
AN EFFICIENT METHODOLOGY FOR THE ANALYSIS AND MODELING OF COMPUTER EXPERIMENTS WITH LARGE NUMBER OF INPUTS

Bertrand Iooss\textsuperscript{1,2} and Amandine Marrel\textsuperscript{3}

\textsuperscript{1}EDF R&D
6 Quai Watier, 78401 Chatou, France
e-mail: bertrand.iooss@edf.fr

\textsuperscript{2} Institut de Mathématiques de Toulouse
31062 Toulouse, France

\textsuperscript{3} CEA, DEN, DER
13108 Saint-Paul-lez-Durance, France
e-mail: amandine.marrel@cea.fr

Keywords: Computer experiments, Uncertainty Quantification, Sensitivity Analysis, Metamodel, Gaussian process

Abstract.

Complex computer codes are often too time expensive to be directly used to perform uncertainty, sensitivity, optimization and robustness analyses. A widely accepted method to circumvent this problem consists in replacing cpu-time expensive computer models by cpu inexpensive mathematical functions, called metamodels. For example, the Gaussian process (Gp) model has shown strong capabilities to solve practical problems, often involving several interlinked issues. However, in case of high dimensional experiments (with typically several tens of inputs), the Gp metamodel building process remains difficult, even unfeasible, and application of variable selection techniques cannot be avoided. In this paper, we present a general methodology allowing to build a Gp metamodel with large number of inputs in a very efficient manner. While our work focused on the Gp metamodel, its principles are fully generic and can be applied to any types of metamodel. The objective is twofold: estimating from a minimal number of computer experiments a highly predictive metamodel. This methodology is successfully applied on an industrial computer code.
1 INTRODUCTION

Quantitative assessment of the uncertainties tainting the results of computer simulations is nowadays a major topic of interest in both industrial and scientific communities. One of the key issues in such studies is to get information about the output when the numerical simulations are expensive to run. For example, in nuclear engineering problems, one often faces up with cpu time consuming numerical models and, in such cases, uncertainty propagation, sensitivity analysis, optimization processing and system robustness analysis become difficult tasks using such models. In order to circumvent this problem, a widely accepted method consists in replacing cpu-time expensive computer models by cpu inexpensive mathematical functions, called metamodels [6]. This solution has been applied extensively and has shown its relevance especially when simulated phenomena are related to a small number of random input variables (see [7] for example).

However, in case of high dimensional numerical experiments (with typically several tens of inputs), depending on the complexity of the underlying numerical model, the metamodel building process remains difficult, even unfeasible. For example, the Gaussian process (Gp) model [22] which has shown strong capabilities to solve practical problems, has some caveats when dealing with high dimensional problems. The main difficulty relies on the estimation of Gp hyperparameters. Manipulating pre-defined or well-adapted Gp kernels (as in [18, 5]) is a current research way, while coupling the estimation procedure with variable selection techniques has been proposed by several authors [24, 16, 25].

In this paper, we pursue the effort on the latter technique by proposing a more rigorous and robust method for building a Gp metamodel with a high-dimensional vector of inputs. First, we clarify the sequence of the different steps of the methodology, while updating their technical core with more relevant statistical techniques. For example, the screening step is raised by the use of recent and powerful techniques in terms of variable selection using a small number of model runs. Second, contrary to the previous works, we do not remove the non-selected inputs from the Gp model, keeping the uncertainty caused by the dimension reduction by using the joint metamodel technique [15]. The integration of this residual uncertainty is important in terms of robustness of subsequent safety studies.

The next section of this paper presents our general methodology. The third, fourth and fifth sections are devoted to a detailed explanation of each of its steps. The last section shows an application of this work on a thermal-hydraulic calculation case simulating accidental scenario in a nuclear reactor. It also gives some prospects of this work.

2 GENERAL METHODOLOGY

The system under study is denoted

\[ Y = g(X_1, \ldots, X_d) \]  

where \( g(\cdot) \) is the numerical model (also called the computer code), whose output \( Y \) and input parameters \( X_1, \ldots, X_d \) belong to some measurable spaces \( \mathcal{Y} \) and \( \mathcal{X}_1, \ldots, \mathcal{X}_d \) respectively. \( X = (X_1, \ldots, X_d) \) is the input vector and we suppose that \( \mathcal{X} = \prod_{k=1}^d \mathcal{X}_k \subset \mathbb{R}^d \) and \( \mathcal{Y} \subset \mathbb{R} \). For a given value of the vector of inputs \( x = (x_1, \ldots, x_d) \in \mathbb{R}^d \), a simulation run of the code yields an observed value \( y = g(x) \).

Our approach consists in four steps:

1. **Step 1: Initial experimental design.** Once the uncertain input variables of the numerical model \( g \) and their variation domain identified, a design of \( n \) experiments is firstly
performed and yields \( n \) model output values. To constitute this learning sample, we use a space-filling design (SFD) of experiments, providing a full coverage of the high-dimensional input space.

2. **Step 2: Screening.** From the learning sample, a screening technique is performed in order to identify the primary influential inputs (PII) on the model output variability. It has been recently shown that screening based on dependence measures [2, 4] or on derivative-based global sensitivity measures [12, 21] are very efficient methods which can be directly applied on a SFD. One of their great interest is that, additionally to their screening job, the sensitivity indices they provide can be quantitatively interpreted. From these screening results, the inputs are then ordered by decreasing PII, for the purpose of the metamodeling step.

3. **Step 3: Joint metamodeling and metamodel validation.** The sorted inputs are successively included in the group of explanatory inputs while the other inputs are considered as a global stochastic (i.e. unknown) input and a joint Gp metamodel is built. At each iteration, a first Gp model, only depending on the explanatory inputs, is built to approximate the mean component of the metamodel. The residual effect of the other inputs is captured using a second Gp model which approximates the variance component as a function of the explanatory inputs. For this, a joint metamodeling procedure is used, as proposed by [15]. Moreover, in order to deal with the large number of inputs, the optimization process, which is required to estimate the hyperparameters of the Gp covariances, uses as a starting point the values estimated at the previous iteration.

The accuracy and prediction capabilities of the metamodel are controlled on a test sample or by cross-validation.

All these steps are described in the next subsections. The obtained metamodel, which requires a negligible calculation time, can then be used to perform global sensitivity analysis, uncertainty propagation (for example through Monte-Carlo simulations) or optimization processing.

3 **STEP 1: INITIAL DESIGN OF EXPERIMENTS**

The objective of the initial sampling step is to investigate the whole variation domain of the uncertain parameters in order to fit a predictive metamodel which approximates as accurately as possible the code in the whole domain of variation of the uncertain parameter, independently from their probabilistic distributions. For this, we use a space-filling design (SFD) of a certain number \( n \) of experiments, providing a full coverage of the high-dimensional input space [6]. This design enables to investigate the domain of variation of the uncertain parameters and provides a learning sample.

Mathematically, this corresponds to the sample \( \{x^{(1)}, \ldots, x^{(n)}\} \) which is performed on the model \( g \). This yields \( n \) model output values denoted \( \{y^{(1)}, \ldots, y^{(n)}\} \) with \( y^{(i)} = g(x^{(i)}) \).

The obtained learning sample is denoted \( (X_s, Y_s) \) with \( X_s = \left[ x^{(1)T}, \ldots, x^{(n)T} \right]^T \) and \( Y_s = \left[ y^{(1)}, \ldots, y^{(n)} \right]^T \). The goal is to build an approximating model of \( g \) using the \( n \)-sample \( (X_s, Y_s) \).

The number \( n \) of simulations is a compromise between the CPU time required for each simulation and the number of input parameters. Some thumb rules propose to choose \( n \) at least as large as 10 times the dimension \( d \) of the input vector [14, 16].
For the SFD type, a Latin Hypercube Sample (LHS) with optimal space-filling and good projection properties [25] would be well adapted. In particular, [6, 3] have shown the importance of ensuring good low-order sub-projection properties. Maximum projection designs [11] or low-centered $L^2$ discrepancy LHS [10] are then particularly well-suited.

Remark 3.1 Note that the input values are sampled uniformly, considering only their variation ranges and not their initial probability distributions. Indeed, our aim is to build a metamodel for a multi-objective purpose (sensitivity analysis, uncertainty propagation, etc...). The input probability distributions will then be used in the sensitivity analysis or uncertainty propagation studies.

4 STEP 2: INITIAL SCREENING

From the learning sample, an initial screening is performed in order to identify the PII and sort them by decreasing order of influence. For this, two possibilities are proposed: one based on dependence measures and another based on derivative-based global sensitivity measures.

4.1 Screening based on dependence measure

[2] and more recently [4] have proposed to use dependence measures for screening purpose, by applying them directly on a SFD. These sensitivity indices are not the classical ones variance-based measures (see [9] for a global review). They consider higher order information about the output behavior in order to provide more detailed information. Among them, the Hilbert-Schmidt independence criterion (HSIC) introduced by [8] builds upon kernel-based approaches for detecting dependence, and more particularly on cross-covariance operators in reproducing kernel Hilbert spaces (RKHS).

If we consider two RKHS $\mathcal{F}_k$ and $\mathcal{G}$ of functions $X_k \rightarrow \mathbb{R}$ and $Y \rightarrow \mathbb{R}$ respectively, the crossed-covariance $C_{X_k, Y}$ operator associated to the joint distribution of $(X_k, Y)$ is the linear operator defined for every $f_{X_k} \in \mathcal{F}_k$ and $g_Y \in \mathcal{G}$ by:

$$\langle f_{X_k}, C_{X_k, Y} g_Y \rangle_{\mathcal{F}_k} = \text{Cov} (f_{X_k}, g_Y).$$

(2)

$C_{X_k, Y}$ generalizes the covariance matrix by representing higher order correlations between $X_k$ and $Y$ through nonlinear kernels. The HSIC criterion is then defined by the Hilbert-Schmidt norm of the cross-covariance operator:

$$\text{HSIC}(X_k, Y)_{\mathcal{F}_k, \mathcal{G}} = \|C_k\|_{HS}^2.$$  

(3)

From this, [2] introduces a normalized version of the HSIC which provides a sensitivity index of $X_k$:

$$R_{\text{HSIC}, k}^2 = \frac{\text{HSIC}(X_k, Y)}{\sqrt{\text{HSIC}(X_k, X_k)\text{HSIC}(Y, Y)}}.$$  

(4)

[8] also propose a Monte Carlo estimator of $\text{HSIC}(X_k, Y)$ and a plug-in estimator can be deduced for $R_{\text{HSIC}, k}^2$. Note that Gaussian kernel functions with empirical estimations of the variance parameter are used in our application (see [8] for details).

Then, from the estimated $R_{\text{HSIC}, k}^2$, independence tests are performed for a screening purpose. The objective is to separate the inputs into two sub-groups, the significant ones and the non-significant ones. For a given input $X_k$, it aims at testing the null hypothesis “$H_0^{(k)}$: $X_k$ and $Y$ are independent”, against its alternative “$H_1^{(k)}$: $X_k$ and $Y$ are dependent”. The significance
level\(^1\) of these tests is hereinafter noted $\alpha$. Several statistical hypothesis tests are available: asymptotic versions, spectral extensions and bootstrap versions for non-asymptotic case. All these tests are described and compared in [4]; a guidance to use them for a screening purpose is also proposed. At the end of the screening step, the inputs selected as significant are also ordered by decreasing $R^2_{\text{HSIC}}$. This order will be used for the sequential metamodel building in step 3.

4.2 Screening based on derivative-based global sensitivity measure

The so-called Derivative-based Global Sensitivity Measures (DGSM) consist in integrating the square derivatives of the model output (with respect to each of the model input) over the domain of the inputs. This kind of indices have been shown to be easily and efficiently estimated by sampling techniques (as Monte Carlo or quasi-Monte Carlo). Several authors have shown the interest of DGSM as a screening technique (see [12] for a review). Indeed, the DGSM interpretation is made easy due to its inequality links with variance-based sensitivity indices, which are easily interpretable [9]. Multiplied by an optimal Poincaré constant, DGSM is a narrow upper bound of the total Sobol’ index [21], whatever the input probability distribution.

One of the main issue for this technique in practical situations is to efficiently estimate the model derivatives as the standard practice based on finite-differences is relatively costly. Indeed, its cost linearly depends on the number of inputs as most of the sensitivity analysis techniques [9]. However, if the reverse (adjoint) mode of the numerical model is available, computing all partial derivatives of the model output has a cost independent on the number of input variables. In this case, the screening step can be performed with a reasonable cpu time cost (with a sample of 100 runs of the adjoint model typically) and is therefore possible even for large-dimensional model. This potentiality has been recently applied in [20] which studies a model with 40 inputs and uses automatic differentiation in order to obtain the adjoint model (which has a cost of two times the direct model). On this example, [21] have shown the relevance of DGSM for a screening purpose, which also provides a quantitative interpretation of sensitivity indices.

5 STEP 3: JOINT GP METAMODEL WITH SEQUENTIAL BUILDING PROCESS

Among all the metamodel-based solutions (polynomials, splines, neural networks, etc.), we focus our attention on the Gaussian process (Gp) regression, which extends the kriging principles of geostatistics to computer experiments by considering the correlation between two responses of a computer code depending on the distance between input variables. The Gp-based metamodel presents some real advantages compared to other metamodels: exact interpolation property, simple analytical formulations of the predictor, availability of the mean squared error of the predictions and the proved efficiency of the model [22].

However, for its application to complex industrial problems, developing a robust implementation methodology is required. Indeed, fitting a Gp model implies the estimation of several hyperparameters involved in the covariance function. In complex situations (e.g. large number of inputs), some difficulties can arise from the parameter estimation procedure (instability, high number of hyperparameters, see [16] for example). To tackle this issue, we propose a progressive estimation procedure which combines the result of the previous screening step and a joint Gp approach [15].

\(^1\)The significance level of a statistical hypothesis test is the rate of the type I error which corresponds to the rejection of the null hypothesis $H_0$ when it is true.
5.1 Successive inclusion of explanatory variables

At the end of the screening step, the inputs selected as significant are ordered by decreasing influence. The sorted inputs thus obtained are successively included in the group of explanatory inputs. At the $j^{th}$ iteration, only the $j$ first sorted inputs are considered as explanatory input variables while all the remaining inputs are included in a single macro-parameter. This macro-parameter is considered as an uncontrollable parameter (i.e. a stochastic parameter, notion detailed in section 5.2).

At this stage, a joint Gp metamodel is then built with the $j$ explanatory inputs, following the procedure described in [15] and summarized in the next subsection. A numerical optimization is performed to estimate the parameters of the joint metamodel (covariance and variance parameters). In order to improve the robustness of the optimization process, the estimated hyperparameters obtained at the $(j-1)^{th}$ iteration are used, as starting points for the optimization algorithm. This procedure is repeated until the inclusion of all the significant input variables. Note that this sequential process is directly adapted from the one proposed by [16].

5.2 Joint Gp metamodel

In the framework of stochastic computer codes, [26] proposed to model the mean and dispersion of the code output by two interlinked Generalized Linear Models (GLM), called “joint GLM”. [15] extends this approach to several nonparametric models and obtains the best results with two interlinked Gp models, called “joint Gp”. In this case, the stochastic input is considered as an uncontrollable parameter denoted $X_\varepsilon$ (i.e. governed by a seed variable).

We extend this approach to a group of non-explanatory variables. More precisely, the input variables $X = (X_1, \ldots, X_d)$ are divided in two subgroups: the explanatory ones denoted $X_{\text{exp}}$ and the others denoted $X_{\varepsilon}$. The output is thus defined by $y = g(X_{\text{exp}}, X_{\varepsilon})$. Under this hypothesis, the joint metamodeling approach yields building two metamodels, one for the mean $Y_m$ and another for the dispersion component $Y_d$:

$$Y_m(X_{\text{exp}}) = \mathbb{E}(Y|X_{\text{exp}})$$  \hspace{1cm} (5)

$$Y_d(X_{\text{exp}}) = \text{Var}(Y|X_{\text{exp}}) = \mathbb{E}[(Y - Y_m(X_{\text{exp}}))^2|X_{\text{exp}}] .$$  \hspace{1cm} (6)

To fit these mean and dispersion components, we propose to use the methodology proposed by [15]. First, an initial Gp denoted $G_{p_{m,1}}$ is estimated for the mean component with homoscedastic nugget effect. A nugget effect is required to relax the interpolation property of the Gp metamodel, which would yield zero residuals for the whole learning sample. Then, a second Gp, denoted $G_{p_{v,1}}$, is built for the dispersion component with, here also, an homoscedastic nugget effect. $G_{p_{v,1}}$ is fitted on the squared residuals from the predictor of $G_{p_{m,1}}$. Its predictor is considered as an estimator of the dispersion component. The predictor of $G_{p_{v,1}}$ provides an estimation of the dispersion at each point. It is thus considered as the value of the heteroscedastic nugget effect: the homoscedastic hypothesis is removed. A new Gp, $G_{p_{m,2}}$, is fitted on data, with the estimated heteroscedastic nugget. Finally, the Gp on the dispersion component is updated from $G_{p_{m,2}}$ following the same methodology as the one $G_{p_{v,1}}$.

Remark 5.1 Note that some parametric choices are made for all the Gp metamodels: a constant trend and a Matén stationary anisotropic covariance are chosen. All the hyperparameters (covariance parameters) and the nugget effect (when homoscedastic hypothesis is done) are estimated by maximum likelihood optimization process.
5.3 Assessment of metamodel accuracy

To evaluate the accuracy of the metamodel, we use the predictivity coefficient $Q^2$:

$$ Q^2 = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} (y^{(i)} - \hat{y}^{(i)})^2}{\sum_{i=1}^{n_{\text{test}}} (y^{(i)} - \frac{1}{n_{\text{test}} \sum_{i=1}^{n_{\text{test}}} y^{(i)}})^2} $$

(7)

where $(x^{(i)})_{1 \leq i \leq n_{\text{test}}}$ is a test sample, $(y^{(i)})_{1 \leq i \leq n_{\text{test}}}$ are the corresponding observed outputs and $(\hat{y}^{(i)})_{1 \leq i \leq n_{\text{test}}}$ are the metamodel predictions. $Q^2$ corresponds to the coefficient of determination in prediction and can be computed on a test sample independent from the learning sample or by cross-validation on the learning sample. The closer to one the $Q^2$, the better the accuracy of the metamodel. Note that, in our sequential building process (cf. Section 5.1), the $Q^2$ coefficient metamodel is computed at each iteration.

In the case where the model provides the adjoint code (see Section 4.2), the gradient evaluations could be integrated in the metamodel building. For this, the co-kriging principle could be adapted to the joint metamodel approach.

6 APPLICATION TO A THERMAL-HYDRAULIC COMPUTER CODE

6.1 Description of the use-case

Our use-case consists in thermal-hydraulic computer experiments, typically used in support of regulatory work and nuclear power plant design and operation. Indeed, some safety analysis considers the so-called “Loss Of Coolant Accident” (LOCA), which takes into account a double-ended guillotine break with a specific size piping rupture. It is modeled with code CATHARE 2.V2.5 which simulated the thermalhydraulic responses during a LOCA in a Pressurized water Reactor [17].

In this use-case, 27 scalar input variables of CATHARE are uncertain. In our problem, they are defined by their minimum and maximum. They correspond to various system parameters as initial conditions, boundary conditions, some critical flowrates, interfacial friction coefficients, condensation coefficients, ... The output variable of interest is a single scalar which is the maximal peak cladding temperature during the accident transient. Our objective with this use-case is to provide a good metamodel to the safety engineers. Indeed, the cpu-time cost of this computer code is too important to develop all the statistical analysis required in a safety study only using direct calculations of the computer code. A metamodel would allow to develop more complete and robust demonstration.

1000 CATHARE simulations of this test case have been provided following a space-filling LHS with good projection properties (see Section 3) as the design of experiments. In this test case, the adjoint model is not available and the derivatives of the model output are therefore not computed because of their costs. The screening step will then be based on HSIC, obtained from the inputs-output sample.

In order to test it, our overall methodology is applied with different sizes of the learning sample: $n = 200, 400, 600$ and $800$. In each case, the remaining simulations (from the 1000 we have) are used as a test sample in order to compute the metamodel predictivity.

6.2 Screening step with HSIC

The normalized HSIC coefficients are computed for the different learning sample sizes. Similar results are obtained. Four variables are identified as the most influential: $X_{10}$ (HSIC
Approximately 30\%, $X_{12}$ and $X_{13}$ (HSIC $\approx$ 14\%) and $X_{22}$ (HSIC $\approx$ 9\%). $X_{14}$, $X_{15}$ and $X_{2}$ have also a significant but lower influence (HSIC around 5\%). Thus, statistical significance tests (asymptotic version with $\alpha = 10\%$) have selected these 7 inputs. The estimated HSIC and the results of significant tests are relatively stable and independent from the learning sample size, only one or two additional variables with a very low HSIC ($< 2\%$) are selected for the smallest sample size. This confirms the robustness of the HSIC indices and the associated significance tests for qualitative sorting and screening purpose.

For each learning sample size, the significant inputs are considered as the explanatory variables in the joint metamodel and will be successively included in the building process. The other variables are joined in the uncontrollable parameter.

### 6.3 Joint Gp

From the HSIC-based screening results, the joint Gp metamodel is built following the sequential process described in Section 5. The simple Gp metamodel with all the 27 inputs as explanatory variables is also built, without any sequential approach. To assess the accuracy of the different metamodels, the predictivity coefficient $Q^2$ is computed by cross-validation (leave-one-out process) and on the test sample composed of the remaining simulations. The $Q^2$ obtained with the different metamodels are presented in Table 1. Note that the same optimizer is used to estimate the hyperparameters by maximum likelihood, in order to allow for a fair comparison.

<table>
<thead>
<tr>
<th></th>
<th>Joint Gp with sequential approach</th>
<th>Simple Gp without sequential approach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Q^2$ on test sample</td>
<td>$Q^2$ by cross-validation</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>0.82</td>
<td>0.81</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>0.82</td>
<td>0.85</td>
</tr>
<tr>
<td>$n = 600$</td>
<td>0.86</td>
<td>0.89</td>
</tr>
<tr>
<td>$n = 600$</td>
<td>0.87</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Table 1: Comparison of Gp metamodel predictivity for different sizes $n$ of learning sample and different building processes.

The joint Gp with a sequential building process outperforms the simple Gp directly built with the 27 input variables, especially for the lower learning sample sizes. On average, the $Q^2$ is improved between 3\% to 9\%. Thus, the proposed methodology allows a more robust metamodel building with a high-dimensional vector of inputs, even with small sample sizes. Moreover, even if it is not used and illustrated in this application, the dispersion component of the joint metamodel takes into account the uncertainty due to the non-significant inputs. This residual uncertainty, although low, is kept by using the joint metamodel technique. It appears in the mean squared error of the metamodel predictions and could be integrated in subsequent sensitivity or uncertainty propagation studies.

### 6.4 Work continuation and prospects

Using the fitted joint Gp metamodel, several statistical analysis, not feasible with the numerical model due to its computational cost, are now accessible. First, variance-based sensitivity analysis using Sobol’ indices can be fully expressed using a Gp model [16, 13]. This would provide a fine determination of the critical parameters whose uncertainty has to be reduced.
Second, we are particularly interested by the estimation of high quantile (at the order of 95% to 99%) of the model output temperature. In nuclear safety, methods of conservative computation of quantiles [19] have been largely studied. However, several complementary information are often useful and are not accessible in a high-dimensional context. Then, we expect the Gp metamodel can help to access this information. For instance, quantile-based sensitivity analysis [1] and quantile robustness analysis (using the sensitivity indices called PLI [23]) are fully devoted to quantile. Their relevance to support safety analysis seems promising.

7 Acknowledgments

We are grateful to Henri Geiser and Thibault Delage who performed the computations of the CATHARE code.

REFERENCES


ADAPTIVE EMULATION-BASED RELIABILITY ANALYSIS

P. O. Hristov\textsuperscript{1}, F. A. DiazDelaO\textsuperscript{1}, K. J. Kubiak\textsuperscript{2} and U. Farooq\textsuperscript{3}

\textsuperscript{1}Institute for Risk and Uncertainty, University of Liverpool
Liverpool L69 7ZF, United Kingdom
e-mail: \{sgphrist, fado\}@liv.ac.uk

\textsuperscript{2}School of Computing and Engineering, University of Huddersfield
Queensgate, Huddersfield, HD1 3DH
e-mail: k.kubiak@hud.ac.uk

\textsuperscript{3}Parker Hannifin Manufacturing (UK) Ltd.
e-mail: umer.farooq@parker.com

\textbf{Keywords:} Efficient reliability analysis, Gaussian process emulation, Subset simulation.

\textbf{Abstract.} This paper presents an approximation method for performing reliability analysis with high fidelity computer codes at a reasonable computational cost. Complex models are common in science and engineering due to their ability to substitute costly and some times infeasible practical experiments. These models, however, suffer from high computational cost. This causes problems when performing sampling-based reliability analysis, since the failure modes of the system typically occupy a small region of the input space and thus relatively large sample sizes are required for the accurate estimation of their characteristics. The sequential sampling method proposed in this article, combines Gaussian process-based optimisation and Subset Simulation. Gaussian process emulators construct a statistical approximation to the output of the original code, which is both affordable to use and has its own measure of predictive uncertainty. Subset Simulation is used to efficiently populate those regions of the initial approximation which are likely to lead to the performance function exceeding a predefined critical threshold. Among all samples, the ones that are likely to contribute most to increasing the quality of the surrogate in the vicinity of the failure regions are selected, using Bayesian optimisation methods. The iterative nature of the method ensures that an arbitrarily accurate approximation of the failure region in performance space is developed at a reasonable computational cost. The presented method is applied to a number of benchmark problems.
1 INTRODUCTION

Reliability analysis, in the most general sense, is concerned with the calculation of a probability of failure, \( p_F \) and the identification of the failure domain, \( F \) of a function \( y = \eta(x) \).

\[
p_F = \int_F \eta(x) dx
\]  

(1)

The most straightforward method to evaluate the multidimensional integral in (1) is direct Monte Carlo (DMC) sampling. The accuracy of DMC estimates increases with the number of samples. For a typical system the failure region, \( F \) is small with respect to its input domain and is considered a rare event. Thus, many thousands of evaluations of the code have to be performed to ensure that \( F \) is populated with sufficient number of samples to allow a reasonably accurate estimation of its properties. However, engineering models are in general computationally expensive to evaluate and thus it becomes infeasible to use DMC methods with them. A method widely used in engineering for performing reliability analysis is Subset Simulation (SuS) [1]. Despite the significant improvement SuS brings over DMC, it could still be quite expensive to use it directly with the code. In order to address the computational cost of the simulator, one can build a surrogate model for the output of the code. There exists a large number of methodologies for approximating the output of expensive codes, see e.g. [2]. A well-established approach is Gaussian process emulation (GPE), which builds a statistical approximation to the output of the code. Once the GPE is built it can be used as an inexpensive substitute for the simulator, on which reliability analysis can be performed. The issue with such use of the surrogate is that it is typically trained via some space-filling sampling plan which aims at exploring the input domain of the simulator with as few points as possible. Such sampling plans are based on Monte Carlo methods and usually select points in high probability regions. This is tantamount to saying that the surrogate that needs to be used to predict rare events is built on frequent events. The topic of using surrogate models to perform reliability analysis is not a new one and there are a number of frameworks proposed to reduce the cost of the computations. It should be stressed that the reliability analysis literature is as vast as the number of methods that feature in it. Therefore a review of the techniques which are related to the procedure presented in this article is provided. Most, if not all reviewed work relies on sequential sampling for the gradual improvement of the surrogate in the vicinity of the failure domain. It is worth mentioning that some authors such as [3][4] have explored those strategies for the accurate estimation of target regions in general. They consider that the emulator should only be refined in the regions of interest and that in order to obtain accurate reliability estimates a good quality surrogate is all that is needed. An opposing view is given by [5] where the authors work directly with the reliability estimates as a measure of analysis quality. The present article agrees with and extends the former idea.

A short literature review focused on reliability analysis is provided below. In general, most approaches follow a framework which is composed of a sampling rule, utility function, stopping criterion and any other specific details. The authors of [6] use constraint boundary sampling to select improvement points. A combination between a MCMC sampling and k-means clustering is used to select new data points in [7]. An approach based on the first order reliability method is used in [8] and a probabilistic classification function is presented in [9]. In [10] the authors derive a stepwise uncertainty reduction (SUR) methodology based on expected improvement (EI) [11], but formulated from a Bayesian risk perspective. They use SUR to select new design points to improve the surrogate. Other methods that use EI or EI-based strategies are [12][13][14]. Other previously used utility functions include, the U-function [13][15], and the
improved U-function [16], least improvement function [17] and an unnamed expression in [18].
All approaches based on a utility function, except [14] search the entire input space for a can-
didate point that maximizes that function and add it to the training plan for the next iteration
of the algorithm. Furthermore, many of the aforementioned strategies rely on a pre-generated
population of samples (e.g. [13]), which could prove to be a suboptimal approach. Last from a
sampling point of view, the majority of the methods sample one point at a time, which, given
the dynamical nature of sequential sampling could either miss important regions of the domain,
or slow down convergence. These issue are addressed in the present article. The other major part
of all adaptive algorithms is the stopping condition. This ranges from the use of reliability ind-
dices [7,8] through error in the estimation of the failure probability [5,13,15,16,17] and forms
of measure of the discrepancy between the GPE predictions and code observations [4,6,9,18] to
thresholds on the learning function [3,12,14]. Most frameworks use some form of statistic re-
lated to the surrogate, which, depending on the use and complexity of the problem, could prove
insufficiently robust. In this paper a stopping condition which relies implicitly on the similarity
between the surrogate and the model is proposed to terminate the learning process.

Finally, Gaussian process emulation is not the only surrogate used in reliability analysis.
Among others, general response surfaces [19,20,21,22], neural networks [23] and support
vector machines [24] have been used.

The aim of this paper is to propose an improved efficient algorithm for performing reliability
analysis with complex computer codes.

The remainder of the article is structured as follows: Section 2 briefly introduces Subset
Simulation. Section 3 summarizes the theory behind Gaussian process emulation. Section 4
introduces the proposed method and Section 5 demonstrates the performance of the algorithm
before any conclusions are drawn in Section 6.

2 SUBSET SIMULATION

One very important problem in engineering is the estimation of the probability of failure, \( p_F \)
of a system given in Eq. (1). In the context of numerical simulations failure can be defined
as the scenario where a response variable (output) of the model, exceeds some threshold of
acceptable system behaviour. The output, \( y \) is related to the input variables, \( x \in \mathcal{X} \subset \mathbb{R}^d \), via
some mapping provided by the model,
\[
y = \eta(x)
\]  
(2)
thus the failure domain is defined as the values of \( x \) which cause the system response, \( y \) to exceed some critical value \( y_c \)
\[
F = \{ x : \eta(x) > y_c \}
\]  
(3)
Estimating \( p_F \) is associated with sampling from \( F \). Usually, for a well designed system the true
value of \( p_F \) is very small, that is, \( F \) is a rare event. Also, a typical model has a high dimensional
input space and often the failure domain of that space is disjoint, so sampling from it poses a
significant challenge. SuS [1] aims to divide the rare event into a series of nested less-rare
events.
\[
F \subset F_m \subset F_{m-1} \subset \ldots \subset F_1
\]  
(4)
In Eq. (4) \( F_1 \) is a relatively frequent event. Given that sequence, it can be shown that the prob-
ability of the rare event \( F \) could be expressed as a product of larger conditional probabilities:
\[
\mathbb{P}(F) = \mathbb{P}(F_1) \cdot \mathbb{P}(F_1|F_2) \cdot \ldots \cdot \mathbb{P}(F|F_m) = p_F
\]  
(5)
Beginning from the unconditional level $F_1$, the algorithm “probes” the input space $\mathcal{X}$ via direct Monte Carlo sampling. Then, based on the values of $y$ in Eq. (2) it constructs the first intermediate failure threshold, $y^*_1 < y_c$, defining a “relaxed failure domain”, $F_1$. SuS then populates $F_1$ using a MCMC algorithm. The generation of intermediate levels continues until a predefined number of samples lie in the true failure domain $F$. At the end of the algorithm an estimate of the complementary CDF (CCDF) of the response function is generated [25].

3 GAUSSIAN PROCESS EMULATION

Simulators used to model complex scientific phenomena are usually very computationally expensive. This is to say that a single evaluation of the code’s output at a given set of input values takes sufficiently long time, as to prohibit any type of analysis which requires a large number of model runs. These include finite element modelling, computational fluid dynamics and other ubiquitously used simulation tools. Even though SuS requires fewer samples than DMC to estimate the failure probability, it still relies on such amounts of simulation runs which result in very large computational times. Clearly, the analysis cannot be carried out using the code directly. In such cases it is common to use a less expensive approximation of the code out-

\[ \eta(x) = h(x)^T \beta + Z(x) \]  

where $\eta(x)$ is the simulator output as a function of its inputs, $h(x)^T$ is a known function of the inputs, $\beta$ is a vector of unknown coefficients and $Z(x)$ is a Gaussian process with zero mean and covariance $\sigma^2c(x, x'; \psi)$. The function $h(x)$ should express any expert opinion about the form of the simulator output and together with the parameter $\beta$ reflects its overall trend. In practice, however, the trend is often taken to be constant as $h(x) = 1$ and $\beta \in \mathbb{R}$, charging the Gaussian process in Eq. (6) with the responsibility of capturing the behaviour of the underlying function [26]. In the formulation above, $\sigma^2$ is a scale parameter and $\psi$ is a parameter specifying the behaviour of the correlation function, $c(\cdot, \cdot; \psi)$.

Using Gaussian process emulation, a posterior probability distribution for the mean of the computer code’s output can be constructed, conditional on a relatively small number of simulator runs, $y$ and the parameter estimators, $\hat{\theta} = \{\hat{\beta}, \hat{\sigma}^2, \hat{\psi}\}$. It can be shown that at any unobserved point, $x^*$ this distribution has the form:

\[ \eta(x^*)|y, \hat{\theta}, \sim \mathcal{N}(m(\cdot), C(\cdot, \cdot)) \]  

with posterior predictive mean (also called a surrogate):

\[ m(x^*) = \hat{\beta} + t(x^*)^T C^{-1}(y - 1\hat{\beta}) \]  

and posterior predictive variance:

\[ C(x^*, x^*) = \hat{\sigma}^2(c(x^*, x^*; \hat{\psi}) - t(x^*)^T C^{-1}t(x^*)) \]  

In Eq. (8) and Eq. (9), $C \in \mathbb{R}^{n \times n}$ is such that $C_{ij} = c(x_i, x_j; \hat{\psi})$; $t(x^*) \in \mathbb{R}^n$ such that $t(x^*) = (c(x^*, x_1; \hat{\psi}), \ldots, c(x^*, x_n; \hat{\psi}))^T$; and $1 \in \mathbb{R}^n$ such that $1 = (1, \ldots, 1)^T$. The process of learning $\theta$ from observed data is referred to as training and is well described in [2] from a classical prospective or in [27, 28] from Bayesian standpoint. Once the emulator is trained, its posterior distribution can be sampled many times at an affordable cost to provide data for various analyses.
4 GAUSSIAN PROCESS SUBSET SIMULATION

4.1 Initialization and sampling

In order to reliably use the emulator it needs to be of sufficient quality around the estimated failure regions. This is not usually the case with GP approximations built using data from the high probability regions. For the physical model it is assumed that the critical failure threshold, $y_c$, is known (given a priori) and sensible (i.e. there is a set of values of $x$ for which $\eta(x) > y_c$). The algorithm starts by building a GP emulator based on a parsimonious set of data points, selected according to a space-filling strategy (e.g. Latin hypercube sampling (LHS) [29]). The GPE can be validated to check if there are any large discrepancies between the emulator and simulator as in [30]. This is done to ensure that the initial approximation is of a reasonable overall quality. It is very likely that the original threshold is not reachable from the first GPE which learned about the model from frequent data. That is to say,

$$F_1 = \{x : \mathbb{E}_1[\eta(x)|y] > y_c\} = \emptyset \quad (10)$$

where $F_j$ is the failure domain according to the $j^{th}$ emulator and $\mathbb{E}_j[\cdot]$ is its predictive mean, whose functional form is given in Eq. (8). In general, for a high-dimensional model it will be feasible, but slow to search the input domain exhaustively, since a lot of samples are needed to uncover the location of the failure modes if such are present. Computationally this translates to the calculation of large number of distances between points. Therefore, it is proposed to use SuS to sample from the posterior predictive mean of the emulator. As outlined in Section 2, SuS converges when a predefined number of data points lie in the failure domain. If it is indeed the case that $F_1 = \emptyset$, the algorithm will be unable to naturally converge due to all candidate samples being rejected. Then, an alternative “failure level”, $y_{c1}^{GP}$ could be set for which $\mathbb{P}(\mathbb{E}_1[\eta(x)|y] > y_{c1}^{GP}) > 0$. This approach gives rise to an intermediate emulator failure domain with respect to $y_{c1}^{GP}$, denoted as $F_1^{GP}$. This domain and its associated probability may still be a rare event and thus SuS is used to populate it efficiently. The purpose here is to sample from $F_1^{GP}$ rather than to estimate the probabilities of failure (conditional or otherwise). The emulator is still a very efficient approximation of the code and thus in theory one can search it exhaustively. However, a preference is given SuS as it leads to a quick and reliable convergence, especially in high dimensions. The algorithm can be run with a fairly large number of samples (a lot less than if exhaustive search is performed) at each level in order to ensure that even truly rare subregions of $F_1^{GP}$ are populated. This raises the question of how to calculate $y_{c1}^{GP}$ such that SuS converges, while at the same time it explores potentially interesting regions of the input space. The candidates vary from a single significant point (e.g. minimum or maximum) through moment estimators (e.g. sample mean) to more sophisticated equiprobable measures (e.g. sample percentiles). The criterion that was found to give satisfactory results is presented in Eq. (11). Essentially, the idea is to select the current threshold as the average of all training responses that are lying above the previous critical level and then take the average between the resulting quantity and the previous level value. This method ensures that the current threshold will lie above the previous one, while also safeguarding against setting it too high and discarding regions that may lead to system failure.

$$y_{c1}^{GP} = \begin{cases} \frac{1}{N} \sum_{i=1}^{N} y_i & \text{if } j = 1 \\ \frac{1}{2} \left[ \sum_{i=1}^{N} y_i \mathbb{I}(y_i > y_{c(j-1)}^{GP}) \right] + y_{c(j-1)}^{GP} & \text{if } j > 1 \end{cases} \quad (11)$$

In Eq. (11), $y_i$ are all $N$ training responses and $\mathbb{I}(\cdot)$ is an indicator function.
4.2 Sample selection

When SuS converges, it populates the space with points whose predicted response lies above the current failure level $y_{c,GP}$. Among all samples $X_{F,j}^{GP}$ that lie in $F_j^{GP}$ some additional point/s, $x_{add}$ at which to sample the model have to be selected before fitting the next emulator prediction. Since the model is expensive, these points have to be prudently selected such that they maximize the amount of information gain from sampling at their locations. A popular choice for selecting the coordinates is the expected improvement (EI) function \cite{11}.

\begin{equation}
E[I(x)]_j = (y_c - E_j[\eta(x)|y]) \Phi \left( \frac{y_c - E_j[\eta(x)|y]}{V_j[\eta(x)|y]} \right) + V_j[\eta(x)|y] \phi \left( \frac{y_c - E_j[\eta(x)|y]}{V_j[\eta(x)|y]} \right) \tag{12}
\end{equation}

The EI expression as used in this framework is given in Eq. (12). Following the notation above, the subscript $j$ denotes information regarding the $j^{th}$ emulator. This includes $V_j[\cdot]$ which is the posterior variance of the $j^{th}$ emulator given in its functional form in Eq. (9). The symbols $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative and probability distribution functions of a standard normal random variable, respectively. Expected improvement is a strategy that balances exploitation of the emulator mean and exploration of the design space based on its variance. Gaussian process emulators are particularly suited for use with EI, since a predictive variance is generated as part of sampling the posterior distribution.

$$x_{add} = \underset{x^{F_j}}{\text{argmax}} E[I(X^{F_j})]_j \tag{13}$$

Expected improvement can be run on each point in $F_j^{GP}$ and the one that maximizes it will be selected as the one that is expected to bring the greatest improvement in the quality of the next level GPE. However, applying expected improvement directly to the data poses the risk of neglecting subregions in the intermediate emulator failure domain, $F_j^{GP}$. Unless $F_j^{GP}$ is definitely not disjoint, the presence of separate modes has to be accounted for. This could be achieved by, first identifying the structure of $F_j^{GP}$, detecting any modes and calculating EI on the samples in each mode. A clustering algorithm (here DBSCAN \cite{31}) is used to discover the separate failure sub-domains. Ultimately, the design plan for the emulator at the next level is composed of the current design plus the new points, formally $D_{j+1} = D_j \cup \{x_{add}, \eta(x_{add})\}$, where $x_{add}$ is given in Eq. (13).

Adaptively improving the approximation from the GP will represent the failure regions with increasing accuracy. However, due to the generic nature of both Subset Simulation and expected improvement, new points can be added in regions where the GPE is already good enough for the purposes of estimation of the failure characteristics. Thus a condition is needed which dictates how to choose improvement points. The posterior predictive variance of the emulator is a good indication of the local quality of the approximation. Since the GPE is an interpolator, at any given training point $V_j[\eta(x)|y] = 0$. Therefore, if at any level of the GPE improvement the value of the posterior predictive variance for a sample point lies below a certain level $\varepsilon$, that point should not be added to the design for the next level. This strategy prevents the algorithm from suggesting new samples in regions where the emulator is already doing well and thus saving potentially appreciable time for code evaluations. The correct value of $\varepsilon$ is problem dependent since it scales with the output of the simulator. Of course it could be set arbitrarily small, but attention needs to be paid to the trade-off between final emulator accuracy and computational cost.
4.3 Stopping condition

As outlined above, the GPE could be made to reflect the failure regions of the true function with an arbitrary precision. However to keep the procedure efficient it needs to be refined just enough for the purposes of the underlying analysis. In order to stop the iterative generation of predictions a rule inspired by SuS is proposed. Consider running subset simulation with the true function. By design SuS will stop generating new levels once a sufficient number of samples from the last level lie in the failure domain, \( F \). It follows that a necessary and sufficient criterion for the accuracy of the emulator is the ability of SuS to generate the same number of samples that belong to the failure domain and have \( \mathbb{V}_j \eta(x) | y | < \varepsilon \). If this condition is satisfied, SuS will not be able to differentiate the model from the emulator.

5 NUMERICAL EXPERIMENTS

In this section, the performance of the algorithm is demonstrated with three benchmark problems. These were chosen to test different aspects of GPSS. The function in Section 5.1 presents a challenging limit state contour for the GPE. The problem in Section 5.2 tests the ability of GPSS to deal with disjoint failure domains and Section 5.3 explores the performance of the algorithm in high dimensions.

5.1 Four branch series system

The four branch series system is popular in the reliability literature \cite{17}. The performance function has the following form:

\[
\eta(x) = \min \left\{ 3 + 0.1(x_1 - x_2)^2 - (x_1 + x_2)/\sqrt{2}, 3 + 0.1(x_1 - x_2)^2 + (x_1 + x_2)/\sqrt{2}, (x_1 - x_2) + 6/\sqrt{2}, (x_2 - x_1) + 6/\sqrt{2} \right\}
\]

(14)

where \( x = [x_1, x_2] \in [-5, 5]^2 \). The two inputs are considered independent and historically have been subject to standard normal distribution. However here a uniformly distributed sample is assumed to reduce the concentration of initial training points in the failure domain. The GPE was trained with \( n = 20 \) LHS samples. GPSS with \( \varepsilon = 10^{-4} \) was used with this function on a failure threshold, \( y_c = 2 \). At this level the associated probability of failure, \( p_F \approx 7.44 \times 10^{-2} \). The challenge associated with this function is the complex limit state contour which has some nearly discontinuous points. The progression of the estimation of the surface via GPSS is shown in Figure 1. The coefficient of variation and relative error were estimated from 100 runs of SuS on the improved surface as \( \delta_{p_F} = 4.2\% \) and \( \Delta_{p_F} = 1.5\% \), respectively. The estimated mean failure probability was \( \bar{p}_F = 7.62 \times 10^{-2} \) based on the same 100 runs. A comparison between the CCDF from SuS and CCDF from GPSS is shown in Figure 2. The nature of the local quality of the surrogate is evident from the variations in the first portion of the curve. Indeed this is as intended, since there is no attempt to improve the surrogate globally.

5.2 Mixture of Gaussians

This function was created to test the robustness of the GPSS algorithm. The choice of a Gaussian function as constructive element is justified by the fact that this work is not trying to challenge the emulator itself, but to improve it such that it enables the accurate prediction of the characteristics of the failure domain.
The function has the form:

\[ \eta(x) = \frac{10^2}{2} \left[ a\phi(x_A) + b\phi(x_B) + c\phi(x_C) + d\phi(x_D) \right] \]  \hspace{1cm} (15)

\[ x_A = 10(x - 1/4) \]
\[ x_B = 10(x - 3/4) \]
\[ x_C = [10(x_1 - 3/4), 10(x_2 - 3/4)] \]
\[ x_D = [10(x_1 - 1/3), 10(x_2 - 5/6)] \]

In Eq. (15) \( x = [x_1, x_2] \in [0, 1]^2 \) and \( \phi(\cdot) \) is the standard normal PDF. The fractions in the expressions for \( x_A \ldots x_D \) are location parameters and can be chosen arbitrarily. The constants \( a, b, c \) and \( d \) are used to select how many peaks belong to the failure domain. For this experiment \( a = b = 1, c = 0.85 \) and \( d = 1.1 \), meaning that peak \( c \) is just outside of failure level set as \( y_c = 7.9 \). The associated failure probability is \( p_F = 7.27 \times 10^{-3} \).
The GPE was initially trained with \( n = 20 \) LHS points. The variance threshold was set to, \( \varepsilon = 10^{-5} \). Figure 3 shows the process of discovery of the failure domain. The plots in the first row show the contour values of the intermediate levels \( y_{c_{(j-1)}}^{GP} \), in Eq. (11) (red line), because \( y_c \) was not accessible at these GPE stages. During most of the process points were added in all three modes. However from level \( j = 16 \) to the end the algorithm could not improve the approximations in modes \( a \) and \( b \) any further and stopped sampling from them. This is reflected in the last row of Figure 3 where the green dots show the new samples in each consecutive frame - it can be seen that samples are only being added to mode \( d \). This feature of the algorithm is useful in the presence of highly disjoint failure domains, where the local quality of the GPE can increase independently and resources will not be wasted where they are not needed. As before, some statistics about the probability of failure were calculated from 100 runs of SuS. The mean was found to be \( \bar{p}_F = 7.31 \times 10^{-3} \). The corresponding c.o.v. and relative error were \( \delta_{p_F} = 14.1\% \) and \( \Delta p_F = 2.7\% \), respectively. It is interesting to note that the same figures based on a 100 runs of direct Monte Carlo simulation with 100000 samples were \( \delta_{DMC}^{p_F} = 3.7\% \) and \( \Delta p_{DMC}^{F} = 0.7\% \). Figure 4 shows the CCDF curve for the failure probability estimated with SuS, the one obtained via GPSS and the one calculated relying on the unimproved GPE. It can be seen, that subset simulation was unable to find points from the GPE that belong to \( F \).

5.3 10D wing weight model

The last experiment that was conducted is based on a high dimensional model for estimating the weight of the wing of a general aviation aircraft. The functional expression is:

\[
0.0365 w^{0.758} W^{0.0035} f_w \left( \frac{A}{\cos^2 \Lambda} \right)^{0.6} q^{0.006} \lambda^{0.04} \left( \frac{100t c}{\cos \Lambda} \right) (NzWdg)^{0.49} + SwWp
\]

(16)

The nomenclature of Eq. (16) is somewhat involved and can be found in [2]. For the purposes of this paper the dimensionality of the problem is of interest. GPSS was ran on the model with 5000 samples per level for SuS and variance threshold, \( \varepsilon = 10^{-5} \). Since it was known a priori
that more points will be added to the training sample for the GPE, the initial DOE consisted of 60 LHS points. The procedure took 108 iterations with a final data point count of 168 samples.

![Figure 3: The performance of GPSS on the mixture function. The title in each tile shows the level of the algorithm. The red contours in the first row correspond to the intermediate levels from Eq. (11) at the specified level. The green diamonds in the last row show the samples added since the previous tile. The number of samples in each tile is \( n = \{20, 22, 24, 44, 50, 59, 62, 63, 64\} \).](image)

Figure 3 (left-centre) shows a set of diagnostics from the first and the last fit of the GPE for 100 LHS data points. The left column shows the observed functional response versus predictions generated from the GPE. The error bars reflect the 95% credible interval for each prediction. In the centre column the individual prediction errors (IPE) are plotted as suggested in [30]. The IPE should have a \textit{student}-t distributions and should therefore lie in the interval \([-2, 2]\), 95% of the time. There is not much of a difference between the two sets of plots which suggests that the global quality of the surrogate has not changed significantly - the GPE is a good global predictor before and after GPSS. The same cannot be said about the local approximation of the failure region.
Figure 4: Complementary CDF of the mixture function according to the GPE (green), GPSS (red) and SuS (black). The blue circles show the intermediate levels of SuS. Each level was populated with 3000 samples.

Figure 5: Diagnostics for the quality of the GPE before (top left-centre) and after (bottom left-centre) GPSS. The correlation between observation and prediction at 100 data points is shown on the left. The distribution of IPE is shown in the centre. The correlation between observation and prediction at points originally lying in $F$ before and after GPSS, top right and bottom right, respectively. In all cases mean and 95% credible interval are given in dark grey dots and light grey error bars, respectively. Dashed lines on top mark $F$ on both axes.

Figure 5 (right) shows a comparison between the prediction of function response, originally in the failure domain, before (top) and after (bottom) GPSS. Results are obvious - the correlation between observed and predicted values is much higher. The mean probability of failure was $\bar{p}_F = 1.97 \times 10^{-5}$. The corresponding c.o.v. and relative error were $\delta p_F = 21.8\%$ and $\Delta p_F = 11.4\%$, respectively. These were calculated based on 100 runs of SuS. Finally, Figure 6 shows a comparison between the complementary CDF curves from the GPE (green), GPSS (red) and
SuS (black).

![Figure 6: Complementary CDF of the mixture function according to the GPE (green), GPSS (red) and SuS (black). The blue circles show the intermediate levels of SuS. Each level was populated with 5000 samples. The global quality of the GPE before GPSS is good only in the high probability regions.](image)

The CCDF curve from the first GPE level summarizes the diagnostics from Figure 5 - the original surrogate provides a good approximation in the high probability regions of the performance function, but fails to capture its behaviour in the rare regions.

6 CONCLUSIONS

In this paper an adaptive algorithm for reliability analysis, called GPSS, was presented. GPSS is a combination between Gaussian process emulation and subset simulation which allows an efficient characterization of rare events and estimation of the associated probabilities of failure. The reduction in the number of samples require by GPSS as compared to SuS and DMC is significant. The algorithm was validated with three benchmark problems. Future work, as planned by the authors includes reliability analysis of computationally expensive computer codes, where the advantages offered by GPSS can be appreciated.

REFERENCES


COSSAN SOFTWARE: A MULTIDISCIPLINARY AND COLLABORATIVE SOFTWARE FOR UNCERTAINTY QUANTIFICATION

Edoardo Patelli¹, Matto Broggi², Silvia Tolo¹, Jonathan Sadeghi¹

¹Institute for Risk and Uncertainty, University of Liverpool (UK)
e-mail: edoardo.patelli@liverpool.ac.uk

²Institut für Risiko und Zuverlässigkeit, Leibniz Universität Hannover (DE)
e-mail: broggi@bauinf.uni-hannover.de

Keywords: COSSAN-X, OpenCOSSAN, Uncertainty Quantification, Risk Assessment

Abstract. Computer-aided modelling and simulation is now widely recognised as the third ‘leg’ of scientific method, alongside theory and experimentation. Many phenomena can be studied only by using computational processes such as complex simulations or analysis of experimental data. In addition, in many engineering fields computational approaches and virtual prototypes are used to support and drive the design of new components, structures and systems.

A general purpose software for uncertainty quantification and risk assessment, named COSSAN, is under continuous development. This is a multi-disciplinary software that satisfies industry requirements regarding numerical efficiency and analysis of detailed models that can be used to solve a wide range of industrial and scientific problems. The main aim of the COSSAN software is to allow the inclusion of non-deterministic analyses as a practice standard routing in scientific computing. This paper presents two recent toolboxes added to the OpenCOSSAN: Credal Networks and Interval Predictive model.
INTRODUCTION

Uncertainty quantification is a key requirement and challenge across various disciplines and industry is fully aware of the importance of non-deterministic analysis. This is particularly true in many engineering fields where computational approaches and virtual prototypes are used to predict and simulate complex systems. The continuous growth of computational power available allows to simulate the effect of the uncertainties on the model under analysis with increased precision. These advancements have allowed engineering practitioners to reduce the number of expensive experimental tests and analyse conditions that can not be reproduced.

However, in order to increase the confidence in the digital models, the effect of unavailable uncertainty needs to be simulate as well. In fact, improved model prediction can only be accomplished when different sources of uncertainty are explicitly included in the analysis [2]. Hence, uncertainty management is necessary to provide support to the decision makers. Unfortunately, industrial design methods are still predominantly deterministic. This is because of limited expertise in uncertainty quantification in industrial design offices together with lack of proper software for uncertainty quantification and stochastic analysis.

Software for uncertainty quantification have received considerably less attention than their deterministic counterparts mainly due to the computational costs associated with non-deterministic analysis.

In general, deterministic analysis provides a map between a single point in the input space (i.e. the model parameters) and a point in the output space (i.e. component or system performance). Stochastic analysis extends this map to a region in the input space and a corresponding region in the output space by repeating the deterministic analysis many times as represented in Figure 1. Generalised probabilistic approaches such as interval and fuzzy methods, imprecise probabilities and any combination thereof (see e.f. [3, 15]) allow to cope with model uncertainties, errors in modelling and measurements and noise in signals require in general the evaluation of different models stochastic models further increasing the computational cost of the analysis.

In order to increase the knowledge transfer between academia and industry and to include stochastic analysis as standard procedure in engineering practice, G.I Schüeller initiated the development of the COSSAN project at the Institute for Engineering Mechanics, University of Innsbruck, Austria (?). The COSSAN project aims at developing a new generation of a general purpose software for non-deterministic analysis that can be used by industry, academics, researchers and for teaching purpose as well. The software incorporates the knowledge, understanding and intellectual property from more than 30 years of research in the field of computational stochastic analysis.

Starting from 2006, the next generation software referred as COSSAN-X is under continuous development and it is intended for a wider range of applications in different fields, which includes optimization analysis, life-cycle management, reliability and risk analysis, sensitivity, optimization and robust design [17]. Since 2012, an open source version of COSSAN-X, called OPENCOSAN, is available under the Lesser GNU licence [9]. This means that the program can be used for free, redistributed and modified under the terms of the GNU General Public License. The OPENCOSAN aims to promote learning and understanding of non-deterministic analysis through the distribution of an intuitive, flexible, powerful and open computational toolbox in Matlab environment [16]. The combination of various algorithms with specific solution sequences permits the analysis of different engineering and scientific problems.
1.1 Software developing approach

Nowaday, COSSAN software is developed collaboratively between the Liverpool Institute for Risk and Uncertainty and the Institut fr Risiko und Zuverlssigkeit from the Leibniz Universitt Hannover (DE) and under continuous improvments [14, 16].

OPENCOSSAN, the computational core of the COSSAN software, is released under the terms of the GNU Lesser General Public License [9]. The source code is accessible via available at the web address [http://www.cossan.co.uk](http://www.cossan.co.uk). The Apache Subversion system is used for software versioning and revision control system distributed to keep tracks of changes and contribution from different developers. OPENCOSSAN is coded exploiting the object-oriented Matlab programming environment, where it is possible to define specialized solution sequences, which include reliability methods, sensitivity analysis, optimization strategies, surrogate models and parallel computing strategies. The computational framework is organized in packages. A package is a namespace for organizing classes and interfaces in a logical manner, which makes large software project OPENCOSSAN easier to manage. A class describes a set of objects with common characteristics such as data structures and methods. Objects, that are instances of classes can be aggregated forming more complex objects and proving solutions for practical problem in a compact, organized and manageable format. OPENCOSSAN provides intuitive, clear, well documented and human readable interfaces to the classes. No acronyms are used to define methods and properties.

The availability of an extended documentation, tutorials and examples is of a key importance for the usability of a software. The COSSAN documentation is written collaboratively by COSSAN developers and end-users using MediaWiki tool where with rare exceptions, articles can be edited by anyone.

COSSAN software is developed as a general purpose software which provides generic functionalities and can be changed and expanded through user-defined routines and additional codes.
This implies that solutions for solving very specific problems are not generally available although they can be added by the users. Instead, methods and tools that can be used for solving a reasonably wide range of engineering and scientific problems are prioritised. One of the challenges developing COSSAN software is maintaining its flexibility and modularity as well its usability for the end-user. Another key element of the software is a comprehensive risk management and uncertainty quantification based on different representation of the uncertainties based on probabilistic approaches, interval and fuzzy methods, imprecise probabilities and any combination thereof.

In the following sections two different toolboxes recently introduced into OpenCOSSAN are shown.

2 Meta Models toolbox

As shown in Figure 1, stochastic analysis requires the repeated evaluation of a (detailed) numerical model. The analysis time can be reduced significantly by using meta-models, which approximate the quantities of interest at low computational costs. In other words, meta-models mimic the behaviour of the original model (e.g. FE model), by means of a mathematical model with negligible computational cost.

3 Interval Predictor Models toolbox

Interval Predictor Models (IPMs) are a class of meta-model which describe the expected spread of the output of a model whilst making very few assumptions about the model. The meta-model finds an upper and lower bound for the predicted output of the model using a function chosen by the user (at present limited to radial or polynomial basis). Therefore, once trained, the meta-model can make predictions very quickly [6].

Interval Predictor Models offer a robust quantification of uncertainty, even when few data points are available. This makes them useful when a very computationally expensive model limits the number of simulations which can be performed. The robust quantification of uncertainty is achieved in two ways. Firstly, a measure of uncertainty is intrinsic to the meta-model since the output is an interval - it’s spread describes our uncertainty. Secondly, the reliability, $R$, of this interval i.e. the probability future data points will fall inside the interval is bounded by

$$\text{Prob}_{\textrm{IPM}}[R \geq R^*] \geq C,$$

(1)

where $R^*$ and $C$ are the bounds on the reliability of the model and our confidence in it respectively [4]. $R^*$ as a function of $C$ is calculated by OpenCOSSAN. For this model it is recommended to use these quantities as a method of measuring calibration rather than the coefficient of determination which is normally used in OpenCOSSAN. In addition, the samples in the training data set must be independently distributed [5].

3.1 Comparison to other meta-models

In one sense Interval Predictor Models are similar to Gaussian Processes. However, Interval Predictor Models are useful in situations where Gaussian Processes may be too computationally expensive, for example where there are many data points and therefore it is too computationally expensive to make predictions. In addition, Interval Predictor Models have the advantage that they often yield tighter bounds than Gaussian Processes.
The Interval Predictor model requires more time to train than a response surface, however it offers a very robust quantification in the epistemic uncertainty resulting from the construction of the meta-model.

The amount of data points which can be used to create the IPM is limited only by the time the user is prepared to spend on the training the meta-model, which is performed by convex optimisation. In fact, for fewer than tens of thousands of data points the time required to train the model is in general negligible.

3.2 Work flow and implementation details

The IPMs implemented into OPENCOSSAN are the linear parameter dependency type-1 IPMs described in [6].

The IPM class is a subclass of the meta-model abstract class in OPENCOSSAN, and therefore inherits most of it’s behaviour in a similar way to the response surface class. However some of the arguments which must be specified to the constructor method are crucially different, and therefore they are described below:

- **type**: a string containing radial or polynomial, which specifies which basis type should be used.
- **maximumexponent**: a positive integer giving the maximum index of polynomial to be used in a full factorial polynomial basis. If the basis is polynomial and this is not specified then exponents must be specified.
- **exponents**: a matrix of indices - one row for each monomial, one column for each input variable. Mandatory for polynomial basis unless maximumexponent is specified.
- **augmented**: An integer (0, 1 or 2) specifying if the IPM is augmented (see [6] for explanation). Optional (default to 0).
- **sigma**: the sigma matrix for radial basis. Mandatory when type is radial.
- **centers**: the centres matrix for radial basis. Mandatory when type is radial.

Once training is complete the IPM parameters are stored in the matrices MIPMParametersU and MIPMParametersL.

The implemented meta-model class IPM provides a method (removeOutliers) to remove the outlier which implements the outlier removal algorithm described in [6]. In simple terms, this method will remove training data which prevents the spread of the IPM being reduced, and then re-calibrate the IPM.

The performance of the IPM model can be used to calculate using the reliabilityPlot method and finally, for models with 1 or 2 inputs it is possible to visualise the IPM using the plotModel method.

3.3 Application

In order to demonstrate the application of the class, an IPM surrogate of the Rosenbrock function is constructed. The created IPM has a polynomial basis of degree 2. Figure 2 shows a plot of $R^*$ as a function of $C$. If 20 more samples are taken from the function 96.25% fall inside the IPM which is in agreement with Fig. 2. In order to create a more robust IPM than this either the degree of the polynomial used must be decreased or more training data must be obtained.
Now the user may choose to use the output from the meta-model in their performance function. For example if the output represented a quantity which must be below a certain value (such as displacement of a cantilever beam) then the user would choose the upper bound of the displacement from the IPM and then calculate an upper bound on the failure probability. This could be repeated with the lower bound in order to calculate a lower bound on the failure probability.

4 Credal Networks toolbox

4.1 General remarks

Bayesian Belief Networks, more commonly known as Bayesian Networks (BNs), are a probabilistic graphical model based on the use of directed acyclic graphs, integrating graph theory with the robustness of Bayesian statistics. The graphical framework of such models consists of nodes, representing the variables of the problem of interest, connected to each other by edges, generally arrows, that depict the dependency link existing between two nodes. Each variable is assumed conditionally independent of its non-descendants given its parent variables: this propriety is known as local Markov propriety. Each node of the network is linked to a Conditional Probability Distribution (CPD) that, according to the local Markov propriety, defines the strength of the probabilistic dependency existing between each individual node and its parents. When a node has no parents, i.e. it is a root of the network, a marginal probability distribution is associated with it.

The main aim of the Bayesian Network approach is to factorize the probability of a complex event exploiting the knowledge regarding the dependencies existing among its sub-parts. BNs are extremely attractive for a large variety of applications providing decomposition procedure and the possibility to update the overall knowledge regarding the event of interest when new
information (i.e. evidence) about any of its sub-elements becomes available.

BNs exact inference algorithms (i.e. adopting an analytical approach) are almost exclusively limited to the use of crisp probability values which are not able to fully capture the data generally available. This often leads to the adoption of discretization procedure which empowerish the quality of the information available. On the other hand, approximate inference algorithms allow the use of continuous probabilistic variables but at the cost of lower accuracy, and they can result often inefficient or have unknown rates of convergence. The limitations of the approximate inference approach result even more significant when considering the field of risk analysis and decision making, or more generally those applications where rare events and near-real-time computation play a crucial role [?].

Generalizing further, even traditional continuous probabilistic variables are often not able to fully capture the uncertainty of data without introducing strong subjective assumptions not justified by the information available. Alternatively, imprecise probabilities theory provides several mathematical frameworks able to depict the uncertainty inevitably affecting the data and hence reducing the introduction of subjective biases [1].

In order to overcome the limitations associated with traditional BNs, the integration of such approach with the imprecise probability theory has attracted increasing attention in the scientific community leading to the formulation and study of Credal Networks (CNs). CNs can be considered as a more general approach that relaxes the requirements of BNs. They still rely on the use of directed acyclic graphs but differently from the former method allow the use of imprecise probabilities as input parameters of the model, ensuring a higher degree of flexibility in the representation of the data uncertainty. In spite of the capabilities and promising potential of this methodology the application of such models is still quite limited, mainly due to the high computational cost associated with CNs inference.

4.2 Work flow and implementation

Further efforts and research are strongly required in order to enhance the attractiveness of CNs outside the academic world and to ensure the reliability and efficiency of their performance in real-world applications. Well known and novel methodologies are integrated in the software in order to provide the implementation, manipulation and analysis of CNs. The main tasks of the toolbox can be listed as:

- Implementation of the CN model defined by the user
- Reduction of the initial user-defined model (to ease the burden of the inference computation)
- Inference computation of the network
- Sensitivity analysis of the reduced model

The following sections offer a brief introduction to each of the mentioned tasks.

Implementation The toolbox provides the definition of initial models allowing a range of possible variables (and subsequently node types) such as:

- events with a discrete and finite number of outcome states associated with crisp probability values (discrete nodes)
events with a discrete and finite number of outcome states associated with interval probability values (*discrete* nodes)

- events whose outcome states are described by continuous probabilistic distributions (*probabilistic* nodes)

- events whose outcome states are described by interval variables (*bounded* nodes)

- events whose outcome states are described by imprecise probability distributions (*hybrid* nodes)

The example in Figure 3 refers to an elementary structural system [11, 8] consisting of one-bay elastoplastic frame subject to a vertical load $V$ and horizontal load $H$. The structure is defined by the plastic moment capacities $R_1, R_2, R_3, R_4, R_5$, and presents three possible failure modes (*FailureMode1*, *FailureMode2*, *FailureMode3* in Fig.3) defined by as many limit-state functions. The occurrence of any of these failure modes is assumed to result in the failure of the whole system (event $E$ in the network). The nodes $M_4$ and $M_5$ refer to measurement of the plastic capacities $R_4$ and $R_5$ and are affected by a measurement error $\epsilon$. The plastic moment capacities are assumed to have a probabilistic distributions, while the measurements $M_4$ and $M_5$, being affected by the error $\epsilon$ represented by an interval variable, are bounded nodes (as indeed $\epsilon$). The vertical load is supposed to have a probabilistic distribution whose mean is not precisely known but represented by an interval: the load $V$ is then represented as a *hybrid* node differently from the load $H$ assumed as probabilistic.

In Fig.3 the rectangular shaped nodes are discrete, the oval shaped ones bounded, the circular ones probabilistic and finally the trapezoidal shaped nodes are hybrid. The graphical representation of the network is realized using the Bioinformatics toolbox for MATLAB [12] and it is designed in order to highlight the different nature of the model variables.
Reduction A key procedure of the methodology implemented in the toolbox consists of reducing the initial model to an equivalent one containing only discrete nodes. The implemented reduction procedure ensures the preservation of the initial information. For instance, the children nodes FailureMode1, FailureMode2 and FailureMode3 can be expressed by the related limit-state functions, which are defined in function of the parents nodes $R_1, R_2, R_3, R_4, R_5, H$ and $V$. The computation of the probability bounds is obtained adopting numerical methods for structural reliability analysis. A range of options are available in the toolbox. The default option relies on the use of the Advanced Line Sampling technique [2].

The two main phases of the reduction procedure are:

- Computation of continuous nodes children of at least one continuous node. When also continuous nodes are expressed as domains in the outcome space of their parents they are automatically computed according to their nature:
  - Probabilistic nodes: probabilistic distributions are constructed over the samples collected from the parent nodes through the use of Monte Carlo methods.
  - Bounded nodes: the bound are obtained through a random search in the outcome space of their parents.
  - Hybrid nodes: the current implementation is limited to the use of parametric imprecise random variables. Conversely, the imprecise parameters of the distribution can be expressed as parent nodes of the hybrid ones, as for the nodes $V_{mean}$ and $V$ in Figure 3.

- Computation of discrete nodes children of at least one continuous node. According to the nature of the nodes involved (this time the parent nodes) the option available are different:
  - Only probabilistic and discrete parent nodes: in this case the approach reduces to the Enhanced Bayesian Network methodology [13] and a wide variety of method-
ologies can be adopted (e.g. FORM, SORM, traditional and Advanced Monte Carlo methods etc.)

- Bounded, probabilistic and discrete parent nodes: two options are available, one consisting of a generalized version of FORM based on the use of convex sets \([10]\), one on the Advanced Line Sampling methods. It is opportune that only this latter method allows to compute probability bounds on the event of interest while the formers one focuses only on the estimation of the worst-case scenario probability.

- Hybrid, bounded, probabilistic or discrete parent nodes: currently the only option available for the computation is the Advanced Line Sampling \([?]\).

The result of the reduction of the model of Fig.3 is shown graphically in Fig.4. As shown by the graph, the nodes \(M_4\) and \(M_5\) have been removed from the reduced network, since non-discrete and not receiving any evidence. In order to process evidence, it is necessary that the node receiving it is present in the reduced model. Subsequently, if a non-discrete node receives evidence, it is necessary to discretize it in order to keep it in the final model. In this case, the first phase of the reduction procedure, i.e. the computation of continuous nodes children of at least one continuous node, is not entirely carried out but the discretization of the continuous nodes receiving evidence is automatically performed, avoiding their elimination. Let assume that some measurement regarding the plastic moments capacities \(R_4\) and \(R_5\) is available and hence the nodes \(M_4\) and \(M_5\) receive evidence in the network. In this case, the modified network containing the discretized information of \(M_4\) and \(M_5\) is shown in Fig.5. The procedure followed for the discretization is similar to that proposed by Straub and Kiureghian \([18]\) and aims to limit the inevitable loss of information quality associated with such procedure. The final network obtained from the reduction of the model in Fig.5 is shown in Fig.6, in this latter the nodes \(M_4_{\text{discrete}}\) and \(M_5_{\text{discrete}}\) refer to the measurements of the plastic moment capacity and can hence receive evidence updating the overall estimate associated with the probability of failure of the structure.

**Inference Computation** Three inference computational methods are available in the toolbox:
Figure 6: Final network resulting from the reduction of the model in Fig.5

- Coarse Approximation: the lowest computational cost associated with the inference computation, implying on the other hand a lower level of accuracy providing an outer approximation of the true value of the output.

- Fine Approximation: based on a original and novel inference computational algorithm which aims to achieve a high degree of accuracy at a lower computational cost than that associated with exact inference. The method provides the true value of the query probability bounds when no evidence is introduced in the network and provides inner approximation values otherwise [19].

- Exact Inference: currently this option is restricted to the adoption of a combinatorial approach to inference computation. This can be performed either using the built in variable elimination algorithm.

**Sensitivity Analysis**  A novel method allows to identify the possible changes of input parameters (i.e. interval conditional probabilities associated with the nodes of the network) that result in the achievement of the user-defined degree of accuracy in output (i.e. probability bounds of the query probability). In the current implementation the method is limited to the use of boolean nodes and relies on a combinatorial approach. The results provide an intuitive understanding of when parameter changes do or do not matter in terms of query robustness. Further detail can be found in [20].

5 Conclusion

The availability of an intuitive and easy to use general purpose software is a key for making the non-deterministic analysis a common practice in computational models and numerical simulations. In fact, uncertainty and imprecision are unavoidable and they must be accounted for in any analyses. Only implementing stochastic analysis, digital (or virtual) design will be credible and applicable in different sectors and fields. Digital design allows to design fast and better, reducing the design costs and provide cost-effective and feasible engineering solutions.
COSSAN-X provides an easy to use yet powerful software for uncertainty quantification and risk management that can be used by analysis without an extensive learning curve and allowing to create a bridge between the academic research and the industrial practice. The engine of the COSSAN software OPENCOSSAN is freely available making the software development more sustainable, continuously updated and avoiding code duplication. It is an excellent collaborative tools for researchers and academics encourages the cross-discipline utilisation of stochastic analysis.

REFERENCES


GAUSSIAN PROCESS RESPONSE SURFACE MODELING AND GLOBAL SENSITIVITY ANALYSIS USING NESSUS

John M. McFarland\textsuperscript{1}, John A. Dimeo\textsuperscript{2}, and Barron J. Bichon\textsuperscript{1}

\textsuperscript{1}Southwest Research Institute
6220 Culebra Road
San Antonio, TX 78238, US
john.mcfarland@swri.org
barron.bichon@swri.org

\textsuperscript{2}Elder Research, Inc.
300 W Main St., Suite 301
Charlottesville, VA 22903, US
dimeo@elderresearch.com

Keywords: Software, Uncertainty Quantification, Response Surface Modeling, Gaussian Process, Reliability Analysis, Global Sensitivity Analysis, Variance Decomposition.

Abstract. NESSUS\textsuperscript{®} is a general-purpose software program for probabilistic analysis that includes state-of-the-art algorithms, flexible methods for interfacing with external numerical models, and a mature graphical user interface. NESSUS was originally developed for NASA under a long-term research and development program to develop methods and tools for reliability analysis of space shuttle main engine components. In the past few years, recent NESSUS development has focused on the incorporation of advanced response surface modeling and global sensitivity analysis methods. NESSUS now includes a variety of tools for building and analyzing Gaussian Process (GP) models. This includes general-purpose GP response surface models as well as the Efficient Global Reliability Analysis (EGRA) method, which uses adaptive sampling to target surrogate model accuracy in the vicinity of the limit state. In addition, several methods have been implemented into NESSUS for the calculation of variance-based sensitivity indices, including sampling-based methods and analytical solutions based on response surface models. This paper gives an overview of these recent enhancements. In particular, we introduce the NESSUS Response Surface Toolkit (RST), which is a recently released standalone software application included with NESSUS for building, visualizing, and assessing response surface models.
# INTRODUCTION

NESSUS (Numerical Evaluation of Stochastic Structures Under Stress) is a general-purpose software program for probabilistic analysis. It was originally created by a team led by Southwest Research Institute (SwRI) as part of a 10-year NASA project started in 1984 to develop a probabilistic design tool for the space shuttle main engine with a focus on probabilistic finite element analysis. The methods and capabilities in NESSUS were designed to support predicting the probabilistic response and/or probability of failure for computational intensive models. The input variations were modeled using probability density functions and propagated using traditional and newly developed probabilistic algorithms. In 1999, SwRI was contracted by Los Alamos National Laboratory to adapt NESSUS for application to extremely large and complex weapon reliability and uncertainty problems in support of its Stockpile Stewardship program. In 2002, SwRI was contracted by the NASA Glenn Research Center to further enhance NESSUS for application to large-scale aero-propulsion system problems. The end result of these large research programs was a completely redesigned software tool that includes a sophisticated graphical user interface (GUI), capabilities for performing design of experiments and sensitivity analysis, a probabilistic input database, a geometric uncertainty modeling tool for perturbing geometry in existing finite element models, and state-of-the-art interfaces to many third-party codes such as Abaqus, ANSYS, LS-DYNA, MSC.NASTRAN, and NASGRO.

NESSUS has seen continuous improvement and application since its beginnings in the late 1980s. The most recent enhancements include extensive capabilities for Gaussian process response surface modeling and variance-based global sensitivity analysis. In particular, the NESSUS Response Surface Toolkit was released in 2015 as a standalone application specifically created to address response surface model fitting, visualization, goodness of fit, and sensitivity analysis.

This paper describes the recent NESSUS enhancements for response surface modeling and sensitivity analysis. Section 2 gives a description of the Gaussian process model formulation used by NESSUS and the NESSUS Response Surface Toolkit. Section 3 gives an overview of the Response Surface Toolkit itself. Section 4 describes three different approaches for reliability analysis with Gaussian process models in NESSUS, including the Efficient Global Reliability Analysis (EGRA) method, which uses adaptive sampling. Section 5 gives an overview of global sensitivity analysis in NESSUS.

## NESSUS GP MODEL FORMULATION

Gaussian Process (GP) regression (also known as kriging) is a powerful technique for developing fast-running response surface or “surrogate” models for computationally expensive computer simulations. The response surface model enables detailed analysis such as Monte Carlo sampling, numerical optimization, model calibration, and sensitivity analysis, that would be intractable if working directly with original simulation. This section gives a summary of the GP model formulation used by the NESSUS software.

NESSUS uses a Gaussian process response surface model formulation employing a low-order trend function and a squared-exponential correlation function. NESSUS supports both constant and linear trend functions. A separable correlation function is employed with unique correlation length parameters for each input variable:

\[
c(a, b) = \exp \left[ -\sum_{i=1}^{d} \left( \frac{a_i - b_i}{\theta_i} \right)^2 \right],
\]  \(1\)
where $\mathbf{a}$ and $\mathbf{b}$ are vectors of dimensionality $d$, and $\theta_i$ denotes a correlation length parameter associated with dimension $i$. The covariance matrix associated with the training data is given by:

$$
\Sigma = \sigma^2 \mathbf{R} + \Sigma_{\text{noise}},
$$

(2)

where $\sigma^2$ is the process variance, $\mathbf{R}$ is the correlation matrix based on Eq. (1), and $\Sigma_{\text{noise}}$ is a diagonal noise covariance matrix with all diagonal entries given by $\sigma_{\text{noise}}^2$. When modeling training data from a deterministic computer simulation, the noise variance is commonly assumed to be zero, resulting in a response surface that directly interpolates the training data.

In NESSUS, the GP parameters (the trend coefficients, process variance, correlation lengths, and optionally noise variance) are estimated using maximum likelihood. The profile log likelihood function is maximized by searching over the correlation parameter space, and the trend coefficients and process variance are set to their optimal values conditional on the correlation parameters. The noise variance can also be estimated as part of the procedure, in which case NESSUS maximizes the restricted likelihood function. Further details can be found in ref. [1]. NESSUS supports both gradient-based and global search methods.

NESSUS automatically performs goodness of fit assessment using leave-one-out cross-validation. This is determined by holding all of the GP model parameters (trend and covariance parameters) constant, deleting one observation from the GP model, predicting its value, and repeating for each observation in the training set. Instead of explicitly creating $n$ GP models, the cross-validation errors can be computed efficiently as described by ref. [2]. NESSUS summarizes the model accuracy by a “cross-validation $R^2$” (also known as “prediction $R^2$” [2]):

$$
R_{CV}^2 = 1 - \frac{SSE_{CV}}{SS_T},
$$

(3)

where $SSE_{CV}$ is the sum of squares of the leave-one-out cross-validation residuals and $SS_T$ is the usual total sum of squares of the training data response values.

3 NESSUS RESPONSE SURFACE TOOLKIT

The NESSUS Response Surface Toolkit (RST), released in 2015, is a standalone application included with NESSUS that provides a suite of tools for designing, fitting, assessing, visualizing, and making predictions with response surface models. Some of the key capabilities include:

- Design of experiments using Latin Hypercube sampling, with options for correlation reduction and point spacing optimization
- Fit polynomial regression and Gaussian process regression response surface models
- Evaluate goodness of fit using leave-one-out cross-validation
- Plot 2D and 3D slices of the model
- Evaluate importance of variables using variance-based (global) sensitivity analysis
- Make model predictions at user-specified input values
- Export models for use in the NESSUS software
- Import models created by the NESSUS software
The RST provides streamlined capability for quickly generating and visualizing a design of computer experiments based on Latin Hypercube sampling (Figure 1). The interface allows the user to declare, name, and assign bounds to the variables. Then the user simply specifies the number of samples and clicks on the “Generate Design” button. By default, 100 random designs are generated and the best is selected as the one that maximizes the minimum distance between points. The number of designs can be configured by the user from the “Iterations” field. An option is also provided for correlation reduction using the Iman-Conover method [3]. Random number seed control is also provided for reproducibility. Once the samples have been generated, pairwise or three-way plots among the variables can quickly be generated using a menu driven system, as shown in Figure 2.

Generated designs can be saved with metadata (such as the variable bounds, random number seed, and other settings) and also exported to a delimited text file. The RST also includes the capability to “augment” an existing design [4]. For an initial design with \( n \) points, augmenting adds \( n \) new points to the design such that the overall \( 2n \)-point design maintains a Latin Hypercube structure.

The RST provides a fast and flexible interface for loading existing training data to create a response surface, as shown in Figure 3. Training data can be read from plain text files using whitespace, comma, and other delimiters. Headers, when available, are used to initially define variable names, and can be overridden by the user. Each column detected in the file can be assigned as an input variable, a response variable, or ignored. The default settings treat the last column as the response variable and all others as input variables.
Once the training data have been loaded, the RST displays a summary of correlations between the inputs and the response variable, and provides the ability to show a scatter plot between any two or three variables (Figure 4).

Next, the user is provided a simple set of options for fitting the response surface model, as shown in Figure 5. Two model types are available: Gaussian Process and polynomial regression. The Gaussian Process model provides options for a constant or linear trend, three settings for observation noise, and several options for the optimization method used during parameter estimation. The options for observation noise include noise-free ($\sigma_{\text{noise}}^2 = 0$), estimation of $\sigma_{\text{noise}}^2$ by restricted likelihood maximization, and user-specification of the noise.
standard deviation, $\sigma_{\text{noise}}$. The option to fit the noise variance is especially useful when the training data are the results of physical experiments, as opposed to computer experiments.

Figure 5. NESSUS RST response surface fitting options

Upon fitting the response surface model, the RST displays a summary of information that includes cross-validation results, input variable sensitivities, and surface visualization (Figure 6). This includes fit and cross-validation $R^2$ values, as well as the noise standard deviation, reciprocal condition number, and log likelihood for GP models. The reciprocal condition number can be used to identify ill-conditioning of the GP correlation matrix, which can adversely impact performance when the value begins to approach roughly $10^{-13}$ or smaller. This can be remedied by removing training points or re-fitting the model with a small amount of noise variance.

Figure 6. NESSUS RST screen for reviewing and visualizing the response surface model

The miniature “Sensitivities” plot in the Review Model screen can be expanded to a full-size plot, as shown in Figure 7. On this plot, the x-axis shows the main and total effect indices obtained using variance-based global sensitivity analysis. In order to do this, each response surface input variable is modeled using a normal distribution with mean and standard
deviation estimated from the training data (the range of the training data is treated as five standard deviations). The sensitivity indices are then computed efficiently using the analytical solution given in [5]. For complete control of how the probability distributions for the input variables are defined, the user can load the response surface model into NESSUS and use the global sensitivity analysis type (see Section 5).

The controls in the bottom of the Review Model screen are geared towards response surface visualization when the Plot tab is selected. By default, a 2D line plot of normalized main effect curves for all inputs, with confidence bounds, is shown, as seen in Figure 6. In this plot, the x-axis represents a scaled version of each input variable, in which the training data values are used to scale each input to the range [-1,1]. In other words, a value of -1 on this scale indicates that the variable is equal to the smallest value of that variable occurring in the training set. Similarly, the largest value in the training set corresponds to +1.

For each curve, the y-axis plots the main effect of the response, which is the expectation of the response conditional on the value of the given input variable. The same probability distributions used for variance decomposition are employed for determination of the main effect curves. By default, the plot also shows a central 95% confidence interval for each main effect curve as a shaded region. The confidence bounds are derived from the GP model fit. Both the main effect curves and their confidence bounds are computed efficiently using a combination of derivations from refs. [5,6].

To the left of the plot, a variety of settings are available to control the behavior of the plot. The “Input variables” section provides control of which variables are shown in the plot. When the plot type is “Multiple”, individual variables can be removed from the plot using these controls. For convenience, the global sensitivity indices are shown next to each variable, making it easy to quickly find and plot the most important variables.

The type of plot can be configured under the “Plot settings” area. The “Single” plot type limits the plot to a single curve. In this mode, the input variable selections control which curve is plotted, as opposed to adding or removing curves, like they do in “Multiple” plot mode.
The “Tiled” plot type presents an array of plots, each of which plots the main effect curve for a single input variable. An example of the tiled plot is shown in Figure 8, where the plotting area has been maximized so that all nine plots can be seen at once. Unlike the “Multiple” plot type, which uses scaling in order to display all variables on a single plot, the tiled plot uses the original units for each input variable, which can make the interpretation more intuitive. Also note that the tiled plot uses the same y-axis range for each individual plot, making it possible to quickly assess the relative influence of each variable.

![Figure 8. NESSUS RST “Tiled” plot mode, showing the main effect curve for each response surface variable.](image)

Under the plot settings area, display of the training data can also be enabled for either the single or tiled plot types. When this is done, the training data are projected onto the two-dimensional plot axes. Some caution is needed when interpreting this kind of display of the data, since not all dimensions of the training data input space are represented in the plot.

Three-dimensional visualization of the response surface can also be enabled using the plot settings. In 3D plotting mode, the user selects two input variables for the horizontal axes, and the response is plotted on the vertical axis. In 3D mode, the surface is conditioned on fixed values of the input variables not being plotted, unlike 2D mode, which averages over the remaining variables to obtain main effect curves.

The RST allows for reading and writing both polynomial and GP response surface models to a platform-independent file. When writing a GP model to a file, the following data are included:

- The training data input and output values
- The Cholesky factor of the GP correlation matrix, $\mathbf{R}$
- Solution of the linear system used for prediction, $\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{F}\mathbf{\beta})$, where $\mathbf{Y}$ are the training data response values and $\mathbf{F}\mathbf{\beta}$ is the trend function evaluated at the training data locations
The correlation parameters determined via maximum likelihood estimation

The trend coefficients, process variance, and noise variance

The response surface model file can then be read back into either NESSUS or NESSUS RST.

In addition to capturing the options used to fit the model such as the trend order and noise variance, this has the advantage that the GP mean and covariance parameters do not need to be estimated again, nor does the correlation matrix need to be reconstructed or factored. Thus, for large data sets, reloading the GP from file provides significant time and computation savings.

4 NESSUS RELIABILITY ANALYSIS

Unlike the Response Surface Toolkit, the NESSUS software is designed for general purpose probabilistic analysis. In particular, NESSUS has several probabilistic methods for reliability analysis, including Monte Carlo sampling, Latin Hypercube sampling, importance sampling, Advanced Mean Value, First Order Reliability Method (FORM), and more. Any of these reliability analysis methods can be used in conjunction with response surface modeling.

There are three approaches for using response surface models for reliability analysis in NESSUS: declaring an equation in the problem statement as a regression model, using a response surface-based reliability method, and using the Efficient Global Reliability Analysis (EGRA) method. Further information about each of these approaches is given in the following sections.

4.1 Regression model type

The first approach is to declare one or more equations in the problem statement as a numerical model, and then define the numerical model as a regression model type. NESSUS provides options for both polynomial and Gaussian Process regression models. The GP model can be defined based on user-specified training data or based on a previously saved GP model file. Thus, this is the way to use in NESSUS a GP model created by RST.

If the GP regression model is specified based on training data, then upon running the analysis, NESSUS will first fit the GP model using maximum likelihood estimation, as described in Section 2. NESSUS will also save the resulting GP model to a file, so that it can be analyzed using the RST or re-used from NESSUS without having to re-fit the model.

Once the regression model is defined, the NESSUS analysis behaves the same as with any other equations or numerical models. Any probabilistic methods can be used, including global sensitivity analysis as described in Section 5. Since the response surfaces are computationally very efficient, it is common to use them in conjunction with exhaustive sampling methods such as Monte Carlo or Latin Hypercube sampling.

This approach also allows one to build up the problem statement hierarchically. For example, a finite element model for predicting stress might be replaced by a response surface, whose output is fed into an analytical equation for crack growth prediction. A sequence of multiple independent response surface models could also be used.

4.2 Response surface method

In the second approach, the three steps of design of computer experiments, response surface fitting, and response surface sampling are considered collectively as a reliability analysis method. With this approach, the user defines the problem statement so that NESSUS interfaces directly with the original (computationally expensive) model or models. For example,
the problem statement may include calls to a finite element model or other external numerical code.

NESSUS provides three probabilistic analysis methods that fit into this category: the “Response Surface Method (RSM)”, “Gaussian Process Response Surface Method (RSM_GP)”, and “Efficient Global Reliability Analysis (EGRA)”. EGRA uses adaptive sampling and will be discussed in the next section.

The NESSUS Response Surface Method uses low-order polynomial fitting and is based on traditional Design of Experiments techniques such as the central composite design. The Gaussian Process Response Surface Method was added in NESSUS version 9 and is preferred when interfacing NESSUS with external models such as finite element simulations. The user has three options when working with this method. The first is the number of response surface training points that NESSUS will use. This determines how many evaluations of the problem statement NESSUS will perform.

The second option is specification of the bounds used for the design of experiments. The bounds are specified in terms of an equivalent standard normal variable (“u-space” in NESSUS terminology). The default bounds for each variable are -3 to +3 in u-space, which will correspond to +/-3 standard deviations for variables that are assigned normal distributions. NESSUS uses the bounds to generate a Latin Hypercube design. Unlike the Latin Hypercube Sampling reliability analysis method, this Latin Hypercube design is based on uniform distributions, regardless of the user-specified distribution types.

Finally, the user can also specify the number of times to sample the response surface when performing the probabilistic analysis. This sampling is performed using basic Monte Carlo simulation. The default is to use 100,000 samples.

4.3 Efficient Global Reliability Analysis (EGRA) method

The Efficient Global Reliability Analysis (EGRA) [7] method is similar to the Gaussian Process Response Surface Method in NESSUS, in that it incorporates all steps of the analysis including design of computer experiments, response surface fitting, and sampling. The difference is that EGRA uses an adaptive sampling approach to iteratively add new training points to the GP model, specifically targeting accuracy of the GP model in the vicinity of the limit state. By ensuring accuracy of the GP model in this critical region, EGRA can efficiently provide an accurate estimate of the reliability. EGRA is an extension of the Efficient Global Optimization (EGO) [8] method for numerical optimization.

EGRA uses the so-called expected feasibility function (EFF) to select the location at which a new training point should be added to the Gaussian process model by maximizing the expectation that the point lies on the limit state contour. A point could be expected to lie on the limit state contour if its predicted value is near the limit state value, or if the uncertainty in its prediction is such that there is a significant probability of its true value being near the limit state value. Because the uncertainty is higher in regions of the space with fewer observations, this provides a balance between exploiting areas that are predicted to be near the limit state, and exploring areas where more information is needed. The general procedure of EGRA is:

1. Build an initial Gaussian process model of the response function.
2. Iteratively add training data until a stopping criterion is met:
   a. Find the point that maximizes the EFF.
   b. Evaluate the true response function at the new point and add it to the training set.
3. Use Monte Carlo sampling or another appropriate probabilistic method to compute the probability of failure by sampling the converged Gaussian process response surface.
To define the expected feasibility function, let the limit state be denoted by the equality \( g(x) = z_0 \). The prediction from the GP model at any point in the design space follows a Gaussian distribution, and can be written:

\[
\hat{g}(x) \sim N[\mu_g(x), \sigma^2_g(x)]
\]

where the mean \( \mu_g(x) \) and variance \( \sigma^2_g(x) \) are the GP mean and variance conditioned on the training data (see, e.g., ref. [1]). Inspired by the contour estimation work in ref. [9], the expected feasibility function is defined as:

\[
EF(x) = E[\varepsilon(x) - \min(\hat{g}(x) - z_0, \varepsilon(x))]
\]

where the \( \varepsilon \) term is used to focus the search in the immediate vicinity of the response threshold. This expectation can be calculated by integrating over \( z_0 \pm \varepsilon \):

\[
EF(x) = \int_{z_0-\varepsilon}^{z_0+\varepsilon} (\varepsilon - |\hat{g} - g|) f_g \, dg
\]

This integral can be expressed analytically in terms of the standard normal probability density function and cumulative density function, as given in ref. [7].

In NESSUS, the initial training set is generated with \( \left[\left(n_{\text{var}} + 1\right) / 2\right] \) samples, where \( n_{\text{var}} \) is the number of random variables. These samples are generated following a Latin Hypercube design that is uniformly spaced with bounds based on critical values from an equivalent standard normal distribution. By default, the design space is defined based on equivalent standard normal values of \( \pm 5 \), but this value can be changed by the user. Alternatively, NESSUS allows the user to prescribe the search bounds for each variable in the original space (x-space).

To specify the expected feasibility, NESSUS uses \( \varepsilon = 2\sigma_g(x) \). The expected feasibility function is often multimodal, so NESSUS uses the DIRECT optimization algorithm [10] to ensure that the global optimum is found. The default stopping criterion for the expected feasibility is 0.00001, but this can also be changed by the user.

Finally, NESSUS uses multimodal adaptive importance sampling [11] by default to sample the final GP model for probability of failure calculation. However, the user can also select basic Monte Carlo sampling or Latin Hypercube sampling. The sample size can also be set by the user.

5 **NESSUS GLOBAL SENSITIVITY ANALYSIS**

Introduced in version 9.6, NESSUS includes variance-based global sensitivity analysis as an alternative analysis type to reliability analysis. When global sensitivity analysis is used, NESSUS decomposes the variance of the top-level response variable from the problem statement with respect to the user-defined input random variables. The main and total effect sensitivity indices [12] are computed for each variable.

Three numerical methods are provided for variance decomposition. The first is referred to in NESSUS as “Structured Monte Carlo” sampling, which implements the efficient sampling scheme outlined by [12]. The default is to use Sobol sequences, but NESSUS also supports Latin Hypercube sampling and basic Monte Carlo sampling. Sample estimates for the main and total effect indices are adapted from the equations proposed in refs. [13,14]. The second method available in NESSUS is the Fourier Amplitude Sensitivity Test (FAST) [15].
The third method available in NESSUS is “GP with analytical sensitivities”, which is similar to the GP Response Surface Method for reliability analysis, in that it performs both the design of computer experiments and response surface model fitting. The training points are generated based on a uniform Latin Hypercube design with bounds set at -3 and +3 in equivalent standard normal space. Once the GP model is constructed, the sensitivity indices are computed exactly, as described in ref. [5]. This method is only supported with random variables that have either normal or uniform distributions.

It is also possible to use sampling-based estimation of the sensitivity indices in conjunction with response surface modeling. This enables one to perform global sensitivity analysis for a GP model for random variable types that are not supported by the analytical calculations. It also allows the user to perform sensitivity analysis on response surface models that have already been built, or to achieve greater control of the sampling scheme used for design of computer experiments. This is done by using the regression model type described in Section 4.1 to define response surface models for one or more of the equations in the problem statement. The response surface model can then be specified based on either existing training data or a previously saved response surface model file.

6 CONCLUSIONS

NESSUS is a general-purpose software program for probabilistic analysis and was originally developed in the 1980’s as part of a long-term research and development program with NASA. This paper describes the most recent enhancements to the NESSUS software, which have focused on Gaussian process response surface modeling and global sensitivity analysis. NESSUS includes several options for employing GP models in reliability analysis, including adaptive sampling via the Efficient Global Reliability Analysis (EGRA) method. NESSUS also includes capabilities for variance-based global sensitivity analysis using sampling-based estimates or response surface-based solutions.

The NESSUS Response Surface Toolkit (RST) was introduced in 2015 as standalone companion application included with NESSUS. The RST provides a suite of tools for designing, fitting, assessing, visualizing, and making predictions with response surface models. The RST includes options for fitting both Gaussian process and polynomial regression models, and it can read and write to an open “.rsm” file format that is compatible with NESSUS.

REFERENCES


The graphical user interface of OpenTURNS, a UQ software in simulation

Michaël Baudin\textsuperscript{1}, Anne Dutfoy\textsuperscript{1}, Anthony Geay\textsuperscript{1}, Anne-Laure Popelin\textsuperscript{1}, Aurélie Ladier\textsuperscript{2}, Julien Schueller\textsuperscript{2}, and Thierry Yalamas\textsuperscript{2}

\textsuperscript{1}Phimeca Engineering. 18/20 boulevard de Reuilly, 75012 Paris - France, yalamas@phimeca.com
\textsuperscript{2}EDF R&D. 6, quai Watier, 78401, Chatou Cedex - France, michael.baudin@edf.fr

June 12, 2017

Abstract
The number of software for uncertainty quantification is already substantial, and this number is growing. OpenTURNS, for example, is a C++ library for uncertainty propagation by probabilistic methods. OpenTURNS is also available as a Python module and has gained maturity thanks to more than 10 years of development. However, there are situations where we do not want to use a programming language such as C++, Python (e.g. OpenTURNS) or Matlab. In this context, providing a graphical user interface (GUI) may allow to greatly increase the use of OpenTURNS and, more generally, of the UQ methodology.

In this talk, we will present how OpenTURNS’s GUI is integrated into SALOME, a platform for pre and post-processing of numerical simulations. Through examples, we will discuss the main features of the tool: central dispersion analysis, global sensitivity analysis and threshold probability estimate. Other advanced features will be presented during the session, with the aim of seeing how the tool can be used during training sessions or within a HPC context for example.

1 Introduction
The number of software for uncertainty quantification is already substantial, and this number is growing. OpenTURNS, for example, is a C++ library for uncertainty propagation by probabilistic methods. OpenTURNS is also available as a Python module and has gained maturity thanks to more than 10 years of development. However, there are situations where the engineer in charge of performing an uncertainty study does not want to use a programming language such as C++, Python (e.g. OpenTURNS) or Matlab. In this context, providing a graphical user interface (GUI) may allow to greatly increase the use of OpenTURNS and, more generally, of the UQ methodology.

2 Methods and tools
2.1 The uncertainty management methodology
The uncertainty management generic methodology consists of the following steps:
• Step A: specify the random inputs $X$, the deterministic inputs $d$, the model $G$, the variable of interest (model output) $Y$ and the quantity of interest on the output. The fundamental relation writes:

$$Y = G(X, d).$$

The model $G$ may be a symbolic function (defined by a character string), a Python function or a complex computer code. The quantity of interest can be, for example, the central dispersion, the probability to exceed a threshold or the whole distribution.

• Step B: quantify the sources of uncertainty. This step consists in modeling the joint probability density function (pdf) of the random input vector by direct methods (e.g. statistical fitting, expert judgment).

• Step B': quantify the sources of uncertainty by indirect methods using some real observations of the model outputs. The calibration process aims to estimate the values or the pdf of the inputs while the validation process aims to model the bias between the model and the real system.

• Step C: propagate uncertainties to estimate the quantity of interest. Depending on this quantity, the computational resources and the CPU time cost of a single model run, different specific methods can be applied such as analytical formula, Taylor expansion approximations, Monte Carlo sampling or metamodel-based techniques.

2.2 OpenTurns

OpenTurns[1, 3] is an open source software, available as a C++ library and a Python interface. It works under the Linux and Windows environments. The key features of OpenTurns are the following:

• open source initiative to secure the transparency of the approach,
• generic to the physical or industrial domains for treating of multi-physical problems,
• high performance computing,
• includes a variety of qualified algorithms in order to manage uncertainties in several situations,

OpenTurns is available under the LGPL license.

The main features of OpenTurns are uncertainty quantification, uncertainty propagation, sensitivity analysis and metamodeling.

Moreover generic wrappers allows to link OpenTurns to any external code $G$.

OpenTurns can be downloaded from its dedicated website www.openturns.org which offers different pre-compiled packages specific to several Windows and Linux environments. It is also possible to download the source files from the SourceForge server and to compile them within another environment: the OpenTurns Developer’s Guide provides advices to help compiling the source files.

2.3 Salome

Salome[4] is a pre and post-processing software which allows to create a 3D geometry, create the mesh, perform the numerical simulation, submit the simulation to a high performance computing resource and visualize the results. It provides tools to connect specific numerical simulators to Salome so that all the steps of the computation can be done in one single environment. Specific versions of Salome are dedicated to more restricted engineering domains: Salome-Meca for mechanical computations based on Code-Aster, Salome-Hydro for hydraulic computations based on TELEMAC / MASCARET and Salome-CFD for CFD studies based on Code-Saturne.
The YACS module allows calculation schemes in Salome to be built, edited, and executed. It provides both a graphical user interface to chain the computations by linking the inputs and outputs of computer codes and then to execute these computations on remote machines.

3 The graphical user interface of OpenTURNS

3.1 The main ideas of the GUI

In this section, we present the main principles of the graphical user interface, its goals and its external and internal design.

In order to propagate the uncertainties through a deterministic computer code, the input data we are required to provide to the graphical user’s interface are:

- the computer code \( G \): this can be a complex finite element software such as Code-Aster, a Python function or a simple one-line symbolic function.
- the joint distribution on the input \( X \) (if required, estimated based on a sample) defined by its marginal distribution and its dependence (defined in terms of a copula).

The main features of OpenTURNS are available in the GUI, but not all features. This does not limit the use of the tool, since advanced studies can be performed by programming the Python layer of the GUI which listens to Python statements, dynamically updating its appearance when required (thanks to a model-view architecture).

The result of the study is made of:

- the quantitative statistical results e.g. numerical values and tables,
- classical graphics such as histograms.

One principle of the GUI is that the computation is made as simple as possible by providing:

- the classical algorithms with a state-of-the-art implementation so that the users is guided in terms of methodology,
- default parameters of the algorithms whenever possible so that the non-expert user can use complex algorithms based on self-speaking parameters,
- an easy access to the high performance computing resources (provided in the JobManager of SALOME), submitting the job with parallel computing if required,
- an automated connection to the computer code (using the API integrated in SALOME).

3.2 Overview

The interface is organized so that user can progress step-by-step in the study. The main features of the GUI are the following.

- Physical model: defines the deterministic numerical simulator \( G \) through which we want to propagate the uncertainties.
- Probabilistic model: defines the joint probability distribution function on the input.

Once these two steps have been performed, the following features are available.

- Central dispersion: estimates the central dispersion of the output \( Y \) (e.g. mean, variance, distribution).
- Threshold probability: estimates the probability that the output exceeds a given threshold \( S \).
- Sensitivity analysis: estimates the importance of the inputs to the variability of the output.
3.3 Internal software architecture

In this section, we present the software architecture of the interface. This hidden part of the software required a lot of attention in order to provide both a flexible interface (for the user) and a powerful way of driving it (for the developer).

The architecture of the GUI provides two different entry points to the GUI:

- an interactive entry point based on keyboard and mouse interactions,
- a Python entry point which allows to program the GUI.

While the interactive entry point is the main goal of the GUI, the second entry point provides two different advantages.

- It allows to make Python unit tests of the GUI, which would otherwise be technically difficult to perform. Indeed, the `otgui` Python module allows to dynamically update the content of the GUI, as a user would do by interactively using the keyboard or the mouse. This feature is essential for the quality of the GUI.
- It allows to go beyond the widgets available in the GUI, by giving access to the internal data structures that can be used in combination with Python (including the OpenTURNS Python module embedded in the GUI) to perform computations not directly available in the GUI. This feature is essential for advanced users.

The figure 1 presents the architecture of the GUI, which is based on a C++ source code, organized in three layers: I0, I1, V0.

- The Model layer I0 only depends on OpenTURNS and can be called in Python (with an interface automatically generated with SWIG). This is the layer used in the unit tests.
- The Model layer I1 is a thin software layer which depends on QtCore and implements the signal/slot classes of the Qt C++ library.
- The View layer (V0) implements the view (i.e. the widgets that the user can interact with) and is made of C++ classes using the Qt C++ library.

The GUI depends on several components including OpenTURNS to perform the computations, YACS to evaluate the G function within SALOME (optional), Qt4 to provide the graphical widgets, Qwt for the Qt graphics and Sphinx to generate the help pages (optional).
4 A tutorial

4.1 Overview

The figure 2 presents the main window of the graphical interface. The work space contains the menu bar (on top), the toolbar (on top), the study tree (on the left) and the other graphical windows (on the right). In the bottom left of the figure, the Graphic configuration tool allows to configure each plot: the user can configure the X and Y labels, the range (i.e. the bounds of the plot) and the title.

On the top left side, the study tree contains the various objects created by the user during a study, as shown in the figure 3. Several studies can be managed at the same time. The name of each element created by can be modified by double-clicking on the element. The items Deterministic study, Probabilistic study and Designs of experiment cannot be renamed because these are the features of the tool and not objects created by the user. Right clicking on these elements gives access the context menu of the elements. Any item can be removed thanks to this context menu.

The bottom of the GUI provides a Python console, which can execute Python statements (including OpenTURNS statements) and dynamically updates the interface when required. This Python console is presented in the figure 4. This feature is possible because of the Model-View architecture of the GUI described previously. It allows to program the interface with Python statements. Furthermore, when a feature is not directly available in the GUI, the user can interact with Python objects in order to make computations that the GUI does not allow directly. For example, we can create a graphics with the Matplotlib Python module, which then appears as a separate graphics window.
4.2 The physical model

The first step in the study is to create a physical model. For some basic studies and during most training sessions, we often use a simple simulator, where the output $Y$ is a simple symbolic expression involving the inputs $X$. The figure 5 shows how the formula can be defined. The dialog box contains two tables:

- The *Inputs* table presents the input variables. Each input variable is described by its name, its description and its default value. The *Name*, generally a short word, is used in the tables or the graphics when a short description of a variable is required in order to save space in the dialog box or in the graphics. The *Description*, generally a full description of the physical meaning of the variable along with its unit, is used in the graphics, when the physical space is less limited. The *Value* is a default value parameter which is used if the variable is deterministic. It is also used to compute the default parameters of the associated marginal distribution. We can add or remove an input variable with the associated buttons.

- The *Outputs* table presents the output variables. Each output variable is described by its name, its full description and a symbolic formula. This formula can involve operators (e.g. $>$, $<$), functions (e.g. sin, sqrt) and constants (e.g. \_pi and \_e). The *Evaluate* button evaluates all the output variables depending on the input variables.

For more complex studies, a Python script can be used, which allows to use all the power of the Python environment (e.g. Numpy and Scipy) and may allow to connect to a software provided in executable form on the command line (e.g. a commercial software).

Perhaps one of the most spectacular feature is the ability to connect to complex simulators available in SALOME. Provided that the simulator comply to the requirements of the SALOME system and provides a YACS schema to evaluate the output $Y$ depending on the inputs $X$, the GUI can read the XML file defining this schema. Once done, the complex...
Figure 5: A physical model based on a symbolic formula.
The simulator can be evaluated without requiring any programming from the user. Moreover, any push to the Evaluate button submits the job to SALOME, using remote HPC resources if available: the data is transferred to and from the computer code without any programming from the user.

The figure 6 presents the physical model associated with a simulation based on Code-ASTER. This simulation computes the temperature $T$ on the top of a geometry depending on its height $H$ and the conductivity coefficient $C$ of the material. In order to propagate the uncertainties through this model, the user loads the XML file which defines the parametric deterministic ASTER study. This reads the input and output variables defined in the study along with the default values of the parameters. Moreover, the parameters associated with the computational resources for the evaluation of the YACS schema are loaded, so that the ASTER computation can be done in parallel.

### 4.3 The probabilistic model

The next step is to configure the joint distribution of the input vector, as shown in the figure 7. In the left part, the interface automatically creates one marginal for each input variable. This variable is enabled by default (with the checkbutton on the left) so that all variables are probabilistic. This can be disabled by unchecking the button if the corresponding variable is deterministic instead of the probabilistic. In this case, the default value, configured in the Physical model, is used. The probability distribution function of the variable can be chosen in the combobox under the Distribution column within a collection of 18 distributions including the most commonly used continuous distributions (e.g., normal, uniform, lognormal, triangular) or more advanced (Weibull, Gumbel, logistic, gamma, beta, exponential). The right part of the dialog box allows to configure the parameters of the distribution. The...
probability distribution function is plotted on the top of the dialog box. The user can plot either the probability distribution function or cumulated distribution function by using a radiobutton not shown in the figure. The bottom of the dialog box allows to select the parameters of the distribution. Any distribution can be optionally truncated with a lower bound or an upper bound (or both).

There are interesting features which are provided along with this important dialog box. When the probabilistic model is created, the default distribution is the normal distribution. Its parameters are computed from the default values of the physical model: its mean is the default value and its standard deviation is computed so that the coefficient of variation is 0.1 (i.e. the standard deviation is equal to the mean times 0.1). Then, if the user selects another distribution, the parameters of the distribution are also computed from the default values available from the physical model. For example, if the uniform distribution is chosen, the bounds are computed so that they are centered on the default value.

When the user changes the value of a parameter, the plot is automatically updated which allows to see how the changes in the parameters reflect in the distribution. The plot can be explored dynamically with the mouse.

- The left click allows to interactively change the X and Y range of the plot.
- The scroll wheel allows to zoom (scroll up) or unzoom (scroll down).
- Moving the pointer in the plot prints the X and Y values.

Moreover, the plot can be exported in various formats, including png and pdf.

For the distributions where several parametrization are available, the Type combobox selects which parametrization is to be used. For the Lognormal distribution for example,
we can configure the mean and standard deviation of the random variable or the mean and standard deviation of the logarithm of the random variable. Furthermore, the interface automatically computes the values of the parameters in the target parametrization from the values of the parameters in the source parametrization, solving nonlinear equations when required.

If there is a statistical dependency, we can optionally configure the Spearman rank which are converted into the covariance matrix of the associated Gaussian copula. The default is to consider a diagonal Spearman’s rank matrix, so that the variables are independent.

Limit state study

Suppose that you want to estimate the probability that the output is below a given threshold. The "Limit State" feature of the GUI allows to select the output, the operator and the threshold as shown in the figure 8.

We can now configure the parameters of the Monte-Carlo algorithm. This can be done by setting the maximum coefficient of variation of the estimated probability as shown in the figure 9. This ensures that the estimate of the threshold probability is at least as accurate as the coefficient of variation states. The default value of the coefficient of variation, 0.1, ensures that the probability estimate roughly has at least 1 significant digit, i.e. the order of magnitude of the probability is correct. In order to restrict the CPU time required to perform the simulation, the interface does not call the function G more than specified in the Maximum outer sampling field. Within these limits, the algorithms calls the function G by blocks which size is defines in the Block size field. This increases the performance of the simulation by vectorization. In the particular situation where the computer code G is parallelized, this parameter can be configured as the number of processors available in the supercomputer.

The Summary tab presented in the figure 10 displays the results of the study:

- the output used for the probability estimation,
- the number of simulations (i.e. calls to the physical model) actually performed,
- the failure probability, its coefficient of variation and the corresponding confidence interval at 95%.

The Histogram tab in the figure 11 presents the histogram of the output sample. The red vertical line in the histogram is the threshold. In the bottom left, the Graph settings dialog box allows to configure the labels of the graphics.

4.4 Central dispersion study

The GUI provides two algorithms for central dispersion studies: with Taylor expansion for nearly linear functions G or with Monte-Carlo. For the Taylor expansion, there is no parameter to adjust. Indeed, the gradient of the function is approximated based either on exact derivatives (if the function is symbolic) or with finite differences (if the function is a Python function or a YACS schema). In the Monte-Carlo dialog box presented in the figure 12, the user can configure its parameters, for example the size of the sample.

The Done button runs the simulation. The richest content is available for Monte-Carlo studies. The figure 13 presents the top part of the summary of the output of such a study. The top combobox allows to select the output that we want to study (here, the height H).
Figure 9: Limit state study: configuration of the Monte-Carlo algorithm.

Figure 10: Limit state study: the results.
Figure 11: Limit state study: histogram of the output.
The Minimum and Maximum section presents the two samples which lead to the minimum and maximum output H.

The figure 14 presents the bottom part of the Summary tab. This presents the moments estimates, including an estimate of the mean and a 95%

The other tabs present different graphics allowing to analyze the results. The Scatter plots tab presented in the figure 15 allows to plot any output versus any input. As with any other graphics in the interface, the "Graph settings" widget on the left allows to configure the graphics.

### 4.5 Sensitivity analysis

The global sensitivity analysis ranks the inputs with respects to their importance in the variability of the output. The interface provides two different methods to perform the sensitivity analysis.

- The standardized regression coefficients indices are based on a linear regression of the output with respect to the input variables. These indices are based on the hypothesis that the relation between the inputs and the output is linear.
- The Sobol’ sensitivity indices are based on a variance decomposition of the output. The interface estimates the first order indices, i.e. the separate influence of each variable, and the total order indices, i.e. the influence of each variable, including its interactions with other variables.

Both methods are based on sampling algorithms: the user must select the size of the sample (the default sample size is 10 000). The figure 16 presents the Sobol’ sensitivity indices of a given physical model. The top part of the dialog presents the plot of the sensitivity indices of each variable. The bottom part of the dialog presents the table of numerical results, where each line contains the first and total indices of a variable.

This dialog box is associated with interesting interactive features. If the user clicks on the "Total order index" column title, the lines of the table are sorted in decreasing or increasing order. The reordering automatically and dynamically updates the graphics, by reordering the labels of the variables in the X-axis. For example, this feature allows to order the variables by decreasing total order indices, so that variables with low indices can...
Figure 13: Central tendency study: the summary results (top part).

Figure 14: Central tendency study: the summary results (bottom part).
Figure 15: Central tendency study: the scatter plot of an output vs an input.
Figure 16: Sensitivity analysis: Sobol’ sensitivity indices.
be identified as non-important variables: these probabilistic variables can be replaced by a constant without changing the variability of the output. Another feature allows to detect an inconsistent sample size. Mathematically, the total order indice is always greater or equal to the first order indice. However, if the sample size is too low, the estimates do not respect this constraints. The GUI signals this by inserting warning icons within the table. In this case, the user should run a new sensitivity analysis with a greater sample size.

5 Current development

5.1 Future features

The next release will provide the following features.

- Step B :
  - given a univariate sample, fits a distribution by estimating its parameters and test the hypothesis based on the KS test,
  - given a multivariate sample, estimates the dependence by estimating the parameters copula which fits best (e.g. Gaussian, Franck, Gumbel),

- Step C :
  - for central tendency and sensitivity analysis, the user can configure the maximum coefficient of variation,
  - estimate a threshold probability with FORM and importance sampling algorithms,

- Metamodel : given a (X,Y) sample, create a metamodel (Generalized Polynomial Chaos or kriging) a convert it into a physical model.

5.2 Interactive uncertainty visualization in Paraview

Paraview[5] is a visualization software which is used in SALOME in order to visualize geometries, meshes and fields. However Paraview also provides multidimensional data analysis tools that are interesting with respect to uncertainty quantification. This is why we are developing the components that will allow the OpenTURNS GUI to embed Paraview graphics to visualize the results of simulations. The figure 17 presents the Plot Matrix View and Parallel Coordinate View for a sample generated by the OpenTURNS GUI. This sample has 1000 experiments and represents the link between the inputs Q and Ks and the output S.

The main advantage of the Plot Matrix View (PMV) and Parallel Coordinate View (PCV) graphics is the interactivity. For example, in the Parallel Coordinate View on the right of the figure, we have selected the output S greater than -5. In the Parallel Coordinate View, the corresponding computations have been drawn in red, the remaining lines being in black. Furthermore, the two graphics (PMV and PCV) are linked so that the corresponding experiments are highlighted in the Plot Matrix View. It allows to see that a large value of Q combined with a small value of Ks both lead to a large value of S. In order to get the graphics, we currently have to export the data in the CSV format and import it in the Paraview Spread Sheet View. In the future release, the graphics will be embedded within the GUI, along with the current Qt graphics.

5.3 Multidimensional uncertainty propagation

The current version of the GUI allows to map an input random vector to an output random vector. However, there are situations when the input or the output cannot be simply described as a vector. More precisely, it happens that the data is better described as a function of time and space. For example, consider the situation where the output is the thermal
Figure 17: Paraview: Plot Matrix View and Parallel Coordinate View.
power of a nuclear power plant. This power may be modeled as the solution of an ordinary
differential equation, so that the power is a scalar function of the time. Another example is
when the output is the height of a river for a given two dimensional spatial domain. In this
case, the central tendency study can be performed by a simple random sampling, leading to
a sample of fields.

Telemac-Mascaret is a set of computational fluid dynamics modules dedicated to free
surface flows and groundwater flows\cite{2}. The figure 18 presents a simulation based on the
hydraulic simulator TELEMAC2D. A sample of 70 000 simulations has been performed for
a 50 km segment of the Garonne river: the empirical mean on each point of the 2D mesh
has been performed. This shows that the mean river height is higher within the minor bed
of the river (as expected), but specific parts of the major bed might have a high mean as
well.

The current computation is essentially based on Python scripting, based on the Open-
TURNS Python module and TELEMACS2D simulations. The goal of this future develop-
ment is to be able to manage this situation in a fully automated way. This involves several
enhancements with respect to the current state.

- Being able to submit and retrieve a set of simulations performed by TELEMAC2D and
driven incrementally by OpenTURNS based on a given accuracy criteria. This is at
the interface between OpenTURNS which generates the inputs, TELEMAC2D which
generates the output and SALOME which submit the jobs on the supercomputer.
- Being able to estimate the central tendency of a field, which includes estimating the
mean and confidence intervals on this mean. This is a pure OpenTURNS computation.
- Being able to visualize the mean field and other statistical results as well. This is at
the interface of the GUI and Paraview.

References

\cite{1} Michaël Baudin, Anne Dutfoy, Bertrand Iooss, and Anne-Laure Popelin. Open-
TURNS: An Industrial Software for Uncertainty Quantification in Simulation, pages

\cite{2} Electricité de France, Sogreah, Hydraulic Research Wallingford, Centre d’Etudes Tech-


AERODYNAMIC SHAPE OPTIMIZATION UNDER FLOW UNCERTAINTIES USING NON-INTRUSIVE POLYNOMIAL CHAOS AND EVOLUTIONARY ALGORITHMS

Athanasios G. Liatsikouras\textsuperscript{1,2*}, Varvara G. Asouti\textsuperscript{1}, Kyriakos C. Giannakoglou\textsuperscript{1}, Guillaume Pierrot\textsuperscript{3}, Mustafa Megahed\textsuperscript{2}

\textsuperscript{1}National Technical University of Athens
School of Mech. Eng., Parallel CFD & Optimization Unit, Athens, Greece
\{liatsi,kgianna\}@central.ntua.gr, vasouti@mail.ntua.gr

\textsuperscript{2}ESI Software Germany GmbH
Kruppstr. 90, ETEC H4, 3.OG, 45145, Essen, Germany
\{athanasios.liatsikouras, mustafa.megahed\}@esi-group.com

\textsuperscript{3}ESI - Group
99 Rue des Solets, 94513 Rungis, France
guillaume.pierrot@esi-group.com

Keywords: Optimization under Uncertainties, Shape Optimization, Non-Intrusive Polynomial Chaos, Radial Basis Functions, Harmonic Coordinates, Shape Smoothing, Mesh Morphing.

Abstract. The existence of uncertainties, associated with the operating conditions or manufacturing imperfections, occur quite often in aerodynamic optimization problems. In this paper, a workflow for shape optimization in the presence of environmental uncertain flow conditions, varying stochastically around an average value with an a–priori known standard deviation, is presented. To do so, the Uncertainty Quantification (UQ) for the objective function needs to be carried out. This is based on the non-intrusive Polynomial Chaos Expansion (niPCE) method \cite{1}, which allows for a controllable number of calls to the CFD tool used to evaluate each candidate solution, compared to other sampling methods such as Monte–Carlo. Within the proposed workflow, PCE is combined with the optimization platform EASY (Evolutionary Algorithms SYstem) \cite{2}, which undertakes the optimization task. The overall process is fully automated. The shape under consideration is parameterized using CAD-free approaches, such as Radial Basis Function techniques or the combined use of two cages and the corresponding Harmonic Coordinates, which are responsible only for surface deformations whereas, as it will be explained below, other morphing/smoothing \cite{3} tool undertakes the adaptation of the CFD mesh to the updated boundaries at each optimization cycle. The PCE method selects the points (Gaussian nodes) in the design space to be evaluated and the computed performance metrics are integrated with weights indicated by the Gauss integration rules, in order to compute the mean value and standard deviation of the objective function of the flow problem under uncertainties. The aforementioned tools are applied in two problems, in which OpenFOAM is the CFD evaluation software.
1 INTRODUCTION

In aerodynamics, shape optimization under uncertainties [4,5] aims at designing a shape that performs efficiently in a range of operating points determined by stochastically varying flow conditions or in the presence of manufacturing imperfections. In this paper, a workflow that performs 2D or 3D aerodynamic shape optimizations under uncertainties, exclusively caused by variations in the flow conditions, is presented and demonstrated in two cases.

Let us denote the function characterizing the performance of an aerodynamic body (at a single operating point) by 

$$F = F(\vec{b}, \vec{U}(\vec{b})),$$

where \( \vec{b} \in \mathbb{R}^N \) is the design vector controlling the shape and \( \vec{U} \) denotes the flow variables. In the presence of uncertain flow conditions, the objective function to be optimized can be generally expressed as 

$$\hat{F} = \hat{F}(\vec{c}, F(\vec{c}, \vec{b}, \vec{U}(\vec{b}, \vec{c})))$$

to denote the dependency of \( \hat{F} \) on the stochastically varying environmental variables \( \vec{c} \in \mathbb{R}^M \). One of the most frequently used forms of \( \hat{F} \) is

$$\hat{F} = \hat{\mu}_F \pm \kappa \hat{\sigma}_F$$

in which \( \hat{F} \) is expressed as a linear combination of the mean value (\( \hat{\mu}_F \)) of \( F \) and its standard deviation (\( \hat{\sigma}_F \)) weighted by \( \kappa \). In the literature, there are some, stochastic or deterministic, methods of computing/estimating the mean value and the standard deviation of a function. A well-known stochastic technique is the Monte–Carlo (MC) [6] one, which may become very expensive in terms of CPU, since a great number of stochastic variable data-sets should be evaluated by the CFD software (in CFD-based shape optimization problems). To decrease the cost of MC, techniques such as Quasi-MC [7] and Latin-Hypercube Sampling [8] have been developed. On the other hand, a rival method to cope with the same problem is the Method of Moments [9], in which the adjoint method [10] and direct differentiation are used to compute up to second–order derivatives of \( F \) with respect to (w.r.t.) the environmental variables \( \vec{c} \) in order to get the first two statistical moments of \( F \) with second-order accuracy. This method proves to be a viable alternative of MC techniques, though its computational cost depends on the number of stochastic variables. In case of a gradient-based optimization method, \( \hat{F} \) needs to be differentiated w.r.t. the design variables \( \vec{b} \), which means that third–order mixed derivatives of \( F \) w.r.t the environmental (twice) and the design variables (once) are required. This has been presented, for the first time in the literature, by the NTUA group in [9,11].

In this work, the computation of the \( \hat{\mu}_F \) and \( \hat{\sigma}_F \) relies on the PCE technique [12]. There are two ways to implement this technique, the intrusive and non-intrusive one. In the intrusive PCE variant, every uncertainty affecting the flow model should be introduced in the governing equations, adding complexity to the problem to be solved. In the non-intrusive variant, which is the one used herein, the evaluation software is used as black–box to compute the objective function values for predefined data-sets of the uncertain variables. The latter are the Gaussian nodes determined by the Gauss integration rules according to which the computation of \( \hat{\mu}_F \) and its \( \hat{\sigma}_F \) is the weighted sum of \( F \) taken on at these nodes.

2 THE PROPOSED WORKFLOW AND ITS CONSTITUENTS

In shape optimization, a technique to parameterize the aerodynamic shapes should be available. In this framework, CAD–free approaches using either Radial Basis Functions (RBF) or control cages based on the Harmonic Coordinates (HC) are employed. The coordinates of the RBF centers or the HC cage knots form the design variables vector \( \vec{b} \). Over and above, a mesh morpher/smoker (Rigid Motion Mesh Morpher; R3M) is used to adapt the deformations in the 3D computational mesh. The latter is generated for the reference geometry using CFD–GEOM.
Regarding the uncertain variables (flow conditions), it is assumed that these follow a normal distribution around a–priori known average value and standard deviation. The general purpose optimization platform EASY (Evolutionary Algorithms SYstem, [2]) undertakes the optimization through Metamodel–Assisted Evolutionary Algorithms, MAEAs [14, 15]. The overall workflow is outlined in Fig. [1].

![Figure 1: Workflow for CFD-based shape optimization under uncertainties.](image)

### 2.1 Optimization Platform

The EASY software [2] can handle single- or multi-objective, constrained or unconstrained, optimization problems by accommodating any evaluation software as a black-box. EASY implements a population–based Evolutionary Algorithm (EA) that handles three populations, namely the parents, offspring and elites and applies evolution operators in conformity with the binary or real encoding of the design variables. For each offspring, the UQ w.r.t. a performance metric \( F \) (such as the drag, lift, losses, etc.) should be carried out. This is based on the niPCE where the evaluation tool (herein the CFD s/w) is used, for different sets of the uncertain variables, along with the Gauss quadrature integration formula. Among the important features of EASY is the smart use of low–cost surrogate evaluation models (metamodels; artificial neural networks) during the optimization giving rise to the so-called Metamodel–Assisted EA (MAEA). This is appropriate for computationally demanding problems such as those with uncertainties, where the niPCE requires many calls to the CFD tool. In the MAEA, local metamodels, on–line trained for each and every new individual generated during the evolution, are used. For all but the first few generations, the metamodels are used to pre–evaluate the offspring population by, practically, interpolating the \( \hat{F} \) value using archived previously evaluated individuals and only the most promising members (according to the pre–evaluation on metamodels) are selected to undergo CFD–based evaluation.
2.2 Control of the Shape and Mesh Morphing

This section describes briefly the two CAD–free approaches utilized in this workflow, based either on RBF networks or HC cages. These tools are responsible only for modifying the shapes whereas the CFD mesh is adapted to the changed shape using a mesh morpher.

2.2.1 Radial Basis Function Model

The use of the RBF model involves two steps, namely training and implementation. Initially, the user has to define the so–called RBF centers. The $K$ RBF centers can either be a subset of the mesh nodes lying on the shape to be optimized or some user–defined nodes around it. The coordinates of the RBF centers form the design vector. The displacement $\Delta \vec{r}$ of any CFD node on the shape, initially located at $\vec{r}$, is given by

$$\Delta \vec{r} = \sum_{i=1}^{K} \vec{w}_i \phi(||\vec{r}_{ci} - \vec{r}||)$$

where $\vec{r}_{ci}$ is the initial position vector of the RBF center $i$, $\phi$ is the RBF activation function and $\vec{w}_i$ weight coefficients, as many as the RBF centers, for each Cartesian direction. To compute them, the model needs to be trained and, to this end, eq.2 is applied $K$ times on the $K$ centers of the RBF model and needs to be satisfied for all Cartesian coordinates. For the training, the displacements $\Delta \vec{r}$, of the RBF centers, generated anew during the evolution, are used.

2.2.2 The Harmonic Coordinates two–cage Model

Cages based on HC, initially proposed for character articulation [16] is a technique to parameterize the points in a 2D or 3D domain. In an HC–cage–based parameterization, surface deformations are controlled by the so–called “cage” (a topologically flexible structure) that encloses the surface under consideration. An HC–based technique that may control both shape deformations and CFD mesh adaptation (not used, though, in the problems studied in this paper) to the new geometry has been proposed in [17], by adopting a two–cage control mechanism. The cages are filled with a quite coarse unstructured mesh and the nodal HC values on the coarse mesh are computed by solving as many Laplace equations as the cage control vertices, with appropriate boundary conditions. The HC fields are, then, interpolated from the cage nodes to the CFD mesh nodes. Thus, the CFD mesh (or its boundary only, as in the examined problems) can be controlled by displacing the HC cage control vertices, which become the design variables in the optimization problem. In the two–cage structure, the inner cage controls both the shape deformation and mesh morphing while the outer cage limits the effect of morphing and guarantees a smooth transition, without inverted elements, between the CFD mesh parts that lie inside and outside the outer cage. Thus, the outer cage vertices should remain still during the optimization and Laplace equations are not solved for them.
2.2.3 Morphing–Smoothing Tool

Usually, whenever a CAD–free parameterization tool, such as HC cages or RBF networks, is used, the same tool can also undertake the adaptation/deformation of the CFD mesh to the updated shapes. However, in the workflow presented herein, the CFD mesh adaptation is controlled by a mesh morpher.

In the literature, several morphing techniques such as those in may be found. Most of them cannot handle mesh anisotropies. The Rigid Motion Mesh Morpher and its adaptive smoother (R3M) [3], used in this work, is a meshless method and belongs to the optimization–based methods. The internal nodes of the CFD mesh are displaced so as to minimize a given distortion metric. Therefore, it can handle mesh anisotropies since it favours rigidity in the critical directions of imminent distortion. The basic idea is to keep some parts/elements of the CFD mesh (referred to as stencils) as-rigid-as-possible. An example of such a stencil handled by the morpher could be a cluster of neighbouring grid nodes. By definition, ‘Rigid Motion’ refers to a movement of a body without changing its shape or size. Obviously, this cannot be kept entirely rigid, since this would be in contrast with the mesh adaptation, even if there are ways to favour rigidity in the critical directions, when distortion of an element/part becomes imminent.

If a stencil \((s)\) in motion was totally rigid, its motion could be associated with a translation and a rotation velocity, \(\vec{\alpha}_s\) and \(\vec{b}_s\), respectively. In this case, the velocity of any node \(i\) belonging to this stencil would be

\[
u_{is} = \vec{\alpha}_s + \vec{b}_s \times (\vec{x}_i - \vec{c}_s)
\]

(3)

where \(\vec{x}_i\) is the position vector of node \(i\) that belongs to stencil \(s\). Since the motion of a stencil cannot always be entirely rigid, the actual velocity of node \(i\) would be \(u_i \neq u_{is}\). R3M and its smoother tends to minimize the difference between these two velocities (with some weighting coefficients \(w_s\) and \(\mu_{js}\)) for each and every stencil node. Moreover, in order to enforce smoothness of the surface, a subset of its nodes are selected as handles which control surface changes. For each handle node, there is also an underlying node which is free to move with a different velocity. By displacing the handle nodes (Fig. 2), the CFD mesh nodes are moved by minimizing the so–called ‘Morpher’s Energy’

\[
E = \sum_s w_s \sum_{j \in s} \mu_{js} \left[ \vec{u}_j - \vec{\alpha}_s - \vec{b}_s \times (\vec{x}_j - \vec{c}_s) \right]^2 + \sum_{i \in H} \lambda_i \left( \vec{u}_i - \vec{V}_t \right)^2
\]

(4)

where \(\vec{u}_j\) is the actual velocity of node \(j\), \(\vec{x}_j\) is its position vector, \(c_s\) the center of the gravity of the stencil \(s\), \(\vec{V}_t\) is the so–called target velocity of the handle (displacement of CFD mesh node \(i\) on the shape) and \(\lambda_i\) a weighting coefficient standing for the stiffness of a spring connecting the handle with its underlying node. If \(N_n\) is the number of all the CFD mesh nodes (including the surface nodes), \(N_s\) the number of stencils and \(N_h\) the number of handles, then the unknowns in eq. (4) are the \(N_n\) nodal velocities \((\vec{u}_i)\), the \(N_s\) translation velocities \((\vec{\alpha}_s)\) and the \(N_s\) rotation velocities \((\vec{b}_s)\) of the stencils. Once the \(N_h\) target velocities of the handles become known, the morpher’s energy, eq. (4) is minimized in the least squares sense, by finding the stationary points w.r.t. the corresponding unknowns by satisfying \(\frac{\partial E}{\partial \vec{u}_j} = 0, \frac{\partial E}{\partial \vec{\alpha}_s} = 0\) and \(\frac{\partial E}{\partial \vec{b}_s} = 0\), (3).
2.3 UQ using Non-Intrusive PCE

Let $F(\xi)$ be a stochastic function, where $\xi$ is stochastic variable and $w(\xi)$ its probability density function (PDF). We also assume a family of orthogonal polynomials $\Psi_i(\xi)$, where $i$ is the maximum degree of each polynomial. According to the PCE theory [1], $F$ can be approximated through the linear combination of orthogonal polynomials $\Psi_i(\xi)$,

$$F(\xi) \simeq f(\xi) = \sum_{i=0}^{q} \alpha_i \Psi_i(\xi)$$

(5)

where $q$ is the user-defined chaos order, truncating eq. [5] which might otherwise have an infinite number of terms. The mean value ($\mu_F$) and variance ($\sigma_F^2$) of $F$ can be expressed as

$$\hat{\mu}_F = \int f(\xi)w(\xi)d\xi \quad \sigma_F^2 = \int (f(\xi) - \hat{\mu}_F)^2 w(\xi)d\xi$$

(6)

By developing eqs. [6] and making use of appropriate Galerkin projections, the final expressions of the first two statistical moments of $F$, become

$$\hat{\mu}_F = \alpha_0, \quad \hat{\sigma}_F^2 = \sum_{i=1}^{q} \alpha_i^2$$

(7)

The PCE coefficients required in eqs. [7] result from the following integrals

$$\alpha_i = \int_D F(\xi)\Psi_i(\xi)w(\xi)d\xi, \quad i = 0, 1, \ldots, q$$

(8)

where $D$ is the design space in which $F$ is defined.

The integrals in eq. [8] can be evaluated by the Gauss quadrature integration formula in the domain $D$. According to this, each integral can be approximated by the weighted sum of the performance metric $F$ values at $n$ (the value of $n$ depends on the desired accuracy) points $z_i$ within the domain of integration, namely

$$\int_D F(\xi)w(\xi)d\xi = \sum_{i=1}^{n} t_i F(z_i)$$

(9)

where $t_i$ are known weights.
3 APPLICATIONS

In this section, the workflow (Fig. 1) is used and results are demonstrated in two external and internal aerodynamic problems. In both, OpenFOAM was used for the evaluation of each aerodynamic shape.

3.1 Optimization of a 2D Isolated Airfoil Under Uncertainties

The first problem is dealing with the design of an airfoil (Fig. 3), in which the goal is to optimize its shape by maximizing the objective function \( \hat{F} = \hat{\mu}_F - \hat{\sigma}_F \). The performance metric \( F \) is the lift coefficient \( C_L \), \( \hat{F} \) is thus defined in the worst case scenario. The initially generated CFD mesh has approximately 88K nodes and 160K elements. The flow conditions are: freestream Mach number \( M_\infty = 0.15 \) and Reynolds number based on the chord \( Re_c = 3.33 \times 10^6 \). The Spalart–Allmaras turbulence model [18] is used. The freestream flow angle \( (\alpha_\infty) \) is considered as the uncertain stochastic variable with mean value \( \mu_{\alpha_\infty} = 5^\circ \) and standard deviation \( \sigma_{\alpha_\infty} = 0.7^\circ \), which is assumed to follow a normal distribution. Before proceeding with the optimization, a UQ for the baseline geometry is carried out in order to decide about the PC order to be used. The results of this parametric study are tabulated in Table 1.

From Table 1 it can been seen that chaos order \( q = 2 \) is a good compromise in terms of accuracy and computational cost (3 CFD evaluations are needed for the UQ of a single airfoil). The airfoil is parameterized using a pair of HC cages, as in Fig. 3. The inner/control cage comprises 16 nodes, 14 of which are allowed to vary in both directions summing up to 28 design variables in total. The leading and trailing edges are not allowed to move. A \((\mu, \lambda) = (8, 12)\) MAEA with \( \mu \) parents and \( \lambda \) offspring was used for the optimization and the metamodels were activated after the first 30 evaluations. In the subsequent generations, all individuals were
pre-evaluated on the metamodels and the top two of them were selected for CFD-based re-evaluation. For a termination criterion of 200 evaluations an increase in the objective function ($\hat{F}$) by approximately 23% is achieved. In Fig. 4 the initial and the optimized geometry of the isolated airfoil are demonstrated along with the convergence of the optimization. The average cost per UQ on an Intel(R) Xeon(R) CPU E5-2630 v2 processor is 8 min. (on 6 cores), 88% of which stands for the cost of running the CFD evaluations.

Figure 4: Initial and optimized shape of the airfoil is presented. Furthermore, the convergence history of the optimization is presented (for $-\hat{F}$).

3.2 Optimization of S-Bend Duct Under Uncertainties

The second case deals with the shape optimization of an S-bend duct for minimum $\hat{F} = \hat{\mu}_F + \hat{\sigma}_F$, with $F$ being the total pressure losses across the duct. The latter is computed at the inlet/outlet cross sections, from the expression

$$\Delta p_t = \frac{\int (p + \frac{1}{2} \rho \bar{u}^2) \bar{u} \cdot \bar{n} dS}{\int \bar{u} \cdot \bar{n} dS}$$

(10)

where $\bar{u}$ is the velocity vector, $p$ is the pressure and $\bar{n}$ is the unit outward normal vector at the boundaries of the flow domain.

The 3D CFD mesh has been generated using CFD-GEOM and consists of hexahedrals at walls, a zone of prisms and tetrahedra everywhere else. The flow is laminar with $Re=550$. As uncertain variables, the inlet velocity and the kinematic viscosity are used. It is assumed that both uncertain variables follow normal distributions with mean values and standard deviations as in Table 2.

<table>
<thead>
<tr>
<th>Uncertain Variable</th>
<th>mean value</th>
<th>standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_{in}(m/s)$</td>
<td>10</td>
<td>0.5</td>
</tr>
<tr>
<td>$\nu(m^2/s)$</td>
<td>$10^{-4}$</td>
<td>$5 \cdot 10^{-5}$</td>
</tr>
</tbody>
</table>

Table 2: Mean values and standard deviations of the two uncertain variables in the S–bend duct optimization problem. Both follow normal distributions.

In Fig. 5 the initial shape of the duct is presented. The grey part of the geometry remains fixed whereas the (green) central curved part is free to deform.
Initially, a UQ study has been performed so as to make a decision on the appropriate chaos order. From Table 3, it is concluded that using chaos order equal to $q=2$ (9 CFD evaluations per UQ) is acceptable in terms of computational cost in order to compute the statistical moments of $F$.

<table>
<thead>
<tr>
<th>Chaos Order</th>
<th>Mean Value</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q=1$</td>
<td>137.89217</td>
<td>11.14798</td>
</tr>
<tr>
<td>$q=2$</td>
<td>137.73546</td>
<td>10.93479</td>
</tr>
<tr>
<td>$q=3$</td>
<td>137.77606</td>
<td>11.08064</td>
</tr>
</tbody>
</table>

Table 3: Optimization of the S–bend duct. Mean value and standard deviation of total pressure losses for each of the user-defined chaos order.

After having decided the chaos order, the central curved part of the duct, which is free to deform, is parameterized using RBF. The coordinates of the 24 RBF centers are selected as design variables for the optimization workflow using EASY. For this case a $(\mu, \lambda) = (8, 12)$ MAEA was used for the optimization along with a termination criterion of 150 evaluations. Metamodels were activated after the first 25 evaluations and, in each subsequent generation, the three top individuals were selected for CFD–based re–evaluation. After the optimization in this duct, the objective function ($\hat{F}$) is reduced approximately by 10% and the changes in shape are shown in Fig. 6. The average cost per UQ on an Intel(R) Xeon(R) CPU E5-2630 v2 processor is 38 min., about 94% of which is the cost of performing the 9 CFD evaluations.

Figure 6: Optimization of the S–bend duct. Total pressure distribution in the optimal duct at each Gaussian node (i.e. set of flow conditions) indicated by the Gauss integration formula. This optimal duct yields about 10% less $\hat{F}$ value than the starting geometry.
4 CONCLUSIONS

This paper presents an automated workflow for performing aerodynamic shape optimization under flow uncertainties. The UQ of the performance metric is based on the non–intrusive PCE technique. CAD–free approaches (RBF or HC cages) are controlling the shape under consideration and R3M with its adaptive smoother are responsible for adapting the CFD mesh to the updated geometry. The shape optimization is performed by an Evolutionary Algorithm assisted by surrogate evaluation models. The new optimization platform has great potential and each and every of its components could readily be replaced by other tools with the same functionalities.

5 ACKNOWLEDGEMENTS

This work has been conducted within the IODA ITN on: "Industrial Optimal Design using Adjoint CFD". The first author is an Early Stage Researcher (ESR) in this project which has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under the Marie Sklodowska–Curie Grant Agreement No. 642959.

REFERENCES


IDENTIFICATION AND QUANTIFICATION OF MULTIVARIATE POSSIBILISTIC MANUFACTURING UNCERTAINTY IN DYNAMIC NUMERICAL MODELS

M. Faes 1, D. Vandepitte 2, D. Moens 1

1KU Leuven - Department of Mechanical Engineering
Jan De Nayerlaan 5, B-2860 St.-Katelijne-Waver
e-mail: matthias.faes,david.moens@kuleuven.be

2KU Leuven, Department of Mechanical Engineering,
Celestijnenlaan 300, B-3001, Leuven, Belgium

Keywords: Uncertainty Quantification, Multivariate interval uncertainty

Abstract. The objective of this work is to validate a novel methodology for the identification and quantification of possibilistic multivariate uncertainty that has been presented by the authors in previous work. The method is based on the convex hull concept for both the representation of uncertainty in the result of an interval finite element computation, as for the variability that was measured on the real-life structure. Identification of the parametric multivariate interval uncertainty is performed by minimisation of a metric describing the discrepancy between these convex hulls. This method has been proven to be able do deliver an accurate identification of multivariate possibilistic uncertainty, which is also robust against certain measurement set metrics, however only on small-scale academic examples. This paper therefore first introduces a generic method for the reduction of the dimensionality of the identification at hand, and shows a validation of the method on a high-dimensional, complicated numerical model. Specifically, a test structure containing uncertainty in the stiffness of several bolted connections, introduced during the assembly of the structure, will be considered. The uncertainty in the stiffness of these connections is identified by using the presented method.
1 INTRODUCTION

In the context of incorporating uncertainty on model parameters into the numerical design models that are increasingly being used in industrial design processes, the interval method has been shown to give an accurate prediction of the occurring uncertainty in the model responses, based on limited data sets. Following this technique, the non-determinism is depicted as an interval on the model parameters thus propagated through the numerical model, thus eliminating the need for the identification of a full probabilistic data description, as needed for the probabilistic counterparts, which may be very cumbersome. Moreover, less expensive numerical procedures are necessary for the description of the variability [4, 11, 8], which makes these techniques highly suitable for early design stages.

However, in order to obtain a realistic assessment of the non-determinism in the responses of the model under consideration, an accurate estimation of the interval uncertainty at the input side of the model is pre-emptive. The authors proposed in this context a generic methodology for the identification of multivariate interval uncertainty [3, 2], based on the computation of the convex hulls over the computed realisations of the input non-determinism and the set of repeated measurement data. However the promising results obtained on simple academic cases, no validation of the method was performed on a model having a realistic amount of degrees of freedom. In order to apply the methodology on high-dimensional datasets, the dimensionality has to be reduced due to the exponential complexity of the computation of the convex hulls. This paper therefore introduces a generic reduction scheme. The method, including the identification procedure, is illustrated using the AIRMOD test structure (see e.g., [6, 5, 7, 13]) in order to show the performance of the method in a realistic model in conjunction with a high-dimensional interval uncertainty associated with the parameters of the model.

2 MULTIVARIATE INTERVAL IDENTIFICATION

By definition, an interval parameter $x$ is indicated using apex $I$: $x^I$. Vectors are expressed as lower-case boldface characters $\mathbf{x}$, whereas matrices are expressed as upper-case boldface characters $\mathbf{X}$. For the remainder of the text, interval parameters are either represented using the bounds of the interval $x^I = [\underline{x}; \bar{x}]$ or the centre point $\hat{x} = \frac{\underline{x} + \bar{x}}{2}$ and interval radius $r_x = \frac{\bar{x} - \underline{x}}{2}$.

2.1 The interval finite element method

2.2 Interval finite elements

The interval field FE method comes down to finding the solution set $\mathbf{y}$, when the model parameter uncertainty is depicted as an interval field $\gamma^I_F(r) \in \mathbb{R}^k$ over the geometrical model domain $\Omega$, with $\mathbb{R}^k$ the $k$-dimensional space of interval scalars and $r \in \Omega \subset \mathbb{R}^t$. $\mathbf{y}$ usually spans a multidimensional non-convex region in $\mathbb{R}^d$, and is therefore commonly approximated by an uncertain realization set $\mathbf{y}_s$, which is obtained by propagating $q$ deterministic realizations $\mathbf{y}_{sj}$ of the interval field $\gamma^I_F(r)$:

$$\mathbf{y}_s = \{ \mathbf{y}_{sj} \mid \mathbf{y}_{sj} = f(\gamma_{F,j}(r)); \gamma_{F,j}(r) \in \gamma^I_F(r) \}$$

(1)

Herein, $\mathbf{y}_{sj}$ is a vector containing the $d$ output responses of the deterministic solution of the propagation of the $j^{th}$ input uncertainty realization, with $j \in [1, q]$:

$$\mathbf{y}_{sj} = [y_{s1}, y_{s2}, \ldots, y_{sd}]$$

(2)
These deterministic propagations should represent the solution set \( \tilde{y} \) as close as possible. In the case of strict monotonicity of \( f() \), also the transformation method [8] can be used.

### 2.3 Multivariate interval identification

This section explains the novel methodology for the identification and quantification of multivariate interval uncertainty, as presented by the authors in [2, 3]. This methodology is based on the comparison of a set of repeated real-live measurements \( \tilde{y}_m \), performed on the physical component under consideration, with the result of the interval computation \( \tilde{y}_s \). For the comparison, a convex polytope is constructed around \( \tilde{y}_m \) and \( \tilde{y}_s \). Specifically, the convex hulls of \( \tilde{y}_m \) and \( \tilde{y}_s \), respectively \( C_m \) and \( C_s \), as well as the corresponding \( d \)-dimensional volumes \( V_m \) and \( V_s \), are hereto computed. All convex hull computations are performed using the QHULL library, which makes use of the Quickhull algorithm [1]. Identification of \( \alpha^I \) is performed by minimising a cost function \( \delta(\alpha^I) \), which described the discrepancy between \( \tilde{y}_s \) and \( \tilde{y}_m \):

\[
\delta(\alpha^I) = \left( \Delta V^2_m + \Delta V^2_o + \Delta c^2 \right)
\]  

(3)

with:

\[
\Delta V_m = 1 - \frac{V_s(\alpha^I)}{V_m}
\]  

(4a)

\[
\Delta V_o = 1 - \frac{V_o(\alpha^I)}{V_m}
\]  

(4b)

\[
\Delta c = \|c_m - c_s(\alpha^I)\|_2
\]  

(4c)

with \( c_m \) and \( c_s \) the geometrical centres of mass of respectively \( \tilde{y}_m \) and \( \tilde{y}_s \). \( V_o \) is the multi-dimensional volume of the overlap \( \tilde{y}_o \) between \( \tilde{y}_m \) and \( \tilde{y}_s \). The interval vector \( \alpha^{I,*} \), used for the construction of \( \gamma^I_F(r) \), is finally determined as:

\[
\alpha^{I,*} = \arg\min (\delta(\alpha^I))
\]  

(5)

### 2.4 Response set dimensionality reduction

The numerical computation of the convex hull and the corresponding multi-dimensional volume is done using the QHULL library, which uses the ”Quickhull” algorithm, as developed by Barber et al. [1]. The time complexity, of the Quickhull computation of a convex hull is in a worst case scenario:

\[
\mathcal{O}(\lfloor v_c \frac{d}{2} \rfloor / \lfloor \frac{d}{2} \rfloor !)
\]  

(6)

with \( v_c \) the number of vertices of \( C_s \) [1]. Therefore, care should be taken when computing convex hulls over large response vectors in \( \mathbb{R}^d \).

In order to prevent the computation of the convex hulls in general \( d \) dimensions, the measurement data set and uncertain realisation set are projected onto a \( d_r \)-dimensional orthogonal basis, with \( d_r << d \). In a first step, an orthogonal basis \( B \) is constructed in \( \mathbb{R}^{d_r} \), which is defined as:

\[
B = \{ \phi_{m,d-d_r}, \phi_{m,d-d_r+1}, \ldots, \phi_{m,d} \}
\]  

(7)
with \( d_r \) chosen as such that all non-zero dimensions are included in \( \mathcal{B} \), and \( \phi_m \) the eigenvectors corresponding to the \( d_r \) largest eigenvalues of the covariance structure of the measurement data set \( \tilde{y}_m \). Subsequently, all \( \frac{d_r!}{d_r!(d_r-d_r^+)!} \) \( d_r^+ \)-dimensional projections from this \( \mathbb{R}^{d_r} \)-dimensional are constructed, and they are gathered in a set \( \mathcal{B}^+ \).

For the actual reduction, the realisations \( y_{sj} \in \tilde{y}_s \) are first normalised to a normalised interval \( \hat{y}_{sj} \), having following properties:

\[
\hat{y}_{sj} = 0
\]

\[
r_{y_{sj}} = 1
\]

These \( \hat{y}_{sj} \) are subsequently concatenated into a matrix \( \hat{Y}_s \in \mathbb{R}^{d_r \times q} \), which is defined as:

\[
\hat{Y}_s = [ \hat{y}_{s1}, \hat{y}_{s2}, ..., \hat{y}_{sq} ]
\]

with \( \hat{y}_{sj} \), \( j = 1, ..., q \) the \( q \) realisations of \( \tilde{y}_s \), and are then projected onto the \( i^{th} \) projection \( \mathcal{B}_i^+ \) of the orthogonal basis \( \mathcal{B} \) such that:

\[
\hat{Y}_s^{n,r} = \hat{Y}_s^T \mathcal{B}_i^+ = [ \hat{y}_{s1}, \hat{y}_{s2}, ..., \hat{y}_{sq} ]
\]

The \( q \) realisations \( \hat{y}_{s}^{n,r} \) are then used to construct the reduced uncertain realisation set \( \hat{y}_s^{r} \). The convex hulls \( C_m \) and \( C_s \) are also computed in the \( d_r^+ \)-dimensional projections of this \( d_r \)-dimensional vector space. Incorporating this in the optimisation problem, presented in eq. (5), this yields:

\[
\delta(\alpha^I) = \sum_{i=1}^{d_r^+} \left( \Delta V_m^2(B_i^+) + \Delta V_o^2(B_i^+) + \Delta c^2(B_i^+) \right)
\]

with:

\[
\Delta V_m = 1 - \frac{V_m(\alpha^I, \mathcal{B}_i^+)}{V_m(\mathcal{B}_i^+)}
\]

\[
\Delta V_o = 1 - \frac{V_o(\alpha^I, \mathcal{B}_i^+)}{V_m(\mathcal{B}_i^+)}
\]

\[
\Delta c = \| \mathbf{c}_m - \mathbf{c}_s(\alpha^I) \|_2
\]

and where \( (\mathcal{B}_i^+) \) is used to indicate that the respective multidimensional volumes are computed using the \( i^{th} \) projection of the \( d_r \) dimensional orthogonal basis. The total error is computed as the sum of squared errors over all considered projections.

3 CASE STUDY

3.1 AIRMOD model

The DLR AIRMOD structure is used in this paper to demonstrate the application of the proposed uncertain set reduction method and the corresponding identification procedure. First, a measurement data set is generated by forward propagation of well-defined intervals that are
defined on model parameters. This high dimensional data set is subsequently reduced by projecting it on a lower-dimensional basis. The identification of the interval uncertainty, based on the method presented in [3, 2], is then performed using well-defined projections of the measurement data set in this lower-dimensional basis.

<table>
<thead>
<tr>
<th>Description</th>
<th>f</th>
<th>n_o</th>
<th>Description</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st wing bend</td>
<td>5.50 Hz</td>
<td>14</td>
<td>7th wing bend</td>
<td>145.91 Hz</td>
</tr>
<tr>
<td>3rd wing bend</td>
<td>15.11 Hz</td>
<td>15</td>
<td>2nd HTP bend</td>
<td>206.73 Hz</td>
</tr>
<tr>
<td>1st ant-sym. wing tors</td>
<td>31.31 Hz</td>
<td>16</td>
<td>1st HTP fre-aft bend</td>
<td>225.73 Hz</td>
</tr>
<tr>
<td>1st sym. wing tors</td>
<td>33.62 Hz</td>
<td>17</td>
<td>1st wing bend right</td>
<td>261.53 Hz</td>
</tr>
<tr>
<td>1st VTP bend</td>
<td>35.39 Hz</td>
<td>18</td>
<td>1st wing bend left</td>
<td>262.64 Hz</td>
</tr>
<tr>
<td>4th wing bend</td>
<td>44.66 Hz</td>
<td>19</td>
<td>3rd wing bend</td>
<td>278.71 Hz</td>
</tr>
<tr>
<td>1st wing fre-aft bend</td>
<td>47.21 Hz</td>
<td>20</td>
<td>1st winglet bend left</td>
<td>320.15 Hz</td>
</tr>
<tr>
<td>2nd wing fre-aft bend</td>
<td>52.91 Hz</td>
<td>21</td>
<td>1st winglet bend right</td>
<td>321.64 Hz</td>
</tr>
<tr>
<td>5th wing bend</td>
<td>60.59 Hz</td>
<td>22</td>
<td>3rd fuselat bend</td>
<td>324.12 Hz</td>
</tr>
<tr>
<td>1st VTP tors</td>
<td>67.69 Hz</td>
<td>23</td>
<td>2nd sym. wing tors</td>
<td>336.31 Hz</td>
</tr>
<tr>
<td>2nd fuse lat bend</td>
<td>102.59 Hz</td>
<td>24</td>
<td>2nd ant-sym. wing tors</td>
<td>341.15 Hz</td>
</tr>
<tr>
<td>2nd VTP bend</td>
<td>128.62 Hz</td>
<td>25</td>
<td>4th wing bend</td>
<td>343.55 Hz</td>
</tr>
<tr>
<td>6th wing bend</td>
<td>132.08 Hz</td>
<td>26</td>
<td>2nd fuse vert bend</td>
<td>359.54 Hz</td>
</tr>
</tbody>
</table>

The DLR AIRMOD structure is a scaled replica of the GARTEUR SM-AG19 benchmark. The physical AIRMOD structure is constructed from six aluminum beams that are connected by five bolted joints and weighs approximately 40 kg to represent the fuselage, wings, winglets, vertical tail plate (VTP) and horizontal tail plate (HTP). It has a wing span of 2.0 m, the fuselage length is 1.5 m and the height is 0.46 m. The complete FE model, constructed in NX Nastran, consists of 1440 CHEXA, 6 CPENTA and 561 CELAS1, 55 CMASS1, 18 CONM2 and 3 CROD elements, and is constructed after e.g., [5, 7, 13, 6]. This model is shown in figure ??

A set of 14 parameters including support and joint stiffness values, as well as mass parameters are selected for the benchmarking. These parameters, together with their deterministic values are shown in table [2]. Identification of the corresponding interval uncertainty is hence an identification problem in 28 dimensions. This model provides a challenge for the identification procedure due to high dimensionality of the problem, as well as the combination of symmetric/anti-symmetric mode combinations, as well as closely spaced eigenfrequencies. Table [2] lists the non-rigid eigenfrequencies that are obtained by solving the deterministic model. As may be noted, closely spaced eigenfrequencies are located around 30 Hz and 320 Hz. From this set, the 1st, 2nd, 4th, 8th – 14th and 26th eigenfrequency are selected for the identification.

Measurement data are generated by sampling a predefined set of intervals on the input parameters. The intervals that were used to construct the set of measurement data is shown in table [3] (indicated as \(\alpha^\star\)). Specifically, 85 samples are taken out of a uniform distribution between these interval boundaries. Special care is taken to track mode shifts between the different realisations. This is done by MAC-based mode correlation.

### 3.2 Surrogate modelling

The interval model is solved using the Transformation Method, leading to \(2^{14} = 16384\) deterministic model evaluations. Since each deterministic model evaluation takes approx. 10
Table 2: Parameters that are used in the identification

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Location</th>
<th>Orientation</th>
<th>Det. value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1$</td>
<td>Stiff. Sensor cable</td>
<td>VTP/HTP joint</td>
<td>$y$</td>
<td>$1.30 \cdot 10^5 N/m$</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>Stiff. Sensor cable</td>
<td>WING/FUSELAGE joint</td>
<td>$y$</td>
<td>$7.00 \cdot 10^6 N/m$</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>Stiff. Joint Stiff.</td>
<td>VTP/HTP joint</td>
<td>$x, y$</td>
<td>$1.00 \cdot 10^7 N/m$</td>
</tr>
<tr>
<td>$\gamma_4$</td>
<td>Stiff. Joint Stiff.</td>
<td>VTP/HTP joint</td>
<td>$z$</td>
<td>$1.00 \cdot 10^9 N/m$</td>
</tr>
<tr>
<td>$\gamma_5$</td>
<td>Mass Sensor cables</td>
<td>VTP/HTP joint</td>
<td>/</td>
<td>$2.00 \cdot 10^{-1} kg$</td>
</tr>
<tr>
<td>$\gamma_6$</td>
<td>Mass Screws and glue</td>
<td>Wingtips</td>
<td>/</td>
<td>$1.86 \cdot 10^{-1} kg$</td>
</tr>
<tr>
<td>$\gamma_7$</td>
<td>Mass Sensor cables</td>
<td>Wingtips</td>
<td>/</td>
<td>$1.50 \cdot 10^{-2} kg$</td>
</tr>
<tr>
<td>$\gamma_8$</td>
<td>Mass Sensor cables</td>
<td>Outer wing</td>
<td>/</td>
<td>$1.50 \cdot 10^{-2} kg$</td>
</tr>
<tr>
<td>$\gamma_9$</td>
<td>Mass Sensor cables</td>
<td>Inner wing</td>
<td>/</td>
<td>$1.50 \cdot 10^{-2} kg$</td>
</tr>
<tr>
<td>$\gamma_{10}$</td>
<td>Stiff. Joint Stiff.</td>
<td>WING/FUSELAGE joint</td>
<td>$x$</td>
<td>$2.00 \cdot 10^7 N/m$</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>Stiff. Joint Stiff.</td>
<td>WING/FUSELAGE joint</td>
<td>$y$</td>
<td>$2.00 \cdot 10^7 N/m$</td>
</tr>
<tr>
<td>$\gamma_{12}$</td>
<td>Stiff. Joint Stiff.</td>
<td>WING/FUSELAGE joint</td>
<td>$z$</td>
<td>$7.00 \cdot 10^6 N/m$</td>
</tr>
<tr>
<td>$\gamma_{13}$</td>
<td>Stiff. Joint Stiff.</td>
<td>VTP/FUSELAGE</td>
<td>$x$</td>
<td>$5.00 \cdot 10^7 N/m$</td>
</tr>
<tr>
<td>$\gamma_{14}$</td>
<td>Stiff. Joint Stiff.</td>
<td>VTP/FUSELAGE</td>
<td>$y$</td>
<td>$1.00 \cdot 10^7 N/m$</td>
</tr>
</tbody>
</table>

seconds of wall-clock time on a Intel Xeon E5-2695 @2.30 GHz, evaluating this interval model would take prohibitively long. Therefore, in order to reduce the computational cost, an Artificial Neural Network (ANN) approach is followed. Hereto, a (14:16:14:1) neural network is trained for each individual eigenfrequency based on a dataset containing 10000 Monte Carlo samples drawn from a uniform distribution between $0.1$ and $10$ times the deterministic values of the model parameters. The size of the network is selected in correspondence with recent work of Patelli et al. \cite{13}. Overtraining is prevented by using Bayesian regularisation for the training of the network \cite{10}. One call to the ANN model (compiled in C++) takes about $0.55 s$, when all 16384 deterministic realisation are propagated through in a vectorised fashion (i.e. simultaneously).
Figure 2 plots 100 randomly drawn samples (which are not included in the training set) from the model against the prediction of the ANN for a selection of eigenfrequencies for the 1st, 2nd, 4th, 6th, 7th and 8th eigenfrequency. As can be noted, a highly accurate prediction of the actual FE model response is obtained by the ANN model.

3.3 Results and discussion

Due to the high dimensionality and general non-convexity of the optimisation problem, introduced in eq. (5), the minimisation is performed using a hybrid optimisation approach. In a first step, a rough estimation for the global minimum is searched with the Particle Swarm Optimisation algorithm [9]. This estimation of the global optimum is then used as an initial estimate for a sequential quadratic programming approach [12]. Specifically, a swarm size of 250 uncertain model evaluations is used, and the algorithm is deemed to be converged to approximately the global optimum when it stalls for 10 generations. As a result of using the Particle Swarm algorithm, no initial assumptions on the constituting uncertainty are needed for the identification.

The identification is performed using respectively $d_r = 7$ and $d_r = 13$ orthogonal base vectors for the reduction of the set of the 26-dimensional measurement data set $\tilde{y}_m$ and the uncertain realisation set $\tilde{y}_s$. For both identification procedures, $d_r^+ = 2$, leading to respectively 21 and 78 projections of the $d_r$ dimensional vector space in which $\tilde{y}_r^+$ is defined. The results of these identification runs are listed in table 3. These results are normalised with respect to the goal multivariate interval uncertainty (indicated with an asteriks in 3).

As can be seen, no identification succeeds in finding the exact parameter set, based on the limited set of measurement data. However, the identification with $d_r = 13$ provides the best results of the two datasets, since more orthogonal base vectors are included herein. A first reason for not obtaining the exact parameters is that not all extreme realisations of the interval
uncertainty at the input of the model are included in the measurement data set, making an exact identification impossible. The root cause however for this, is that the used optimisation algorithms are not capable of finding the exact global minimum of eq. (12) due to the high dimensionality of the search space (28 independent parameters) and the general non-convexity of eq. (12). This inherently leads to inaccuracies in the identification.

The result of the identification with \( d_r = 13 \) are illustrated in figure 3. This figure shows \( d_r^+ = 2 \)-dimensional cross-sections of the convex hull of physical (non-transformed) model responses for the first 5 eigenfrequencies of the model. As may be noted, the convex hull of the model response (in blue) perfectly circumscribes the set of measurement data points (indicated as black crosses), albeit the result is somewhat over-conservative. However, some over-conservatism still is present in the identification, which is due to the fact that the optimiser did not converge to the exact global minimum. The latter is explained by the high dimensionality of the optimisation problem, as 28 uncertain parameters need to be jointly optimised in order to exactly identify the uncertainty in the model.

### Table 3: Parameters that are used in the identification

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \alpha_1^* )</th>
<th>( \alpha_2^* )</th>
<th>( \alpha_1 - 7 )</th>
<th>( \alpha_2 - 7 )</th>
<th>( \alpha_1 - 13 )</th>
<th>( \alpha_2 - 13 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_1 )</td>
<td>1.25 \cdot 10^{02}</td>
<td>1.32 \cdot 10^{02}</td>
<td>0.88</td>
<td>0.9</td>
<td>1.01</td>
<td>0.99</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>6.30 \cdot 10^{01}</td>
<td>7.50 \cdot 10^{01}</td>
<td>0.01</td>
<td>1.26</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>( \gamma_3 )</td>
<td>9.00 \cdot 10^{06}</td>
<td>1.30 \cdot 10^{07}</td>
<td>0.98</td>
<td>1.04</td>
<td>0.98</td>
<td>0.92</td>
</tr>
<tr>
<td>( \gamma_4 )</td>
<td>1.50 \cdot 10^{09}</td>
<td>2.10 \cdot 10^{09}</td>
<td>0.75</td>
<td>1.05</td>
<td>1</td>
<td>0.98</td>
</tr>
<tr>
<td>( \gamma_5 )</td>
<td>1.50 \cdot 10^{-01}</td>
<td>2.30 \cdot 10^{-01}</td>
<td>0.98</td>
<td>1.02</td>
<td>1.03</td>
<td>0.99</td>
</tr>
<tr>
<td>( \gamma_6 )</td>
<td>1.75 \cdot 10^{-01}</td>
<td>2.12 \cdot 10^{-01}</td>
<td>1.01</td>
<td>1.03</td>
<td>0.95</td>
<td>1.02</td>
</tr>
<tr>
<td>( \gamma_7 )</td>
<td>1.00 \cdot 10^{-02}</td>
<td>1.35 \cdot 10^{-02}</td>
<td>1.05</td>
<td>0.81</td>
<td>0.89</td>
<td>0.93</td>
</tr>
<tr>
<td>( \gamma_8 )</td>
<td>1.10 \cdot 10^{-02}</td>
<td>1.58 \cdot 10^{-02}</td>
<td>1.05</td>
<td>0.95</td>
<td>0.98</td>
<td>1.11</td>
</tr>
<tr>
<td>( \gamma_9 )</td>
<td>1.12 \cdot 10^{-02}</td>
<td>1.70 \cdot 10^{-02}</td>
<td>1.05</td>
<td>0.97</td>
<td>1.13</td>
<td>0.98</td>
</tr>
<tr>
<td>( \gamma_{10} )</td>
<td>1.85 \cdot 10^{-07}</td>
<td>2.00 \cdot 10^{-07}</td>
<td>0.99</td>
<td>1.05</td>
<td>1.02</td>
<td>0.99</td>
</tr>
<tr>
<td>( \gamma_{11} )</td>
<td>1.95 \cdot 10^{-07}</td>
<td>2.35 \cdot 10^{-07}</td>
<td>0.98</td>
<td>1.00</td>
<td>1.01</td>
<td>0.99</td>
</tr>
<tr>
<td>( \gamma_{12} )</td>
<td>6.35 \cdot 10^{-06}</td>
<td>7.10 \cdot 10^{-06}</td>
<td>1.00</td>
<td>0.89</td>
<td>0.93</td>
<td>1.20</td>
</tr>
<tr>
<td>( \gamma_{13} )</td>
<td>4.45 \cdot 10^{-07}</td>
<td>5.00 \cdot 10^{-07}</td>
<td>1.00</td>
<td>0.98</td>
<td>1.10</td>
<td>0.98</td>
</tr>
<tr>
<td>( \gamma_{14} )</td>
<td>0.75 \cdot 10^{-07}</td>
<td>1.02 \cdot 10^{-07}</td>
<td>1.05</td>
<td>0.90</td>
<td>0.75</td>
<td>1.21</td>
</tr>
</tbody>
</table>

4 CONCLUSIONS

The interval method has been proven to deliver an accurate, objective representation of the uncertainty that is present in the parameters of a numerical model, even when the datasets on this uncertainty are limited. However, in order to apply these powerful techniques, a realistic assessment of the uncertainty has to be made, based on measurement data. In this context, the authors presented a novel methodology for the identification and quantification of multivariate interval uncertainty in FE Models [3][2]. However, the application of this novel methodology to high-dimensional datasets still presents a challenge. This paper therefore introduced a methodology for the reduction of a high-dimensional data set. First, an orthogonal basis is constructed, based on the singular value decomposition of the covariance structure of the measurement data set. This basis is then used to reduce the dimensionality of both the measurement data as the result of the interval FE model. The quantification of the multivariate interval uncertainty, as presented by the authors in [3][2], is then performed using projections of the datasets on subspaces of this orthogonal basis in order to reduce the computational burden of the analysis.
Figure 3: 2-dimensional cross-sections of the convex hull of physical (non-transformed) model responses for the first 5 eigenfrequencies of the model. The response of the identified interval model is indicated in blue, whereas the measurement data points are indicated as black crosses.

The method was illustrated on the AIRMOD test structure, which proved to be a highly challenging example due to the high dimensionality at the input side of the model. As the used optimisation solver failed to converge to the global optimum, over-conservatism was introduced into the identification. Future work will therefore focus on improving this convergence, possibly by applying appropriate scaling and/or using more performing optimisation algorithms. Moreover the identification shall be performed using experimentally obtained modal data.

ACKNOWLEDGEMENTS

The authors would like to gratefully acknowledge dr. Yves Govers from DLR Germany for sharing the Finite Element model of the AIRMOD structure.

REFERENCES


AN APPROACH TO ROBUST DESIGN OF A CRUMPLE-ZONE STRUCTURE USING FUZZY ARITHMETIC

Markus Mäck and Michael Hanss

Institute of Engineering and Computational Mechanics
University of Stuttgart
Pfaffenwaldring 9, 70569 Stuttgart, Germany
e-mail: {markus.maeck, michael.hanss}@itm.uni-stuttgart.de

Keywords: Fuzzy arithmetic, crash simulation, robust design, non-probabilistic uncertainty, surrogate modeling

Abstract. Mathematical modeling and numerical simulation of crashworthy structures are state-of-the-art tools in automotive design. During the design phase, system parameters are only partially known or even unknown, leading only to a limited predictive significance of the simulation results. In this paper, an approach to robust design of a crumple-zone structure using fuzzy arithmetic is presented. The crumple zone is extracted from a full scale finite-element model of a Ford Taurus. In order to reduce the computation time, a surrogate model based on sparse-grid interpolation is derived. Using a sampling approach for fuzzy arithmetic, the simulation and analysis of the fuzzy-parametrized system is realized. The multivariate fuzzy output includes dependencies of the parameters and is therefore suitable for defining a robustness criterion. This leads to a multi-objective optimization problem including fuzzy-valued uncertainties to achieve a robust design of the crumple-zone structure.
1 INTRODUCTION

Traffic accident is still one of the most prevalent causes of death. According to data of the Global Health Observatory (GHO), provided by the World Health Organization (WHO), about 1.25 million deaths were caused by traffic accidents in the year 2013 [1]. In Germany alone, about 3500 people died by traffic accidents in 2015, and the recent statistics show an equal trend for the year 2016 [2]. Against this background, passenger safety plays a vital role in automotive research and development. Besides active safety systems, e.g. traction control systems and autonomous emergency braking, which are designed to avoid accidents, the crashworthiness of vehicles, and therefore, passive safety systems are crucial factors in protecting driver and passengers. In case of a frontal crash, the crumple-zone structure of the car plays an important role for the passenger safety and, thus, to prevent fatal and lethal injuries. Its main function is the absorption of the kinetic energy of a driving car by plastic deformation. In this respect, the resulting acceleration, which affects the passengers and acts as a loading quantity, needs to be kept below a critical value. Furthermore, the deformation of the crumple zone must not be too large, for example to prevent the penetration of the driver cabin by the engine mount.

Nowadays, mathematical modeling and numerical simulation of crashworthy structures are state-of-the-art tools in automotive design. However, the steadily increasing refinement and complexity of the models lead to a significant increase in computing time. To manage this drawback, reduction techniques, simplification procedures and surrogate models are used to obtain acceptable results within a reasonable time. In return, however, these simplification and approximation procedures give rise to uncertainty in the models, potentially supplemented by uncertainty due to insufficient or vague knowledge about parameters or crash conditions. Additionally, during the initial design phase, some parameters may be kept intentionally vague because of unknown requirements and specifications to achieve a later optimal design.

Common approaches to deal with uncertainties in a numerical way are based on probability theory. It is able to treat stochastic processes and models with parametric uncertainties triggered by randomness. The outcome usually provides a probability distribution based on a large sample size. However, if there is no variability and randomness of quantities, probability theory cannot handle these uncertainties appropriately. To cope with imprecision or a lack of information, interval analysis and possibility theory are best practice [3]. One way to handle these uncertainties in a possibilistic way is the use of fuzzy-valued model parameters instead of crisp ones and the application of fuzzy arithmetic to perform the uncertainty analysis.

In this paper, the crumple-zone structure, as an extracted substructure of a full-scale finite element model of a car, is modeled by uncertain, fuzzy-valued design parameters. With this fuzzy arithmetical approach, the influence of the design parameters on the acceleration and the absorbed kinetic energy during a frontal crash is investigated. With the help of the introduced methods, a fuzzy two-dimensional fuzzy set, i.e. a binary fuzzy relation, of the acceleration and the absorbed energy can be determined, showing that the resulting values are not independent of each other. On this basis, a robustness measure can be defined which can be used for further investigations.

To accomplish robust design, a trade-off problem between nominal optimality and robustness against uncertainty, caused by fuzzy-valued quantities for objective functions and constraints, must be solved. On this basis, a reliable and robust design for the crumple-zone structure can be achieved, depending on some predefined level of confidence.
2 CRUMPLE-ZONE MODEL

The crumple-zone structure is the most relevant part for passenger safety of a vehicle in case of a frontal crash. Due to high loads in the event of an impact, highly nonlinear processes have to be taken into account, including contacts, large deformations and nonlinear material behavior. To model the dynamic behavior of the structure, the finite-element method proves to be a suitable tool. In this work, the commercial finite-element code LS-DYNA is used [4]. In the following sections, the model structure and the model uncertainties during the design process are introduced and discussed.

2.1 Model Structure

In this work, the finite-element model of a Ford Taurus is used. The full-scale model is provided by the National Crash Analysis Center (NCAC) of the George Washington University under a contract with the FHWA and NHTSA of the US DOT. It consists of approximately 2.7 million degrees of freedom [5] and is, even with modern high performance cluster, too big for performing a large number of evaluations within a reasonable time. Therefore, only the crumple-zone structure, as the most relevant part of the full model, will be considered. The identified, relevant parts for this work are shown in Figure 1.

To obtain reasonable results, the extracted parts have to be preprocessed by adding additional constraints, boundary conditions and lumped masses. To model the inertia properties of the full vehicle, a lumped mass \( m_L = 500 \text{ kg} \) is added to the structure. The nodes on the backside of the structure are constrained, restricting the translation in vertical and lateral direction. To take account of the lateral stiffness of the structure resulting from the engine mount, additional rigid links are placed in lateral direction. The properties of the extracted and modified crumple-zone structure and the full-scale model are summarized in Table 1.

<table>
<thead>
<tr>
<th>Model</th>
<th>Elements</th>
<th>Nodes</th>
<th>Parts</th>
<th>Total mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full model</td>
<td>973416</td>
<td>922007</td>
<td>804</td>
<td>1739 kg</td>
</tr>
<tr>
<td>Crumple-zone structure</td>
<td>33717</td>
<td>32067</td>
<td>17</td>
<td>546 kg</td>
</tr>
</tbody>
</table>

Table 1: Properties of the finite-element model.

2.2 Model Uncertainties

During product development, some design parameters are still vague or only partially known, which implies some uncertainty to the model parameters. As an example for uncertain design parameters, the sheet thickness of three different parts of the crumple-zone structure are selected. The chosen parts are colored in blue in Figure 1. Hence, the system equations, which need to be solved, reformulates to

\[
\begin{align*}
\mathbf{M}(\mathbf{\tilde{p}})\ddot{\mathbf{u}}(t) + \mathbf{K}(\mathbf{u}(t), \mathbf{\tilde{p}}) &= \mathbf{f}(t) \\
\mathbf{z}(t) &= \mathbf{C}\mathbf{u}(t),
\end{align*}
\]

with neglected damping, \( \mathbf{u}(t) \) being the vector of state variables and \( \mathbf{\tilde{p}} \) denoting the vector of uncertain parameters. In this example, there exist no precisely studied random observations for the sheet thicknesses, but a lack of knowledge and therefore, missing information. To incorporate
those uncertainties of epistemic kind, appropriate methods are needed. Common probabilistic methods are not suitable to deal with uncertainties of the mentioned kind, because they rely on precise frequentist data and are not capable of modeling the occurring imprecision. As a possibilistic approach to incorporate those epistemic uncertainties, fuzzy arithmetic can be used. In fuzzy arithmetic, uncertain parameters are modeled in terms of fuzzy numbers $\tilde{p}_i$ with the entries $\tilde{p}_i$, $i = 1, 2, \ldots, n$, of the form $\tilde{p}_i = \{(x, \mu_{\tilde{p}_i}(x)) | x_i \in \mathbb{R}, \mu_{\tilde{p}_i}(x_i) \in [0, 1]\}$ and with $\mu_{\tilde{p}_i}$ being the membership functions. As an example, a triangular fuzzy number is shown in Figure 2 with $\bar{x}_i$ denoting the nominal value, $a_i$ the left-side deviation and $b_i$ the right-side deviation, respectively. The fuzzy-valued sheet thicknesses are modeled by symmetric triangular fuzzy numbers with worst-case deviations $a_i$ and $b_i$ of $\pm 10\%$ of their nominal values $\bar{x}_i$. 

![Triangular fuzzy number $\tilde{p}_i$.](image)
3 Fuzzy Arithmetic

The analysis of dynamical systems can be extended to the inclusion of uncertain parameters by the use of fuzzy numbers. Without loss of generality, the fuzzy-parametrized system of Equation (1) can be described as

\[ \tilde{q}(t) = F(\tilde{p}, u, t) \]  

(2)

with the fuzzy-valued parameters \( \tilde{p} \). The membership value of the result \( \mu_{\tilde{q}}(z) \) can theoretically be determined based on the extension principle introduced by Zadeh [6]. The membership function of the, in general, multivariate fuzzy output reads as

\[ \mu_{\tilde{q}}(z) = \begin{cases} \sup_{z=F(x,u,t)} \mu_{\tilde{p}}(x) & \text{if } \exists z = F(x,u,t), \\ 0 & \text{else.} \end{cases} \]  

(3)

It provides a possibilitic distribution of the outcomes with respect to their membership values, including worst-case bounds and nominal values. Hereinafter, two different implementations will be introduced to practically perform an uncertainty analysis by the use of fuzzy arithmetic.

3.1 The Transformation Method

One problem-independent implementation of fuzzy arithmetic is the Transformation Method introduced in [7]. Based on an \( \alpha \)-cut approach, the membership functions of the input fuzzy parameters \( \tilde{p} \) are decomposed into sets of nested intervals \( X_i^{(j)} = [a_i^{(j)}, b_i^{(j)}] \) related to \( m + 1 \) equally spaced levels of membership \( \mu_j = j/m \) with \( j = 0, 1, \ldots, m \). The obtained intervals are then transformed into arrays \( \hat{X}_i^{(j)} \) of the input samples containing the lower and upper interval bounds in a predefined scheme. After the evaluation of the model with the parameter combinations based on the arrays, the results are comprised by the arrays \( \hat{Z}_i^{(j)} \) and are retransformed into output intervals \( Z_i^{(j)} = [a_i^{(j)}, b_i^{(j)}] \). In the last step, the output intervals are recomposed into the output fuzzy number \( \tilde{q} \). The Transformation Method resorts to a cartesian grid for the evaluation steps. The union of the single-input combinations results in a non-regular grid for both, the reduced and the general Transformation Method.

3.2 The Sampling Approach of Fuzzy Arithmetic

In order to handle multi-dimensional fuzzy outputs and to solve implicit problems, the so-called sampling method, as introduced in [8], is presented and illustrated. For this sampling approach, a discretization of the membership functions is not needed. In order to receive the proper fuzzy arithmetical solution, the generation of appropriate sample points is moved to the very beginning of the evaluation. This is contrary to the common \( \alpha \)-cut approaches. In case of an explicit and non-adaptive sampling, this can be achieved in a straightforward manner. The multivariate membership function of \( n \) fuzzy numbers is obtained by

\[ \mu_{\tilde{p}} = t(\cdots t(\mu_{\tilde{p}_1}, \mu_{\tilde{p}_2}) \cdots , \mu_{\tilde{p}_n}) \]  

(4)

with the operator \( t(\cdot) \) representing a \( t \)-norm. In case of completely independent parameters, the minimum operator can be used. It yields a conservative estimation of the multivariate membership function and is therefore suitable if no information about the interdependency is present. In this work, the fuzzy parameters are considered as completely independent. In the following,
In the explicit case, the system is described by the mapping $f : X \to Z$ with the parameter domain $X \in \mathbb{R}^n$ and the output domain $Z \in \mathbb{R}^l$. The sampling sequence of $N$ points is

$$[x] = [x^{[i]}]_{i=1}^N = \left[ \left( x_1^{[i]}, \ldots, x_n^{[i]} \right) \right]_{i=1}^N, x^{[i]} \in \mathbb{R}^n,$$

and has to be constructed from the parameter space. With a pointwise evaluation of the function $f(x^{[i]})$, the corresponding pairs $z^{[i]}$ in the fuzzy output domain are computed. Additionally, the given joint membership function is to be evaluated, yielding potential membership values of $z$, and thus, potential boundary points of the membership function

$$\mu = \left[ \mu^{[i]} \right]_{i=1}^N = \left[ \mu_p (x^{[i]}) \right]_{i=1}^N$$

of the result. This is in contrast to the above-mentioned Transformation Method, which already provides points that lie on the membership curve. Because not all obtained points are valid candidates for the membership curve, an additional step, namely the reconstruction of the membership function, has to be performed, where the valid pairs $[\hat{z}, \hat{\mu}]$ are extracted from the results. For $z \in \mathbb{R}$, the simplest way to obtain the valid points of the membership curve is to construct a convex hull with piecewise linear functions. This is already a sufficient approximation in most cases. However, there are many additional possibilities to obtain an appropriate boundary curve, e.g. by approximating of the $\alpha$-cut bounds or the membership function [8].

In order to receive appropriate samples, the sampling of the parameter domain can be done in multiple ways. However, there exist some specific points which should definitely be included in the sampling sequence, independent of the chosen method: the nominal value of the input fuzzy relation with the joint membership $\mu_p$ to obtain the nominal value of the output domain, and the vertex points of the nested multi-dimensional hyper cuboids ($\bar{X} \subset X$) for constant $\alpha$-cuts, resulting in the same pattern as in the above-mentioned reduced Transformation Method. This leads to the proper fuzzy-valued results in case of monotonic behavior and includes the nominal value of the input fuzzy relation ($\mu_p = 1$), as a degenerated hypercuboid. Additionally, uniformly distributed random points can be included in the sampling sequence to handle even non-monotonic functions. Besides the random pattern, also regular patterns schemes or irregular pattern, like the general Transformation Method, can be used.

**Reconstruction of a multi-dimensional fuzzy output**

For a multi-dimensional fuzzy output, as e.g. the two-dimensional eigenvalues in the complex plane, the reconstruction step of the sampling method has to be adapted. One way is to project the obtained tuples $[z_1, z_2, \ldots, z_l, \hat{\mu}]$ on a predefined, cartesian grid. The output domain can be decomposed using hierarchical, recursive decomposition techniques. In the two-dimensional case, this can be achieved by using the quadtree decomposition [9]. The decomposition is completed, if a predefined minimum size of the quadtree elements is reached, or no output value is found. Using the extension principle by ZADEH, the supremum of the membership values of all fuzzy outputs lying in a quadtree element is determined. With this formulation, a piecewise constant membership function for the multi-dimensional fuzzy sets can be constracted, which is already a good approximation and contains valuable information. Observe that in the multi-dimensional case the output is in general not a fuzzy number because the convexity criterion does not necessarily hold.
4 RESPONSE SURFACE MODELING

Despite the increasing performance capacities of modern computer systems and high-perform-
ance clusters the above-mentioned methods still need an enormous number of system evalu-
atations to provide reasonable results. Even moderately complex systems with only a few fuzzy
parameters turn out to be challenging in terms of computation time. The two major influences
on the overall computation time are on the one hand the complexity of the actual model and,
thus, the computation time of a single evaluation, and on the other hand the number of system
evaluations based on the chosen fuzzy arithmetical approach. The latter is strongly dependent
on the number of fuzzy-valued parameters. Hence, the number of system evaluations, for a
given resolution, grows exponentionally with the dimension of the fuzzy-parameter space. The
so-called course of dimensionality impedes the straightforward evaluation of high-dimensional
models. In order to reduce the computational cost, one idea is to set up a surrogate model which
is able to represent the response of the system for the range of the fuzzy-valued parameters in
a decent manner. There exist several approaches to obtain a suitable surrogate model, such as
Krigin, polynomial basis or radial basis functions. In this paper, sparse-grid interpolation will
be used, as introduced in [10].

5 ROBUST DESIGN APPROACH

The fuzzy arithmetical analysis of the crumple-zone structure is performed with three fuzzy-
valued parameters as mentioned above. Instead of a perfect frontal crash, a quasi-frontal crash
with a wall angle of \( \alpha = 5^\circ \) is simulated, to enforce a more assymmetric behavior of the
crumple-zone deformation.
The influence of the fuzzy input parameters on the maximal absolute value of the acceleration
and the maximal absorbed energy is to be investigated. The acceleration is measured at the
mounting point. The translation and rotation of the mounting point are restricted, so that only
a translation along the driving axis is allowed. The acceleration serves as a loading measure
for the occupants of the car. Because LS-DYNA does not provide direct access to the plastic
deformation energy, the internal energy is chosen as a measure for the energy absorption. It also
takes into account the elastic deformation energy.
For the surrogate model, an adaptive sparse-grid model with 2000 evaluation points is con-
structed. The normalized, fuzzy-valued results of the maximal acceleration and internal energy
are shown in Figure 3. The results show a good conformity between the Transformation Method
and the sampling approach for an equally chosen number of evaluations. For the sampling ap-
proach, a combined random and pattern-based scheme is chosen which is able to recover outliers
of the solution (see Figure 3b). The two-dimensional fuzzy output is shown in Figure 4 on a
256 \times 256 grid. The result of the Transformation Method leads to an overestimation and a too
conservative result in the multivariate case. With the used sampling approach, the dependencies
of the two output parameters are evident. Observe that for the two-dimensional output signifi-
cantly more evaluations are needed to obtain an acceptable resolution.
To deduce an appropriate measure to rate the robustness of the model in respect of fuzzy-valued
uncertainty, the fuzzy-valued output and input are to be put in relation to each other. Such a
robustness measure can be introduced by

\[
\rho (\tilde{q}, \tilde{p}) = \frac{\int_{z \in Z \neq \emptyset} \mu_\tilde{q} \, dz}{\int_{z \in Z \neq \emptyset} dz}, \quad \frac{\int_{x \in X \neq \emptyset} dx}{\int_{x \in X \neq \emptyset} \mu_\tilde{p} \, dx}.
\]

(7)
It refers to the (relative) cardinality of fuzzy sets as described in [7] and describes the sensitivity of the dependent fuzzy output variables with respect to the fuzzy-valued input parameters. For the considered crumple-zone structure, the robustness measure results in $\rho = 0.0188$. Observe that the normalization is redundant in case of already normalized fuzzy sets. Also, only measures with equal dimensions in their input and output domain are comparable.

In order to obtain a robust optimal design, not only the absorbed energy of the crumple-zone structure has to be maximized and the maximal acceleration has to be minimized, but also the introduced robustness measure needs to be minimized. The problem can be reformulated as a multi-objective optimization problem. There exist multiple concepts to obtain an optimal solution in the presence of conflicting objectives. A short overview of gradient-free methods is given in [11]. One way of attaining a formulation for multi-objective optimization is the weighted-sum method. By this means, the different objectives are aggregated into one scalar optimization problem (see e.g. [12]). Some of the drawbacks of this method are outlined in [13]. The problem is formulated according to

$$\begin{align*}
\text{minimize} \quad & F(\mathbf{x}) = \left( \sum_{k=1}^{l} (\lambda_k f_k(\mathbf{x}))^p \right)^{\frac{1}{p}}, \quad \lambda_k \in [0, 1], \; p \in \mathbb{N} \\
\text{subject to} \quad & g_j(\mathbf{x}) \leq 0, \; j = 1, \ldots, m, \\
& x_i^l \leq x_i \leq x_i^u, \; i = 1, \ldots, n, \\
& \sum_{k=1}^{l} \lambda_k = 1,
\end{align*}$$

with the single, reformulated objective functions $f_k(\mathbf{x})$, the weighting factors $\lambda_k$, and $p = 1$.

A generalization of the weighted-sum approach is formulated as a weighted $L_p$-norm problem. The weights, and therefore, the preferences of the designer are determined a priori.

![Figure 3: Fuzzy-valued output parameters for the Transformation Method (TM) and the sampling approach (SA).](image)
6 CONCLUSIONS AND OUTLOOK

In this paper, a fuzzy arithmetical approach to robust design, regarding uncertainties of primarily epistemic type, is performed on an exemplary application, namely the crumple-zone structure of a Ford Taurus. The uncertainties arising from design parameters which are still to be determined are modeled in terms of fuzzy numbers. Explicitly, the sheet thicknesses of three different parts of the crumple-zone structure are treated as design parameters. In order to perform a fuzzy arithmetical analysis, two different methods are outlined and discussed. A surrogate model, based on sparse-grid interpolation, is used to significantly reduce the computation time. As a result, the fuzzy-valued acceleration and internal energy of the system are shown. With the sampling approach of fuzzy arithmetic, a two-dimensional output is constructed, showing the dependencies of the output parameters, and thus, obtaining a tighter bound of the fuzzy-valued output than e.g. the Transformation Method. The decomposition of the output domain is performed using the quadtree decomposition. Using the presented methods to perform a fuzzy arithmetical analysis, a new robustness criterion for the multivariate case can be deduced, which represents an appropriate measure for the sensitivity of the system output with respect to the fuzzy-valued input parameters. By this means, a multi-objective optimization problem for robust optimal design based on fuzzy arithmetic is formulated. Apparently, the presented paper only shows the principle approach for robust design. In the next step, the actual optimization of the crumple-zone structure is to be performed. For this purpose, appropriate design parameters, e.g. material properties, sheet thickness or geometry, need to be chosen. Since the optimization needs additional model evaluations, it can be included in the sparse-grid surrogate model. The addition of supplementary parts, as e.g. the engine mount, provides more accurate results. Using model order reduction techniques, the full-scale model is manageable in terms of computation time. Finally, regression on sparse grids can be used to
obtain a more suitable surrogate model of the crumple-zone structure.

Acknowledgements

The authors gratefully acknowledge the support of this research by the German Research Foundation DFG in the framework of the priority programme 1886 Polymorphic Uncertainty Modelling for the Numerical Design of Structures.

REFERENCES

INTERVAL DYNAMIC RESPONSE OF DISCRETIZED STRUCTURES WITH UNCERTAIN-BUT-BOUNDED PARAMETERS

F. Giunta¹, G. Muscolino¹ and A. Sofi²

¹ Department of Engineering, University of Messina
Villaggio S. Agata, 98166 Messina Italy
e-mail: fgiunta@unime.it, gmuscolino@unime.it

² Department of Architecture and Territory, University “Mediterranea” of Reggio Calabria, Salita Melissari, Feo di Vito, Reggio Calabria 89124, Italy
e-mail: alba.sofi@unirc.it

Keywords: Uncertain-but-bounded parameters, Interval Analysis, Vibration Analysis, Structural Dynamic Response, Lower and Upper Bounds of time-history responses.

Abstract. The dynamic analysis of structures plays an important role in the design of structural systems. Unfortunately, mechanical properties are usually uncertain due to physical imperfections and model inaccuracies. Therefore, it is important to estimate the effects of these uncertainties on the structural dynamic response. Since information regarding structural parameters is often quite limited, the use of non-probabilistic approaches is deemed advisable to realistically characterize uncertainties. In the framework of these approaches, the interval method seems today the most widely adopted.

The interval dynamic analysis of structures with slight parameter fluctuations is usually performed by applying the Interval Perturbation Method (IPM). However, the effectiveness of the IPM is limited to small uncertainties since the effect of neglecting higher-order terms is unpredictable.

In this paper, a novel procedure, able to overcome the main limitations of the IPM is proposed for evaluating the lower and upper bounds of the response of linear structural systems, with uncertain-but-bounded properties, subjected to dynamic deterministic excitations. The accuracy of the proposed method is assessed by evaluating the bounds of the response of a truss structure and a shear type frame subjected to seismic acceleration.
1 INTRODUCTION

The dynamic analysis of structures plays an important role in the design of structural systems. Unfortunately, the structural mechanical properties are usually uncertain due to physical imperfections and model inaccuracies. Therefore, it is important to estimate the effects of these uncertainties on the structural dynamic response. However, while the numerous available data permit to model with good accuracy the excitations as stochastic processes, unfortunately the data about structural parameters are often quite limited. It follows that the probabilistic approach cannot always be realistically applied to represent structural uncertainties; indeed, it requires a wealth of data, often unavailable, to define the probability density function of the fluctuating structural parameters. Non-probabilistic approaches can be alternatively used to treat these uncertainties. In this framework, the interval model seems today the most suitable analytical tool [1, 2]. The main advantage of the interval model is that it provides analytically rigorous enclosures of the solution, but its application to practical engineering problems is not an easy task due to two main drawbacks commonly faced in the development of interval-based procedures for structural analysis: i) the drastic overestimation of the interval solution range due to the so-called dependency phenomenon [3]; ii) the high computational costs when the exact combinatorial approach, known as Vertex Method, is adopted [3, 4].

Based on the matrix perturbation theory and interval extension of functions, the upper and lower bounds of the dynamic response were obtained by Chen et al. [5] using Taylor series expansion. Subsequently, the Interval Perturbation Method (IPM), which is based on Taylor series expansion and parameter perturbation, has been introduced to evaluate the interval dynamic response of structures subjected to deterministic [6-9] or stochastic excitations [10]. More recently, Gao et al. [11] presented the interval factor method to calculate the dynamic response of truss structures. Yang et al. [12] proposed an interval analysis method for dynamical systems using Laplace transform to solve the equations of motion. Xia and Yu [13-15] developed a modified IPM based on the modified Neumann expansion for the response analysis of interval structures and interval structural-acoustic systems.

Although other methods are more accurate, the IPM or, equivalently, the First-Order Interval Taylor Series Expansion, is the most widely used to evaluate the interval structural dynamic response. The main advantages of the IPM are the flexibility and the simplicity of the mathematical formulation. However, since the effect of neglecting higher-order terms is unpredictable, the effectiveness of this method is limited to uncertainties with small intervals.

In this paper, a novel procedure, which exhibits the same advantages of the IPM, is presented. The proposed method evaluates the bounds of the interval structural dynamic response by using the same numerical procedures traditionally adopted in the dynamics of structural systems without uncertainties. Moreover, the drawbacks of the IPM are overcome, as shown in the Numerical Application section, where the lower and upper bounds of the dynamic response of a truss structure and a shear-type frame are evaluated.

2 PROBLEM STATEMENT

Without loss of generality, the attention is herein focused on a quiescent $n$-DOF classically damped linear structural system subjected to a deterministic excitation $f(t)$. The equations of motion governing the dynamic response can be cast in the form:

$$M \ddot{u}(t) + C \dot{u}(t) + K u(t) = f(t)$$

(1)

where $M$, $C$ and $K$ are $n \times n$ mass, damping and stiffness matrices, respectively; $u(t)$ and $f(t)$ are $n \times 1$ deterministic vectors listing the nodal displacements and the external loads,
respectively; finally a dot over a variable denotes differentiation with respect to time $t$. In the following, the Rayleigh model is adopted for the damping matrix.

The elements of the stiffness matrix are assumed to be affected by uncertainties which are described by the dimensionless uncertain-but-bounded parameters $\alpha'_i$ ($i = 1, 2, \ldots, r$) with the apex $I$ meaning interval variable. In fact, a realistic situation is herein considered in which available information on the structural parameters is not enough to justify an assumption on their probabilistic distribution. According to the classical interval analysis [16,17], the $r$ uncertain structural parameters $\alpha'_i$ ($i = 1, 2, \ldots, r$), introduced before, are assumed to be independent. By applying the interval algebra formalism, the $i$-th uncertain parameter can be defined as $\alpha'_i \doteq [\underline{\alpha}_i, \overline{\alpha}_i] \in \mathbb{IR}$, where $\mathbb{IR}$ denotes the set of all closed real interval numbers, while $\underline{\alpha}_i$ and $\overline{\alpha}_i$ are the lower bound (LB) and upper bound (UB), respectively. Let the uncertain parameters $\alpha'_i$ be collected into the interval vector $\alpha' = [\alpha'_1, \alpha'_2, \ldots, \alpha'_r]^T$, with the apex $T$ meaning transpose operator, which is a bounded set-interval vector of real numbers $\alpha' \doteq [\underline{\alpha}, \overline{\alpha}] \in \mathbb{IR}^r$, such that $\underline{\alpha} \leq \alpha \leq \overline{\alpha}$, with the symbols $\underline{\alpha}$ and $\overline{\alpha}$ denoting the LB and UB vectors.

By taking into account the structural mechanical uncertainties, Eq.(1) can be rewritten as:

$$M \ddot{\mathbf{u}}(\alpha,t) + C(\alpha) \dot{\mathbf{u}}(\alpha,t) + K(\alpha) \mathbf{u}(\alpha,t) = \mathbf{f}(t), \quad \alpha \in \alpha' = [\underline{\alpha}, \overline{\alpha}]$$

(2)

where the stiffness matrix $K(\alpha')$ as well as the displacement vector $\mathbf{u}(\alpha',t)$ depend on the interval parameters $\alpha'_i$ collected into the vector $\alpha'$. Moreover, since the Rayleigh model is adopted for the damping matrix, the following relationship holds:

$$C(\alpha') = c_u M + c_k K(\alpha'),$$

(3)

where $c_u$ and $c_k$ are the Rayleigh damping constants having units $s^{-1}$ and $s$, respectively.

According to the improved interval analysis [10], the $i$-th real interval variable $\alpha'_i$ can be written in the following affine form [18]:

$$\alpha'_i = \alpha_{0,i} + \Delta \alpha_i \hat{\alpha}_i, \quad (i = 1, 2, \ldots, r)$$

(4)

where $\hat{\alpha}_i \doteq [-1,1]$ is the so-called Extra Unitary Interval (EUI) [10]; $\alpha_{0,i}$ is the midpoint value (or mean) and $\Delta \alpha_i$ denotes the deviation amplitude (or radius), defined, respectively, as:

$$\alpha_{0,i} = \frac{1}{2}(\underline{\alpha}_i + \overline{\alpha}_i); \quad \Delta \alpha_i = \frac{1}{2}(\overline{\alpha}_i - \underline{\alpha}_i).$$

(5a,b)

However, in structural engineering, the uncertain-but-bounded parameters can be reasonably assumed to be symmetric, i.e. $\overline{\alpha}_i = -\underline{\alpha}_i \equiv \alpha_i$, so that:

$$\alpha_{0,i} = \frac{\overline{\alpha}_i + \underline{\alpha}_i}{2} = 0; \quad \Delta \alpha_i = \frac{\overline{\alpha}_i - \underline{\alpha}_i}{2} = \alpha_i > 0.$$ \hspace{1cm} (6a,b)

Under this assumption, the generic interval variable can be written in affine form as:

$$\alpha'_i = \Delta \alpha_i \hat{\alpha}_i.$$ \hspace{1cm} (7)
Following the interval formalism above introduced, the interval stiffness matrix $K(a')$ can be expressed as a linear function of the dimensionless interval parameters $a'_i$, i.e.:

$$K(a) = K_0 + \Delta K(a), \quad a = a' = [\underline{a}, \overline{a}]$$

where $K_0$ is the nominal stiffness matrix, which is a positive definite symmetric matrix of order $n \times n$ and $\Delta K(a')$ is the interval deviation of the stiffness matrix with respect to the nominal one.

To solve the dynamical problem (2) involving the interval parameters (7), it is required to find at each time instant an interval vector containing the dynamical response set, i.e.

$$u(a', t) = [\underline{u}(t), \overline{u}(t)]$$

or in component form

$$u_j(a', t) = [\underline{u}_j(t), \overline{u}_j(t)], \quad j = 1, 2, \ldots, n$$

with

$$\underline{u}_j(t) = \min \{u_j(a, t) \mid u_j(a, t) \in \mathbb{R}, a \in \mathbb{R}^r\},$$
$$\overline{u}_j(t) = \max \{u_j(a, t) \mid u_j(a, t) \in \mathbb{R}, a \in \mathbb{R}^r\}; \quad j = 1, 2, \ldots, n$$

where the symbols $\min \{\ast\}$ and $\max \{\ast\}$ denote minimum (inferior) and maximum (superior) value, respectively, while $\{S(a) \mid P(a)\}$ means “the set of quantities $S(a)$ such that the proposition $P(a)$ holds”.

In the next section, a method to evaluate in approximate form the interval dynamic response is described.

### 3 BOUNDS OF INTERVAL DYNAMIC RESPONSES

The first step in vibration analysis of structures is the solution of an eigenproblem for evaluating the natural frequencies and the associated mode shapes. For classically damped linear discretized structures with $r$ uncertain-but-bounded parameters a generalized interval eigenvalue problem must be solved [19-22]:

$$K(a)\phi_j(a) = \lambda_j(a)M \phi_j(a); \quad a \in a' = [\underline{a}, \overline{a}], \quad (j = 1, 2, \ldots, n)$$

where $K(a')$ is the $n \times n$ stiffness matrix of the structural system which depends on the dimensionless uncertain parameters collected into the interval vector $a' \in \mathbb{R}^r$; $M$ is the $n \times n$ mass matrix; $\lambda_j(a') = \omega_j^2(a')$ is the $j$-th interval eigenvalue, equivalent to the squared interval natural frequency, and $\phi_j(a')$ is the associated interval eigenvector.

According to the classical interval analysis [16, 17], the interval stiffness matrix satisfies the following relationship:

$$K(a') = [K, \overline{K}] = \{K(a) \mid \underline{k}_i \leq k_y \leq \overline{k}_y\}$$

where $k_{ij}$ and $\overline{k}_{ij}$ are the $(i, j)$-th element of the stiffness matrices $K$ and $\overline{K}$, respectively. In vibration problems, $K \in K(a')$ is a symmetric positive definite matrix.
The solution of the generalized interval eigenvalue problem involves the evaluation of all possible eigenvalues satisfying Eq.(12) as the matrix $K(a')$ assumes all possible values inside the intervals defined in Eq.(13). The solutions constitute a complicated region in the real number field $\mathbb{R}$. Therefore, the objective is to evaluate, for each eigensolution, the narrowest interval enclosing all possible eigenvalues satisfying Eq.(12), i.e. [19-22]:

$$
\lambda_j(a) = \omega^2_j(a) = \left[ \lambda_j, \bar{\lambda}_j \right], \quad a \in a' = [a, \bar{a}]
$$

where $\lambda_j$ and $\bar{\lambda}_j$, \(j = 1, 2, \ldots, n\), are the LB and UB of the $j$-th interval eigenvalue.

The eigenvectors associated with the interval eigenvalues are also affected by the uncertainties and turn out to be bounded by interval vectors $\Phi_j \in \Phi_j(a')$.

Since the eigenvalues are monotonic functions of the uncertain parameters $\alpha_j \in \alpha'_j = [\alpha_j, \bar{\alpha}_j]$, \(j = 1, 2, \ldots, r\), then the bounds of the eigenvalues can be evaluated solving the following two deterministic eigenvalue problems [22]:

$$
K(a)\Phi_j^{(LB)} = \lambda_j M \Phi_j^{(LB)}; \quad \Phi_j^{(LB)T} M \Phi_j^{(LB)} = \Delta_{jk} \\
K(\bar{a})\Phi_j^{(UB)} = \bar{\lambda}_j M \Phi_j^{(UB)}; \quad \Phi_j^{(UB)T} M \Phi_j^{(UB)} = \Delta_{jk}, \quad (j = 1, 2, \ldots n).
$$

where $\Delta_{jk}$ is the Kronecker delta; $\Phi_j^{(LB)}$ and $\Phi_j^{(UB)}$ are the eigenvectors associated to the eigenproblem in which $a = a$ and $a = \bar{a}$, respectively. Notice that the two stiffness matrices $K(a)$ and $K(\bar{a})$ as well as the mass matrix $M$ are real, symmetric and positive definite matrices. Then, the eigenvectors of both eigenproblems are real vectors while the eigenvalues are real and positive quantities.

Introducing the diagonal matrices $\Omega^2$ and $\bar{\Omega}^2$ whose $j$-th element is $\lambda_j$ and $\bar{\lambda}_j$, respectively, and the matrices $\Phi^{(LB)}$ and $\Phi^{(UB)}$ whose $j$-th column is $\Phi_j^{(LB)}$ and $\Phi_j^{(UB)}$, respectively, it is possible to evaluate the response of quiescent structural systems in integral form as follows [23, 24]:

$$
y(a, t) = \int_0^t \Theta(a, t - \tau) V f(\tau) d\tau; \quad \Theta(a, t) = \begin{bmatrix} \Theta(a, t) \\ \Theta(\bar{a}, t) \end{bmatrix}
$$

where $y(a, t)$ and $y(\bar{a}, t)$ are the state variable vectors defined as:

$$
y(a, t) = \begin{bmatrix} u(a, t) \\ \dot{u}(a, t) \end{bmatrix}; \quad y(\bar{a}, t) = \begin{bmatrix} u(\bar{a}, t) \\ \dot{u}(\bar{a}, t) \end{bmatrix}
$$

and $V$ is the following $2n \times n$ matrix:

$$
V = \begin{bmatrix} 0 \\ M^{-1} \end{bmatrix}.
$$

In Eqs.(16), $\Theta(a, t)$ and $\Theta(\bar{a}, t)$ are the $2n \times 2n$ transition matrices given, respectively, by:
In the previous equations, \( g(a,t) \) and \( g(\bar{a},t) \) are two diagonal matrices whose \( j \)-th element can be evaluated, respectively, as:

\[
g_j(a,t) = -\frac{1}{\omega_j} \exp(-\xi_j \omega_j t) \left[ \cos\left(\omega_j \sqrt{1-\xi_j^2} \right) + \frac{\xi_j}{\omega_j} \sqrt{1-\xi_j^2} \right]
\]

\[
g_j(\bar{a},t) = -\frac{1}{\bar{\omega}_j} \exp(-\bar{\xi}_j \bar{\omega}_j t) \left[ \cos\left(\bar{\omega}_j \sqrt{1-\bar{\xi}_j^2} \right) + \frac{\bar{\xi}_j}{\bar{\omega}_j} \sqrt{1-\bar{\xi}_j^2} \right]
\]

where \( \omega_j \) and \( \bar{\omega}_j \) denote the \( j \)-th element of the diagonal matrices \( \Omega \) and \( \bar{\Omega} \), respectively; \( \xi_j = (c_M + c_K \omega_j^2)/2\omega_j \) and \( \bar{\xi}_j = (c_M + c_K \bar{\omega}_j^2)/2\bar{\omega}_j \) are the LB and UB of the \( j \)-th damping ratio under the Rayleigh condition (3).

Finally, the LB, \( \underline{y}(t) \), and UB, \( \overline{y}(t) \), of the state variable dynamic response vectors can be obtained as:

\[
\underline{y}(t) = \min \{y(a,t), y(\bar{a},t)\}; \quad \overline{y}(t) = \max \{y(a,t), y(\bar{a},t)\}.
\]

In the previous equations, the symbols \( \min \{\ast\} \) and \( \max \{\ast\} \) mean minimum (inferior) and maximum (superior) value component wise, respectively.

It is worth mentioning that the exact bounds of the dynamic response are not generally obtained setting all the uncertain parameters simultaneously to their LB and UB. Hence, more accurate results may be achieved by introducing in the present formulation the most common combinations of the extreme values of the interval variables \( \alpha_i^r \) among those detected at each time instant by performing a preliminary sensitivity analysis.

### 4 NUMERICAL APPLICATIONS

The first application concerns the 3D 26-bar truss structure with 18 DOFs, subjected to an impulsive load \( f(t) = 1000 \delta(t) \) N, as shown in Figure 1. The following geometrical and mechanical properties are assumed: nominal cross-sectional area of the bars \( A_0 = A_{0,j} = 4.27 \times 10^{-4} \text{ m}^2 \) and nominal Young’s moduli \( E_0 = E_{0,i} = 2.1 \times 10^8 \text{ kN/m}^2 \), \( i = 1,2,\ldots,26 \). Young’s moduli of \( r = 13 \) bars are modeled as interval variables, \( E_i^r = E_0 (1 + \Delta \alpha \hat{\alpha}_i) \), \( i = 1,2,\ldots,13 \) (see bar numbering in Figure 1).

According to Eq. (16), the two dynamical responses under the impulsive load, pertaining to the LB and UB of the uncertain parameters, can be evaluated as:

\[
y(a,t) = \Theta(a,t) v; \quad y(\bar{a},t) = \Theta(\bar{a},t) v
\]
where \( \mathbf{v} \) is a 18×1 vector having all elements equal to zero except the 18-th one which is equal to 1000 N.

![3D 26-bar truss structure with uncertain Young’s moduli of \( r = 13 \) bars.](image)

In Figures 2 and 3, the LB and UB time histories of the horizontal displacement \( u_{18}(t) \) are plotted for two deviation amplitudes of the interval uncertainties, say \( \Delta \alpha = 0.1 \) and \( \Delta \alpha = 0.2 \), respectively. The proposed bounds are compared with the ones obtained by applying the IPM as well as with the exact bounds provided by the Vertex Method. The latter requires to perform the deterministic analysis in the time domain for all possible combinations of the bounds of the interval Young’s moduli, say \( \mathcal{Z} \), and then take at each time instant the minimum and maximum among the corresponding responses. By inspection of these figures, it can be observed that the proposed method is able to predict the bounds of the dynamic responses with great accuracy. Furthermore, it is worth remarking that the proposed method is much more accurate than the IPM as the degree of uncertainty increases.

![Time-histories of the a) UB and b) LB of the horizontal displacement \( u_{18}(t) \) for a deviation amplitude of the uncertain parameters \( \Delta \alpha = 0.1 \) : comparison between IPM, Proposed method and Vertex Method.](image)
As second application the shear-type frame depicted in Figure 4 is analyzed.

This frame has a uniform story height $H = 3.0 \text{ m}$ and a bay width $L = 6.0 \text{ m}$, as shown in Figure 4. The beams are considered rigid to enforce a typical shear building behaviour. Under this assumptions, the shear-frame is modelled as a three DOFs linear system. The tributary mass per story, $M$, accounting for the structure’s own weight, as well as for permanent and live loads, is equal to $M = 81520 \text{ kg}$. Young’s moduli of columns are modelled as uncertain-but-bounded parameters, $E_i = E_0 (1 + \Delta \alpha \hat{e}_i) , \, i = 1,2,\ldots,9$, with $\Delta \alpha$ denoting the deviation amplitude and $E_0 = E_{0j} = 3.15 \times 10^7 \text{ kN/m}^2$. The frame is subjected to the seismic acceleration recorded at El Centro (1940).

The accuracy of the presented method can be detected by inspection of Figures 5 and 6, where the proposed time-histories of the UB and LB of the displacements of the third floor, $u_3(t)$ and $\bar{u}_3(t)$, are compared with the exact and IPM solutions for two different deviation amplitudes of the uncertain parameters, say $\Delta \alpha = 0.1$ and $\Delta \alpha = 0.2$. The exact bounds are evaluated following the philosophy of the Vertex Method. Notice that, contrary to the IPM, the proposed method gives accurate estimates of the LB and UB of the response even when relatively large uncertainties are involved. It is worth mentioning that the presented procedure
is much more efficient from a computational point of view than the Vertex Method, especially when a large number of uncertain parameters is considered.

![Figure 5](image5.png)

Figure 5. Time-histories of the a) UB and b) LB of the horizontal displacement $u_3(t)$ for a deviation amplitude of the uncertain parameters $\Delta \alpha = 0.1$: comparison between IPM, Proposed method and Vertex Method.

![Figure 6](image6.png)

Figure 6. Time-histories of the a) UB and b) LB of the horizontal displacement $u_3(t)$ for a deviation amplitude of the uncertain parameters $\Delta \alpha = 0.2$: comparison between IPM, Proposed method and Vertex Method.

## 5 CONCLUSIONS

In this paper, a novel procedure for evaluating the bounds of the response of linear structural systems with uncertain-but-bounded properties, subjected to dynamic deterministic excitations has been proposed. The proposed procedure is able to overcome the main limitations of the IPM, as shown by numerical results.

The main steps required by the proposed approach are: i) to solve two deterministic eigenvalue problems, where the uncertainties are set to the LB, $\underline{\alpha}$, and UB, $\bar{\alpha}$, in order to obtain the LB and UB of the eigenvalues; ii) to compute two transition matrices; iii) to compute the two state variable responses $y(\underline{\alpha},t)$ and $y(\bar{\alpha},t)$; iv) to evaluate the LB and UB of the structural response by handy formulas.

Numerical results have demonstrated that the proposed method provides estimates of the bounds of the dynamic response very close to the exact ones, even for relatively large uncertainty levels.
REFERENCES


RESPONSE OF BEAMS WITH CRACK OF UNCERTAIN-BUT BOUNDED DEPTH SUBJECTED TO DETERMINISTIC OR STOCHASTIC LOADS

G. Muscolino\textsuperscript{1} and R. Santoro\textsuperscript{1}

\textsuperscript{1}Department of Engineering, University of Messina
Villaggio S.Agata, 98166 Messina Italy
e-mail: gmuscolino@unime.it, roberta.santoro@unime.it

Keywords: Damaged beam, uncertain crack depth, interval deterministic response, interval stochastic response.

Abstract. Detection of cracks in structural components and identification of their size for structures having beam form is of crucial importance in many engineering applications. For damaged structures the dynamic response changes with respect to the undamaged ones due to the changes produced on their mechanical properties by the presence of the crack.

In this paper the deterministic behavior of a beam with a transverse on edge non-propagating crack is first studied. Moreover the deterministic and stochastic setting pertaining the case in which the crack has an uncertain depth is investigated. Undamaged elements of the beam are modeled by Euler-type finite elements. The uncertain crack depth is modeled as an interval variable and the cracked beam is subjected to both deterministic and zero-mean non-stationary Gaussian random excitations. In the latter case the equation governing the evolution of the main statistics of the response are derived by means of Kronecker algebra.

Once the mathematical model of the beam is defined, the dynamic response is evaluated by applying a numerical procedure based on the philosophy the Improved Interval Analysis via Extra Unitary Interval. In particular the proposed procedure is based on the following main steps: i) to define a finite element model of the beam in which the model of fully open crack is used to represent the damaged element; ii) to model the crack depth as an interval variable; iii) to evaluate in time domain the response for deterministic and stochastic excitation, by adopting an unified approach.
1 INTRODUCTION

The fracture behavior of structural components subjected to various loading and environmental conditions is of relevance in assessing structural integrity. When a structure is subjected to damage its dynamic response changes due to the change of its mechanical characteristics [1,2].

In this framework an interesting issue is the effect of a single crack on the structural response [3-6]. For the dynamic analysis the presence of the crack is usually neglected in the evaluation of the mass and damping, so only the stiffness is affected by the crack. This study is performed adopting a finite element model for the damaged beam with an on-edge non-propagating crack. Undamaged elements of the beam are modeled by Euler-type finite elements.

It is impossible to be certain of the crack depth, consequently this quantity should be assumed uncertain. Therefore, it is important to estimate the effect of this uncertainty on the structural dynamic response [7]. However, while the numerous available data permit to model with good accuracy the excitations as stochastic processes, unfortunately the data about the structural parameters are quite limited. It follows that the probabilistic approach cannot be realistically applied to represent structural uncertainties; indeed, it requires a wealth of data, often unavailable, to define the probability distribution density of the fluctuating structural parameters. For this reason the uncertain crack depth is here modeled as an interval variable [8]: in this model the uncertainty is defined by the knowledge of the lower and upper bound only.

The aim of this paper is to determine the lower and upper bounds of the response of damaged beams with uncertain-but-bounded crack subjected to both deterministic and stochastic excitations modeled as zero mean Gaussian random processes. To this aim the method proposed by Muscolino and Sofi [9] is extended to cracked beams. Specifically, the method adopts an improvement of the classical interval analysis [8], introducing a particular unitary interval, called Extra Unitary Interval (EUI) to split both the time dependent deterministic and random response, in modal subspace, as sum of two aliquots: the midpoint or nominal solution and the deviation. For stochastic excitations the Kronecker algebra [10] is adopted to derive the differential equations governing the time-evolution of the midpoint and deviation covariance vectors of the response under zero-mean Gaussian stochastic input process.

Once such equations are solved, the upper and lower bounds of the deterministic response as well as the covariance vector are evaluated by applying handy formulas.

In particular the proposed procedure is based on the following main steps: i) to define a finite element model of the beam in which the model of fully open crack is used to represent the damaged element; ii) to model the crack depth as an interval variable; iii) to evaluate in time domain the response for deterministic and stochastic excitation, by adopting an unified approach based on the Improved Interval Analysis via EUI [9].

Numerical results concerning a cracked beam with uncertain-but-bounded stiffness properties under deterministic and uniformly modulated white noise excitation are presented to show the effectiveness of the proposed method.

2 CRACK MODEL DESCRIPTION

The adopted mathematical model used for the damaged beam with a transverse on-edge non-propagating crack is based on the finite element model proposed in Refs. [1,2].

According to Saint-Venant principle only the element that contains a central crack is modified, being the stress field affected only in the region adjacent to the crack. Such a perturba-
The examination of the stress field is relevant especially when the crack is open and determines a local reduction of the flexural rigidity.

It follows that the element stiffness matrix, with the exception of the terms which represent the cracked element, may be regarded as unchanged under a certain limitation of the element size.

Undamaged parts of the beam are modelled by Euler type finite elements with two nodes and two degrees of freedom (transverse displacement and rotation) at each node.

Neglecting shear action, the strain energy of an element without a crack can be written as

\[
W^{(0)} = \frac{1}{2EI} \int_0^\ell (M + Pz)^2 \, dz = \frac{1}{2EI} \left( M^2 \ell + \frac{P^2 \ell^3}{3} + MP \ell^2 \right)
\]

where \( E \) is the Young’s modulus, \( \ell \) the length of the finite element, \( P \) and \( M \) are the shear and bending internal forces at the right node of the element and \( I \) is the moment of inertia.

Fracture mechanics studied the calculation of the additional stress energy of a crack providing flexibility coefficients expressed by a stress intensity factor in the linear elastic range, using Castigliano’s theorem.

Concerning a rectangular beam having width \( b \) and thickness \( h \) the additional energy due to the crack can be expressed as

\[
W^{(1)} = b \int_0^a \left( \frac{K_{ij}^2 + K_{ij}^2}{E'} + \frac{(1+\nu)K_{ij}^2}{E} \right) \, da
\]

where \( a \) is the crack depth, \( E' = E \) for plane stress and \( E = E/(1-\nu^2) \) for plane strain, \( \nu \) is the Poisson ratio and \( K_i, K_{ii}, K_{ii} \) are stress intensity factors for opening type, sliding type and tearing cracks, respectively.

Taking into account only bending, Eq.(2) leads to

\[
W^{(1)} = b \int_0^a \left( \frac{K_{ij}^2}{E'} + \frac{K_{ij}^2}{E} \right) \, da
\]

where

\[
K_{ij} = \frac{6M}{bh^2} \sqrt{\pi a F_i(a)}; \quad K_{ij} = \frac{3P \ell}{bh^2} \sqrt{\pi a F_i(a)}; \quad K_{ij} = \frac{P}{bh} \sqrt{\pi a F_i(a)}
\]

are stress intensity factors for opening-type and sliding-type cracks due to \( M \) and \( P \), respectively, and

\[
F_i(a) = \frac{\sqrt{2}}{\sqrt{a}} \tan \left( \frac{\pi s}{2} \right) 0.923 + 0.199 \left[ 1 - \sin \left( \frac{\pi s}{2} \right) \right]^{\frac{1}{4}}
\]

\[
F_{ii}(a) = \left( 3s - 2s^2 \right) \frac{1.122 - 0.561 s + 0.085 s^2 + 0.18 s^3}{\sqrt{1-s}}
\]

with \( s = a/h \) is the ratio between the crack depth and the height of the element.

The generic component \( d_{ij}^{(0)} \) of the compliance (or flexibility) matrix \( D^{(0)}_e \) of the undamaged element can be derived as
whereas the terms \( d_{ij}^{(1)} \) of the additional flexibility matrix \( D_e^{(1)} \) due to the crack can be formulated by:

\[
d_{ij}^{(1)} = \frac{\partial^2 W^{(1)}}{\partial P_i \partial P_j}, \quad i, j = 1, 2; \quad P_1 = P, \quad P_2 = M
\]  

Finally the total flexibility matrix for the element with an open crack is:

\[
D_e = D_e^{(0)} + D_e^{(1)}
\]  

From the equilibrium condition it follows

\[
(P_i M_i P_{i+1} M_{i+1})^T = T(P_{i+1} M_{i+1})^T
\]  

where the apex \( T \) means transpose matrix and

\[
T^T = \begin{bmatrix} -1 & -\ell & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}
\]  

By the principle of virtual work the stiffness matrix of the undamaged element takes the following form:

\[
K_e = T D_e^{(0)} T^T
\]  

while the stiffness matrix of the cracked element may be derived as

\[
K_{e,c} = T D_e^{(1)} T^T
\]  

Once the stiffness matrices of the undamaged and cracked elements are defined, for the beam discretized in \( N_e \) finite elements the stiffness matrix \( K \) of order \( n \times n \) with \( n = 2N_e \) can be straightforwardly evaluated following the classical assembly rules.

Moreover, in the framework of the finite element approximation, it is usually assumed that the crack does not modify the mass distribution.

### 3 Governing Equations for Uncertain-But-Bounded Crack Depth

The equation of motion of a quiescent damaged beam discretized by \( N_e \) finite elements subjected to an external excitation \( f(t) \) with uncertain crack depth modeled as \( \text{interval} \) variable can be written as:

\[
M \ddot{u}(\alpha, t) + C(a_0, \alpha) \dot{u}(\alpha, t) + K(a_0, \alpha) u(\alpha, t) = f(t), \quad \alpha \in \alpha^l = [\alpha, \bar{\alpha}]
\]  

where \( M \) is the \( n \times n \) mass matrix of the structure, \( C(a_0, \alpha) \) is the \( n \times n \) damping matrix, \( K(a_0, \alpha) \) is the \( n \times n \) stiffness matrix and \( f(t) \) is the vector function of order \( n \times 1 \); \( u(\alpha, t) \) is the \( \text{interval} \) vector of nodal displacements of order \( n \times 1 \) and a dot over a variable denotes differentiation with respect to time \( t \).

The \( n \times n \) interval stiffness matrix \( K(a_0, \alpha) \) is here expressed as a function of the uncertain structural parameter \( \alpha \) as follows:
\[
\mathbf{K}(a_0, \alpha) = \mathbf{K}_c(a_0) + \alpha \mathbf{K}_1(a_0); \quad \mathbf{K}_1(a_0) = \left. \frac{\partial \mathbf{K}(a_0, \alpha)}{\partial \alpha} \right|_{\alpha=0} \tag{14}
\]

where \(\alpha\) is the dimensionless fluctuation of the uncertain crack depth \(a = a_0 (1 + \alpha)\) with \(a_0\) its mean value. In Eq.(14) \(\mathbf{K}_c(a_0)\) is the *midpoint stiffness matrix*. It is a positive definite symmetric matrix of order \(n \times n\), while \(\mathbf{K}_1(a_0)\) is a symmetric matrix of order \(n \times n\) and rank \(r\). The Rayleigh model is herein adopted for the interval damping matrix, i.e.:

\[
\mathbf{C}(a_0, \alpha) = c_0 \mathbf{M} + c_1 \mathbf{K}(a_0, \alpha) = c_0 \mathbf{M} + c_1 \mathbf{K}_c(a_0) + c_1 \alpha \mathbf{K}_1(a_0) = \mathbf{C}_c(a_0) + \alpha \mathbf{C}_1(a_0) \tag{15}
\]

where \(c_0\) and \(c_1\) are the Rayleigh damping constants having units \(s^{-1}\) and \(s\), respectively. Hereafter we indicate \(\mathbf{K}_c(a_0) = \mathbf{K}_c\), \(\mathbf{K}_1(a_0) = \mathbf{K}_1\) and \(\mathbf{C}_c(a_0) = \mathbf{C}_c\) for sake of notation compactness.

According to the *Interval Analysis* [8], the interval real number \(\alpha' \in [\alpha, \bar{\alpha}] \subseteq \mathbb{R}\) such that \(\alpha \leq \alpha' \leq \bar{\alpha}\), with \(\alpha\) and \(\bar{\alpha}\) denoting the lower and upper bound of \(\alpha\), is introduced. In this case the midpoint (or mean) and deviation amplitude (or radius) can be written as:

\[
\alpha_0 = \frac{\bar{\alpha} + \alpha}{2}; \quad \Delta \alpha = \frac{\bar{\alpha} - \alpha}{2} > 0 \tag{16}
\]

Introducing the so-called extra symmetric unitary interval (EUI) [9] variable \(\hat{e}_\alpha \in [-1,1]\) the following affine form definition of the interval variable \(\alpha\) is provided:

\[
\alpha = \alpha_0 + \Delta \alpha \hat{e}_\alpha \tag{17}
\]

Moreover the dimensionless fluctuation \(\alpha\) of the uncertain-but-bounded crack depth \(a = a_0 (1 + \alpha)\) around its mean value \(a_0\) can be modeled as symmetric interval in such a way \(\alpha \in [\alpha, \bar{\alpha}]\) with \(\bar{\alpha} = -\alpha\). Under this assumption Eq.(17) reduces to \(\alpha = \Delta \alpha \hat{e}_\alpha\) and the \(n \times n\) interval stiffness matrix \(\mathbf{K}'(a_0, \alpha)\) can be rewritten in the form:

\[
\mathbf{K}'(a_0, \alpha) = \mathbf{K}_c + \Delta \alpha \hat{e}_\alpha \mathbf{K}_1 \tag{18}
\]

where the matrices \(\mathbf{K}_c\) and \(\mathbf{K}_1\) have been previously defined in Eq.(14).

In the framework of the traditional modal analysis, the solution of the equations of motion (13) may be pursued by introducing the following coordinate transformation:

\[
\mathbf{u}(\alpha, t) = \Phi_c \mathbf{q}(\alpha, t), \quad \alpha \in \alpha' = [\alpha, \bar{\alpha}] \tag{19}
\]

where \(\mathbf{q}(\alpha, t)\) is the interval vector gathering the first \(m\) modal coordinates \(q_j(\alpha, t)\) \((j = 1, 2, \ldots, m)\); \(\Phi_c\) is the modal matrix, of order \(n \times m\), pertaining to the midpoint configuration. Specifically, the modal matrix \(\Phi_c\), collecting the first \(m\) eigenvectors normalized with respect to the mass matrix \(\mathbf{M}\), is evaluated as solution of the following eigenproblem:

\[
\mathbf{K}_c \Phi_c = \mathbf{M} \Phi_c \Omega_c^2; \quad \Phi_c^T \mathbf{M} \Phi_c = \mathbf{I}_m; \quad \Phi_c^T \mathbf{K}_c \Phi_c = \Omega_c^2 \tag{20}
\]
\( \mathbf{I}_m \) being the identity matrix of order \( m \) and \( \Omega_c^2 \) the spectral matrix listing the squares of the natural circular frequencies of the structure related to the mean value of the uncertain crack depth and the apex \( T \) means transpose matrix.

By applying the coordinate transformation (19), the equations of motion can be projected in the modal space:

\[
\ddot{\mathbf{q}}(\alpha,t) + \mathbf{\Xi}(\alpha)\dot{\mathbf{q}}(\alpha,t) + \Omega^2(\alpha)\mathbf{q}(\alpha,t) = \mathbf{\Phi}_c^T \mathbf{f}(t), \quad \alpha \in \alpha' = [\alpha, \bar{\alpha}]
\]

where \( \Omega^2(\alpha) = \mathbf{\Phi}_c^T \mathbf{K}(\alpha) \mathbf{\Phi}_c \) and \( \mathbf{\Xi}(\alpha) = \mathbf{\Phi}_c^T \mathbf{C}(\alpha) \mathbf{\Phi}_c \) is the generalized damping matrix. Based on Eq. (18), the matrix \( \Omega^2(\alpha) \) can be split as

\[
\Omega^2(\alpha) = \mathbf{\Phi}_c^T \mathbf{K}(\alpha) \mathbf{\Phi}_c = \Omega^2_c + \Delta \alpha \bar{e}^l_c \Omega^2_c, \quad \alpha \in \alpha' = [\alpha, \bar{\alpha}];
\]

where the matrix \( \Omega^2_c \) is not a diagonal matrix.

Analogously the matrix \( \mathbf{\Xi}(\alpha) \) can be split as

\[
\mathbf{\Xi}(\alpha) = \mathbf{\Phi}_c^T \mathbf{C}(\alpha) \mathbf{\Phi}_c = \mathbf{\Xi}_c + \Delta \alpha \bar{e}^l_c \mathbf{\Xi}_c, \quad \alpha \in \alpha' = [\alpha, \bar{\alpha}];
\]

where only the matrix \( \mathbf{\Xi}_c \) is a diagonal one.

### 4 TIME DOMAIN RESPONSE

#### 4.1 Deterministic excitation

The key idea is to evaluate two response vectors [9] associated respectively to LB of \( \hat{e}^l_c \) and to UB of \( \hat{e}^l_c \).

The first one, denoted by \( \mathbf{q}^- (\alpha, t) \), is evaluated by solving the following differential set of equations:

\[
\ddot{\mathbf{q}}^- (\alpha, t) + [\mathbf{\Xi}_c - \Delta \alpha \mathbf{\Xi}_c] \dot{\mathbf{q}}^- (\alpha, t) + \left[ \Omega^2_c - \Delta \alpha \Omega^2_c \right] \mathbf{q}^- (\alpha, t) = \mathbf{\Phi}_c^T \mathbf{f}(t)
\]

while the second, denoted by \( \mathbf{q}^+ (\alpha, t) \), is evaluated by solving the following differential set of equations:

\[
\ddot{\mathbf{q}}^+ (\alpha, t) + [\mathbf{\Xi}_c + \Delta \alpha \mathbf{\Xi}_c] \dot{\mathbf{q}}^+ (\alpha, t) + \left[ \Omega^2_c + \Delta \alpha \Omega^2_c \right] \mathbf{q}^+ (\alpha, t) = \mathbf{\Phi}_c^T \mathbf{f}(t)
\]

Notice that the two set of Eqs. (24) and (25) cannot decoupled by a real coordinate of transformation, then they are referred in literature as non-classically damped systems. It has been recognized that in order to evaluate the response of non-classically damped systems the state variables have to be introduced [11].

In state variables the previous equations can be rewritten as:

\[
\ddot{\mathbf{y}}^\pm (\alpha, t) = \mathbf{D}^\pm (\alpha) \mathbf{y}^\pm (\alpha, t) + \mathbf{V} \mathbf{f}(t)
\]

being
where $\mathbf{0}$ is the zero matrix.

The solution of these equations can be obtained by applying the numerical procedure proposed by [11] writing:

$$
y^+(\alpha, t_{k+1}) = \Theta^+(\alpha, \Delta t) y^+(\alpha, t_k) + \mathbf{y}_0^+(\alpha, \Delta t) \mathbf{V} f(t_k) + \mathbf{y}_1^+(\alpha, \Delta t) \mathbf{V} f(t_{k+1})$$

(28)

where

$$
\Theta^+(\alpha, \Delta t) = \exp[\Delta t \mathbf{D}^+(\alpha)]; \\
\mathbf{L}^+(\alpha, \Delta t) = \left[ \Theta^+(\alpha, \Delta t) - \mathbf{I}_{2m} \right] \left( \mathbf{D}^+(\alpha) \right)^{-1}; \\
\mathbf{y}_0^+(\alpha, \Delta t) = \left[ \Theta^+(\alpha, \Delta t) - \frac{1}{\Delta t} \mathbf{L}^+(\alpha, \Delta t) \right] \left( \mathbf{D}^+(\alpha) \right)^{-1}; \\
\mathbf{y}_1^+(\alpha, \Delta t) = \left[ \frac{1}{\Delta t} \mathbf{L}^+(\alpha, \Delta t) - \mathbf{I}_{2m} \right] \left( \mathbf{D}^+(\alpha) \right)^{-1}
$$

(29)

Once the vectors $\mathbf{q}^-(\alpha, t)$ and $\mathbf{q}^+(\alpha, t)$ have been evaluated, the corresponding nodal responses can be written as:

$$
\mathbf{u}^-(\alpha, t) = \mathbf{C}_c \mathbf{q}^-(\alpha, t); \quad \mathbf{u}^+(\alpha, t) = \mathbf{C}_c \mathbf{q}^+(\alpha, t)
$$

(30)

In order to evaluate the lower bound (LB), $\mathbf{u}(\alpha, t)$, and upper bound (UB), $\bar{\mathbf{u}}(\alpha, t)$, of the interval dynamic response, $\mathbf{u}(\alpha, t)$, let denote with $u_k^-(t)$ the $k$-th element of the vector $\mathbf{u}(\alpha, t)$, and with $u_k^-(t)$ and $u_k^+(t)$ the $k$-th elements of the vectors $\mathbf{u}^-(\alpha, t)$ and $\mathbf{u}^+(\alpha, t)$, respectively. Then the lower and upper bounds of the $k$-th nodal dynamic response in the time domain can be written as:

$$
\underline{u}_k(t) = \min \left\{ u_k^-(t), u_k^+(t) \right\}; \quad \overline{u}_k(t) = \max \left\{ u_k^-(t), u_k^+(t) \right\}
$$

(31)

where $\min \{ \bullet \}$ and $\max \{ \bullet \}$ means minimum and maximum values of the quantity into curly parentheses, respectively.

Once the LB and UB functions are evaluated, the midpoint function can be evaluated as:

$$
u_{k, \text{mid}}(t) = \frac{\underline{u}_k(t) + \overline{u}_k(t)}{2}
$$

(32)

### 4.2 Stochastic excitation

The statistics of the response of a damaged beam with uncertain-but-bounded crack depth under random excitation turn out to be represented by time-dependent stochastic interval vectors. In the following, a method for determining the upper and lower bounds of second-order statistics will be developed.
Without loss of generality, the forcing term in Eq.(13) is herein assumed to be a mono-correlated zero-mean Gaussian random vector process, appropriately defined as follows:

\[ f(t) = F(t)s \]  \hspace{1cm} (33)

where \( s \) is a \( n \)-order vector listing the spatial distribution of loads and \( F(t) \) is a zero-mean Gaussian non-stationary uniformly modulated random process, fully characterized by the autocorrelation function \( R_{FF}(t_1, t_2) \).

As known, under the assumptions of linear elastic behaviour and uncertain-but-bounded structural parameter, namely the crack depth, the response process of a structural system excited by a zero-mean Gaussian random process is zero-mean Gaussian too. The covariance matrix is defined as:

\[ \Sigma_{YY}(\alpha, t) = E\{Y(\alpha, t)Y^T(\alpha, t)\} = E\left\{\begin{pmatrix} Q(\alpha, t)Q^T(\alpha, t) & Q(\alpha, t)Q^T(\alpha, t) \\
Q(\alpha, t)Q^T(\alpha, t) & Q(\alpha, t)Q^T(\alpha, t) \end{pmatrix}\right\} \]  \hspace{1cm} (34)

\( E\{\cdot\} \) being the mathematical expectation operator.

Alternatively, the stochastic response can be defined in a more suitable form by introducing the covariance vector \( \sigma_Y(\alpha, t) \):

\[ \sigma_Y(\alpha, t) = E\{Y(\alpha, t) \otimes Y(\alpha, t)\} = \text{Vec}\{E\{Y(\alpha, t)Y^T(\alpha, t)\}\} = \text{Vec}\{\Sigma_{YY}(\alpha, t)\} \]  \hspace{1cm} (35)

where the symbol \( \otimes \) denotes the block Kronecker product [10]. The vector \( \sigma_Y(\alpha, t) \) represents the block vectorialized form (“Vec\{\cdot\}”) of the covariance matrix, that is a column vector defined in such a way that the columns of the sub-matrices forming the matrix in parentheses are written one below each other.

After some algebra, the differential equations governing the time-evolution of the vector \( \sigma_Y(\alpha, t) \) can be written as:

\[ \dot{\sigma}_Y(\alpha, t) = D_{\alpha}^\top(\alpha)\sigma_Y(\alpha, t) + F_Y(\alpha, t) \]  \hspace{1cm} (36)

where

\[ \sigma_Y(\alpha, t) = \begin{bmatrix} \sigma_{QQ}(\alpha, t) \\ \sigma_{\alpha Q}(\alpha, t) \\ \sigma_{Q \alpha}(\alpha, t) \\ \sigma_{\alpha \alpha}(\alpha, t) \end{bmatrix} ; \\
D_{\alpha}^\top(\alpha) = D^\top(\alpha) \otimes I_{2m} + I_{2m} \otimes D^\top(\alpha) ; \\
F_Y(\alpha, t) = V_2 E\{Y^\top(\alpha, t)F(t)\} \]

with

\[ V_2 = \left[ (Vs \otimes I_{2m}) + (I_{2m} \otimes Vs) \right] \]  \hspace{1cm} (38)
The matrices $\mathbf{D}^{-}(\alpha)$ and $\mathbf{D}^{+}(\alpha)$ required to define $\mathbf{D}_{2}^{\pm}(\alpha)$ have been previously defined in Eq.(27). Taking into account the formal analogy between Eq.(36) and the equation of motion in the modal state variable space (26), the response can be evaluate numerically as:

$$\mathbf{\sigma}_{\pm}(\alpha, t_{k+1}) = \mathbf{\Theta}_{2}(\alpha, \Delta t) \mathbf{\sigma}_{\pm}(\alpha, t_{k}) + \mathbf{\gamma}_{0,2}(\alpha, \Delta t) \mathbf{F}_{2}^{\pm}(\alpha, t_{k}) + \mathbf{\gamma}_{1,2}(\alpha, \Delta t) \mathbf{F}_{2}^{\pm}(\alpha, t_{k+1})$$

(39)

where

$$\mathbf{\Theta}_{2}(\alpha, \Delta t) = \mathbf{\Theta}^{T}(\alpha, \Delta t) \odot \mathbf{\Theta}(\alpha, \Delta t)$$

$$\mathbf{L}_{2}(\alpha, \Delta t) = \left[ \mathbf{\Theta}_{2}(\alpha, \Delta t) - \mathbf{I}_{(2m)^2} \right] \left( \mathbf{D}_{2}^{\pm}(\alpha) \right)^{-1}$$

$$\mathbf{\gamma}_{0,2}(\alpha, \Delta t) = \left[ \mathbf{\Theta}_{2}(\alpha, \Delta t) - \frac{1}{\Delta t} \mathbf{L}_{2}(\alpha, \Delta t) \right] \left( \mathbf{D}_{2}^{\pm}(\alpha) \right)^{-1}$$

$$\mathbf{\gamma}_{1,2}(\alpha, \Delta t) = \frac{1}{\Delta t} \left[ \mathbf{L}_{2}(\alpha, \Delta t) - \mathbf{I}_{(2m)^2} \right] \left( \mathbf{D}_{2}^{\pm}(\alpha) \right)^{-1}$$

$$\mathbf{F}_{2}^{\pm}(\alpha, t) = \mathbf{V}_{2} \mathbf{E} \left( \mathbf{Y}^{\mp}(\alpha, t) \mathbf{F}(t) \right) = \mathbf{V}_{2} \int_{0}^{t} \mathbf{\Theta}^{\mp}(\alpha, t - \tau) \mathbf{V} \mathbf{s} \mathbf{R}_{FF}(t, \tau) d \tau$$

A procedure proposed to evaluate the inverse of the matrices $\mathbf{D}_{2}^{\pm}(\alpha)$ in Eq.(40), namely $(\mathbf{D}_{2}^{\pm}(\alpha))^{-1}$, moves from the evaluation of $\mathbf{\Psi}_{j}^{\pm}$ and $\lambda_{j}^{\pm}$ being the $j$-th eigenvector and the associated eigenvalue, solutions of the following eigenproblems:

$$\mathbf{D}^{\mp} \mathbf{\Psi}^{\mp} = \lambda^{\mp} \mathbf{\Psi}^{\mp}; \quad \mathbf{\Psi}^{T} \mathbf{A}^{\mp} \mathbf{\Psi}^{\mp} = \mathbf{I}_{j}$$

(41)

solved respectively for the matrices $\mathbf{D}^{-}(\alpha)$ and $\mathbf{D}^{+}(\alpha)$ with

$$\mathbf{\Psi}^{\mp} = \begin{bmatrix} \psi_{1}^{\pm} & \psi_{2}^{\pm} & \cdots & \psi_{r}^{\pm} \end{bmatrix}; \quad \mathbf{A}^{\mp} = \text{Diag} \left[ \lambda_{1}^{\mp}, \lambda_{2}^{\mp}, \cdots, \lambda_{r}^{\mp} \right]$$

(42)

and

$$\mathbf{A}^{\mp}(\alpha) = \begin{bmatrix} \mathbf{\Xi}^{\mp} + \Delta \alpha \mathbf{\Xi} \end{bmatrix}$$

(43)

The evaluation of the complex diagonal matrices $\mathbf{A}_{2}^{\mp}$ defined as

$$\mathbf{A}_{2}^{\mp} = \mathbf{A}^{\mp} \odot \mathbf{I}_{2m} + \mathbf{I}_{2m} \odot \mathbf{A}^{\mp}$$

(44)

allows to calculate the matrices $(\mathbf{D}_{2}^{\mp}(\alpha))^{-1}$ as follows:

$$(\mathbf{D}_{2}^{\mp}(\alpha))^{-1} = (\mathbf{\Psi}^{T} \odot \mathbf{\Psi}^{\mp}) \left( \mathbf{A}_{2}^{\mp} \right)^{-1} \left( \mathbf{\Psi}^{T} \mathbf{A}^{\pm}(\alpha) \odot \mathbf{\Psi}^{T} \mathbf{A}^{\mp}(\alpha) \right)$$

(45)

Once the modal covariances vectors $\mathbf{\sigma}_{\pm}(\alpha, t)$ and $\mathbf{\sigma}_{\mp}(\alpha, t)$ have been evaluated, the nodal corresponding quantities can be evaluated as follows:
The lower bound (LB) and upper bound (UB) of the \( k \)-th nodal interval variance of the dynamic response can be evaluated as a function of the \((k-1)m+k\)-th element of vectors \( \mathbf{\sigma}_{\mathbf{u}u}(\alpha,t) \) and \( \mathbf{\sigma}_{\mathbf{u}u}^+(\alpha,t) \), that is:

\[
\sigma_{\mathbf{u}u}^2(t) = \min \left\{ \mathbf{\sigma}_{\mathbf{u}u}(\alpha,t), \mathbf{\sigma}_{\mathbf{u}u}^+(\alpha,t) \right\}_{(k-1)m+k} \\
\bar{\sigma}_{\mathbf{u}u}^2(t) = \max \left\{ \mathbf{\sigma}_{\mathbf{u}u}(\alpha,t), \mathbf{\sigma}_{\mathbf{u}u}^+(\alpha,t) \right\}_{(k-1)m+k}
\]

where \( \min \{ \bullet \}_{(k-1)m+k} \) and \( \max \{ \bullet \}_{(k-1)m+k} \) means minimum and maximum values of the \((k-1)m+k\)-th element of the vectors into curly parentheses, respectively. Once the LB and UB functions are evaluated the midpoint vector can be evaluated as:

\[
\mathbf{\sigma}_{\mathbf{u}u,mid}(\alpha,t) = \frac{\mathbf{\sigma}_{\mathbf{u}u}^+(\alpha,t) + \mathbf{\sigma}_{\mathbf{u}u}^2(\alpha,t)}{2}
\]

In order to highlight the potential of the proposed formulation, in the following of this section the non-stationary random excitation is modelled as a uniformly modulated white noise process. As known, such a model is successfully adopted in many engineering applications since it allows substantial simplifications in the mathematical formulation.

In this case, the stochastic process \( F(t) \) in Eq.(33) is expressed multiplying a stationary Gaussian white noise process, \( W(t) \), by an amplitude modulating deterministic function \( \phi(t) \), i.e. : \( F(t) = \phi(t)W(t) \). Accordingly, the autocorrelation function of \( F(t) \) reads:

\[
R_{FF}(t_1,t_2) = 2\pi S_W \phi(t_1)\phi(t_2)\delta(t_2-t_1)
\]

where \( S_W \) is the Power Spectral Density (PSD) of \( W(t) \) and \( \delta(t) \) denotes the Dirac delta function. By substituting Eq.(50) into the last of Eqs.(40) the following forcing vector is obtained:

\[
\mathbf{F}_F(\alpha,t) \equiv \mathbf{F}_F(\alpha,t) = 2\pi S_W \phi^2(t)(\mathbf{V}_s) \odot (\mathbf{V}_s)
\]

It follows that, when the input process is modelled as a uniformly modulated white noise, the proposed method involves much simpler mathematical derivations.
5 NUMERICAL APPLICATION

In order to demonstrate the efficiency of the procedure presented in the previous section, a cantilever damaged steel beam studied in [1] subjected to an action \( f(t) \) applied to the free-end is examined. Its geometry is reported in Fig.1.

![Figure 1: Geometry of the cantilever steel beam](image)

The beam presents a rectangular cross-section with width \( b=1\,\text{mm} \) and height \( h=7.8\,\text{mm} \) and length \( L=200\,\text{mm} \). The Young’s modulus and the material mass density are assumed \( E=207000\,\text{N/mm}^2 \) and \( \rho=7860\,\text{Kg/m}^3 \), respectively. The cantilever beam has been modelled by \( N_e = 5 \) finite elements (see Fig.1), assuming positive upward vertical displacements and counter-clock wise rotations. Two different locations for the crack have been considered. The crack is supposed to be located in one case in the middle of the second element and in a second case study in the middle of the fourth element. Therefore the damaged element stiffness matrix is evaluated as \( \mathbf{K}_{c,2} = \mathbf{T} \mathbf{D}_2 \mathbf{T}^T \) or \( \mathbf{K}_{c,4} = \mathbf{T} \mathbf{D}_4 \mathbf{T}^T \) (see Eq.12) respectively. The remaining four elements are modelled as undamaged beam finite elements.

The mean value \( a_0 \) of the uncertain-but-bounded crack depth \( a \) defined as \( a = a_0 \left( 1 + \Delta \alpha \hat{e}_a \right) \) is assumed to be \( a_0 = 0.4h \).

The different change into stiffness related to the location of the crack provides different values for the Rayleigh damping constants in Eq.(15). For the case of the crack located in the second element, the constants takes the values \( c_0=83.927\,\text{s}^{-1} \) and \( c_1=0.0000138\,\text{s} \), while for the case of damaged fourth element the constant assume the values \( c_0=86.939\,\text{s}^{-1} \) and \( c_1=0.0000139\,\text{s} \), respectively. In both cases they are evaluated in such a way that the modal damping ratio for the first and second modes of the structure referred to the midpoint stiffness matrix \( \mathbf{K}_c \) is \( \zeta = 0.05 \).

The analysis has been conducted for both cases of deterministic and stochastic action \( f(t) \). Two different types of deterministic action have been considered: the first one represented by the unit step function \( f(t) = U(t) \) while the second expressed by an harmonic force \( f(t) = \sin(200t) \).

Via the proposed procedure resumed in section 4.1 and applying Eqs.(31), the lower and upper bounds of the nodal displacement \( u(t) \) of the free end of the cantilever beam, namely the 9-th nodal dynamic response, have been straightforwardly evaluated. The zero mean deviation has been assumed \( \Delta \alpha = 0.3 \).

Figs.2a and 2b show the time varying lower \( \underline{u}_9(t) \) and upper \( \overline{u}_9(t) \) bounds of the displacement (in mm) of the cantilever tip subjected to \( U(t) \) considering the two crack locations, re-
spectively in the elements 2 and 4 (as specified in the apex in brackets). In Fig.3 similar results are reported for the damaged beam under harmonic action.

Both bounds results coincide with the exact bounds in presence of crack with uncertain-but-bounded depth.

Comparing the response bounds trend it is worth to note that for both cases of acting load there is not correspondence among the peaks of the response which result shifted in time.

The stochastic excitation acting at the free end of the beam has been defined as in Eq.(33) with \( F(t) = \varphi(t)W(t) \). The stationary White Noise process has spectral density function \( S_W = 6 \times 10^{-4} \text{N}^2\text{s} \) and the modulating deterministic function \( \varphi(t) \) has been expressed by \( \varphi(t) = t \left( \exp(1)/\varepsilon \right) \exp(-t/\varepsilon) \) with \( \varepsilon = 0.3 \).

Taking advantage of the mathematical simplifications, Eqs.(48) allow to evaluate the bounds of the nodal interval variance of the dynamic response in terms of displacement under uniformly modulated white noise.

In Figs.4a and 4b, as for the previous cases of deterministic actions, the bounds of the time varying displacement variances of the cantilever tip, \( \sigma^2 \) and \( \sigma^2 \) respectively, are reported and compared for the two crack locations considered (middle of second and fourth
element). Also for this application the deviation amplitude is fixed $\Delta \alpha = 0.3$. Same results are obtained performing the Monte Carlo simulation.

Figs. 4 show that for stochastic action the bounds peaks of the variance of the response in terms of free-end displacement are both lower when the crack is located at the middle of the fourth element respect to the case of crack located in the second element.

Moreover it is worth to note that the gap is larger between the upper bounds of the variances $\bar{\sigma}_{U(t)}^2(2)$ and $\bar{\sigma}_{U(t)}^2(4)$, respectively (see Fig. 4b).

Lastly in Figs. 5 results in terms of bounds of the variances of the free-end displacement are reported for different deviation amplitude values, respectively $\Delta \alpha = 0.3$ and $\Delta \alpha = 0.1$.

Results shown in Fig. 5a are referred to the case of crack located in the middle of the second element while those in Fig. 5b are referred to the crack shifted in the fourth element.

Comparisons between the bounds of the variances of the free-end displacement show that larger intervals occur when the crack is located in the second element (see Fig. 5a) while the intervals are narrow when the crack is located at the fourth element also for the larger deviation amplitude $\Delta \alpha = 0.3$. It follows that the dynamic response of the damaged cantilever beam is more sensitive to the presence of the crack when the latter is located closer to the joint end, as expected.

Figure 4: (a) Lower and (b) upper bound of the variance of the free-end displacement (in mm$^2$) of the damaged cantilever beam with crack located at the middle of second element (continuous line) or fourth element (dashed line) under non-stationary random excitation.

Figure 5: Lower and upper bounds of the variance of the free-end displacement (in mm$^2$) of the damaged cantilever beam with crack located at the middle of (a) second element or (b) fourth element under non-stationary random excitation for deviation amplitude $\Delta \alpha = 0.3$ (continuous line) and $\Delta \alpha = 0.1$ (dot-dashed line).
6 CONCLUSIONS

In this paper lower and upper bounds of the response of damaged beams with uncertain-but-bounded crack subjected to both deterministic and stochastic excitations modeled as zero mean Gaussian random processes are evaluated.

The proposed method allows to split both the time dependent deterministic and random response, in modal subspace, as sum of two aliquots: the midpoint or nominal solution and the deviation. For stochastic excitations, the differential equations governing the time-evolution of the midpoint and deviation covariance vectors of the response under zero-mean Gaussian stochastic input process are derived taking full advantage of the Kronecker algebra.

Main steps of the provided procedure summarize as follows: i) to define a finite element model of the beam in which the model of fully open crack is used to represent the damaged element; ii) to model the crack depth as an interval variable; iii) to evaluate in time domain the response for deterministic and stochastic excitation, by adopting an unified approach based on the Improved Interval Analysis via EUI.

Effectiveness and efficiency of the present method are shown in the numerical results, coinciding with the exact ones, concerning a cracked beam with uncertain-but-bounded stiffness properties under deterministic and uniformly modulated white noise excitation.

REFERENCES

MODELLING OF MATERIAL PARAMETER UNCERTAINTY OF RESONANCE WOOD USED IN VIOLINS USING THE INTERVAL FIELD

Maurice Imholz\textsuperscript{1}, Dirk Vandepitte\textsuperscript{1} and David Moens\textsuperscript{1}

\textsuperscript{1}KU Leuven, Department of Mechanical Engineering
Celestijnenlaan 300, B-3001, Leuven

e-mail: maurice.imholz@kuleuven.be - dirk.vandepitte@kuleuven.be - david.moens@kuleuven.be

Keywords: uncertainty analysis, interval methods, interval fields

Abstract. The interval field method can be used to model spatial uncertainty in finite element models. When analysing spatial uncertainty, dependency is an important quantity to consider, as it is physically implausible to assume a quantity is independent in locations that are close to each other. The interval field method was introduced to join this concept of dependency with classical interval analysis, omitting the need to define probability distributions. This paper will apply the theory of interval field (IF) to the dynamical properties of a violin top plate, with uncertain stiffness and density.
1 INTRODUCTION

The construction of musical instruments is an interesting field for numerical modelling. Not only because numerical modelling is still almost non-existent in this field, but also because due to the sensitivity of the human eye, almost all aspects than can be modelled accurately from the raw material up to the response of the final product are relevant. On the other hand, many effort has already been taken in defining numerical models for all kinds of instruments, and there is still much to do, because accurately modelling sound generated by instruments is notoriously difficult.

Some instruments can be mass produced, of which guitars are the best example, but in most instruments, a lot of handwork is still needed to obtain the desired end result. The raw material used in the construction is of vital importance to the sound, and much effort goes into selecting the best material. However, this too is mostly done using only the naked eye to assess material quality. This leaves a large amount of uncertainty on the material’s structural properties, which in turn affects the structural output of the material, and ultimately the sound generated.

This paper considers the violin as instrument of interest, and attempts to quantify the uncertainty on relevant natural frequencies, originating from the uncertainty on the wood stiffness and density. Both parameters are assumed to vary over the domain of a violin top plate, leading to a field-based approach of the uncertain parameters. The reader should consider this work as part of a full sound model of a violin, in which much research has been conducted already. Important for subsequent analysis is that the uncertainty of the frequencies is captured as accurately as possible. If not, possible conservatism in the results could multiply, leading to useless results.

2 VIOLIN BODY CONSTRUCTION

The violin body accounts for the primary source of sound. The wooden corpus consists of two panels, the violins back and top panel. The panels are connected by the ribs of the instrument, enclosing a body of air. The f-shaped holes in the top plate allow air to flow in and out of the body, producing the core sound of the violin.

The top plate is made from quarter cut spruce (picea abies), sawn as a single piece, and is about 3 mm thick at the thickest point. The quarter-cut sawing technique produces planks with the growth rings almost perpendicular to the plank plane. This produces strong and light planks, but is more expensive to make as relatively few quarter-cut planks can be made from a single tree. The wood grains lie along the violins longitudinal axis.

The back plate is generally made from figured maple, and can be both a single piece or a two-pieced part. Two-pieced plates are made by sawing a single wedge from a tree into two slender wedges and gluing them together at the wide side. Because they originate from the same wedge, the two halves are visually almost identical, producing a symmetrical back plate. Single-pieced back plates can be both quarter-cut or slab-cut. Slab cut wood originates from sawing a tree all the way through. Because the alignment of the material axis vary over the plank, slab-cut backs are more sensitive and tend to bend more easily during the violins life.

The ribs are made from 1 mm thick maple wood, bended to follow the violins characteristic shape.

Within the violin body, a slender bar of spruce called the bass bar is attached to the top plate adding stiffness in the longitudinal direction. The bass bar is always attached with a slight offset from the middle, removing the symmetrical behaviour of the top plate.

Lastly, a small piece of spruce, the sound post, is put between the f-holes that connects the
Maurice Imholz, Dirk Vandepitte and David Moens

Figure 1: Tuning modes of the violin top plate

top and back plate, usually not glued in place, but clamped in between the plates, accounting for extra support of the top plate. Also placed eccentrically with respect to the top plates line of symmetry, the symmetry of the violin is further distorted.

For the purpose of this paper, only the top plate will be considered. Research has proven that for the characteristic modes of the violin body, the top plate is usually exhibiting the largest deformation.

3 MODAL ANALYSIS

The model is used in a modal analysis to characterise the structural behaviour of the violin top plate. The chain between violin body characteristics and sound production is notoriously hard to model accurately and has been subject of much research already, including work on bridge and snare mobility ([1, 2]), analysis of damping factors ([4, 5]) and modal characterisation of complete string instruments ([7, 6]). Earlier work on characterising the mechanics of the violin body already showed that not some structural modes are more effective at radiating sound than others ([3]). The effectiveness of a mode can be characterized in two ways:

- **Modal displacement at the bridge**: the string vibration is passed to the body through the bridge, placed in between the f-holes. If the mode shape exhibits a nodal line at the bridge, it is difficult to directly excite through exciting the strings, reducing sound radiation capability.

- **Degree of asymmetry of the mode shape**: with every mode shape a volume difference of the captured body of air is associated. Due to their symmetry, some heavily vibrating modes only lead to local motion of the air within the cavity, without actual air motion through the f-holes. Typically, highly symmetrical modes show this behaviour. More asymmetrical modes with more relative volume difference are therefore more effective at radiating sound.

Based on this fact, only a small set of modes is truly interesting to look at. During the construction process, most luthiers perform a plate tuning of both the top and back plate, specifically targeting one mode and tuning it to a desired value ([8]). Most luthiers use a manual process for this, holding the plate between thumb and index finger at a node and then gently tapping the plate at an antinode, producing an audible frequency associated to a single mode. By locally carving away some wood, the right frequency can be obtained. The modes mostly used to tune the top plate are the first, second and fifth structural mode, images of which are included in figure [1].

The first mode corresponds to the first torsional mode of a flat plate, and is situated at about 60 Hz. The second mode corresponds to the first bending mode. This mode is highly influenced by the anisotropic character of the wood, which can be seen from the curved nodal lines running...
across the plate. This mode’s frequency is about 100 Hz. The fifth mode is a bending mode with
synchronic bending in x- and y-direction, creating a trampoline-like movement. It is situated
around 220 Hz. This mode is closely related to the A0-mode of the full violin body. This
important mode of the full violin exhibits opposite movement of top and back plate, leading to
a large volume difference of the enclosed air, leading to large air movement through the f-holes.
This mode is often referred to as the 'breathing'-mode. This mode is given in figure 2.

Figure 2: breathing mode of the violin body.

The model of the violin plate measures 210 mm at its widest and 356 mm at its longest point
and consists of 15000 shell elements. The thickness is taken as 2.5 mm throughout.

With respect to possible automatisation of violin construction, the manual process of plate
tuning is obviously the hardest to do. This analysis therefore assumes a constant thickness
but allows a certain variation of the structural properties of the wood, as these are hard to
determine for a violin maker. The output will be the frequency variation of the three tuning
modes described above. However, not only the variation of the modes individuially is important
to observe, but also the shape of the combined uncertainty region in the \((f_1, f_2, f_5)\)-domain,
as the sound produced by the violin results of an interaction of these and probably even more
modes. It is therefore very important to, as a start, accurately estimate an uncertain region of
these mode frequencies, which can then be used in subsequent analysis of a full violin body
analysis.

4 WOOD VARIABILITY

The primary use of wood in engineering purposes is in construction, for which only the
structural properties are of interest. For use in musical instruments, the acoustic properties are
the main source of interest. The term resonance wood is the term for all wood that is used for
acoustical purposes, i.e. in musical instruments. The characterisation of resonance wood in this
chapter mainly focusses on its orthotropic nature, and the different sources of variability. These
sources will then be translated to a physically inspired representation of material uncertainty in the wood.

Wood is considered a nominally orthotropic material with cylindrical orthotropy. It is characterized by a density $\rho$, and three Moduli of Elasticity $E_L$, $E_R$ and $E_T$, representing the stiffness in the longitudinal, radial and tangential direction. Assuming quarter cut plates, the tangential direction is located perpendicular to the plate and is therefore of no interest in this analysis. Derived quantities are the speed of sound in longitudinal and radial direction, given by eq. 1 and 2. Specifically for resonance wood, $V_{LL}$ is quite high with respect to construction wood, indicating a large desired value for $E_L$ combined with a low density ([9]). On the radial stiffness, less conclusive results are known, but a relatively low value is generally preferred.

\begin{align*}
V_{LL} &= \sqrt{\frac{E_L}{\rho}} \\
V_{RR} &= \sqrt{\frac{E_R}{\rho}}
\end{align*}

If we look closely to the microstructure of wood, we observe tubular cells with a very large length-to-width ratio. In the case of spruce this is even larger than 100 [10]. The cell walls make up for almost all of the cells stiffness and are essentially a fibre-reinforced matrix. These fibres form a helix-shape around the cell with a varying angle called the microfibril angle (MFA). If the angle approaches zero, the fibres align with the cells main axis, maximizing the longitudinal stiffness and minimizing the radial stiffness, whereas if the angle increases, the stiffness of the matrix becomes increasingly dominant, reducing the longitudinal stiffness but increasing the radial stiffness. As a result, resonance wood is characterized by a small average MFA of 10 or less. Moreover, research has showed that a small difference between the average MFA of latewood compared to earlywood is beneficial for the acoustic properties, as this leads to a more homogeneous sound velocity ([11, 12]). A certain degree of continuity of this quantity over the violin plate is therefore desired.

5 WOOD SELECTION

Mostly, the selection process is based on visual techniques combined with experience to identify the optimal wood to be used in their instruments ([10, 13, 14]). Their interest specifically goes out towards the annual ring pattern and wood colour. Narrow rings with good consistency over the length of the grains is desired, also because of aesthetic reasons. Also, a bright colour indicates lighter wood which is preferred as well. Overall, one can say that the violin makers selection process favors a low density and an impurity-free wood specimen over acoustical properties, which are much more difficult to assess by using only the senses. It has been shown that $E_R$ is proportional to the percentage of latewood present in the plate ([14]). Latewood is defined as the wood the tree grows in summer/autumn. Improving the trees strength, the density is more than twice the density of so called earlywood. Resonance wood usually has a latewood percentage of under 20 %, suggesting the preference of a low $E_R$.

6 SPATIAL VARIABILITY

Many literature exists with data on the variability of structural parameters of spruce ([18, 17, 16]), but most of these are defined on macro scale and derived from measuring entire wood
planks. Focus on meso-level (i.e. variability between grains or along a single grain) is less abundant. We know wood consists of slender-shaped cells with very stiff cell walls, composed of a matrix-reinforced composite. The MFA is key in this definition as this determines the longitudinal and transverse stiffness of the cells and therefore of the whole material. Good correlation has been observed between the continuity of this parameter over the material and the acoustical quality of the resonance wood ([15]). Moreover, a large degree of anisotropy has the same positive effect on the quality, suggesting a small overall value for the MFA.

Research indicates that the stiffness is heavily dependent of the MFA, showing a linear decrease of longitudinal stiffness and an increase of radial stiffness with increasing MFA. Based on this reasoning, we assume the following dependency of the stiffnesses with respect to density and MFA, in each point on the top plate (eq. 3):

\[
E_L = E_{L,0} \cdot (1 + C_\rho (\rho - \rho_0)) \cdot (1 - C_\theta (MFA - MFA_0)) \\
E_R = E_{R,0} \cdot (1 + C_\rho (\rho - \rho_0)) \cdot (1 + C_\theta (MFA - MFA_0))
\] (3)

We will assume the density and the MFA are not correlated, an assumption validated by literature ([19]), in which strong correlation between these quantities could not be proven. Note that for positive \( C_\theta \), \( E_L \) decreases as MFA increases, whereas the radial stiffness \( E_R \) increases. For the purpose of this numerical exercise, we assume the density and the microfibrillar angle have similar effect on the final value of the stiffness (literature shows values between 1/3 and 2/3). Based on the ranges of \( \rho \) and MFA, the values of the constants for spruce in equation 3 are given below in equation 4:

\[
E_L = 12 \cdot 10^9 \frac{N}{m^2} \cdot \left( 1 + 2.78 \cdot 10^6 \frac{mm^3}{kg} (\rho - \rho_0) \right) \cdot (1 - 0.0417(MFA - MFA_0)) \\
E_R = 0.9 \cdot 10^9 \frac{N}{m^2} \cdot \left( 1 + 2.78 \cdot 10^6 \frac{mm^3}{kg} (\rho - \rho_0) \right) \cdot (1 - 0.0417(MFA - MFA_0))
\] (4)

with MFA expressed in degrees. The density is the quantity on which the selection process of luthiers is mostly focused. As previously stated, light wood with a low contribution of latewood is preferred. The percentage of latewood can be relatively accurately estimated by the trained eye, so we assume no more than 20 % intra-variability on the density in all directions. However, the density of the wood is expected to be more constant in the longitudinal direction than in the radial direction, as changes in weather conditions manifest themselves in the radial direction. We will therefore account for a higher dependency in the longitudinal direction than in the radial direction. The MFA is obviously much more difficult to notice with the blind eye. Fact is that Norway spruce (\textit{Picea abies}) has a very small average MFA with respect to other wood species, to which it largely owes its preferred use in violin building. The MFA is known to decrease with the age of the growth ring, and also with tree height [11]. The decrease rate is however subject to uncertainty. To express the MFA in each point, the following equation is used (equation 5):

\[
MFA I(x, y) = (\theta_2 - \theta_1) \cdot \left( \frac{x}{210 mm} \right)^{p_l} + \Delta \theta I \cdot \left( \frac{y}{356 mm} \right)^{p_l} + \theta_1 
\] (5)

In this equation, \( \theta_1 \) and \( \theta_2 \) are the MFA at the lower corners of the plank from which the plate is made, \( \Delta \theta \) is the change in MPA across the longitudinal axis of the plate. To all of
these parameters independent intervals are assigned. The exponents $p_1$ and $p_2$ are intervals that model the uncertainty on the decrease rate of the MFA across the plate. The interval bounds for the spruce wood used in the top plate are given in table [6]. The absolute bounds of MFA across the wood are $[2, 14]$.

For the modelling of the uncertainty on the density, we use an interval representation introduced by the authors in [20], which includes the spatial dependency of the density by setting bounds on the maximal gradient of the density. The general definition is given in eq. [6]

$$\rho^I(x, y) = \rho_0 \cdot \left(1 + \sum_{i=1}^{n} \alpha_i^I \cdot \phi_i(x, y)\right)$$ (6)

In this definition, $\phi_i(s, y)$ are radial basis functions. Each function is associated with a separate interval parameter, due to the overlap of the basis functions the interval parameters can be assumed independent while still obeying a constraint on the maximum gradient of the field. Their orientation is illustrated in figure [3]. Figure [4] shows such a basis function.
Maurice Imholz, Dirk Vandepitte and David Moens

<table>
<thead>
<tr>
<th>parameter in MFA model</th>
<th>lower bound</th>
<th>upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>$\Delta \theta$</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>$p_1$</td>
<td>-0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>$p_2$</td>
<td>-0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Table 1: interval bounds in the MFA model

![a single shape function for R=0.2 m](image)

Figure 4: Example of a basis function.

For the $z$-value, a piecewise continuous second order polynomial function is used:

$$z(r, \theta) = a \cdot \begin{cases} 
2 \left( 1 - \frac{r}{R_{\max}(\theta)} \right)^2 & 1 > \frac{r}{R_{\max}(\theta)} > 0.5 \\
1 - 2 \left( \frac{r}{R_{\max}(\theta)} \right)^2 & \frac{r}{R_{\max}(\theta)} \leq 0.5 \\
0 & \frac{r}{R_{\max}(\theta)} > 1 
\end{cases} \quad (7)$$

with $r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$ the distance from the central point and $R_{\max}(\theta)$ the maximum radius of the basis function in a certain direction. For the top plate, a definition of 56 basis functions is used. The base functions used have a different radius in $x$- and $y$-direction, as can be seen from figure 3. The variation of the density is known to vary more in the radial direction as in the longitudinal direction, simply due to the fact that changing weather conditions manifest themselves across the tree’s ageing rings in the radial direction. For this, we will define separate radii $R_L$ and $R_r$ for the longitudinal and radial directions, leading to oval shaped basis functions, the bases of which are illustrated in figure 3. By choosing the scaling factor $a$, the function radii $R_L$ and $R_r$ and the distances between the center points $\Delta x$ and $\Delta y$, we can set the maximum gradient $\left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)$ as well as the maximum deviation $U$ in each point. The relations are given below in equation 8. The $x$-axis is set along the radial direction, the $y$-axis is set along the longitudinal direction.
Maurice Imholz, Dirk Vandepitte and David Moens

Figure 5: some examples of density distributions over the violin top plate

Figure 6: some examples of longitudinal stiffness distributions over the violin top plate

\[
U = \frac{7\pi}{24} \cdot \frac{R_r R_L}{\Delta x \Delta y}
\]

\[
\frac{\partial f}{\partial y} = 2aR_r \frac{\partial f}{\Delta x \Delta y} \Delta x
\]

\[
\frac{\partial f}{\partial x} = 2aR_L \frac{\partial f}{\Delta x \Delta y} \Delta y
\]

(8)

the choice of \(\Delta x\) and \(\Delta y\) depend on the choice of \(R_L\) and \(R_r\), as equation [8] only holds for sufficiently large values of \(\frac{R_L}{\Delta y}\) and \(\frac{R_r}{\Delta x}\). A 20\% deviation on \(\rho\) corresponds to \(U = 0.2\), a maximum (relative) gradient of \((0.00182, 0.00365) mm^{-1}\) corresponds to \(R_L = 240 mm, R_r = 120 mm\) and \(a = 0.038\). \(\Delta x\) and \(\Delta y\) are set at 50 mm and 100 mm respectively.

The bases of the basis functions are illustrated in figure [8]. Note that the central point of some functions are located outside of the domain of the plate, but are still included as a term in equation [6] because a part of their nonzero domain is located on the plate. Figures [5], [6] and [7] show some random realisations of the top plate by randomly sampling the intervals in equations [5] and [6] showing the dependency present in both the stiffness and the density.

7 PROPAGATION

In order to accurately propagate the effect of the density, we will use the spatial information of the density interval parameters. Although we assume the interval parameters can vary independently, their effects on the violin body natural frequencies are probably highly similar. To account for this, we make use of so called coefficient fields. We know from the basic me-
mechanics of plates that the square of the eigenfrequency is proportional to the specific stiffness $\frac{E}{\rho}$ (equation 9).

$$\omega_i^2 = a_L \cdot \frac{E_L}{\rho} + a_R \cdot \frac{E_R}{\rho}$$

In the case of varying properties, we can again create a similar expression for the density in each field point:

$$d(\omega_i^2) = a_L(x,y) \cdot \frac{E_L(x,y)}{\rho(x,y)} + a_R(x,y) \cdot \frac{E_R(x,y)}{\rho(x,y)}$$

in this equation $d(\omega_i^2)$ represents the change in eigenfrequency due to a varying density in a single. Simply integrating this equation over the entire plate then leads to the following expression:

$$\omega_i = \sqrt{\int_{x,y} d(\omega_i^2)} = \sqrt{\int_{x,y} a_L(x,y) \cdot \frac{E_L(x,y)}{\rho(x,y)} dx dy + \int_{x,y} a_R(x,y) \cdot \frac{E_R(x,y)}{\rho(x,y)} dx dy}$$

In this expression, the coefficients of the polynomial model are given a spatial character. Previous research [21] indicated that for low order polynomials applied to plate based models, these coefficient fields exhibit a high degree of continuity, because of the correlated behaviour of the eigenfrequencies with respect to changing density in different points on the plate. This allows us to in turn model these coefficient fields by low order polynomials:

$$a(x,y) = q_{1122}x^2y^2 + q_{11}x^2 + q_{12}xy + q_{123}x + q_{22}y^2 + q_{22}y + q_0$$

This allows us to drastically reduce the true dimension of the problem by immediately focussing on the coefficients of equation [12] rather than identifying a relationship for each field point separately. For this problem, the 15000 discrete coefficients that would have to be determined for the specific stiffness variation are then replaced by only 9 coefficients that define the continuous coefficient field. Once the model of equation [11] is known, the relationship of

Figure 7: some examples of radial stiffness distributions over the violin top plate
the parameters governing the MFA can then be inferred from equation 3. This leads to two coefficient fields with 9 coefficients each, so a total of 18 parameters have to be determined.

A very important remark to make here is that as the order of the polynomial in eq. 9 increases, the applicability of this approach decreases, as the continuity of the coefficient fields rapidly decreases for higher order terms. More on this can be found in [21, 22].

8 RESULTS

For the training of the metamodel with the coefficient fields, 600 random experiments are used. After the creation of the metamodel, 50000 samples are generated to estimate the uncertainty on the frequencies. The coefficient fields are shown in figures 8 and 9 for both the specific longitudinal stiffness and the specific radial stiffness. Both are in the same order of magnitude, indicating that both the longitudinal and radial stiﬀnesses are of important influence on the natural violin plate frequencies. The high degree of symmetry corresponds to our expectations as the associated mode shapes are symmetrical as well. On figures 10, 11 and 12 the uncertain domain is shown as the 2D projections of the \( (f_1, f_2, f_5) \)-domain. The samples of the FE-solver are marked in red (600 in total), the 50000 samples from the coefficient field model are marked in blue and serve as an estimate of the uncertain region. We see that 20 % density variation combined with the uncertain MFA lead to a maximum frequency variation of about 10 %. We also see positive correlation of the first and second eigenfrequency, which from the mode shapes seems feasible. But also the coefficient fields of figure 8 support this, as the spatial effect of the longitudinal stiffness on both the first and second eigenfrequency is similar. However, striking is the negative correlating between mode shape 2 and 5, which we believe is due to the way the longitudinal and radial stiffnesses are related. Because of the model used in equation 3 both have opposite relations with respect to the MFA, and the frequency of the 5th mode appears to decrease drastically with increasing MFA, much more than mode 2.

Figure 8: coefficient fields for the three output frequencies for the longitudinal stiffness
Figure 9: coefficient fields for the three output frequencies for the radial stiffness

Figure 10: uncertainty region projected on the \((f_1, f_2)\)-plane
9 CONCLUSIONS

In this paper, the interval field method has been applied on a large model of a violin top plate to assess the natural frequency variation due to uncertain structural properties of the wood. The
uncertainty was estimated based on the wood selection process in violin construction. Literature data was used to feed an interval field model that accounts for dependency of the density parameter, which was different in radial and longitudinal direction. The stiffnesses were related to the MFA through a simple linear relationship.

Future steps will be taken to feed this model with more accurate estimations of the density and the MFA through acquiring experimental data with the emphasis on spatial variability. Also, the relation of MFA and stiffness shall be further investigated as the linear relationship is likely to be too simple. However, the theory of interval fields provides an effective tool to include whatever information there may be on dependency and spatial uncertainty.

10 ACKNOWLEDGEMENTS

The research fund of the KU Leuven as well as the fund for scientific research (F.W.O.) are gracefully acknowledged for their support.

REFERENCES


REAL-TIME FUZZY ANALYSIS OF MACHINE DRIVEN TUNNELING

Ba Trung Cao¹, Steffen Freitag¹, and Günther Meschke¹

¹Institute for Structural Mechanics, Ruhr University Bochum
Universitätstrasse 150, 44801, Bochum, Germany
e-mail: ba.cao@rub.de; steffen.freitag@sd.rub.de and guenther.meschke@rub.de

Keywords: Surrogate Model, Mechanized Tunneling, Uncertainty Quantification, Fuzzy Analysis.

Abstract. Reliability assessment in mechanized tunneling requires to take into account limited information describing the local geology and the corresponding geotechnical parameters. The geotechnical data are often quite limited and generally not available in the form of precise models and parameter values. In this case, epistemic uncertainty should be considered within the reliability assessment. The concept of fuzzy numbers is applied to predict tunneling induced settlements in real-time. An advanced numerical simulation is utilized as forward model for the settlement prediction. However, to achieve real-time capabilities, surrogate models are required. Deterministic surrogate models can be used together with an $\alpha$-cut optimization approach to compute fuzzy data. In this paper, a surrogate modeling strategy is introduced to directly process fuzzy input-output data. Within this approach, the time-consuming optimization procedure is replaced by a surrogate model to obtain the fuzzy settlement field prediction in real-time. The significant reduction in computation time maintaining similar prediction performance leads to potential applications in steering of mechanized tunneling processes.
1 INTRODUCTION

In tunnel projects, the information of the local geology and the associated soil conditions are often limited and imprecise. Therefore, a reliability assessment procedure is required, which also considers epistemic source of uncertainty, i.e. by means of intervals, fuzzy numbers, and imprecise probability approaches. General concepts for reliability analyses in mechanized tunneling are presented e.g. in [1].

To perform these analyses a forward model is required, which is able to capture the system response and the various interactions between individual components during the tunneling process. Currently, simulations based on the Finite Element (FE) method are used to investigate and predict the soil-structure interactions in mechanized tunneling. In this paper, a three dimensional FE model for simulations of shield-driven tunnels in soft soil, see [2], is applied. Taking into account fuzzy numbers for the geotechnical parameters, multiple runs of the simulation model are required within the $\alpha$-cut optimization. The FE simulation must be replaced by surrogate models to reduce the computation time significantly, especially for real-time applications while maintaining the prediction performance of the original FE simulation model.

In [3], a hybrid surrogate model based on a combination of Recurrent Neural Networks (RNN) and Proper Orthogonal Decomposition (POD) is proposed for deterministic input-output mapping with high-dimensional outputs. The hybrid surrogate model has been utilized together with Particle Swarm Optimization to perform optimization based interval analyses for computing the time variant interval settlement field with more than 100 surface points. To obtained results with an optimization based interval analysis takes several hours, which is not practical for real-time applications in mechanized tunneling. To enable real-time capabilities, an extension of the hybrid surrogate model is presented in [4] to directly map interval input-output data, which are expressed by a midpoint-radius representation. The key idea of the strategy is to also replace the complete time-consuming optimization loop by surrogate models and not only the deterministic FE simulation model. The results of an application example show a significant reduction in computation time with similar accuracy as the optimization based approach. In this paper, this approach is further extended for time variant fuzzy settlement field predictions in real-time. Fuzzy numbers are expressed by $\alpha$-cuts and a $\Delta$-representation of the bounds, which requires use the Non-Negative Matrix Factorization (NNMF) technique instead of the POD within some sub-surrogate models. The performance of the proposed strategy is illustrated through a fuzzy analyses and also within a fuzzy probability box analyses of a mechanized tunneling process.

2 UNCERTAINTY QUANTIFICATION

2.1 Intervals and fuzzy numbers

The epistemic source of uncertainty can be quantified by intervals or fuzzy numbers. An interval

$$\bar{x} = [l_x, u_x]$$

(1)

is defined by its lower bound $l_x$ and upper bound $u_x$ without further assumptions of a distribution within this possible range. By defining a membership function $\mu(x)$ of the uncertain set, the concept of intervals is extended to fuzzy numbers. If the membership function is discretised by $\alpha$-cuts, a set of nested intervals is obtained, see Fig. [1]. In general, computations with fuzzy numbers can be performed similar to interval analyses, i.e. by means of fuzzy arithmetic [5] or optimization based approaches [6], see also [7] for an overview.
A typical fuzzy number with D \( \alpha \)-cuts, as shown in Fig. 1, can be represented by a sorted sequence containing all lower and upper bounds

\[
\tilde{x} = \langle \underbrace{l_1 x, \ldots, l_D x, 12 \Delta x, \ldots, (D - 1)D \Delta x, D \Delta x, D(D - 1) \Delta x, \ldots, 21 \Delta x}_{l_D x} \rangle,
\]

(2)

The fuzzy number can also be expressed by defining a reference point, e.g. \( l_D x \), and incremental differences \( \Delta \) to all other bounds, see e.g. [3].

\[
\tilde{x} = \langle l_D x, 12 \Delta x, \ldots, (D - 1)D \Delta x, D \Delta x, D(D - 1) \Delta x, \ldots, 21 \Delta x \rangle.
\]

(3)

Whereas the reference point \( l_D x \) is defined in \( \mathbb{R} \), all \( \Delta \) values in the above equation must be positive numbers. In this paper, the fuzzy settlement field, i.e. the vector \( \tilde{S} \) containing all settlements of a surface area, will be described by the \( \Delta \)-representation as

\[
\tilde{S} = \langle l_D S, 12 \Delta S, \ldots, (D - 1)D \Delta S, D \Delta S, D(D - 1) \Delta S, \ldots, 21 \Delta S \rangle.
\]

(4)

The type of input-output mapping used in this paper

\[
P(t) \mapsto \tilde{S}(t)
\]

(5)

means that the time variant deterministic steering parameters \( P(t) \) are defined as inputs and mapped with fuzzy model parameters onto the time variant fuzzy settlement field \( \tilde{S}(t) \). The fuzzy model parameters of the mapping model contain the influence of the time constant geotechnical fuzzy parameters.

### 2.2 Probability box approach

In case of limited statistical information, e.g. estimation of stochastic models with small sample sizes, imprecise probability concepts can be adopted to quantify the variability of uncertain parameters. The probability box (p-box) approach, see e.g. [9], can be used to define imprecise stochastic numbers by its lower bound \( \underline{F}(x) \) and upper bound \( \overline{F}(x) \) cumulative distribution.
function (cdf). In general, arbitrary stochastic models, including empirical distributions, can be used for the lower and upper bound cdf.

Here, a fuzzy stochastic approach for real-time predictions of time variant settlements in mechanized tunneling is developed. This is realized by running a stochastic analysis (i.e. Monte Carlo simulations) with fuzzy samples. At each \( \alpha \)-cut, the lower bound \( ^\text{l}F(x) \) and upper bound \( ^\text{u}F(x) \) cdfs of the structural responses are computed. As can be seen in the left part of Fig. 2 for each Monte Carlo run, a deterministic tunneling simulation model has to be computed taking fuzzy soil parameters into account, i.e. for each settlement component, two optimization problems have to be solved at each \( \alpha \)-cut. Due to the high computational effort, a deterministic surrogate model is created to replace the FE simulation model, see middle part of Fig. 2. In order to achieve real time performance, also the optimization loops are replaced by a fuzzy surrogate model, see right part of Fig. 2. This allows to significantly reduce the computation time within the p-box approach.

![Figure 2: Reliability analyses with polymorphic uncertain data (p-box approach).](image)

### 3 HYBRID SURROGATE MODEL STRATEGY

The new surrogate modeling scheme for real-time predictions with fuzzy numbers in mechanized tunneling is depicted in Fig. 3. In the offline stage, a surrogate model of a tunnel section is generated using FE simulation results based on deterministic inputs (realizations \( X \) of the fuzzy geotechnical parameters \( \tilde{X} \) and time variant steering parameters \( P(t) \)) and deterministic outputs (time variant surface settlement field \( S(t) \)). By varying the input parameters of the numerical model, a set of deterministic input-output data is collected. Afterwards, the time variant fuzzy settlement fields corresponding to a fixed fuzzy number of \( \tilde{X} \) and different scenarios of \( P(t) \) are computed within fuzzy analyses using an optimization approach, e.g. particle swarm optimization. For real-time application, the obtained results are used to create a hybrid RNN-GPOD surrogate model for the mapping according to Eq. (5) based on the \( \Delta \)-representation, see Eq. (4).

In the online stage, the surrogate model is operated to predict the fuzzy bounds of the expected surface settlement field of the next time step \( n + 1 \). Trained RNNs are employed to predict the fuzzy bounds of settlements at several monitoring points for the next time step \( n + 1 \). The complete time variant fuzzy surface settlement field from time step 1 to \( n \) is approximated by trained POD-Radial Basis Functions (POD-RBF) for \( \Delta S \) and NNMF-RBF surrogate models for the non-negative \( \Delta S \) values, respectively. Finally, the GPOD and the NNMF approaches are adopted to reproduce and predict the complete fuzzy settlement field based on a combination of the results from the two previous methods. The predicted results are then included into the available fuzzy data set and the procedure is repeated for the subsequent time steps.
3.1 Recurrent neural networks for fuzzy data

Fuzzy data can directly be processed with RNNs by fuzzy arithmetic operations, see e.g. [10]. Here, the $\Delta$-representation for fuzzy data is used. For the prediction of the time variant fuzzy settlements at selected monitoring points, two types of RNNs for predicting $l_D S$ and the $\Delta S$ values are generated. For the training procedure, the deterministic steering parameters $[n] P$ are used as inputs in all RNNs but the target data types are different, i.e. real numbers for $l_D S$ and positive real numbers for all $\Delta S$ values. This is realized by different activation functions in the neurons. For the RNN to predict the lower bound of $\alpha$-cut $D (l_D S)$, the hyperbolic tangent function and the linear activation function are used in the hidden and output neurons, respectively. For the $\Delta S$ RNNs, the logistic sigmoid function or the positive linear function is used in the output neurons to satisfy the constraint of non-negative outputs. The Levenberg-Marquardt back-propagation algorithm is employed to train the RNNs.

3.2 POD-RBF and NNMF-RBF networks for fuzzy data

The used POD approach, see [11] for an overview, and how to combine the method with RBFs to form a surrogate model are described in detail in [12]. The POD-RBF approach is used to predict $l_D S$ from time step 1 to $n$. To ensure the non-negativity constraint of the prediction of the $\Delta S$ values, the POD approach is replaced by the NNMF technique.

Basically, a high-dimensional matrix $S$ and a single column of $S$ can be approximated as a linear combination of the truncated basis vectors $\hat{\Phi}$ as $S \approx \hat{\Phi} \cdot \hat{A}$ and $S_i \approx \hat{\Phi} \cdot \hat{A}_i$. The truncated basis vectors are obtained from solving an eigenvalue problem of the covariance matrix $C = S^T \cdot S$. At this step, the truncated coefficient matrix $\hat{A}$ contains constant values associated with the given matrix $S$. Hence, it is only an approximation of the snapshots generated in the original high-dimensional snapshots matrix $S$.

To obtain a rather continuous approximation, each coefficient vector is expressed as a non-linear function of input parameters on which the system depends. The coefficient matrix $\hat{A}$ can
be related to the interpolation functions by an unknown matrix of constant coefficients \( B \) as \( \hat{A} = B \cdot F \). In this equation, \( F \) contains a set of predefined interpolation functions \( f_j(z) \) of input parameters \( z \). The choice of \( f_j(z) \) can be arbitrary and in this study an inverse multiquadric radial function, a type of RBF (see [13] for a description), is chosen as interpolation function. Finally, an approximation of the output system response corresponding to an arbitrary set of input parameters is obtained by \( \Delta S^\text{a} \approx \hat{F} \cdot B \cdot F^\ast \).

The NNMF, see [14], is utilized to ensure the positive sign of the predicted results for the \( \Delta S^\text{a} \) values. Given a non-negative matrix \( \Delta S \), the NNMF algorithm is searching for two non-negative matrices \( W \) and \( A^+ \) satisfying the following optimisation problem \( \min \| \Delta S - W \cdot A^+ \|_F^2 \) subject to \( W, A^+ \geq 0 \). Similar to the POD approach, \( W \) and \( A^+ \) are denoted as the basis matrix and coefficient matrix, respectively. The alternating non-negative least squares algorithm proposed in [15], which ensures the convergence of the minimization problem, is implemented in this paper. The prediction for non-negative outputs \( \Delta S^\text{a} \) is computed similar to the POD-RBF method as \( \Delta S^\text{a} \approx W \cdot B^+ \cdot F^\ast \). The coefficient matrix \( B^+ \) is the non-negative solution of the following minimization problem \( \min \| A^+ - B^+ \cdot F \|_F \).

### 3.3 GPOD and Gappy NNMF for fuzzy data

The basic POD method is combined with a linear regression called GPOD, see [16], to reconstruct the complete \( \Delta S \) values of the fuzzy settlement field. More details about the explanation and implementation of the method is given in [12] and [4]. The GPOD approach is applied to reconstruct missing elements of a given vector \( S^\ast \). It employs the concept of a gappy norm based on available data since the full norm cannot be evaluated correctly due to missing elements. The intermediate repaired vector \( S^\ast_{\text{int}} \) can be expressed in terms of truncated POD basis vectors \( \hat{F} \) as follows \( S^\ast_{\text{int}} \approx \hat{F} \cdot \hat{A}^\ast \). The coefficient vector \( \hat{A}^\ast \) can be computed by minimizing the error \( E = \| S^\ast - S^\ast_{\text{int}} \|_n^2 \). A solution to this least squares problem is given by a linear system of equations \( M \cdot \hat{A}^\ast = R \) with \( M = (\hat{F}^T, \hat{F}) \) and \( R = (\hat{F}^T, S^\ast) \).

The reconstruction procedure for a non-negative vector \( \Delta S^\ast \) follows the steps of the GPOD method with some minor modifications. The corresponding objective function \( E = \| \Delta S^\ast - W \cdot \Delta A^\ast \|_n^2 \) to be minimized contains the distances between the available incomplete data vector and the predicted vector. The non-negative basis matrix \( W \) is assumed to be known from the available non-negative data matrix \( \Delta S \). The coefficient vector \( \Delta A^\ast \) is obtained considering the non-negativity constraint by solving the non-negative least squares problem \( \min \| M^+ \cdot \Delta A^\ast - R^+ \| \) with \( M^+ = (W^T, W) \) and \( R^+ = (W^T, \Delta S^\ast) \).

Finally, by replacing the missing elements in \( S^\ast \) and \( \Delta S^\ast \) by those in the corresponding reconstructed vectors, the complete fuzzy settlement field of time step \( n + 1 \) is predicted.

### 4 APPLICATION IN MECHANIZED TUNNELING

The proposed surrogate modeling strategy for fuzzy data is applied to predict the time variant fuzzy settlement field of a mechanized tunneling process. The results, which are computed based on the presented \( \Delta \)-representation, are compared to the reference solution (optimization approach) in terms of prediction performance and computation time.

Figure 4(a) shows the simulation model with dimensions of 144m, 220m and 67m (in x, y, z directions). The length of each excavation step is 2m, i.e the tunnel section consists of 72 steps. The tunnel diameter and the cover depth are 10.97m and 6.5m, respectively. In this study, the elastic modulus \( E_1 \) of soil layer 1 is considered as an uncertain geotechnical parameter defined as a trapezoidal fuzzy number \( \tilde{E}_1 = \langle 52, 60, 70, 75 \rangle \) MPa. The grouting pressure \( \text{n}GP \) and the
support pressure \([n]SP\) in each excavation step \(n\) are chosen as deterministic steering parameter, which can vary in each step in the corresponding ranges of 120 to 220 kPa and 100 to 200 kPa. It is assumed that the current state of the TBM advance corresponds to the \(36^{th}\) step of the process. The fuzzy analysis performed in this paper is used to support the selection of \([n]GP\) and \([n]SP\) in the next time steps with the purpose to reduce tunneling induced settlements when the machine advances under an existing railway track.

An effective surface area of 110m in y-direction is investigated considering the z-displacements of 154 surface points, see Fig. 4(b), as the outputs. In the example, the fuzzy settlements of 18 selected monitoring points among these 154 surface points are predicted by the RNNs. Figure 5 shows the fuzzy settlement process of a selected surface point computed by \(\alpha\)-cut optimization and by the proposed surrogate model for fuzzy data. The relative error of the proposed method is 3.8% in average compared to the optimization based reference solution. The most important benefit of the proposed approach is that the computation time (3 seconds) is significantly reduced compared to the optimization approach, which often required many hours.

Additionally to the fuzzy analysis, a reliability analysis adopting the p-box approach is performed with a Monte Carlo simulation using 1000 samples and the fixed \(\tilde{E}_1\). The pressures \([n]GP\) and \([n]SP\) are treated as stochastic processes assuming a Gaussian distribution with the mean values of 170 kN and 150 kN for \([n]GP\) and \([n]SP\), respectively. The standard deviations in the distribution for both pressures are considered the same at the value of \(\sigma = 30\) kN. The minimum and maximum cdfs of a chosen point settlement in time step 37 corresponding to two nested intervals of \(\tilde{E}_1\) obtained from classical optimization approach and the proposed surrogate model are depicted in Fig. 6. The probability boxes obtained from the surrogate model are reasonable compared to the optimization approach. Nevertheless, the \(\Delta\)-representation leads to an accumulative error for the bounds of the lower \(\alpha\)-cuts. However, the computation time drops dramatically from around 1 day to just 20 minutes for an analysis with two \(\alpha\)-cuts.
5 CONCLUSIONS

In this paper, the hybrid surrogate modeling for interval data has been extended to process fuzzy data by means of a $\Delta$-representation. The time-dependent behaviour of several selected points in future steps is predicted by Recurrent Neural Networks (RNNs), whereas order reduction techniques (Proper Orthogonal Decomposition and Non-Negative Matrix Factorization) are utilised to approximate the complete surface field based on the RNN predictions. The proposed method is used to predict the fuzzy surface settlements during the construction phase in mechanized tunneling for selected scenarios of the TBM steering parameters taking fuzzy geotechnical parameters into account. In comparison to the $\alpha$-cut optimization approach, the computational time is significantly reduced by the proposed strategy. The new approach takes only 2 to 3 seconds to compute the fuzzy bounds of a settlement field with 154 settlement components with a similar accuracy compared to the optimization approach which requires many hours. Future developments of the proposed approach include the implementation of additional steering objectives, e.g. the tunnel face stability, controlling the tunnel lining forces and reducing the tunneling induced damage of existing infrastructure and buildings.

6 ACKNOWLEDGEMENT

Financial support was provided by the German Research Foundation (DFG) in the framework of project C1 of the Collaborative Research Center SFB 837 "Interaction Modeling in Mechanized Tunneling". This support is gratefully acknowledged.

REFERENCES

Figure 6: P-boxes for the settlement of the surface point 44 for time step 37 considering $\tilde{\mathbf{E}} = \langle 52, 60, 70, 75 \rangle$.


BAYESIAN UPDATING FOR PROBABILISTIC CLASSIFICATION USING RELIABILITY METHODS

P.G. Byrnes and F.A. DiazDelaO

1Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK
e-mail: \{paul.byrnes,f.a.diazdelao\}@liverpool.ac.uk

Keywords: Bayesian Updating, Reliability Analysis, Subset Simulation, Machine Learning, Gaussian Process Classification

Abstract. Probabilistic classification requires the computation of the posterior probability distribution of a class given a data observation. In order to generate posterior samples, an analogy has recently been established between the Bayesian updating problem and the engineering reliability problem which allows reliability methods to be applied to the former. The modification of the BUS (Bayesian Updating with Structural Reliability Methods) formulation is based on the conventional rejection principle and suggests the application of Subset Simulation (SuS) from reliability engineering to sample from posterior distributions. Under the original BUS framework a likelihood multiplier is required to be calculated before the implementation of SuS. A recently proposed algorithm learns the likelihood multiplier automatically. This research proposes the utilization of BUS for Gaussian Process classification. The above framework is illustrated using a benchmark Machine Learning dataset from an engineering application.
1 INTRODUCTION

Classification is a branch of supervised Machine Learning which allocates test instances to a class based on a training dataset. Let $C_k$ be the $k^{th}$ class from a family $\{C_1, ..., C_k\}$ and $D$ be a data set. Define the posterior probability distribution of interest as $P(C_k|D)$. Various techniques exist for sampling from and approximating a posterior distribution. Popular methods include Markov Chain Monte Carlo (MCMC) [1], Laplace Approximation [2], Expectation Propagation (EP) [3] and Rejection sampling [4]. Rejection sampling consists of drawing a random sample from the prior distribution, computing the likelihood and accepting the sample proportional to the likelihood. However, the disadvantage of this method is that the acceptance rate can be extremely low. This opens the possibility as viewing posterior sampling as rare event simulation.

An analogy has recently been established between the Bayesian updating problem and the engineering reliability problem. The formulation, called BUS (Bayesian Updating with Structural reliability methods) [5] is based on the conventional rejection principle and allows reliability methods to sample from posterior distributions. Through the realisation that the probability of acceptance in rejection sampling is equivalent to the probability of failure in a reliability problem, it is concluded that reliability methods may be applied to Bayesian updating problems. Subset Simulation (SuS) [6] is an advanced Monte Carlo technique used in reliability engineering which estimates the probability of a rare event. SuS expresses the (rare) failure event $F$ as contained in a nested sequence of more frequent events. This enables the algorithm to calculate the probability of failure. SuS has thus been shown to be a robust technique which is suitable for Bayesian computations.

One problem which stems from the original BUS framework is that SuS is dependent on the choice of a constant referred to as the likelihood multiplier. Some suggestions regarding the calculation of the multiplier have been given [5]. However, correctly choosing the value of this multiplier has in general remained an open question. A revised BUS formulation [7] allows SuS to be implemented to sample from the posterior distribution whilst learning the multiplier automatically. The objective of this paper is to apply the modified BUS framework to sample from a posterior distribution in a binary classification setting. The classification framework implemented depends on a Gaussian Process (GP) model. The data set used is the widely applied 'Ionosphere' data set [8].

The organization of this paper is as follows. SuS along with the BUS methodologies are introduced in section 2. Section 3 presents the GP classifier along with experimental results. Section 4 contains conclusions and comments on future work.

2 METHODOLOGY

Given that the BUS methodology identifies a relationship between the Bayesian updating problem and the engineering reliability problem, the following section will first provide a brief overview of reliability analysis before introducing SuS and the BUS framework.

2.1 Reliability Analysis

The behaviour of a system may be represented by a response variable $Y$ which is dependent on input variables $x = (x_1, ..., x_d)$ such that

$$Y = g(x_1, ..., x_d)$$ (1)
where $d$ represents the dimension of the problem and $g(x)$ the performance function. In this setting, the failure event $F$ occurs when the output of $Y$ exceeds a critical threshold $b$. This may be expressed as

$$F = \{x : g(x) > b\}$$  \hspace{1cm} (2)

Let $\pi(x)$ denote the joint PDF for $x$. The engineering reliability problem is to compute the probability of failure $P(F)$, given by

$$P(F) = P(x \in F) = \int_{F} \pi(x) dx$$  \hspace{1cm} (3)

### 2.2 Subset Simulation

SuS is a widely used simulation technique in reliability analysis for simulating rare events and estimating failure probabilities. By expressing the rare event $F$ as an intersection of nested events, $F = F_m \subset F_{m-1} \subset ... \subset F_1$, where $F_m$ is rare event whilst $F_1$ may be viewed as a rather frequent event, the algorithm estimates the value of $P(F)$. Given this decomposition, it is easily shown that

$$P(F) = P(F_m|F_{m-1}) * P(F_{m-1}|F_{m-2}) * ... * P(F_2|F1) * P(F_1)$$  \hspace{1cm} (4)

By an appropriate selection, the probabilities $P(F_1)$ and $P(F_{i+1}|F_i)$ can be estimated by direct Monte Carlo. Hence, the original rare event problem is broken down into a series of intermediate subproblems which define a series of intermediate thresholds. SuS seeks to estimate the complementary cumulative distribution function (CCDF) of a response quantity, that is $P(Y > b)$. The CCDF is viewed simply as the tail of the distribution of an ‘exceedance’ area. This CCDF can be used directly for estimating the failure probability that the response exceeds a specified threshold $b$. Let $p_0 \in [0, 1]$ be the level probability which in essence governs how many intermediate failure thresholds are required to reach the failure domain $F$. In the reliability literature, a sensible choice is $p_0 \in [0.1, 0.3]$ [9]. Let $N \in \mathbb{N}$ be the total number of generated samples. While $n_s = p_0N \in \mathbb{N}$ governs the required number of accepted samples at each level. SuS requires the parameters $p_0$ and $N$ be determined before beginning the simulation.

Beginning at level $0$, the algorithm probes the input space generating $N$ independent and identically distributed (i.i.d) samples by direct Monte Carlo methods. Based on the values of $Y$ computed by the performance function, the first intermediate failure threshold $b_1$ is calculated. The $n_s$ samples which exceed $b_1$ from level $0$ are thus stored as seeds for generating additional samples conditional on $F_1 = \{Y > b_1\}$ at level $1$. For level $1$ a MCMC algorithm is utilized to populate $F_1$. Similar to level $0$, the samples which exceed $b_2$ are used as seeds for the next level. The generation of intermediate levels is continued in the same manner until $F$ is populated with the pre-defined number of samples where the probability of failure is approximated by

$$P(F) \approx p_l^{n_F} \frac{n_F}{N}$$  \hspace{1cm} (5)

where $l$ represents the termination level and $n_F$ the number of failure samples at that level. For more details on the implementation of SuS, refer to [6].
2.3 BUS

The BUS formulation builds a relationship between the Bayesian updating problem and the engineering reliability problem, thereby allowing reliability methods (in this case SuS) to be applied to the former. Consider a continuous random variable $X$. Let $L(x)$ denote the likelihood function, $q(x)$ denote the prior PDF, $P_D$ denote the normalizing constant $\int L(x)q(x)dx^{-1}$ and $P(x)$ denote the posterior PDF. Thus,

$$P(x) = P^{-1}_D L(x)q(x) \tag{6}$$

The Rejection sampling scheme generates a sample from $P(x)$ as follows:

1. Generate $U$ uniformly distributed on $[0,1]$ and $x$ with the prior PDF $q(x)$.
2. If $U < cL(x)$, return $x$ as the sample. Otherwise repeat step 1.

where $c$ is a constant such that the rejection principle inequality ($cL(x) \leq 1$) holds. In the context of the Bayesian updating problem, let the driving variable of the engineering reliability problem be defined as follows:

$$Y = cL(x) - u \tag{7}$$

where $u$ is a standard uniform random variable in $[0,1]$. The corresponding failure event

$$F = \{ Y > 0 \} \tag{8}$$

can be sampled from using SuS. It is a well-known fact that in the calculation of $c$ by the rejection principle, the largest admissible value of $c$ is given by

$$c_{\text{max}} = \frac{1}{\max_x L(x)} \tag{9}$$

In the case of the value being greater than $c_{\text{max}}$, the posterior generated samples contain bias. A smaller value will still produce the correct samples but will be less efficient [7].

2.4 Modified Bus Framework

A modification of the original BUS formulation has been proposed [7] which isolates the effect of the multiplier on the simulation. The modification involves a reexpression of the failure event such that

$$F = \left\{ \ln \left[ \frac{L(x)}{u} \right] > -\ln c \right\} \tag{10}$$

The above formulation results in the driving variable $Y$ in SuS being defined as

$$Y = \ln \left[ \frac{L(x)}{u} \right] \tag{11}$$

Adjustments to how $F$ and $Y$ are defined allow the target failure event to be expressed as $F = \{ Y > b \}$, where $b = -\ln c$. It is evident that as $Y$ is no longer dependent on $c$ it is not necessary to choose the value of $c$ before SuS runs. The multiplier only affects the target threshold level $b$ beyond which the samples can be collected as posterior samples. Under the new framework the distribution of the samples conditional on $\{ Y > b \}$ will remain unchanged for a sufficiently large $b$. The minimum value of $b$ beyond which the distribution of the samples will settle at the posterior PDF can be shown to be

$$b_{\text{min}} = -\ln c_{\text{max}} = \ln[\max_x L(x)] \tag{12}$$
2.4.1 Identification of Minimum Threshold Level

Similar to $c_{\text{max}}$ the value of $b_{\text{min}}$ is unknown. The behaviour of $P(Y > b)$ as $b$ varies can be examined during a SuS run to determine whether the threshold value of a particular level has passed $b_{\text{min}}$. Regarding the CCDF, when $b$ is at the left tail of distribution then $P(Y > b) \approx 1$. The value of $P(Y > b)$ typically decreases with $b$ equal to $P_D$ at $b = b_{\text{min}}$. When $b > b_{\text{min}}$ it can be shown that $P(Y > b) = e^{-b}P_D$ [7] where $P_D = \int q(x)L(x)dx$.

It can be expected that as $b$ increases from the left tail and to a value greater than $b_{\text{min}}$, the CCDF of $Y$ typically changes from a decreasing function to an exponentially decaying function. Correspondingly, the function $\ln P(Y > b)$ changes from a slowly decreasing function to a straight line with a slope of -1. Additionally, consider the following:

$$V(b) = b + \ln P(Y > b)$$  \hspace{1cm} (13)

This function can be used for computing the log-evidence in $\ln P_D$ as it can be observed that

$$V(b) = \ln P_D \quad b > b_{\text{min}}$$  \hspace{1cm} (14)

When $b$ is at the left tail of the CCDF, $\ln P(Y > b) \approx 0$ and so $V(b) \approx b$ increases linearly with $b$. In other words as $b$ increases from the left tail of the CCDF of $Y$ the function $V(b)$ increases linearly, going through a transition until it settles at $\ln P_D$ after $b > b_{\text{min}}$. Figure 1 contains a graphical representation of how $\ln P(Y > b)$ and $V(b)$ change when SuS has surpassed $b_{\text{min}}$. Both above quantities $b$ can only be estimated on a sample basis.

![Figure 1: Theoretical characteristic trends of $\ln P(Y > b)$ and $V(b)$](image)
2.4.2 Automatic Stopping Condition

On the basis of the above characteristic trends an automatic stopping condition can be implemented once the algorithm detects that the transition has occurred. Consider an inadmissible level \( m \) such that \( b_m < b_{\min} \). There exists a constant \( e^{-b_m} \) and a monotone decreasing sequence \( a_m \)

\[
\lim_{m \to \infty} a_m = 0
\]

where \( a_m \) is the prior probability of the inadmissible set \( B_m = \{ x : e^{-b_m}L(x) > 1 \} \). Through the expression of the marginal distribution of the target variable as

\[
p(x|F_m) \propto \begin{cases} q(x), & \text{if } x \in B_m \\ e^{-b_m}q(x)L(x), & \text{if } x \in B^c_m \end{cases}
\]

it is observed that for \( x \in B_m \) the marginal is proportional to the prior distribution. Given that the prior distribution is a probability measure it satisfies the monotonicity property. Therefore it follows that \( a_m \) is a monotone decreasing sequence of value which will converge to zero. Through the expression of the prior probability as

\[
a_m = P_x(B_m) = P_x(L(x) > e^{b_m})
\]

the stopping condition takes the form of a reliability problem that in turn may also be solved by SuS.

3 GAUSSIAN PROCESS BINARY CLASSIFICATION

Due to the non-parametric nature of Gaussian processes, their use for Machine Learning applications has grown greatly in recent years. By focussing on processes which are Gaussian, it turns out that the computations required for inference and Machine Learning become relatively easy \[10\] in comparison to other methods e.g Neural Networks. A GP model is defined by its mean and covariance functions respectively such that

\[
GP \sim N(m(x), k(x, x'))
\]

Bayesian inference in a GP classification model is performed about the latent function \( f \) having observed data \( D = \{ (y_i, x_i) \mid i, ..., n \} \). Let \( f = [f_1, ..., f_m]^T \) represent the values of the latent function, \( d = [d_1, ..., d_m]^T \) the model inputs and \( y = [y_1, ..., y_m]^T \) the class labels where \( y \in \{-1, 1\} \). Given the latent function, the class labels are independent Bernoulli random variables. It can be shown \[11\] that the likelihood can be factorized as

\[
p(y|f) = \prod_{i=1}^{m} p(y_i|f_i) = \prod_{i=1}^{m} \Phi(y_i, f_i)
\]

where \( \Phi \) represents the CDF of the standard Gaussian distribution. The latent function is given a GP prior which implies that any finite subset of latent variables has a multivariate Gaussian
distribution [10]. Often in a binary setting since neither of the class labels is more probable the mean of the prior over \( f \) is set to zero. A covariance function of the form \( k(x, x'|\theta) \), where \( \theta \) represents the functions hyperparameters is combined with the zero mean function to define the GP. Through the application of Bayes rule the posterior distribution over the latent function \( f \) for given hyperparameters \( \theta \) is expressed as

\[
p(f|D, \theta) = \frac{p(y|f)p(f|X, \theta)}{p(D|\theta)} = \mathcal{N}(f|0, K) \prod_{i=1}^{m} \Phi(y_i f_i)
\]  

(20)

In order to predict \( y^* \) from \( d^* \) the distribution of the latent function may be computed by marginalising as follows

\[
p(f^*|D, \theta, d) = \int p(f^*|f, X, \theta, d)p(f|D, \theta)df
\]

(21)

which in turn allows the predictive distribution to be obtained by taking the expectation of the marginal

\[
p(y^*|D, \theta, d^*) = \int p(y^*|f^*)p(f^*|D, \theta, d^*)df^*
\]

(22)

In this work, the computation of the above posterior is carried out through the modified BUS framework.

4 EXPERIMENTAL RESULTS

4.1 Ionosphere Data Set

The Ionosphere data set [8] consists of 351 radar observations in 34 dimensions. The training set consists of 200 instances and the test set 151. In a binary classification setting 'Good' radar returns are those showing evidence of some type of structure in the ionosphere and are given the class label \( y = 1 \). On the other hand 'Bad' returns are those which do not show evidence of a structure type and are given the class label \( y = -1 \). The inputs are standardized to a zero mean and unit variance.

Expectation Propagation is chosen as a benchmark comparison for this study. Similar to previous benchmark studies [11,12] the covariance function chosen is the squared exponential

\[
k(x, x'|\theta) = \sigma^2 \exp \left( \frac{-||x - x'||^2}{-2l^2} \right)
\]  

(23)

where \( \theta = [\sigma, l] \), \( \sigma^2 \) refers to the signal variance and \( l \) is the characteristic length scale. The likelihood function used is the CDF of the standard Gaussian distribution or the 'probit' function of the form

\[
\Phi(x) = \int_{-\infty}^{x} \mathcal{N}(x|0, 1)dx
\]

(24)

4.2 Results

The parameters for SuS were set to \( N = 5,000 \) and \( p_0 = 0.2 \) respectively. A one dimensional proposal PDF was chosen to be uniform in [0,1]. Figure 2 shows the resulting log-CCDF and
log-evidence respectively. The general nature of the curves follows the characteristic trends predicted in Figure 1.

![Figure 2: Log-evidence of characteristic trends for SuS sampling. Both $\ln P(Y > b)$ (left) and $V(b) = b + \ln P(Y > b)$ (right) exhibit behaviour which coincides with the theory. The vertical lines represent the threshold value at each intermediate level.](image)

For the automatic stopping condition a tolerance of $a_m = 10^{-6}$ was set as the threshold for the probability of inadmissibility in Eq.(17). The evolution of the threshold and the values of the probability of inadmissibility are presented in Table 1. The probability of inadmissibility has converged to a value smaller than the set tolerance at level 3. Thus, the samples generated at this level may be assumed to be drawn from the desired posterior distribution.

<table>
<thead>
<tr>
<th>Level</th>
<th>$b_m$</th>
<th>$c_m$</th>
<th>$a_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>138.5642</td>
<td>6.6425e-61</td>
<td>0.9</td>
</tr>
<tr>
<td>2</td>
<td>151.6193</td>
<td>1.4209e-66</td>
<td>0.0015</td>
</tr>
<tr>
<td>3</td>
<td>159.2372</td>
<td>6.9849e-70</td>
<td>4.89e-07</td>
</tr>
</tbody>
</table>

Table 1: Evolution of the threshold and the probability of inadmissibility

Having computed the posterior distribution over the latent function, the samples generated at level 3 are used to produce predictive latent function values $f^*$. An investigation on a regular 21x21 grid of values for the log hyper-parameters was carried out whereby for each value of $\theta$ on the grid the approximate log marginal likelihood is calculated. Figure 3 shows a contour plot of the approximate log marginal likelihood for both BUS and EP.
Classification performance on the test data set in terms of predictive probabilities is calculated through the scaled information score, given by

\[ I = H + \frac{1}{2n} \sum_{i=1}^{n} (1 + y_i) \log_2(p_i) + (1 - y_i) \log_2(1 - p_i) \]  

(25)

where \( n \) is the number of test observations. The value of \( I \) may range from 0 to 1 where 1 bit indicates perfect prediction and 0 bits random guessing. The information score expresses the difference between the baseline entropy \( H \) of the training set labels and the average negative log probabilities. The entropy for the training set labels is given by

\[ H = - \sum_{y=+1,-1} \frac{n_{y_{\text{test}}}}{n_{\text{test}}} \log_2 \frac{n_{y_{\text{train}}}}{n_{\text{train}}} \leq 1 \]  

(26)

where \( n_{y_{\text{test}}} \) and \( n_{y_{\text{train}}} \) denote the number of observations in the test and training data sets respectively with the given target class label. In this case \( y = 1 \) and \( H = 0.9908 \). Had the classes, training and test sets been perfectly balanced \( H \) would equal 1. Figure 4 shows a contour plot for BUS and EP where the maximum \( I \) values produced were 0.6123 and 0.656 respectively.
The misclassification rate for 35 independent runs of the modified BUS algorithm using the optimal hyperparameter values from the log marginal likelihood is presented in figure 5. The error threshold used for each test point was 0.5. The average error rate produced was 10.77%. This figure follows closely with other studies [11].

5 Conclusion

This paper has presented an implementation of the modified BUS framework as an inference method for GP classifiers. The modified BUS algorithm has been shown to produce samples from the relevant posterior distribution. From the comparison with Expectation Propagation, scaled information scores and misclassification error rates the framework appears to perform comparably well. Future work includes extending the framework to more complex scenarios, in particular multi-class classification.

REFERENCES


A BAYESIAN APPROACH TO MODEL CALIBRATION AND PARAMETER ESTIMATION IN POTENTIAL DROP MEASURING

Thomas Berg\(^1\), Sven von Ende\(^2\) and Rolf Lammering\(^1\)

\(^1\)Institute of Mechanics, Helmut Schmidt University/University of the Federal Armed Forces Hamburg
Holstenhofweg 85, 22043 Hamburg, GERMANY
e-mail: {thomas.berg, rolf.lammering}@hsu-hh.de

\(^2\)Rolls-Royce Deutschland Ltd & Co KG
Eschenweg 11, Dahlewitz, 15827 Blankenfelde-Mahlow, GERMANY
e-mail: sven.vonende@rolls-royce.com

Keywords: Probabilistic analysis, Fatigue crack growth, Fatigue test methods, Bayesian model calibration, Parameter estimation.

Abstract. The direct current potential drop method is a widespread technique used to monitor fatigue-related crack initiation and growth. However, to determine the size of cracks in a structure indirectly via potential drop measurements in the presence of unknown material-dependent parameters bears not only the challenge of model calibration but of parameter estimation as well.

A solution to this inverse problem is accomplished by a Bayesian approach. Therefore, physics and measuring characterising models as well as prior information on the unknown crack size and unknown parameters are specified. Within this thus defined probabilistic state space model, Bayesian inference is performed to infer the crack size of fatigue-tested specimens from potential drop measurements while taking into account persistent variabilities and uncertainties. The results obtained in the form of joint conditional posterior distributions are utilised to calibrate the potential drop measuring and to estimate the unknown parameters on the one hand as well as to estimate the crack size on the other hand. By comparison with experimentally acquired actual crack size data, the proposed Bayesian approach is found to yield accurate results, rendering the methodology highly promising.
1 INTRODUCTION

Within the direct current potential drop (DCPD) method, the difference in electric potential between two points with an intermediate defect is measured. Considering a crack or corrosion as the defect, the measurements can be utilised to obtain accurate information about the shape and size in situ without the necessity of optical assessment, constituting the DCPD method as widely accepted approach to non-destructive testing in aviation, civil engineering, petrochemical and power generation industries [1]. Monitoring fatigue-related crack initiation and growth, a main challenge in applying DCPD measuring lies in facilitating a linkage between the observed potential drop and the underlying crack geometry, e.g. the crack size. However, the linkage between both quantities can be accomplished by a calibration curve as conceptually shown in Figure 1. A new methodology to address this problem by Bayesian filtering and smoothing has been presented in [2] and is carried on and enhanced with the present contribution.

The various ways to obtain calibration curves encompass experimental, analytical and numerical approaches: By means of manually introduced and pre-defined defects, a reference test serves as experimental calibration for subsequent specimens with a consistent probe and electrode setup [3]. In other experimental approaches, crack front marking techniques like heat tints [4] and beachmarks (frequency or stress ratio shifts) [5, 6] can be either utilised to allow a reference-based calibration as well or, when accepting the loss of the specimen, be harnessed for a specific calibration subsequent to the test. By solving the Laplace equation of an electrical potential for certain geometry and boundary conditions, theoretical calibrations can be derived. Though this is not generally applicable, [7] provides analytical solutions for simple geometries like plane single-edge, double-edge and centre cracks. For more complex geometries numerical methods like FEM have been employed [8, 9].

The challenge of obtaining a calibration curve can be ascribed to finding a solution to the inverse problem of deriving unobservable states from a set of observations and can be reasonably addressed by a Bayesian approach [10]. Bayesian inference has been widely used in
model-based structural health monitoring and prognostics in the context of damage diagnosis and damage prognosis, where sensor data of any kind is used in conjunction with physics and measuring characterising models to assess the state of a single component or a whole structure while taking into account system-inherent uncertainties and variabilities, e.g. [11, 12]. In the presence of unknown parameters that govern the physics-based models, for example parameters of a damage evolution model, the inverse problem becomes even more complex.

Herein, the model calibration challenge in DCPD measuring is addressed in conjunction with parameter estimation by a Bayesian approach for fatigue-tested corner crack specimens. As opposed to [2] where solely the model calibration and state estimation is dealt with, the recursive Bayesian filtering and smoothing framework is extended to additionally infer the unknown dynamic model parameters from the cycle-dependent potential drop observations. Furthermore, two different dynamic models that depict the underlying physical process with diverging precision are utilised. As a consequence, neither are marking techniques and the accompanying destruction of the specimen or cumbersome specimen preparations necessary nor have restrictions regarding the specific specimen geometry to be imposed. Furthermore, unlike in numerical simulation approaches, specimen variability can be accounted for by considering material-dependent parameters as uncertain and by using actual, specimen-specific measurement data.

The paper is organised as follows: Section 2 recalls the fundamentals of Bayesian inference and gives theoretical background on the utilised recursive filtering and smoothing methods. In Section 3, the experimental procedure is presented and harnessed to derive the subsequently applied state space definition. The Section is completed with insight into the adopted algorithm operation. In Section 4, the results obtained by the Bayesian approach are presented and discussed whereas Section 5 concludes the paper.

2 BAYESIAN APPROACH

The modelling of dynamic processes with imperfect or unknown information gives rise to numerous sources of uncertainty and variability that have to be accounted for [13–16]. These sources contribute to the system’s state uncertainty, the modelling uncertainty and - in terms of prognosis - the future uncertainty. The state of a system is uncertain since it is determined indirectly via observations where uncertainty can be recognised in the measurement method itself as well as in the governing measuring parameters that are fixed, yet unknown. Furthermore, the model representing the dynamic process incorporates uncertainty as the true phenomenon under consideration can only be depicted adequately to a certain degree. This includes the mathematical expressions as well as the uncertainty in unknown fixed and variable model parameters. In addition, predictions about future states entail uncertainty associated with not controllable quantities influencing the states. All these uncertainties have to be represented, quantified and propagated.

The inverse problem of identifying unknown states given a set of observations is herein addressed in a probabilistic setting by a Bayesian approach that exhibits three main features:

(I) Such inverse problems are typically ill-posed as the states are often overfitted or underdetermined. By means of regularisation, additional assumptions on the states can be introduced to circumvent the issue. This approach corresponds to incorporating prior information in Bayesian inference, which is why many inverse problems become well-posed when formulated in a Bayesian fashion [10, 17].

(II) Defining a probabilistic state space model, all quantities of interest are modelled as random variables with corresponding probability distributions. The randomness expresses the degree of belief or amount of information about their true values and represents the uncertainty
associated with these unknown quantities.

(III) The objective in the context of solving an inverse problem is not limited to identifying a particular set of states and parameters that optimally satisfies an arbitrary condition, like for example minimising the expected value of a loss function. Utilising a Bayesian approach, the unknown states are inferred from given observations, i.e. new available data is used to adjust the prior belief or knowledge about the states in a probabilistic manner, thereby allowing for systematic uncertainty quantification and uncertainty propagation [18].

2.1 Bayesian filtering and smoothing in parameter estimation

Let \( x_{0:T} = \{x_0, \ldots, x_T \} \) and \( y_{1:T} = \{y_1, \ldots, y_T \} \) denote a vector-valued time series of hidden states \( x_k \in \mathbb{R}^{d_x} \) and observed measurements \( y_k \in \mathbb{R}^{d_y} \) at time step \( k \) up to a last time step \( T \), \( \theta \in \mathbb{R}^{d_\theta} \) a vector containing the unknown parameters and \( p(\cdot) \) the probability distribution of a continuous random variable. In Bayesian inference the joint posterior distribution of the states and parameters given the measurements can then be computed via Bayes’ rule as

\[
p(x_{0:T}, \theta | y_{1:T}) = \frac{p(y_{1:T}|x_{0:T}, \theta)p(x_{0:T}|\theta)p(\theta)}{p(y_{1:T})}, \tag{1}
\]

Since discrete-time observations provide sequential data, it is more efficient and generally more convenient to only infer the current state and parameters on the basis of the preceding ones. Applying a recursive prediction and update scheme with due regard to the Markov property of states and the conditional independence of measurements and the thus ensuing implications for the choice of dynamic models [2] yields the desired effect and ensures a constant number of computations per time step. The predictive distribution of the state \( x_k \) at time step \( k \) given the measurements up to time step \( k - 1 \) and the unknown parameters is given by the Chapman-Kolmogorov equation as

\[
p(x_k | y_{1:k-1}, \theta) = \int p(x_k | x_{k-1}, \theta)p(x_{k-1} | y_{1:k-1}, \theta)dx_{k-1}, \tag{2}
\]

where \( p(x_k | x_{k-1}, \theta) \) is the dynamic model. It characterises probabilistically the propagation of the states in time. The posterior distribution of the state \( x_k \) given the measurements up to time step \( k \) and the unknown parameters follows from Bayes’ rule as

\[
p(x_k | y_{1:k}, \theta) = \frac{p(y_k | x_k, \theta)p(x_k | y_{1:k-1}, \theta)}{p(y_k | y_{1:k-1}, \theta)}, \tag{3}
\]

where \( p(y_k | x_k, \theta) \) is the measurement model. The prediction step in Eq. (2) and update step in Eq. (3) together constitute the recursive Bayesian filtering equations [19] and allow for the inference of the states \( x_k \). In order to infer the unknown parameters \( \theta \), the filtering expressions have to be complemented by the marginalised joint posterior distribution of the parameters given the measurements:

\[
p(\theta | y_{1:k}) = p(\theta | y_{1:k-1}) \int \frac{p(y_k | x_k, \theta)p(x_k | y_{1:k-1}, \theta)}{p(y_k | y_{1:k-1})}dx_k. \tag{4}
\]

Thereby, the inference on the fixed but unknown parameters is enhanced at every time step in the presence of new observations. Beyond incorporating new observations to update the current state and to predict future ones, to adjust preceding states might be beneficial given the context.
as well. By conditioning previous states on all observations (including future observations) as in
\[ p(x_k|y_{1:T}, \theta) = p(x_k|y_{1:k}, \theta) \int \frac{p(x_{k+1}|x_k, \theta)p(x_{k+1}|y_{1:T}, \theta)}{p(x_{k+1}|y_{1:k}, \theta)} \, dx_{k+1}, \tag{5} \]
the marginal posterior distribution of the state \( x_k \) given all observations up to time step \( T \) with \( T > k \) can be obtained in a smoothing step. Finally, by formally modelling an observable dynamic process as probabilistic state space model of the following form
\[
\begin{align*}
\theta & \sim p(\theta), \\
x_0 & \sim p(x_0|\theta), \\
x_k & \sim p(x_k|x_{k-1}, \theta), \\
y_k & \sim p(y_k|x_k, \theta),
\end{align*}
\tag{6}
\]
the issue of solving the inverse problem can be approached by Bayesian filtering and smoothing.

Closed form solutions for Eq. (2) to (5) generally exist only for a few classes of filtering and smoothing problems. For linear Gaussian state space models, optimal Bayesian solutions can be obtained by means of the Kalman filter \([20]\) and Rauch-Tung-Striebel smoother \([21]\). In terms of finite state space models, grid-based methods can be utilised to provide an optimal solution \([22,23]\). However, a variety of approximate filtering and smoothing methods for nonlinear and infinite state space models exist, which encompass inter alia Gaussian filtering and smoothing methods \([19]\), Monte Carlo Sampling approaches \([23]\) and approximate grid-based methods \([24]\). The former can be considered as local approaches, the two latter as global approaches \([25]\).

In the present paper, a combination of local and global approaches is utilised which is why further insight into these methods is given hereafter.

2.2 Unscented transform for filtering and smoothing

The unscented Kalman filter (UKF) \([26–28]\) as well as the unscented Rauch-Tung-Striebel smoother (URTSS) \([19]\) address the filtering and smoothing problem by utilising the unscented transform as a way of deterministic sampling \([14]\) in a Gaussian framework. Considering two random variables \( \xi \) and \( \nu \) with a nonlinear model function \( \nu = g(\xi) \), an approximation to the joint distribution can be obtained by estimating the mean \( \mu_\nu \) and covariance \( \Sigma_{\nu\nu} \) by means of a minimal set of weighted samples. These so called sigma points \( \mathcal{X}^i \) are chosen deterministically and propagated through the respective model to form new sigma points \( \mathcal{Y}^i = g(\mathcal{X}^i) \) and compute mean and covariance:
\[
\begin{align*}
\mu_\nu & = \sum_i W^i \mathcal{Y}^i, \\
\Sigma_{\nu\nu} & = \sum_i W^i (\mathcal{Y}^i - \mu_\nu)(\mathcal{Y}^i - \mu_\nu)^T.
\end{align*}
\tag{7, 8}
\]
Assuming an a priori Gaussian state distribution as well as additive Gaussian noise in the dynamic and measurement models
\[
\begin{align*}
x_k &= f(x_{k-1}, \theta) + q_{k-1}, \\
y_k &= h(x_k, \theta) + r_k,
\end{align*}
\tag{9, 10}
\]
where \( f(\cdot) \) is the dynamic transition function, \( h(\cdot) \) is the measurement function, \( q_k \sim \mathcal{N}(0, Q_k) \) is the process noise and \( r_k \sim \mathcal{N}(0, R_k) \) the noise of measurements, Gaussian approximations to the filtering distributions in Eq. (2) and (3) can be obtained as

\[
p(x_k | y_{1:k-1}, \theta) \approx \mathcal{N}(x_k | m_k^-(\theta), P_k^-) ,
\]

\[
p(x_k | y_{1:k}, \theta) \approx \mathcal{N}(x_k | m_k^+(\theta), P_k^+) ,
\]

by utilising the UKF. In the prediction step, the mean \( m_k^- \) and covariance \( P_k^- \) which depend on \( \theta \) are computed by

- forming the sigma points

\[
\mathcal{X}_{k-1}^i(\theta) = \begin{cases} 
m_{k-1}(\theta), & i = 0 \\
m_{k-1}(\theta) + \lambda \sqrt{P_{k-1}(\theta)}, & i = 1, \ldots, d_x \\
m_{k-1}(\theta) - \lambda \sqrt{P_{k-1}(\theta)}, & i = d_x + 1, \ldots, 2d_x, 
\end{cases}
\]  

(12)

- propagating the sigma points

\[
\hat{\mathcal{X}}_k^i(\theta) = f(\mathcal{X}_{k-1}^i(\theta)), \quad i = 0, \ldots, 2d_x,
\]  

(13)

which yields

\[
m_k^-(\theta) = \sum_{i=0}^{2d_x} W^i \hat{\mathcal{X}}_k^i(\theta),
\]

\[
P_k^-(\theta) = \sum_{i=0}^{2d_x} W^i \left( \hat{\mathcal{X}}_k^i(\theta) - m_k^-(\theta) \right) \left( \hat{\mathcal{X}}_k^i(\theta) - m_k^-(\theta) \right)^T + Q_{k-1}.
\]  

(14)

In the update step, the mean \( m_k^+ \) and covariance \( P_k^+ \) are computed by

- forming the sigma points

\[
\mathcal{X}_{k}^{-i}(\theta) = \begin{cases} 
m_k^-(\theta), & i = 0 \\
m_k^+(\theta) + \lambda \sqrt{P_k^+(\theta)}, & i = 1, \ldots, d_x \\
m_k^-(\theta) - \lambda \sqrt{P_k^+}, & i = d_x + 1, \ldots, 2d_x, 
\end{cases}
\]  

(15)

- propagating the sigma points

\[
\hat{\mathcal{Y}}_k^i(\theta) = h(\mathcal{X}_{k}^{-i}(\theta)), \quad i = 0, \ldots, 2d_x,
\]  

(16)

- calculating

\[
\mu_k(\theta) = \sum_{i=0}^{2d_x} W^i \hat{\mathcal{Y}}_k^i(\theta),
\]

\[
S_k(\theta) = \sum_{i=0}^{2d_x} W^i \left( \hat{\mathcal{Y}}_k^i(\theta) - \mu_k(\theta) \right) \left( \hat{\mathcal{Y}}_k^i(\theta) - \mu_k(\theta) \right)^T + R_k,
\]

\[
C_k(\theta) = \sum_{i=0}^{2d_x} W^i \left( \hat{\mathcal{X}}_k^{-i}(\theta) - m_k^-(\theta) \right) \left( \hat{\mathcal{Y}}_k^i(\theta) - \mu_k(\theta) \right)^T ,
\]

\[
K_k(\theta) = C_k S_k^{-1}(\theta),
\]  

(17)
which yields

\[
\begin{align*}
m_k(\theta) &= m_k^{-}(\theta) + K_k(\theta) \left[ y_k(\theta) - \mu_k(\theta) \right], \\
P_k(\theta) &= P_k^{-}(\theta) - K_k(\theta) S_k(\theta) K_k^T(\theta).
\end{align*}
\]

(18)

The free parameter \( \lambda \) is a scaling quantity that should be set to \( \lambda = 3 - d_x \) for Gaussian state variables on a heuristic bases [26]. The weights \( W_i \) can be computed as

\[
W_i = \begin{cases} 
\frac{\lambda}{d_x + \lambda}, & i = 0 \\
\frac{\lambda}{2(d_x + \lambda)}, & i = 1, \ldots, 2d_x.
\end{cases}
\]

(19)

Recalling the complementation of the Bayesian filtering distributions by the marginalised joint posterior distribution of the parameters given the measurements \( p(\theta|y_{1:k}) \) in Eq. (14), the estimation of the unknown parameters has to be addressed as well. The equation is rewritten [29] as

\[
p(\theta|y_{1:k}) = p(\theta|y_{1:k-1}) \int p(y_k|x_k, \theta)p(x_k|y_{1:k-1}, \theta) \frac{p(y_k|y_{1:k-1})}{p(y_k|y_{1:k-1})} dx_k \\
\propto p(\theta|y_{1:k-1})p(y_k|y_{1:k-1}, \theta),
\]

(20)

whereas the second term is obtained as Gaussian approximation

\[
p(y_k|y_{1:k-1}, \theta) \approx N(y_k|\mu_k(\theta), S_k(\theta)).
\]

(21)

The corresponding mean \( \mu_k(\theta) \) and covariance \( S_k(\theta) \) are the results of the calculation in Eq. (17).

By means of the unscented Rauch-Tung-Striebel smoother (URTSS), a Gaussian approximations to the smoothing distribution in Eq. (5) can be obtained as

\[
p(x_k|y_{1:T}, \theta) \approx N(x_k|m_k^*(\theta), P_k^*(\theta)).
\]

(22)

In the smoothing step, the mean \( m_k^*(\theta) \) and covariance \( P_k^*(\theta) \) are computed by

- forming the sigma points

\[
\mathcal{X}_k^i(\theta) = \begin{cases} 
m_k(\theta), & i = 0 \\
m_k(\theta) + \sqrt{d_x + \lambda \lambda P_k(\theta)}, & i = 1, \ldots, d_x \\
m_k(\theta) - \sqrt{d_x + \lambda \lambda P_k(\theta)}, & i = d_x + 1, \ldots, 2d_x.
\end{cases}
\]

(23)

- propagating the sigma points

\[
\mathcal{X}_{k+1}^i(\theta) = f(\mathcal{X}_k^i(\theta)), \quad i = 0, \ldots, 2d_x,
\]

(24)

- calculating

\[
\begin{align*}
m_{k+1}^{-}(\theta) &= \sum_{i=0}^{2d_x} W_i \mathcal{X}_{k+1}^i(\theta), \\
P_{k+1}^{-}(\theta) &= \sum_{i=0}^{2d_x} W_i \left( \mathcal{X}_{k+1}^i(\theta) - m_{k+1}^-(\theta) \right) \left( \mathcal{X}_{k+1}^i(\theta) - m_{k+1}^-(\theta) \right)^T + Q_k, \\
D_{k+1}(\theta) &= \sum_{i=0}^{2d_x} W_i \left( \mathcal{X}_k^i(\theta) - m_k(\theta) \right) \left( \mathcal{X}_k^i(\theta) - m_k(\theta) \right)^T,
\end{align*}
\]

(25)
which yields

\[ G_k(\theta) = D_{k+1}(\theta) \left[ P_{k+1}(\theta) \right]^{-1}, \]

\[ m^*_k(\theta) = m_k(\theta) + G_k(\theta) \left[ m^*_{k+1}(\theta) - m^*_{k+1}(\theta) \right], \]

\[ P_k(\theta) = P_k(\theta) + G_k(\theta) \left[ P^*_{k+1}(\theta) - P^*_{k+1}(\theta) \right] G^T_k(\theta). \]

2.3 Approximate grid-based filtering

In approximate grid-based filtering [22,23], the continuous state space is decomposed into pre-defined cells and the filtering distributions are approximated by a weighted sum of \( \delta \) functions. Discretising the state space into a finite number of states \( \{x^j : j = 1, \ldots, N_x\} \), the filtering solutions in Eq. (2) and (3) can be obtained as

\[ p(x_k|y_{1:k-1}, \theta) \approx \sum_{j=1}^{N_x} w^j_{k|k-1} \delta(x_k - x^j_k), \]

\[ p(x_k|y_{1:k}, \theta) \approx \sum_{j=1}^{N_x} w^j_{k|k} \delta(x_k - x^j_k) \]

where the weights \( w^j_{k|k-1} \) and \( w^j_{k|k} \) represent the conditional probabilities of \( x^j \) given the measurements \( y_{1:k-1} \) respectively \( y_{1:k} \) and the parameters \( \theta \). They are computed as

\[ w^j_{k|k-1} = \sum_{l=1}^{N_x} w^j_{k-1|k-1} p(x^j_k|x^l_k, \theta), \]

\[ w^j_{k|k} = \frac{w^j_{k|k-1} p(y_k|x^j_k, \theta)}{\sum_{l=1}^{N_x} w^j_{k|k-1} p(y_k|x^l_k, \theta)}, \]

where the initial weights are denoted as \( w^j_{0|0} = p(x^j_0|\theta) / \sum_{l=1}^{N_x} p(x^l_0|\theta) \). The approximations arise due to the fact that the probabilities are not integrated over the regions of the continuous state space but averaged for the discretised grid points \( x^j \). In order to obtain accurate approximations to the filtering distributions, the resolution of the grid has to be sufficiently fine which - in combination with a high dimension of the state - renders the approach computational burdensome.

3 PROBLEM STATEMENT

3.1 Experimental procedure

The fatigue testing is performed by subjecting Udimet 720Li superalloy specimens to cyclic uniaxial tensile stress within a convection oven at a temperature of 400 °C, see Figure 2. The quadratic bar specimens (side \( W \)) are pre-notched at a single corner to act as crack initiation point under a sinusoidal loading at \( f = 10 \) Hz with the aim of overcoming non-continuum mechanics-governed and unstable crack growth [30]. A stable Mode-I crack propagation is then accomplished by a trapezoidal shape loading at \( f = 0.25 \) Hz. The crack surface is assumed to evolve in a quarter-circular shape which renders the equidistant length \( r(\phi) \) from pre-notched corner to crack front the single quantity \( a \) characterising the crack size. This simplification is reasonable as long as crack tunnelling can be ruled out [31].
The crack size $a$ can indirectly be measured conveniently by means of the DCPD method. Therefore, a direct current is supplied by electrodes near the lower and upper ends of the specimen (3 and 4 in Figure 2). Subsequently, the potential drop $U$ is determined by measuring the potential difference between the respective crack surfaces via probes that are spot-welded in close proximity to the notch (1 and 2 in Figure 2). With increasing crack size $a$ the obtained potential drop $U$ scales accordingly, yielding time-dependent or more conveniently cycle-dependent values $U_k$ at specific cycles $N_k$. With only roughly pre-defined probe positions, neither symmetry nor accuracy can be assumed, rendering the relationship between the cycle-dependent crack size $a_k$ and potential drop $U_k$ indeed quantifiable, however unknown. The relation can then be depicted as

$$U_k = c(a_k; b) + \varepsilon_k$$

(30)

where $c$ is an arbitrary function depending on the vector of unknown parameters $b$ representing the calibration curve and $\varepsilon_k$ is the error term for the particular observation at time step $k$.

For the purpose of assessing the proposed Bayesian model calibration and parameter estimation, reference damage propagation data have to be obtained. This is done by performing an additional experimental calibration by investigating the crack front of each specimen subsequent to the fatigue testing. Therefore, the specimens are cut in the respective crack propagation cross section where the fracture plane $A_{fp}$ can be measured and utilised to average the crack sizes $a_0$ at the beginning and $a_{end}$ at the end of the crack propagation as

$$a = \sqrt{\frac{4A_{fp}}{\pi}}.$$  

(31)

With the assumption of $c$ being linear [23, 32], the 3-tuples $(N_0, U_0, a_0)$ and $(N_{end}, U_{end}, a_{end})$ can be employed to determine $b$ and subsequently interpolate $a_k$. The reference data are hereafter denoted as actual or true parameter and crack size values. The obtained data consists of six sample trajectories $\{U_k : k = 1, \ldots, n_{ms}\}$ for specimens #1 to #6 (each subjected to various stress ranges and stress ratios) with 26 to 41 total measurements $n_{ms}$.
3.2 State space definition

Recalling the general form of the probabilistic SSM for observable dynamic processes in Eq. (6), different quantities have to be defined probabilistically in order to allow for the application of Bayesian filtering and smoothing. This definition encompasses the modelling of dynamic and measurement processes as well as prior information on the unknown states and parameters and is carried out hereafter for the corner crack fatigue testing presented in the preceding section.

When considering crack growth under fatigue loading, the damage evolution is generally characterised by three stages: (I) crack initiation, (II) stable crack growth, (III) instable crack growth and failure. It can be visualised by plotting the crack growth rate \( \frac{da}{dN} \) over the stress intensity factor (SIF) range \( \Delta K \) as shown in Figure 3. The former describes the crack size increment per cycle. The latter depicts the range of stress intensity at the crack front tip which is dependent on the applied stress range \( \Delta \sigma = \sigma_{\text{max}} - \sigma_{\text{min}} \) as well as the specific specimen geometry. A widely used formula utilised to model sub-critical, stable stage II crack growth is given by the Paris-Erdogan law \([33, 34]\)

\[
\frac{da}{dN} = C(\Delta K(a))^n, \tag{32}
\]

where \( C \) and \( n \) are material-dependent scaling constants. The SIF range \( \Delta K \) can be calculated for corner crack specimens \([30]\) as

\[
\Delta K(a) = \Delta \sigma g(a) \sqrt{\pi a}, \tag{33}
\]

\[
g(a) = M_G(a) M_B(a) M_S(a) \frac{2}{\pi}, \tag{34}
\]

where \( g(a) \) is a crack-geometric correction factor that can be derived numerically for different specimen geometries and crack shapes and is given by \([35]\). The precise expressions for the individual shape functions \( M_G(a), M_B(a) \) and \( M_S(a) \) are omitted for the sake of clarity. Since only the stress range \( \Delta \sigma \) is utilised in Eq. (32), mean stress effects are not taken into account \([36]\). In order to circumvent this problem, Paris’ law can be improved by Walker’s law \([37]\), which is one among other modifications, e.g. \([38-40]\). It reads

\[
\frac{da}{dN} = C \left( \frac{\Delta K(a)}{(1 - R_\sigma)^{1 - \kappa}} \right)^n, \tag{35}
\]
where $R_\sigma = \sigma_{\text{min}} / \sigma_{\text{max}}$ is the stress ratio and $\kappa$ is an empirical constant weighing the influence of the mean stress on the crack growth rate (typically around 0.5 for metals). Where the aforementioned models fail to catch the sigmoidal shape of the crack growth rate plotted over the SIF range as depicted in Figure 3, a modification according to Forman/Mettu [41] allows for stage I and stage III crack growth depiction as well. It is given by

$$\frac{da}{dN} = C \cdot F \cdot \Delta K(a)^n \left( \frac{1 - \frac{\Delta K_{\text{th}}}{\Delta K(a)}}{1 - \frac{1}{1 - R_\sigma \frac{\Delta K(a)}{K_c}}} \right)^q,$$

where $F$ is a crack velocity factor, $\Delta K_{\text{th}}$ is the threshold of SIF range for crack propagation, $K_c$ is the fracture toughness and $p, q$ are empirical constants [42]. In order to allow for a highly accurate representation of the crack propagation process of the available pre-notched specimens, a simplified and further modified Paris’ law is obtained by setting $F = 1, q = 1, p = n$ [42] and introducing an additional exponent $m = 10$:

$$\frac{da}{dN} = C \cdot \Delta K(a)^n \left( 1 - \left( \frac{\Delta K_{\text{th}}}{\Delta K(a)} \right)^m \right)^n.$$  

(37)

The rationale behind this approach is the following: In [2] where the model calibration was accomplished by means of Walker’s law with fixed dynamic model parameters, the contribution of an appropriately represented test-initiation phase with a suitable dynamic model to the accuracy improvement was not investigated. By modifying the Forman/Mettu equation to a degree where an almost exact compliance with the reference data is feasible depending on the material-dependent parameters, the performance of the two dynamic models can be compared.

Both Walker’s law in Eq. (35) and the Forman/Mettu equation in Eq. (36) are first order ordinary differential equations that can be numerically approximated for sufficiently small $\Delta N = N_k - N_{k-1}$ using finite differences and be rewritten as recurrence relation, yielding

$$a_k = a_{k-1} + \left. \frac{da}{dN} \right|_{a=a_{k-1}} \Delta N$$

(38)

at time step $k$. Both equations are used to define the dynamic model and discussed later on: Where the Forman/Mettu equation allows a more precise depiction of the damage evolution, Walker’s law comprises fewer unknown parameters.

As emphasised in Section 3.1, the relation between the measured potential drop and the crack size is assumed to be linear for corner crack specimens [3, 32], yielding

$$U_k = b_1 a_k + b_2 + \varepsilon_k,$$

where $b_1, b_2$ are unknown and $\{\varepsilon_k, k = 1, \ldots, n_{\text{ms}}\}$ has to be accounted for stochastically. If the assumption of linearity does not hold, for example in the case of specimens with a different crack geometry or more complex crack front shapes, the approach can be extended by using other forms, e.g. polynomial functions.

Let $x_k \in \mathbb{R}$ denote the crack size $a_k$, $y_k \in \mathbb{R}$ the potential drop measurement $U_k$ at time step $k$ and the vector $\theta \in \mathbb{R}^{d_\theta}$ the collection of unknown parameters $C, n, b_1, b_2$ with $d_\theta = 4$ respectively $C, n, b_1, b_2, \Delta K_{\text{th}}$ with $d_\theta = 5$. In order to account for existent uncertainties and variabilities as mentioned in Section 4, the hidden states $x_{0:n_{\text{ms}}}$, the measurements $y_{1:n_{\text{ms}}}$ and
the unknown parameters $\theta$ are viewed as random variables that are assigned probability distributions, thereby expressing the degree of information concerning their realisations and the incorporated uncertainty. The initial state $x_0$ is supposed to be normally distributed with mean $\mu_0$ and variance $\sigma_0^2$, since the process of notching the specimens is manageable to a certain degree. Since the threshold SIF range $\Delta K_{th}$ is dependent on the crack size as given in Eq. (33), the prior belief on $\theta_s$ can be obtained by means of $x_0 \sim \mathcal{N}(\mu_0, \sigma_0^2)$. The remaining unknown parameters are initially assumed to be mutually independent random variables uniformly distributed on $[\theta_1, \theta_3] \times \ldots \times [\theta_4, \theta_4]$ because no prior information other than approximate boundaries are available. By further assuming constant additive Gaussian process noise $q \sim \mathcal{N}(0, Q)$ and Gaussian noise of measurements $r \sim \mathcal{N}(0, R)$ within the dynamic models and measurement model, the probabilistic state space model can then be defined as

$$x_k = f^{(I,I)}(x_{k-1}, \theta) + q, \quad (40)$$

$$y_k = h(x_k, \theta) + r, \quad (41)$$

where the dynamic and measurement functions are

$$f^{(I)}(x_{k-1}, \theta) = x_{k-1} + \theta_1 \left( \frac{\Delta K(x_{k-1})}{(1-R_q)^{1-\kappa}} \right)^{\theta_2} \Delta N, \quad (42)$$

$$f^{(II)}(x_{k-1}, \theta) = x_{k-1} + \theta_1 \Delta K(x_{k-1})^{\theta_2} \left( 1 - \frac{\theta_c}{\Delta K(x_{k-1})} \right)^{\theta_3} \Delta N, \quad (43)$$

$$h(x_k, \theta) = \theta_3 x_k + \theta_4. \quad (44)$$

### 3.3 Algorithm operation

With the unscented Kalman filter and the unscented Rauch-Tung-Striebel smoother as recursive approximations to the Bayesian inference in Section 2 and the ensuing definition of the probabilistic state space model in Section 3, the objective of finding a solution to the inverse problem of crack size estimation, model calibration and parameter estimation by means of potential drop measurements can be addressed as of yet only formally. Recalling the prediction, update and smoothing steps, the dependency of the state $x_k$ on the unknown parameters $\theta$ has to be propagated through every time step $k$ resulting in nested functions that become ever more complex with each additional iteration. Therefore, a hybrid of local and global approaches in the form of the UKF/URTSS and approximate grid-based methods as depicted in Algorithm 1 is proposed. Similar methods have been used for example in [43–45] where particle filters instead of Gaussian filters are utilised in differing frameworks to compute the nodes in randomly or fixedly created grids in the parameter space, however with significant higher computational efforts.

As conceptually shown in Section 2.3 for the state space, the parameter space is now discretised into a finite number of cells $\{\theta^j : j = 1, \ldots, N_\theta\}$, leading to recast Eq. (20) as

$$p(\theta | y_{1:k}) \approx \sum_{j=1}^{N_\theta} w_{\theta}^{k,j} \delta(\theta - \theta^j), \quad (45)$$

where the weights $w_{\theta}^{k,j}$ can be computed by

$$w_{\theta}^{k,j} \approx \frac{w_{\theta}^{k-1,j} p(y_k | y_{1:k-1}, \theta^j)}{\sum_{l=1}^{N_\theta} w_{\theta}^{k-1,l} p(y_k | y_{1:k-1}, \theta^l)}. \quad (46)$$
Algorithm 1 State and parameter estimation

1: set \(x_0 \sim \mathcal{N}(\mu_0, \sigma^2_0)\)
2: generate initial grid \(\theta^j\) and weights \(w^0_j\) for \(\theta \sim p(\theta)\)
3: set \(f(x_{k-1}, \theta)\)
4: set \(h(x_k, \theta)\)
5: for \(k = 1 : n_{ms}\) do
   6: \(\text{for } j = 1 : N_\theta \text{ do} \quad \triangleright \text{UKF steps}\)
      7: compute \(p(x_k|y_{1:k-1}, \theta^j)\)
      8: compute \(p(x_k|y_{1:k}, \theta^j)\)
      9: compute \(p(y_k|y_{1:k-1}, \theta^j)\)
     10: compute \(w^k_{\theta,j}\) by updating \(w^k_{\theta,j}\) \(\triangleright \text{Grid-based update}\)
   11: \(\text{end for}\)
   12: compute \(p(\theta|y_{1:k})\) \(\triangleright \text{MAP estimates}\)
13: \(\text{evaluate } \hat{\theta}^{\text{MAP}}\)
14: set \(\hat{x}_k = m^r_k\)
15: \(\text{for } \alpha = 1 : d_\theta \text{ do} \quad \triangleright \text{Marginalised posterior distributions}\)
16: compute \(p(\theta_\alpha|y_{1:k})\)
17: \(\text{end for}\)
18: compute \(p(x_k|y_{1:k})\) \(\triangleright \text{MMSE estimates}\)
19: \(\text{evaluate } \hat{\theta}^{\text{MMSE}}\)
20: \(\text{evaluate } \hat{x}_k^{\text{MMSE}}\)
21: \(\text{end for}\)
22: for \(k = n_{ms} - 1 : 1\) do
23: compute \(p(x_k|y_{1:T}, \hat{\theta}^{\text{MAP}})\) \(\triangleright \text{URTSS steps}\)
24: compute \(p(x_k|y_{1:T}, \hat{\theta}^{\text{MMSE}})\)
25: \(\text{evaluate } \hat{x}_k^{\text{MAP}}\)
26: \(\text{evaluate } \hat{x}_k^{\text{MMSE}}\)
27: \(\text{end for}\)

with the initial weights denoted as \(w^0_j = p(\theta^j) / \sum_{\theta} p(\theta^j)\). The recursive Bayesian filtering and smoothing expressions in Eq. (2), (3) and (5) then become

\[
p(x_k|y_{1:k-1}, \theta^j) = \int p(x_k|x_{k-1}, \theta^j)p(x_{k-1}|y_{1:k-1}, \theta^j)dx_{k-1}
\]

\[
p(x_k|y_{1:k}, \theta^j) = \frac{p(y_k|x_{k}, \theta^j)p(x_{k}|y_{1:k-1}, \theta^j)}{p(y_k|y_{1:k-1}, \theta^j)}
\]

\[
p(x_k|y_{1:T}, \theta^j) = \int p(x_{k+1}|x_k, \theta^j)p(x_{k+1}|y_{1:T}, \theta^j)dx_{k+1}
\]

and have to be evaluated \(N_\theta\) times for each \(\theta^j\) per time step.

In order to assess the performance of the Bayesian approach, different point estimates are obtained. With the proposed approach, the posterior knowledge inferred from the measurements and the prior information can be depicted in a more extensive manner, including variances and higher order moments. However, within the scope of this paper only point estimates are
considered. A natural choice is to maximise the posterior distribution of the parameters given the measurements to determine the parameter set with the highest probability as

$$\hat{\theta}^{\text{MAP}} = \arg \max_{\theta} p(\theta | y_{1:k}),$$  \hspace{1cm} (50)

where $\hat{\theta}^{\text{MAP}} = \theta^*$ is the maximum a posteriori (MAP) estimate of the parameters and $j^*$ the corresponding index of the discretisation. The associated MAP estimate of the crack size can then be identified as

$$\hat{x}_k^{\text{MAP}} = m_{j^*}.$$

Another obvious point estimate can be found in the minimum mean square error (MMSE) estimate of the marginal conditional posterior distributions,

$$\hat{\theta}_\alpha^{\text{MMSE}} = \int \theta_\alpha p(\theta_\alpha | y_{1:k}) d\theta_\alpha.$$  \hspace{1cm} (52)

The marginal conditional posterior distribution of the parameter given the measurements $p(\theta_\alpha | y_{1:k})$ can be obtained by summing over the respective other parameters in Eq. (3) for $\alpha = 1, \ldots, d_\theta$, i.e.

\begin{align*}
p(\theta_1 | y_{1:k}) & \approx \sum_{i=1}^{N_{\theta_1}} w_{\theta_1}^{k,i} \delta(\theta_1 - \theta_1^i), \quad w_{\theta_1}^{k,i} = \sum_{l=1}^{N_{\theta_1}/N_{\theta_1}} w_\theta^{k,(l-1)N_{\theta_1} + i}, \\
p(\theta_2 | y_{1:k}) & \approx \sum_{i=1}^{N_{\theta_2}} w_{\theta_2}^{k,i} \delta(\theta_2 - \theta_2^i), \quad w_{\theta_2}^{k,i} = \sum_{l=1}^{N_{\theta_2}/(N_{\theta_1}N_{\theta_2})} \sum_{j=1}^{N_{\theta_2}} w_\theta^{k,j + (l-1)N_{\theta_1}N_{\theta_2} + (i-1)N_{\theta_1}}, \\
p(\theta_3 | y_{1:k}) & \approx \sum_{i=1}^{N_{\theta_3}} w_{\theta_3}^{k,i} \delta(\theta_3 - \theta_3^i), \quad w_{\theta_3}^{k,i} = \sum_{l=1}^{N_{\theta_3}/(N_{\theta_1}N_{\theta_2}N_{\theta_3})} \sum_{j=1}^{N_{\theta_3}} w_\theta^{k,j + (l-1)N_{\theta_1}N_{\theta_2}N_{\theta_3} + (i-1)N_{\theta_1}N_{\theta_2}}, \\
& \vdots \\
p(\theta_\alpha | y_{1:k}) & \approx \sum_{i=1}^{N_{\theta_\alpha}} w_{\theta_\alpha}^{k,i} \delta(\theta_\alpha - \theta_\alpha^i), \quad w_{\theta_\alpha}^{k,i} = \sum_{l=1}^{N_{\theta_\alpha}/N_{\theta_\beta}} \sum_{j=1}^{N_{\theta_\alpha}} w_\theta^{k,j + (l-1)N_\beta + (i-1)N_{\theta_\beta}}, \hspace{1cm} (53)
\end{align*}

where

$$N_i = \prod_{\beta=1}^{\alpha} N_{\theta_\beta}, \quad N_j = \prod_{\beta=1}^{\alpha-1} N_{\theta_\beta}.$$

The MMSE estimate of the parameters $\hat{\theta}_\alpha^{\text{MMSE}}$ can then be computed by averaging over the weights for each of the $d_\theta$ parameters:

$$\hat{\theta}_\alpha^{\text{MMSE}} \approx \sum_{i=1}^{N_{\theta_\alpha}} w_{\theta_\alpha}^{k,i} \theta_\alpha^i.$$  \hspace{1cm} (54)

The marginalised conditional posterior distribution of the state given the measurements can be derived by integrating out the unknown parameters in Eq. (3) which - in terms of the discretised
parameter space - means summing over the product of the conditional posterior distribution of the state and the posterior distribution of the parameters:

\[ p(x_k|y_{1:k}) = \int p(x_k|y_{1:k}, \theta)p(\theta|y_{1:k})d\theta, \tag{55} \]

\[ p(x_k|y_{1:k}) \approx \sum_{j=1}^{N_\theta} p(x_k|y_{1:k}, \theta^j)w_{\theta}^{k,j}. \tag{56} \]

By discretising the state space into a finite number of states \( \{x^i : i = 1, \ldots, N_x\} \), the discrete approximation to Eq. (56) reads

\[ p(x_k|y_{1:k}) \approx \sum_{i=1}^{N_x} w_x^{k,i} \delta(x_k - x_k^i), \quad w_x^{k,i} = \frac{p(x_k^i|y_{1:k})}{\sum_{l=1}^{N_x} p(x_k^l|y_{1:k})}. \tag{57} \]

It can be utilised to compute the marginalised conditional posterior MMSE of the state analogously to Eq.(52) as

\[ \hat{x}_k^{\text{MMSE}} = \int x_k p(x_k|y_{1:k})dx_k, \]

\[ \hat{x}_k^{\text{MMSE}} \approx \sum_{i=1}^{N_x} w_x^{k,i} x_k^i. \tag{58} \]

4 RESULTS AND DISCUSSION

To begin with, the model calibration of the DCPD method is addressed, i.e. the estimation of the parameters \( \{\theta_\alpha, \alpha = 3, 4\} \). In Figure 4, the normalised estimates of the measurement model parameters are plotted over the normalised cycles as to allow an accurate comparison for both dynamic model functions \( f^{(I,II)}(x_k, \theta) \) for all specimens of the test series. This direct comparison is possible since the actual parameter values are available due to the experimental calibration mentioned in Section 3.1. In addition to the MMSE and MAP estimates \( \hat{\theta}_X \text{MMSE}, \alpha \) and \( \hat{\theta}_X \text{MAP}, \alpha \), the ratio of relative deviation of the MMSE estimates and true parameter values of both dynamic models \( \frac{\hat{\theta}_X \text{MMSE}(I,II)}{\theta_\text{true}} - 1 \) is depicted for several intervals by means of stacked bars in corresponding colors (blue for dynamic model I, red for II). Initially assigned a uniform distribution at \( N_0 = 0 \), both parameter estimates vary noticeably at the beginning for both cases in Figure 4 (a) and (b). Where the MAP estimates remain almost constant for the first half of the cycles, the MMSE estimates diverge from their initial values towards the MAP estimates. In the latter half both parameter estimates approach in conjunction the actual underlying parameter values, albeit, not exhibiting convergence in the sense of an asymptotic behaviour. As pointed out in [2], different reasons can account for this shortcoming which encompass the variability of the material-dependent constants of the respective dynamic model and the possible inability of the dynamic model to cover specific physical behaviour. However, by introducing the dynamic model parameters \( \theta_1 \) and \( \theta_2 \) as additional random variables that are updated from step to step and therefore adjustable to the true underlying parameters instead of assigning them constant values, the asymptotic behaviour of the measurement model parameter estimates is not improved. This can partially be explained by the additional uncertainty incorporated into the SSM without compensation by more observations, i.e. a higher measuring rate. Likewise, the introduction
of a further dynamic model that better accounts for the underlying physics yields no improved convergence since yet another parameter \( \theta_5 \) is required to be modelled as random variable. An additional explanation stems from the inverse problem formulation itself: Considering the exponential crack growth characterised through the dynamic models, the inverse problem is not well-posed for the majority of the cycles where the crack growth is almost linear over time, yielding not a unique solution but various. This implies that the inference results improve the further the crack propagates and the more exponential the crack growth over cycles becomes. Beyond the question of asymptotic behaviour, it is apparent that the MMSE estimates reveal a higher accuracy than the corresponding MAP estimates. Furthermore, the aforementioned relative error bars show lower deviations of the MMSE estimates and true parameter values in the case of dynamic model \( \text{II} \) over \( \text{I} \) which is purchased by higher computational complexity.

The actual fatigue crack growth rate (reference data) is plotted over the stress intensity factor range for all specimens in Figure 5 in logarithmic scales. As elaborated on in Section 3.2, the two stages of test initiation and stable crack growth can be identified easily in Figure 5 (a). It is however apparent, that there are individual outliers especially in the initiation stage as well as deviations from the power growth behaviour in the stage of stable crack growth. In Figure 5 (b), the same plot is displayed in conjunction with both utilised dynamic models, the modified Paris’ laws from Walker in Eq. (35) (solid lines) and from Forman/Mettu in Eq. (37) (dashed lines) where the MMSE estimates obtained at time step \( k = n_{\text{ms}} \) are utilised as realisations of the parameters. Naturally, Walker’s law allows only for the depiction of stable crack growth where the modified Paris’ law by Forman/Mettu is able to cover the test-initiation phase of the entire crack propagation process as well. Nonetheless, the graphs of both dynamic models coincide for the phase of stable crack growth almost exactly, meaning the estimates for \( \theta_1 \) and \( \theta_2 \) do so as well. The MMSE parameter estimates \( \hat{\theta}_1^{\text{MMSE}} \) and \( \hat{\theta}_2^{\text{MMSE}} \) used in both dynamic models allow for a very good coverage of the crack growth rate in stage II of specimens \#1, \#2 and \#6 whereas only the latter measurements of specimens \#3 and \#4 show good compliance. Additionally, the test initiation phase can well be displayed using the obtained MMSE parameter
estimates $\hat{\theta}_\alpha^{MMSE(I)}$ for $\alpha = 1, 2, 5$ in the Forman/Mettu modification for specimens #3, #4 and #6. With the estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ almost coinciding, the main difference in the usage of both dynamic models can therefore be ascribed to covering test-initiation as expected.

In Figure 6, the normalised filtering (a) and smoothing (b) estimates of the crack size are plotted over the normalised cycles for dynamic model functions $f^{(I,II)}(x_k, \theta)$ for all six specimens. Again, the ratio of relative deviation of the MMSE estimates and true crack size of both dynamic models $|\hat{x}_k^{MMSE(I,II)} / x_{k, true} - 1|$ and $|\hat{x}_k^{s,MMSE(I,II)} / x_{k, true} - 1|$ is depicted for several intervals by means of stacked bars in corresponding colors (blue for dynamic model I, red for II). The distinction of the plots in Figure 6 (a) and (b) lies in the conditioning of the respective estimates: In (a), the crack size estimates $\hat{x}_k$ are conditioned on all measurements $y_1:k$ up to time step $k$. The estimates $\hat{x}_k$ therefore correspond to the information available at time step $k$ throughout the testing. In (b), the crack size estimates $\hat{x}_k^s$ are - at every time step $k$ - conditioned on all measurements $y_{1:n_{ms}}$ which represents a refinement of the previously inferred estimates subsequent to the test. Both plots emphasise the better results that can be obtained via the MMSE estimates in comparison with the MAP estimates. Furthermore, the estimates inferred by utilisation of dynamic model II display a better performance in qualitative (graphs) as in quantitative (deviation error bars) regard. Considering that the absolute deviations decrease over the cycles when the relative errors remain constant (due to the normalising with the true crack size $x_{k, true}$ at the specific cycle $k$) and taking into account that the crack size estimates approach the true values at the end of the tests, accurate results by the Bayesian approach can be observed. Considering the filtering estimates in Figure 6 (a), it is however apparent that, as with the estimates of $\theta_3$ and $\theta_4$ in Figure 4, no convergence is discernible. As mentioned before, this shortcoming might be circumvented by a higher measuring rate in general and especially a higher weighting of and increased emphasis on the latter part of the stable crack growth stage. In doing so, the computational expense must be kept in mind to allow for a compromise between accuracy (number of measurements, number of grid points) and timely availability in real-time crack size assessment.
Figure 6: Normalised crack size estimates over normalised cycles $N_k$ for dynamic models I (blue) and II (red) for all specimens as well as ratio of absolute relative deviation of MMSE estimates $\hat{x}_k^{\text{MMSE}}$ and $\hat{x}_k^{\text{MAP}}$ from true crack size (stacked bars in relative scale in corresponding colors).

5 CONCLUSION

A Bayesian approach to the challenge of model calibration in DCPD measuring in conjunction with unknown material-dependent parameter estimation has been presented. Fatigue-tested corner crack specimens are monitored by means of the DCPD method without further knowledge of the calibration curve and the parameters governing the utilised physical models. The obtained potential drop measurements are then used to infer the unknown quantities of crack size, calibration curve and dynamic model parameters and compared with reference data.

1. The proposed methodology is found to provide accurate estimates for both model calibration and parameter estimation while allowing for an accurate crack size growth monitoring, even though no clear convergence of parameter and crack size estimates but only the approach of the respective true underlying values is observable.

2. The utilisation of the modified Forman/Mettu law allows for visibly better estimates in comparison with Walker’s law, implying that the test-initiation phase is to be depicted by the dynamic model as precise as possible. However, this presupposes the introduction of additional parameters as random variables, thereby increasing the computational complexity.

3. For the calibration curve, linearity is assumed. More complex relations between potential drop and crack size are conceivable, thereby allowing for other crack propagation geometries as well, albeit resulting in a higher dimensionality of the parameter vector and consequently the computational complexity.

4. The convergence behaviour mentioned in Enumeration 1. should be addressed with an increased measuring rate especially in the latter part of the stable crack growth stage, where the non-linearity of the crack propagation intensifies.
REFERENCES


PARAMETER IDENTIFIABILITY THROUGH INFORMATION THEORY

G. Capellari\textsuperscript{1}, E. Chatzi\textsuperscript{2}, and S. Mariani\textsuperscript{1}

\textsuperscript{1} Politecnico di Milano, Dipartimento di Ingegneria Civile e Ambientale
Piazza Leonardo da Vinci, 32, 20133 Milano (IT)
e-mail: \{giovanni.capellari, stefano.mariani\}@polimi.it

\textsuperscript{2} ETH Zürich, Institut für Baustatik und Konstruktion
Stefano-Franscini-Platz, 5, 8093 Zürich (CH)
e-mail: chatzi@ibk.baug.ethz.ch

Keywords: Identifiability, Information Theory, Parameter Estimation, Uncertainty Quantification.

Abstract. In this paper, we address the problem of assessing the identifiability of model parameters in a mechanical system, i.e., whether unknown parameters can be estimated given a set of measurements collected through sensor networks. Practical identifiability can arise due to either a lack of sensitivity or a joint effect of the parameters on the measurements. Information theory can be used to detect the sources of non-identifiability, with the purpose of establishing an efficient sensor network design. Mutual Information between the parameter and the measured outputs, and Conditional Mutual Information of each parameter couple, conditioned on the measurements, are considered. Adoption of these indices is overviewed for practically assessing the identifiability of the mechanical properties of a non-linear structural model.
1 INTRODUCTION

In designing structural health monitoring (SHM) systems, the identifiability and observability of the quantities to be estimated ought to be taken into account [1]. Any set of model parameters is deemed as identifiable if it is possible to uniquely estimate this on the basis of the measured data. The concept of identifiability in mechanical systems has been defined in [2] and [3]. While model identifiability is related to the model mathematical structure only, practical identifiability is referred to the relation between measurements and estimated parameters. Therefore, unlike its purely theoretical counterpart, practical identifiability does take into account uncertainties in the considered variables. In [4], a thorough review of existing methods to study structural identifiability and observability for nonlinear structural models is provided. Practical identifiability is usually addressed through the Fisher information matrix, i.e., the sensitivity of the measured quantities with respect to the parameters [5]. These methods allow to study the identifiability of a certain set of parameters, but they do not highlight the relations among the parameters and the link between each of them and the measured data. By means of such methods, non-identifiability may basically be attributed to either the lack of sensitivity of measured quantities or the compensation of the effects of the parameters on the model response. Information theory has been employed in [6] and [7] to study the relations among parameters; in [8], the concepts of stability and observability have been highlighted within an information theory approach. In this paper, we propose to employ mutual information and conditional mutual information to address the two aforementioned causes of non-identifiability.

In the remainder of this paper, a brief review of information measures and associated definitions is first provided (Section 2.1). Then, the identifiability problem is discussed (Section 2.2) and the employed numerical methods are described (Section 2.3). Finally, the application of the mentioned methods to a non-linear system is presented (Section 3).

2 IDENTIFIABILITY AND INFORMATION THEORY

2.1 Preliminary definitions

In the present section, only the main definitions of the information measures that will be used hereinafter are briefly revisited.

The Mutual Information (MI) between two random variables $X$ and $Y$ is defined as:

$$I(X; Y) = \int_X \int_Y p(x, y) \log \left[ \frac{p(x, y)}{p(x)p(y)} \right] dx dy$$

(1)

where $p(x, y)$ is the joint probability distribution function, while $p(x)$ and $p(y)$ denote marginals. The mutual information may be also interpreted as the Kullback-Leibler divergence (expressed as $D_{KL}[-| -]$) of the product of the marginal distributions from the joint distributions:

$$I(X; Y) = D_{KL}[p(x, y)||p(x)p(y)]$$

(2)

In this sense, the MI gives a measure of the difference (or similarity) in information between the joint probability distribution and the marginals, i.e., the degree of correlation between $X$ and $Y$. If $X$ and $Y$ are independent, then $p(x, y) = p(x)p(y)$ and therefore $I(X; Y) = 0$.

A slightly different interpretation relies on the definition of Shannon entropy, according to:

$$I(X; Y) = H(X) - H(X|Y)$$

(3)

where $H(X)$ is the Shannon entropy of $p(x)$ and $H(X|Y)$ is the conditional Shannon entropy of $X$, conditioned on $Y$. Within the Bayesian inference framework, by considering $X$ as the
Let us consider now three random variables $X$, $Y$ and $Z$. The Conditional Mutual Information (CMI) is defined as follows:

$$I(X; Y|Z) = \mathbb{E}_Z[I(X; Y)|Z] = \int_Z p(z) \int_X \int_Y p(x, y|z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} \, dx \, dy \, dz \quad (4)$$

where $\mathbb{E}_Z[\cdot]$ is the conditional expectation with respect to the variable $Z$. The CMI is the variant of the MI when the aforementioned probability distributions are conditioned with respect to the additional random variable $Z$. Analogously to Eq. (2), the CMI can be written as:

$$I(X; Y|Z) = \mathbb{E}_Z[D_{KL}[p(x, y|z)||p(x|z)p(y|z)]] \quad (5)$$

The interaction information is instead defined as:

$$I(X; Y; Z) = I(X; Y) - I(X; Y|Z) \quad (6)$$

Recalling the previous definitions of MI (Eq. (1)) and CMI (Eq. (4)), the interaction information represents the difference between the information shared by $X$ and $Y$ when $Z$ is given, and the same quantity when $Z$ is not given. It can be proved that while $I(X; Y)$ and $I(X; Y|Z)$ are both strictly non-negative quantities, $I(X; Y; Z)$ may assume any value. If $I(X; Y; Z) > 0$, then the random variable $Z$ enhances the correlation between $X$ and $Y$ (redundancy), otherwise it reduces it (synergy). For further details on the interpretation of interaction information, the reader may refer to [9].

### 2.2 Identifiability

Let us consider a structural system, and assume we want to assess its relative mechanical properties through measured data, as obtained from a set of $n_{sens}$ sensors. The vectorial random variable $\mathbf{y} \in \mathbb{R}^{n_{sens}}$ is referred to the measurements, while $\Theta \in \mathbb{R}^{n_{\Theta}}$ represents the $n_{\Theta}$ parameters to be estimated. The relation between parameters and measurements is assumed to be described by the model $\mathcal{M} : \mathbb{R}^{n_{\Theta}} \times \mathbb{R}^{n_{dof}} \rightarrow \mathbb{R}^{n_{sens}}$ as follows:

$$\mathbf{y} = \mathcal{M}(\theta, \mathbf{f}) + \mathbf{\epsilon} \quad (7)$$

where $\mathbf{f} \in \mathbb{R}^{n_{dof}}$ are the model inputs, namely the loads applied on the structure; $\mathbf{\epsilon} \in \mathbb{R}^{n_{sens}}$ represents the measurement error and is relevant to the sensors accuracy. We assume $\mathbf{\epsilon}$ to be a zero-mean Gaussian noise, sampled from the probability density function $p(\mathbf{\epsilon}) = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is the relative covariance matrix.

Model identifiability for structural model updating has been defined in [3]: a set of parameters are said to be identifiable if they can be uniquely determined by the input-output data. In the Bayesian inference framework, this statement holds if the posterior probability distribution $p(\theta|\mathbf{y})$ presents a finite number of maxima in the $\Theta$ space [2]. The parameters are: globally identifiable if there is a unique maximum point $\theta^*$, locally identifiable if there is more than one maximum, and non-identifiable if there are infinite maxima.

Most of the methods to study practical identifiability are based on measures of the local curvature of the likelihood function in the neighbourhood of $\theta^*$, e.g., Hessian or Fisher information.
Compensation of a parameter by others (also known as collinearity): this happens whenever some parameters have the same effect on the model response. In this case, the parameters are somehow redundant and therefore they may not be separately estimated.

Lack of sensitivity of the measurements to the parameters. For instance, this can occur if the model input $f$ is such that the measurements do not depend on the parameters, i.e., the parameters are not "activated".

In order to address (a), we can point out that any couple of parameters $\{\Theta^i, \Theta^j\}$ can be considered as a common influence of the measurement $Y$ if they are highly correlated, having fixed $y$. Therefore, the higher the difference in information between the conditional joint probability distribution $p(\theta^i, \theta^j|y)$ and the product of the conditional marginal distributions $p(\theta^i|y)p(\theta^j|y)$, the harder it is to estimate these parameters separately. As suggested in [7], a natural measure of the aforementioned occurrence is the CMI $I(\Theta^i; \Theta^j|Y)$ between any parameter couple $\{\Theta^i, \Theta^j\}$: the higher this is, the more correlated the parameters are, given the measurements, and therefore the less they are identifiable together.

On the other hand, the interaction information $I(\Theta^i; \Theta^j; Y)$ can be employed to assess the extent of correlation between the parameters which may be attributed to the measurements, or, in other words, if the correlation increases or decreases as the model response is measured. In our applications, we assume that the parameters are not correlated prior to acquiring the measurements. Therefore, the prior joint probability distribution is $p(\theta^i, \theta^j) = p(\theta^i)p(\theta^j)$ and the MI is $I(\Theta^i; \Theta^j) = 0$. From Eq. (6), it turns out that $I(\Theta^i; \Theta^j; Y) = -I(\Theta^i; \Theta^j|Y)$ and we can conclude that, in this special case, the interaction information yields the same results as the MI. Nevertheless, for a general case, it cannot be used to assess identifiability.

Problem (b) is addressed by exploiting $I(\Theta^i; Y)$, i.e., the MI between each parameter and the measurements. If $I(\Theta^i; Y) = 0$, then $Y$ does not depend on $\Theta^i$, and therefore the latter cannot be identified. The dependency of the measurements on the parameters to be estimated can be affected by several factors, such as the model input, the sensor placement and the type of physical quantities to be measured. Regarding the sensor placement, in [5, 12] a method to optimally place them is presented, by maximizing the relative information content with respect to the parameters to be estimated and, hence, reduce problems of non-identifiability.

It is important to underline that the method places no assumptions on the model linearity or on the types of prior distributions.

### 2.3 Numerical solution

If the probability distribution functions defined in the previous sections are not available analytically, the CMI $I(\Theta^i; \Theta^j|Y)$ and the MI $I(\Theta^i; Y)$ cannot be computed analytically from Eqs. (4) and (1), thereby creating the need for an numerical approach. There are essentially three classes of methods for evaluating the MI: methods based on Monte Carlo approximations [13]; Kernel Density Estimation (KDE) based methods [14]; and k-Nearest Neighbors (kNN) based approaches [15].

Here, due to its ease of implementation, we employ the Gaussian KDE method proposed in [16].
The kernel density estimator \( \hat{p}(x) \) of the probability density function \( p(x) \) is defined as:

\[
\hat{p}(x) = \frac{1}{Nh^d} \sum_{i=1}^{N} K \left( \frac{||x - x_i||}{h} \right)
\]

where \( K(\cdot) \) is the Gaussian kernel, \( d \) is the dimension of the random variable \( X \), \( h \) is the kernel bandwidth and \( N \) is the sample size. The estimated MI reads:

\[
\hat{I}(X; Y) = \frac{1}{N} \sum_{i=1}^{N} \log \frac{\hat{p}(x_i, y_i)}{\hat{p}(x_i)\hat{p}(y_i)}
\]

where \( \hat{p}(x_i, y_i), \hat{p}(x_i) \) and \( \hat{p}(y_i) \) are computed according to Eq. (8). In order to reduce the allocated memory required for the computation, an ensemble estimator has been used, as suggested in [17]. The \( N \) samples are thus divided into \( M \) groups and the MI is simply computed as:

\[
\hat{I}(X; Y) = \frac{1}{M} \sum_{i=1}^{M} \hat{I}_i(X; Y)
\]

where \( \hat{I}_i(X; Y) \) is the MI estimation related to the \( i \)-th group of samples.

In [13], [12], MI has been computed through a Monte Carlo approximation of Eq. (1). Despite a faster convergence rate, the latter approach is practically unsuitable for the computation of the CMI, because of the high computational cost of the multi-dimensional numerical integration. For this reason, the computation of the CMI is performed through the same KDE approach adopted for the MI.

3 APPLICATION TO A STRUCTURAL MODEL

In this section a simple example is chosen so that the practical non-identifiability issues described in Section 2.2 are clearly manifested from the model and, hence, the working of the

Figure 1: (a) Shear-type 8-stories building [18]; (b) Relation between non-linear inter-storey drift and shear force, as defined in Eq. (11).
The approach discussed in the previous section is now applied to a shear-type 8-storey building model (Figure 1a). We assume the flexural rigidity of all the horizontal members to be much higher than that of the column elements, so that the only relevant degrees of freedom are the horizontal displacements of each floor. According to this assumption, the floor stiffness at the $i$-th storey is 
\[ k = \frac{c_{12} E_i I_i}{h^3}, \]
where $c$ is the number of columns per floor, $E_i$ is the material elastic modulus, $I_i$ is the moment of inertia in the storey columns and $h$ the floor height.

The inter-story drifts $\Delta u_i$ depend on the shear force $S_i$ according to the following relation (Figure 1b):

\[ \Delta u_i = \begin{cases} 
\frac{h^3}{c_{12} E_i h^3} S_i & \text{if } S_i < S_i^* \\
\frac{h^3}{c_{12} E_i h^3} S_i^* + \frac{h^3}{c_{12} E_i h^3} (S_i - S_i^*) & \text{if } S_i \geq S_i^*
\end{cases} \]  

where we have assumed a simple bi-linear rule for the modulus $E_i$:

\[ E_i = \begin{cases} 
E_i^e & \text{if } S_i < S_i^* \\
E_i^t & \text{if } S_i \geq S_i^*
\end{cases} \]  

where $E_i^e$ is the elastic modulus, $E_i^t < E_i^e$ is the tangent modulus and $S_i^*$ is the shear yield capacity.

We assume availability of displacement measurements $u_8$ at the top floor (the corresponding random variable is named as $Y$) and the aim is to study the identifiability of the parameters relevant to the first two floors, i.e. $\Theta = [E_1^e; E_1^t; I_1; E_2^e; E_2^t; I_2]$. The parameters are assumed as

<table>
<thead>
<tr>
<th>$S_i &lt; S_i^*$</th>
<th>$I(E_1^e; Y)$</th>
<th>$I(E_1^t; Y)$</th>
<th>$I(I_1; Y)$</th>
<th>$I(E_2^e; Y)$</th>
<th>$I(E_2^t; Y)$</th>
<th>$I(I_2; Y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2280</td>
<td>0.0419</td>
<td>0.2256</td>
<td>0.2240</td>
<td>0.0427</td>
<td>0.2233</td>
<td></td>
</tr>
<tr>
<td>0.1040</td>
<td>0.1336</td>
<td>0.2238</td>
<td>0.1052</td>
<td>0.1353</td>
<td>0.2239</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Mutual Information $I(\Theta_i; Y)$ of each parameter in $\Theta = [E_1^e; E_1^t; I_1; E_2^e; E_2^t; I_2]$ and the measured top-floor displacement $Y$, considering the cases $S_i < S_i^*$ and $S_i > S_i^*$. 

The method may be checked and validated.

We assume availability of displacement measurements $u_8$ at the top floor (the corresponding random variable is named as $Y$) and the aim is to study the identifiability of the parameters relevant to the first two floors, i.e. $\Theta = [E_1^e; E_1^t; I_1; E_2^e; E_2^t; I_2]$. The parameters are assumed as
uniformly distributed.

First, we consider the case in which $S_i < S_i^*$ $\forall i$: from Eq. (11), we can point out that the displacements and, hence, the measurement, do not depend on $E_t^i$. The non-identifiability of $E_t^i$ shows up in the values of the MI $I(\Theta; Y)$ in Table 1: $I(E_t^1; Y)$ and $I(E_t^2; Y)$ are one order of magnitude lower than the other MI value. The MI is not exactly zero as one may expect from the definition in Eq. (1), because of the round-off error of the KDE method. The measured displacement depends only on the flexural stiffness $E_e^i I_i$: these two parameters cannot be estimated separately, since they offer a joint influence to the model response. This is highlighted by the CMI values $I(\Theta_i ; \Theta_j | Y)$ reported in Figure 2a: as expected, the maximum values are reached for the couples $\{E_e^1, I_1\}$ and $\{E_e^2, I_2\}$.

On the other hand, if $S_i > S_i^*$ $\forall i$, the non-linear mechanical behaviour described in Eq. (11) affects the solution. As noted from Table 1, there are no parameters for which $I(\Theta_i ; Y) \approx 0$. However, from Eq. (11), we can point out that the model responses $\Delta u_i$ depends with the same relationship on $E_e^i / E_t^i$ and $I_i$, and therefore this prevents identifiability. The related CMI values stem from the latter fact: in Figure 2b, $I(\Theta_i ; \Theta_j | Y)$ is maximum for the couples $\{E_e^1, I_1\}$ and $\{E_e^2, I_2\}$. Moreover, $I(E_e^1 / E_t^1; I_1, I_2 | Y) > I(E_e^2 / E_t^2; I_1, I_2 | Y)$ as, since $E_t^1 I_1 < E_e^1 I_1$, the resulting displacement is heavily dependent on $E_e^i$. The same fact can be underlined in Table 1 as $I(E_e^i; Y) < I(E_t^i; Y)$.

4 CONCLUSIONS

In the present paper, we have addressed the problem of parameter-identifiability in mechanical systems. The non-identifiability of parameters can arise due to either lack of sensitivity or compensations in the dependency of measurements on the parameters. Within an information theoretic approach, identifiability is here detected in terms of these two occurrences using the mutual information between each parameter and the measurements, and the conditional mutual information between each couple of parameters, conditioned on the measurements.

The methodology has been applied to a non-linear mechanical model, namely a shear-type 8-storeys building model, where the former causes of non-identifiability are easily recognizable. The MI and the CMI proved able to detect and quantify identifiability, and to reveal the relations between parameters. Moreover, no unnecessary assumptions on the model linearity or on distributions gaussianity were placed.

Acknowledgements

G. C. wishes to acknowledge the partial financial support by IDEA League through a Student Grant.

REFERENCES


INFEREN CE AND SENSITIVITY IN STOCHASTIC WIND POWER FORECAST MODELS.

Soumaya Elkantassi¹, Evangelia Kalli giannaki¹ and Raul Tempone¹

¹Applied Mathematics and Computational Science, CEMSE, King Abdullah University of Science & Technology
Thuwal 23955, Saudi Arabia
e-mail: {soumaya.kantassi, evangelia.kalli giannaki, raul.tempone}@kaust.edu.sa

Keywords: Indirect inference, wind power, probabilistic forecasting, model selection, sensitivity.

Abstract. Reliable forecasting of wind power generation is crucial to optimal control of costs in generation of electricity with respect to the electricity demand. Here, we propose and analyze stochastic wind power forecast models described by parametrized stochastic differential equations, which introduce appropriate fluctuations in numerical forecast outputs. We use an approximate maximum likelihood method to infer the model parameters taking into account the time correlated sets of data. Furthermore, we study the validity and sensitivity of the parameters for each model. We applied our models to Uruguayan wind power production as determined by historical data and corresponding numerical forecasts for the period of March 1 to May 31, 2016.
1 Introduction

Over the past 10 years, Uruguay has seen a spectacular transformation in production of wind-based power. In 2007, Uruguay did not produce energy from wind; by 2016, the installed capacity for wind power generation was 1210 MWatts satisfying more than 12% of the total electricity demand of the country. Moreover, renewable energy production in Uruguay provides 94.5% of the country’s electricity and 55% of the country’s total energy mix. To incorporate the increased amount of generated power into the supply system, reliable forecasting models are crucial.

Many statistical forecast methods have been developed, for example persistence models [1], Kalman filters [4], and Autoregressive and Moving Average Model (ARMA) [7]. The main feature that distinguishes the forecast methods, and the corresponding time-horizon they are valid, is the purpose of its use (operation scheduling, electrical grid management, maintenance). Stochastic forecast models are adequate for relative long-term applications because of their dependence on Numerical Weather Prediction (NWP) models. They consider the uncertainty of the deterministic forecast by using historical power production data. Probabilistic forecast models can also be used for stochastic optimization problems for optimizing the wind farms operation and energy trading. Taking into account the production costs of each energy source (fossil fuel, thermal hydropower, biomass, wind) and the storage capacity, such a model makes it possible for decision makers to control electricity costs in the market.

Recently, a probabilistic forecast model based on stochastic differential equations (SDEs), [7] was introduced. Our goal is to extend the previous work to include a parametric form for the drift coefficient that controls the quality of the forecast, to explore the need for Lampertti transformation in [7] by considering inferences directly from state-dependent diffusion, and to, find the best model fit by pursuing variability analysis and examining sensitivity of the models parameters.

In Section 2 we formulate the SDEs and propose two models. In Section 3 we describe the available datasets and the preprocessing procedure. Then, numerical results for each model are presented in Section 4. We conclude in Section 5 by discussing our results.

2 Indirect Inference method with Stochastic Differential Equations

In this section, we introduce a stochastic process, $X(t) \in [0, 1]$, that serves as the stochastic analogue of the normalized deterministic forecast, $p(t) \in [0, 1]$, for $t \in [0, T]$. We consider that the stochastic process, $X(t)$, is the solution the following parametrized Ito diffusion SDE

$$\begin{cases} 
  dX(t) = b(X(t), t; \theta)dt + \sigma(X(t), t; \theta)dW(t), & t > 0, \\
  X(0) = x_0, 
\end{cases}$$

and assume that the drift, $b(X(t), t; \theta)$, and the diffusion, $\sigma(X(t), t; \theta)$, coefficients are such that a solution exists and is unique, [2]. Here, $\theta$ denotes the set of parameters for each model. $W(t)$ is a one-dimensional Brownian motion. As we expect the process, $X(t)$, to align with the numerical forecast, $p(t)$, we choose the drift to be a linear polynomial of the state

$$b(x, t; \theta) = -\theta(t)(x - p(t)),$$

where $\theta(t)$ represents the rate by which the variable reverts to $p(t)$ in time.  

1The energy mix statistics are sourced from the Global Wind Energy Council and the World Resources Institute.
We distinguish two major models (Model 1 and Model 2) based on the choice of the diffusion coefficient. The purpose of this distinction is to explore the need for the transformation to have a state independent diffusion coefficient.

**Model 1**

We want the diffusion at the boundaries to vanish (i.e., if \( X(t) \in \{0, 1\} \), then \( \sigma(0) = \sigma(1) = 0 \)) such that the stochastic forecast does not exceed the total capacity. Thus, we choose the diffusion to be a second order polynomial with respect to the state of the process of the form

\[
\sigma(X(t), t; \theta) = \sqrt{2\theta(t)\alpha X(t)(1 - X(t))}.
\]

The first proposed model–Model 1–is

\[
\begin{align*}
\frac{dX(t)}{dt} &= -\theta(t)(X(t) - p(t)) + \sqrt{2\alpha\theta(t)X(t)(1 - X(t))} dW_t, \quad t > 0, \\
X(0) &= x_0,
\end{align*}
\]

and

\[
\theta = (\theta(t), \alpha), \quad \alpha > 0, \theta(t) > 0, \quad \text{for all} \ t > 0.
\]

The inference approach for the parameters is based on the two moment equations for the stochastic process, \( X(t) \). For that, we denote \( \mu_X(t) = \mathbb{E}[X(t)], v_X(t) = \text{var}[X(t)] = \mathbb{E}[(X(t) - \mu_X(t))^2] \) and \( v_X(t, s) = \text{cov}(X(t), X(s)) = \mathbb{E}[(X(t) - \mu_X(t))(X(s) - \mu_X(s))] \) as the mean, variance and co-variance functions respectively. The equation of the mean is

\[
\begin{align*}
\frac{d\mu_X(t)}{dt} &= -\theta(t)(\mu_X(t) - p(t)) \quad t > 0, \\
\mu_X(0) &= \mu_0 = x_0,
\end{align*}
\]

with the solution

\[
\mu_X(t) = e^{-\int_0^t \theta(s)ds} \left( \int_0^t \theta(s)p(s)e^{\int_0^s \theta(u)du} ds + \mu_0 \right).
\]  

The equation of the variance is

\[
\begin{align*}
\frac{dv_X(t)}{dt} &= -2\theta(t)((1 + \alpha)v_X(t) - \alpha\mu_X(t)(1 - \mu_X(t))) \quad t > 0, \\
v_X(0) &= 0,
\end{align*}
\]

from which we have that

\[
v_X^\theta(t) = e^{-2(1+\alpha)\int_0^t \theta(s)ds} \left( \int_0^t 2\theta(s)\alpha\mu_X(s)(1 - \mu_X(s)) e^{2(1+\alpha)\int_0^s \theta(u)du} ds \right).
\]  

The co-variance between two times of \( X(t) \) is given by

\[
v_X^\theta(t, s) = v_X^\theta(s)e^{-\int_0^s \theta(u)du}, \quad \forall t > s.
\]
Model 2

In this section, we transform equation (2) by means of the Lamperti transform such that the diffusion term of the transformed system is independent of the state. This transformation is recommended in both simulation and estimation, [9]. Numerical schemes sampled from an SDE with additive noise are more stable, and usually of higher order than those sampled from an SDE with multiplicative noise. Moreover, [10] and [7] suggested that the transformed process is better approximated by a Gaussian distribution.

Let us denote by

\[ Z(t) = \Psi(X(t)) = \int \frac{1}{\sqrt{x(1-x)}} dx \bigg|_{x=X_t} \]

the transformed process that satisfies an SDE with state-independent diffusion. For the definition of the Lamperti transform see [9]. To ensure the uniqueness of the solution, \( \mu_Z(t) \), for the mean equation of the transformed process, \( Z(t) \), the following condition should be satisfied

\[ \alpha \leq \min(p(t), 1 - p(t)) . \]

One way to satisfy this condition is by choosing the parameter \( \alpha \)

\[ \alpha(t) = \alpha p(t)(1 - p(t)) . \]

Here we denote the multiplicative constant with the same letter as in the function \( \alpha(t) \), because in the sequel we will refer only to the constant \( \alpha \). This suggests the form of the Model 2

\[
\begin{cases}
 dY(t) = -\theta \left( Y(t) - p(t) \right) dt + \sqrt{2 \theta \alpha p(t)(1 - p(t))} Y(t) \left( 1 - Y(t) \right) dW_t, & t > 0, \\
 Y(0) = x_0,
\end{cases}
\]

and

\[ \theta = (\theta(t), \alpha), \quad \alpha > 0, \theta(t) > 0, \quad \text{for all } t > 0. \]

The Lamperti transformation of the process \( Y(t) \) defines the process

\[ Z(t) = \Psi(Y(t)) = \int \frac{1}{\sqrt{x(1-x)}} dx \bigg|_{x=Y(t)} = \arcsin(2Y(t) - 1), \]

which satisfies the following SDE

\[ dZ(t) = \tilde{b}(Z(t); \theta) dt + \tilde{\sigma}(t; \theta) dW_t , \]

where

\[ \tilde{b}(x; \theta) = \frac{-\theta(1 + \sin(x) - 2p(t)) + \frac{1}{2} \sin(x) \tilde{\sigma}^2(t; \theta)}{\cos(x)} , \]

and

\[ \tilde{\sigma}(t; \theta) = \sqrt{2 \theta(t) \alpha p(t)(1 - p(t))} . \]

The mean \( \mu_Z(t) = \mathbb{E}[Z(t)] \) of \( Z(t) \) is

\[ \mu_Z(t) = \arcsin \left\{ e^{-\int_0^t b(s) ds} \left( \int_0^t \theta(s)(2p(s) - 1)e^{\int_0^s h(u) du} ds + \sin(Z(0)) \right) \right\} , \]
where \( h(t) = \theta(t)[1 - 2\alpha p(t)(1 - p(t))] \). Using the inverse of \( \Psi \) given by \( Y(t) = \Psi^{-1}(Z(t)) = \frac{1}{2}\left(1 + \sin(Z(t))\right) \), we obtain the following approximation for the mean \( \mu_Y(t) = \mathbb{E}[Y(t)] \) after using the Lamperti transform
\[
\tilde{\mu}_Y(t) = \Psi^{-1}(\mu_Z(t)) = \frac{1}{2}(1 + \sin(\mu_Z(t))).
\] (8)

The equation of the variance is given by
\[
\frac{dv_Z(t)}{dt} = 2A(t; \theta)v_Z(t) + \sigma^2(t; \theta), \quad v_Z(0) = 0,
\]
and an exact solution for this first-order ordinary differential equation (ODE) is given by
\[
v_Z(t) = e^{\int_0^t A(s; \theta)ds} \int_0^t 2\frac{1}{2}\sigma^2(t; \theta)e^{-\int_s^t A(u; \theta)du} ds\]
where \( A(t; \theta) = \partial_x \tilde{b}(x; \theta)|_{x=\mu_Z(t)} \). The covariance function is
\[
v_Z(t, s) = v_Z(s)e^{\int_s^t A(u; \theta)du}, \quad \text{for all } t > s.
\]
Thus, the approximate variance and covariance functions for \( Y(t) \) are
\[
\tilde{\sigma}_Y(t) = \frac{1}{4} \cos^2(\mu_z(t))v_Z(t),
\] (9)
and
\[
\tilde{\sigma}_Y(t, s) = \frac{1}{4} \cos(\mu_z(t))v_Z(t, s) \cos(\mu_z(s)).
\] (10)

### 2.1 Approximate Likelihood

We assume that the observation variables \( D^{(j)} = (D^{(j)}(t_1), \ldots, D^{(j)}(t_N)) \), \( j = 1, \ldots, J \), satisfy
\[
D^{(j)} = X^{(j)} + e^{(j)}
\] (11)
where \( X^{(j)} = (X^{(j)}(t_1), \ldots, X^{(j)}(t_N)) \) is the solution of the SDE in equations (2) or (6) for \( p(t) = p^{(j)}(t) \), and \( e^{(j)} \) are the measurement errors that are generated from a Gaussian distribution \( \mathcal{N}(0, \Sigma^{(j)}) \).

The approximate likelihood based on the two-moment expansion of the process \( X(t) \) and \( Y(t) \) is
\[
L(\theta; D_{nm}) = \prod_{j=1}^J (2\pi)^{-N/2}|V^{(j)}(\theta)|^{-1/2}\exp\left[\left(\mathbf{d}^{(j)} - \mu^{(j)}(\theta)\right)^T V^{(j)}(\theta)^{-1}\left(\mathbf{d}^{(j)} - \mu^{(j)}(\theta)\right)\right]\]
for a given set of observations \( d_{N,J} = \{d^{(j)} = (d^{(j)}(t_1), \ldots, d^{(j)}(t_N)), j = 1, \ldots, J\} \). Here \( \mu^{(j)}(\theta) = (\mu(t_1), \ldots, \mu(t_N)), V^{(j)}(\theta) = V^{(j)}(\theta) + \Sigma^{(j)}, \) and \( V^{(j)}(\theta) \) is an \( n \times n \) matrix with elements \((V^{(j)}(\theta))_{kl} = v(t_k, t_l), k, l = 1, \ldots, N\). Note that \( \mu(t), v(t) \) and \( v(t, s) \) are given in equations (3) - (5) of Model 1 and in equations (8) - (10) of Model 2. Finally, the maximum likelihood estimates (MLE) of the parameters, \( \theta \), are obtained by maximizing the function in equation (12).
3 Description of the data

Application of our models to the wind power problem is built on two datasets. The first dataset consists of 72 hours of numerical predictions of wind power production and the second dataset includes the actual wind power produced during the same 72 hours (one observation every one hour) in Uruguay. The data correspond to the aggregate power (MWatts) in the following wind farms: ARTI, CAPE, FACE, CPPP, FLO1, JPTE, KENT, LRLM, MIN1, MWIN, RSUR and TDMA. The nominal capacity of this set of parks is 793.7 MWatts. Before proceeding, we normalize each of the datasets with the maximum installed capacity to obtain values between \([0, 1]\). We denote by \(p_j(t_n), d^{(j)}(t_n), j = 1, \ldots, J, J = 81, n = 1, \ldots, 72\) the normalized numerical prediction and historical prediction at the \(j\)th set at time \(t_n\).

![Figure 1: Numerical forecast and real production comparison](image)

In our model, we expect the deterministic forecast to capture some features of the observed real production. Thus, we opt to discard the time series where a big difference between \(x(t)\) and \(p(t)\) (i.e., more than 20%) is observed especially at time \((t = 1)\). We consider them as outliers with respect to our models. Including them affects the quality of the parameter estimates and produces instabilities in the likelihood maximization problem. If we choose a time horizon that is greater than 24 hours, we observe redundancy between successive sets of actual power production, which contradicts the independence assumption while computing the likelihood function. For this reason, some sets are discarded \((J = 73)\).

4 Results

In this section, we present the application of the models described in Section 2 to the datasets described in Section 3. For both models, we consider a parametrization with respect to time-independent parameters. Next, we introduce potential trends of \(\theta\) and select the best model according to the Akaike Information Criterion (AIC) and the Bayesian Information Criterion.
(BIC). Then, to check variability in the estimates from the selected model, we use the bootstrapping method to compute optimal parameters with respect to randomly selected subsets of data and present the bootstrap confidence intervals. To validate the models, we split the data into training and testing sets. We use the training sets for the optimization problem and use the optimal values to generate empirical confidence bands. Finally, we validate the best model by verifying that the test set observations fall within the generated confidence bands. We also present the sensitivity of the model parameters based the Fisher Information Matrix (FIM) for the approximate likelihood.

### 4.1 Proposed models and model selection

The different parametrizations of Model 1 (2) and Model 2 (6) are summarized in Table 4.1.

<table>
<thead>
<tr>
<th>Parametric Model</th>
<th>θ(t)</th>
<th>Number of parameters (θ(t), α, φ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDE0</td>
<td>θ₀</td>
<td>3</td>
</tr>
<tr>
<td>SDE1</td>
<td>θ₁(t) = θ₀e⁻θ₁t</td>
<td>4</td>
</tr>
<tr>
<td>SDE2</td>
<td>θ₂(t) = θ₀e⁻θ₁t + θ₂</td>
<td>5</td>
</tr>
<tr>
<td>SDE3</td>
<td>θ₀e⁻θ₁t + θ₂e⁻θ₂t</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1: Proposed models with parametrizations of the rate θ(t)

We assume that the observed production has non-correlated Gaussian measurement error for each set, j = 1, \ldots, J, ε⁽j⁾ ∼ N(0, φ²I), and we consider φ a parameter of the model.

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Parameters Without noise</th>
<th>φ = 0.05</th>
<th>φ = 0.02</th>
<th>φ = 0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 h</td>
<td>(θ₀, α, φ)</td>
<td>(0.177, 0.082)</td>
<td>(0.256, 0.106, 0.036)</td>
<td>(0.200, 0.094, 0.004)</td>
</tr>
<tr>
<td></td>
<td>standard deviation</td>
<td>10⁻²(0.83, 0.41)</td>
<td>10⁻²(1.75, 0.7, 0.25)</td>
<td>10⁻²(0.91, 0.44, 0.44)</td>
</tr>
<tr>
<td>18 h</td>
<td>(θ₀, α, φ)</td>
<td>(0.140, 0.089, 0)</td>
<td>(0.213, 0.102, 0.038)</td>
<td>(0.163, 0.102, 0.010)</td>
</tr>
<tr>
<td></td>
<td>standard deviation</td>
<td>10⁻²(0.83, 0.41)</td>
<td>10⁻²(1.2, 0.59, 0.19)</td>
<td>10⁻²(0.68, 0.45, 0.19)</td>
</tr>
<tr>
<td>24h</td>
<td>(θ₀, α, φ)</td>
<td>(0.140, 0.087)</td>
<td>(0.196, 0.118, 0.036)</td>
<td>(0.160, 0.098, 0.002)</td>
</tr>
<tr>
<td></td>
<td>standard deviation</td>
<td>10⁻²(0.47, 0.31)</td>
<td>10⁻²(0.96, 0.57, 0.17)</td>
<td>10⁻²(0.57, 0.37, 0.48)</td>
</tr>
<tr>
<td>36h</td>
<td>(θ₀, α, φ)</td>
<td>(0.121, 0.109)</td>
<td>(0.172, 0.147, 0.034)</td>
<td>(0.139, 0.125, 0.004)</td>
</tr>
<tr>
<td></td>
<td>standard deviation</td>
<td>10⁻²(0.47, 0.47)</td>
<td>10⁻²(0.96, 0.57, 0.17)</td>
<td>10⁻²(0.52, 0.51, 0.36)</td>
</tr>
<tr>
<td>48h</td>
<td>(θ₀, α, φ)</td>
<td>(0.114, 0.105)</td>
<td>(0.154, 0.156, 0.035)</td>
<td>(0.126, 0.124, 0.004)</td>
</tr>
<tr>
<td></td>
<td>standard deviation</td>
<td>10⁻²(0.39, 0.39)</td>
<td>10⁻²(0.73, 0.76, 0.17)</td>
<td>10⁻²(0.44, 0.47, 0.32)</td>
</tr>
</tbody>
</table>

Table 2: Model 1. Optimal parameter values for actual data and different time horizons.

In Table 2 we present the optimal values of the parametric model SDE0 (θ, α, φ) for the time horizons \{6, 12, 18, 24, 36, 48\} based on Model 1. In the column termed 'Without noise', we report the output of the minimization of the negative logarithm of the likelihood function (12) using the actual data. For Model 1 is not able to recover the value of φ, and the output is always zero (thus we simply report values of θ and α). To investigate the effect of this observation, we added an artificial Gaussian noise to the data with φ = \{0.05, 0.02, 0.01\}. We can see that the values of θ and α change noticeably and appear to be sensitive to the measurement error. For φ = 0.05, the likelihood is 0.036 while α jumps from 0.082 to 0.106, which suggests that it is difficult for our model to distinguish between the measurement error and the noise coming from the SDE.
through $\alpha$. Part of the noise, $\phi$, is indeed absorbed by $\alpha$, which explains the jump in its value. This observation is confirmed in the last two columns of Table 2, where the likelihood cannot retrieve $\phi$ for noise on the order of 0.02 and 0.01.

We observe that $\theta$ is decreasing with time, which indicates that the predictability of the model becomes less reliable with increasing time horizons. This motivates the use of the exponentially decreasing parametric functions for $\theta$ in Table 4.1, that define the models SDE0, SDE1, SDE2, and SDE3. The models are parametrized by $(\theta_0(t), \alpha, \phi)$ such that $\theta_0(t) = \theta_0$ is a constant rate, $\theta_1(t)$ is a one-term exponential function, $\theta_2(t)$ has an asymptotic lower bound given by $\theta_2$ to avoid converging to zero and $\theta_3(t)$ is a linear combination of exponentially decreasing and increasing functions to better fit the data (see table 4.1).

To obtain good initial values for the optimization for the time varying $\theta(t)$, in SDE1, SDE2, and SDE3 models, we fit the optimal values for the constant $\theta^*$ of the SDE0 model obtained in the first column of Table 2, using the ‘Curve Fitting Toolbox’ of MATLAB. We then use the fitted coefficients of $\{\theta_0, \theta_1, \theta_2, \theta_3\}$ as initial points for the maximum likelihood estimators of $(\theta(t), \alpha, \phi)$. The MLE parameters for all models are presented in Table 3 for the $T = 24$-hour time horizon.

<table>
<thead>
<tr>
<th>Model</th>
<th>Initial values</th>
<th>Optimal values</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDE0</td>
<td>(0.2, 0.4)</td>
<td>(0.140, 0.087)</td>
<td>$10^{-2}(0.47, 0.31)$</td>
</tr>
<tr>
<td>SDE1</td>
<td>(0.202, 0.013, 0.4)</td>
<td>(0.199, 0.034, 0.141)</td>
<td>$10^{-2}(0.68, 0.15, 0.53)$</td>
</tr>
<tr>
<td>SDE2</td>
<td>(0.124, 0.047, 0.099, 0.4)</td>
<td>(0.137, 0.108, 0.087, 0.125)</td>
<td>$10^{-2}(2.09, 1.33, 0.34, 0.46)$</td>
</tr>
<tr>
<td>SDE3</td>
<td>(0.2, 0.01, 2·$10^{-3}$, 0.02, 0.1)</td>
<td>(0.28, 0.103, $10^{-3}$, 0.363, 0.018)</td>
<td>$10^{-2}(2.06, 1.04, 4·10^{-4}, 0.16, 0.05)$</td>
</tr>
</tbody>
</table>

Table 3: Model 1. Optimal parameters for a 24-hour time horizon for the parametric SDE’s in Table 4.1

Among the four candidate models, data in Table 4.1 confirm that SDE1 is the best model as it has the smallest AIC and BIC values.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDE0</td>
<td>3</td>
<td>-8418</td>
<td>-8411</td>
</tr>
<tr>
<td>SDE1</td>
<td>4</td>
<td>-8427</td>
<td>-8418</td>
</tr>
<tr>
<td>SDE2</td>
<td>5</td>
<td>-8426</td>
<td>-8414</td>
</tr>
<tr>
<td>SDE3</td>
<td>6</td>
<td>-8273</td>
<td>-8259</td>
</tr>
</tbody>
</table>

Table 4: Model selection

**Model 2**

Next, we present the best model among those described in Table 4.1 based on Model 2, described in equation (6), following the same steps as for Model 1.
Table 5: Optimal parameter values for actual real data and different time horizons for Model 2 using the Lamperti-transformed process.

From Table 5, we see that \( \theta \) exhibits the same behavior as in Model 1, it decreases with longer time horizons. We also observe that Model 2 is able to capture the measurement error using the same datasets, and the optimal values are consistent with our findings from Model 1.

Table 6: Optimal parameters for 24-hour time horizon using the Lamperti transformed process.

Table 7 confirms that SDE1 is the best model with respect to AIC and BIC.

Table 7: Model selection

4.2 Variability, validation and sensitivity analysis

An important issue to explore is the variability of the datasets. Here, we use the bootstrapping method to select randomly 80% of the data and compute statistics for the parameters from the bootstrapped samples. We compute the sample mean and the bootstrapped confidence intervals of each parameter. Figures 2 and 3 show that our data predict a reliable set of estimates that both models agree on.
Figure 2: Model 1. Bootstrap histograms of $\theta_0$, $\theta_1$ and $\alpha$ based on $n = 100$ iterations.

Figure 3: Model 2. Bootstrap histograms of $\theta_0$, $\theta_1$, $\alpha$ and $\phi$ based on $n = 100$ iterations.

Figure 4 shows that the prediction interval of $\theta(t)$ of Model 1 falls inside the prediction interval of Model 2, suggesting that both predict the same rate.

To validate the selected models we present the empirical path density and confidence bands for two days in the training set, as shown in figures 5-8. We generate the empirical path density from simulated paths of SDE’s (2) and (6) for the MLE optimal parameters for which we use a Euler-Maruyama discretization scheme. On the same graphs, we present the actual wind energy production and the available numerical forecast. The figures verify that the actual production falls into the 95% confidence band for most times.
Next, we analyze the sensitivity of both models to the parameters $(\theta_0, \theta_1, \alpha)$. The sensitivity index is given by the diagonal elements of the FIM that corresponds to the approximate likelihood for each model, [14]. A direct sensitivity analysis of the models based on (2), and (6) is the subject ongoing work, [16]. Figures 9 and 10 demonstrate that the most sensitive parameter for both models is $\theta_1$, which controls the rate of decay of $\theta(t)$ with respect to the time horizon. Note that the least sensitive parameter in the two models is different; for Model 1, it is $\alpha$ while for Model 2, it is $\theta_0$. 

Figure 5: Empirical Confidence Interval (CI) and real wind energy production corresponding to May 16, 2016 (12 hours)

Figure 6: Empirical CI and real wind energy production corresponding to May 22, 2016 (24 hours)

Figure 7: Empirical CI and real wind energy production corresponding to May 16, 2016 (12 hours)

Figure 8: Empirical CI and real wind energy production corresponding to May 22, 2016 (Horizon 24 hours)
Figure 9: Sensitivity of parameters in Model 1. The most sensitive parameter is $\theta_1$ and the less sensitive is $\alpha$.

Figure 10: Sensitivity of parameters in Model 2. The most sensitive parameter is $\theta_1$ and the less sensitive is $\theta_0$, in contrast to Model 1.

5 Discussion and Conclusions

It is important to point out that we based our study on a benchmark example in which we generate synthetic data from real deterministic forecast paths, $p_i(t), i = 1, \ldots, J$ using specific values of $(\theta(t), \alpha, \phi)$. Then, for each $p_i(t)$, we sample one path, $X_i(t)$, to mimic the actual data and check whether the MLE parameters are close to the true parameters. The numerical results of the benchmark example reveal that although both models are able to retrieve the true parameter values, Model 2 is numerically more stable (parameters have smaller confidence intervals, and the rate of convergence is more stable)

With actual data, Model 1 is not able to capture the value of $\phi$ for all $\theta_i(t), i \in \{1, 2, 3\}$, and its parameters are very sensitive to an artificial measurement error. On the other hand, Model 2 provides an estimate of the measurement error and we assume that this is related to the stability of Model 2 due to the use of the Lamperti-transformed process. However, it is remarkable that both models predict the same rate.

The variability analysis verifies that the available datasets are rich enough to predict an informative model. Moreover, the two models predict the same rate, (see Figure 4) and the most sensitive parameter for both models is $\theta_1$, which controls the decay. The least sensitive parameter is different.

To avoid removing datasets that have poor deterministic forecasts or that have extreme observations (i.e., sudden dramatic changes in wind power production that can not be predicted by the deterministic forecast), it may be possible to use regime-switching models.

REFERENCES


FROM ATOMISTIC TO SYSTEMATIC COARSE-GRAINED MODELS FOR MOLECULAR SYSTEMS

Vagelis Harmandaris\textsuperscript{1,2}, Evangelia Kalligiannaki\textsuperscript{3}, Markos Katsoulakis\textsuperscript{3} and Petr Plecháč\textsuperscript{4}

\textsuperscript{1} Department of Mathematics and Applied Mathematics, University of Crete
Heraklion, GR-70013 Crete, Greece
e-mail: name@e-mail.address

\textsuperscript{2} Institute of Applied and Computational Mathematics, Foundation for Research and Technology
Hellas, IACM/FORTH,
Heraklion, GR-70013 Crete, Greece

\textsuperscript{3} Applied Mathematics and Computational Science, CEMSE, King Abdullah University of Science & Technology
Thuwal 23955, Saudi Arabia
e-mail: evangelia.kalligiannaki@kaust.edu.sa

\textsuperscript{4} Department of Mathematics and Statistics, University of Massachusetts at Amherst
Amherst, MA 01003, USA
e-mail: markos@math.umass.edu

\textsuperscript{5} Department of Mathematical Sciences, University of Delaware
Newark, DE 19716, USA
e-mail: plechac@math.udel.edu

Keywords: coarse-graining, potential of mean force, relative entropy, force matching

Abstract. The development of systematic (rigorous) coarse-grained mesoscopic models for complex molecular systems is an intense research area. Here we first give an overview of methods for obtaining optimal parametrized coarse-grained models, starting from detailed atomistic representation for high dimensional molecular systems. Different methods are described based on (a) structural properties (inverse Boltzmann approaches), (b) forces (force matching), and (c) path-space information (relative entropy). Next, we present a detailed investigation concerning the application of these methods in systems under equilibrium and non-equilibrium conditions. Finally, we present results from the application of these methods to model molecular systems.
1 INTRODUCTION

Complex molecular systems characterize materials (e.g. plastics, rubbers, gels) that develop complex, multiphase morphologies through equilibrium self-assembly. Their properties, such as mixing behavior; phase diagrams; melting points, depend on the combination of molecular and macroscopic variables. On the microscopic level, all-atom simulations allow direct quantitative predictions of the properties of molecular systems over a range of length and time scales. However, due to the broad spectrum of characteristic lengths and times involved in complex molecular systems, it is not feasible to apply them to large realistic systems or molecules of complex structure. On the mesoscopic level, coarse-grained (CG) models have proven to be very efficient means in order to increase the length and time scales accessible by simulations.

Here we first give an overview of different algorithms for developing CG models of molecular systems. Methods such as inverse Monte Carlo [1], inverse Boltzmann [2], force matching [3], relative entropy [4], provide parameterizations of coarse-grained effective potentials at equilibrium by minimizing a fitting functional over a parameter space. Then, we further extend these studies using path-space methods (relative entropy rate) for coarse-graining and uncertainty quantification for non-equilibrium processes, [5, 6, 7, 8].

All the above methods mentioned in principle are employed to approximate a many-body potential, the (n-body) potential of mean force, describing the equilibrium distribution of coarse-grained sites observed in simulations of atomically detailed models. We present two main results of our latest studies, [9, 5]. Firstly, on optimizing coarse-grained models in equilibrium, we: (a) reveal the connection of the force matching method with thermodynamic integration. This connection provides us with information on how to construct a local mean force for equilibrium force matching implementations, to best approximate the potential of mean force. (b) We present in a mathematically consistent way the entropy and force matching methods and their equivalence, which we derive for general nonlinear coarse-graining maps, [9]. Secondly, we adopt the use of path-space methods, to define a dynamical analogue of the relative entropy minimization method. We provide a systematic derivation of Langevin type coarse-grained dynamics from fine-scale molecular simulations, based on the minimization of the relative entropy rate. It is shown that this minimization problem is equivalent to a weighted least squares problem, with weights that depend on the diffusion coefficient of the proposed stochastic dynamics for the coarse-grained system. At least for constant diffusion coefficient, it is nothing but the widely applied force matching method used in computational coarse-graining which, however, is restricted to equilibrium processes. Finally, we apply, and compare the above-described methodologies in several molecular systems: gas and fluid methane and water [10].

1.1 Atomistic and Coarse Grained Models

In the following we shortly describe the microscopic (atomistic) and mesoscopic (coarse-grained) representation of a prototypical molecular system. Assume the problem of \( N \) (classical) molecules in a box of volume \( V \) at temperature \( T \). Let \( \mathbf{q} = (q_1, \ldots, q_N) \in \mathbb{R}^{3N} \) describe the position of the \( N \) particles in the atomistic description, with potential energy \( U(\mathbf{q}) \). The probability that the system has a state \( \mathbf{q} \) at the temperature \( T \) is given by the Gibbs canonical measure

\[
\mu(d\mathbf{q}) = Z^{-1} \exp\{-\beta U(\mathbf{q})\} d\mathbf{q},
\]

where \( Z = \int_{\mathbb{R}^{3N}} e^{-\beta U(\mathbf{q})} d\mathbf{q} \) is the partition function, \( \beta = \frac{1}{k_B T} \) and \( k_B \) is the Boltzmann constant. We denote \( f : \mathbb{R}^{3N} \to \mathbb{R}^{3N} \) the force corresponding to the potential \( U(\mathbf{q}) \), i.e., \( f_j(\mathbf{q}) = -\nabla_{q_j} U(\mathbf{q}), \quad j = 1, \ldots, N \), is the force exerted to the \( j \)-th particle.
Coarse-graining reduces the extended system’s complexity by lumping together degrees of freedom into coarse-grained variables and investigates the mesoscopic range and the importance of atomistic detail across scales. Coarse-graining is considered as the application of a mapping (CG mapping) \( \Pi : \mathbb{R}^{3N} \rightarrow \mathbb{R}^{3M} \)

\[
\mathbf{q} \mapsto \Pi(\mathbf{q}) \in \mathbb{R}^{3M}
\]
on the microscopic state space, determining the \( M (< N) \) CG particles as a function of the atomic configuration \( \mathbf{q} \). We denote by \( \mathbf{Q} = (Q_1, \ldots, Q_M) \) any point in the CG configuration space \( \mathbb{R}^{3M} \) and use the bar ” \( \bar{\cdot} \) ” notation for quantities on the CG space. We call atoms the elements of the microscopic space with positions \( \mathbf{q}_j \in \mathbb{R}^3, j = 1, \ldots, N \) and ‘CG particles’ the elements of the coarse space with positions \( Q_i \in \mathbb{R}^3, i = 1, \ldots, M \).

The mappings most commonly considered in coarse graining of molecular systems are linear mappings represented by a set of non-negative real constants \( \{\zeta_{ij}\}_{j=1}^N \), for which

\[
\Pi_i(\mathbf{q}) = \sum_{j=1}^N \zeta_{ij} q_j \in \mathbb{R}^3, \ i = 1, \ldots, M.
\]

In this work we consider CG maps such that a CG particle is the center of mass of a group of atoms for which an atom contributes only to one CG particle, see for example figures 1a and b.

![Figure 1: All-atom and CG representations of: (a) (left) a typical macromolecular chain, (b) (right) water bulk system.](image)

The probability that the CG system has configuration \( \mathbf{Q} \), is given by

\[
\bar{\mu}(\mathbf{Q}) = \int_{\Omega(\mathbf{Q})} \mu(\mathbf{q}) d\mathbf{q}, \quad \Omega(\mathbf{Q}) = \{ \mathbf{q} \in \mathbb{R}^{3N} : \Pi(\mathbf{q}) = \mathbf{Q} \}.
\]

If we require that it is of the canonical Gibbs form then

\[
\bar{\mu}(d\mathbf{Q}) = Z^{-1} \exp\{-\beta \bar{U}^{PMF}(\mathbf{Q})\} d\mathbf{Q},
\]

and the corresponding free energy defines the \( M \)-body potential of the mean force (PMF),

\[
\bar{U}^{PMF}(\mathbf{Q}) = -\frac{1}{\beta} \log \int_{\Omega(\mathbf{Q})} e^{-\beta U(\mathbf{q})} d\mathbf{q}.
\]
2 Parametrizations at equilibrium and Potential of Mean Force

The calculation of the PMF (5) is a task as difficult and costly as is calculating expectations of quantities of interest on the microscopic space. Therefore, one introduces parametric or non-parametric approximations

$$\bar{U}_{\text{eff}}(\mathbf{Q}; \theta), \quad \theta \in \Theta \subseteq \mathbb{R}^k.$$ (6)

Methods such as inverse Monte Carlo (IMC), direct inverse Boltzmann (DBI) and iterative inverse Boltzmann (IBI) \([11, 12, 1]\), force-matching \([13, 14]\), and relative entropy minimization \([15]\) provide optimal parameterizations of approximate coarse-grained models by considering a pre-selected set of observables \(\phi\) and by then minimizing a cost functional over the parameter space,

$$\min_{\theta \in \Theta} \bar{C}(\phi; \theta).$$ (7)

The relative entropy (RE) minimization method is defined as the optimization problem

$$\min_{\theta} \mathcal{R}(\mu|\mu^0),$$ (8)

where

$$\mathcal{R}(\mu|\mu^0) = \mathbb{E}_\mu \left[ \log \frac{\mu(q)}{\mu^0(q)} \right],$$ (9)

is the RE (or Kullback-Leibler divergence) between the microscopic Gibbs measure \(\mu(q)\) and and a back-mapping \(\mu^0(q)\) of the approximate CG measure \(\bar{\mu}^0(\mathbf{Q}) \propto \exp(-\beta \bar{U}_{\text{eff}}(\mathbf{Q}; \theta)), [16, 4]\). \(\mathbb{E}_\mu[\cdot]\) denotes averages with respect to the probability measure \(d\mu(q)\). The minimization of RE is thus equivalent to

$$\min_{\theta \in \Theta} \left\{ \beta \mathbb{E}_\mu \left[ \bar{U}_{\text{eff}}(\Pi(q); \theta) - U(q) \right] - \left[ \log Z^0 - \log Z \right] \right\}, $$ (10)

where \(Z^0 = \int_{\mathbb{R}^{3M}} e^{-\beta \bar{U}_{\text{eff}}(\mathbf{Q}; \theta)} d\mathbf{Q}\) and \(Z = \int_{\mathbb{R}^{3N}} e^{-\beta U(q)}\).

The force-matching method determines a CG effective force \(\bar{F}(\mathbf{Q}; \theta)\) –and thus an effective potential– from atomistic force information as the solution of the mean least-square minimization problem

$$\min_{\theta \in \Theta} \mathbb{E}_\mu \left[ ||h(q) - \bar{F}(\Pi(q); \theta)||^2 \right],$$ (11)

where \(|| \cdot ||\) denotes the Euclidean norm in \(\mathbb{R}^{3M}\). The reference field \(h(q) \in \mathbb{R}^{3M}\) is the local mean force whose component \(h_i(q), i = 1, \ldots, M\) is the force exerted at the \(i\)-th CG particle that is a function of the microscopic forces. For example, if the CG mapping is the one that defines the CG particles as the center of mass of a group of atoms then \(h_i(q) = \sum_{j \in \text{group}_i} f_j(q), i = 1, \ldots, M\). In work \([2]\), we presented a rigorous probabilistic formulation and a generalization of the traditional force matching approach that applies to more complex and non-linear coarse-graining maps. For example, for

$$h(q) = J^{-1}_\Pi(q)D_\Pi(q)f(q) + \frac{1}{\beta} \nabla_q \cdot J^{-1}_\Pi(q)D_\Pi(q),$$

we prove that the solution of (11) is a best approximation of the PMF, which holds both for linear and non-linear mappings \(\Pi(q)\). Here \(J_\Pi(q) = D_\Pi(q)D_\Pi^T(q), D_\Pi \in \mathbb{R}^{m \times 3N}\) (\(D_\Pi)_{ij}(q) =...
the process parametrized with \( \theta \) where \( \bar{\sigma} \). Find the most effective among the proposed CG models such that \( \{ v \}_{i=1}^n \), approximation of the coarse-grained process. We consider the optimization principle

\[
\text{The (pair) CG potential is refined at the iteration } i \text{ according to the following scheme:}
\]

\[
\begin{align*}
\bar{U}_{\text{eff}}(R) &= \frac{1}{\beta} \log \bar{g}(R). \\
\bar{U}_{\text{eff}}^{(i+1)}(R) &= \bar{U}_{\text{eff}}^{(i)}(R) + c k_B T \log \frac{\bar{g}^{(i)}(R)}{\bar{g}^{(\text{ref})}(R)},
\end{align*}
\]

DBI employs directly relation (12) to infer the interaction potential \( \bar{U}_{\text{eff}}(R) \) from a reference CG (pair) distribution function \( \bar{g}^{(\text{ref})}(R) \) obtained from the analysis of the all-atom configurations. In IBI methods, \cite{2}, an iterative numerical minimization problem is introduced based on \( \bar{g}(R) \). The (pair) CG potential is refined at the iteration \( (i + 1) \) according to the following scheme:

\[
\text{where } c \text{ is a constant to ensure stability of the iterative process.}
\]

### 3 Parametrizations away from equilibrium

We present an extension of the RE minimization approach to systems with non-equilibrium steady states as well as dynamics in finite times. The presented method also allows for approximation of dynamical observables, i.e., quantities that are averaged over the path distribution instead of over a distribution at a terminal time.

We consider the evolution of the \( N \) particles described by a diffusion process \( \{ X_t \}_{t \geq 0} \), a continuous time Markov process satisfying the stochastic differential equation (SDE)

\[
\begin{align*}
\frac{dX_t}{dt} &= b(X_t, t) + \sigma(X_t) dB_t, \quad t > 0, \\
X_0 &= \mu_0,
\end{align*}
\]

where \( b(x) \in \mathbb{R}^{6N} \) and \( \sigma(x) \in \mathbb{R}^{6N \times k} \), \( k \leq 6N \) are the drift and diffusion coefficients and \( B_t \) denotes the standard \( k \)-dimensional Brownian motion.

The proposed coarse space dynamics are described by a Markov process \( \{ \bar{X}_t \}_{t \geq 0} \) in \( \mathbb{R}^m \) approximating the process \( \{ \Pi X_t \}_{t \geq 0} \) which is, in principle, non-Markovian. The Markov process \( \{ \bar{X}_t \}_{t \geq 0} \) is given as the solution of the parametrized stochastic differential equations

\[
\begin{align*}
\frac{d\bar{X}_t}{dt} &= \bar{b}(\bar{X}_t, \theta) + \bar{\sigma}(\bar{X}_t) dB_t, \quad t > 0, \\
\bar{X}_0 &= \bar{\mu}_0,
\end{align*}
\]

where \( \bar{\sigma}(x) \in \mathbb{R}^{m \times l}, \quad l \leq m \) is the diffusion and \( \bar{b}(x; \theta) \in \mathbb{R}^m \) the drift coefficient is parametrized with \( \theta \in \Theta. \bar{B}_t \) is an \( l \)-dimensional standard Brownian motion. The goal is to find the most effective among the proposed CG models such that \( \{ \bar{X}_t \}_{t \geq 0} \) “best approximates” the process \( \{ \Pi X_t \}_{t \geq 0} \), that is to find optimal \( \bar{b}(x; \theta) \) and \( \bar{\sigma}(x; \theta) \).

The best approximation is fitted using entropy based criteria in order to find the best Markovian approximation of the coarse-grained process. We consider the optimization principle

\[
\min_{\theta \in \Theta} \mathcal{R} \left( P_{\{0,T\}} | \Pi_{\{0,T\}} Q_{\{0,T\}}^\theta \right),
\]

\[
\nabla_{\theta_i} \Pi_t(q). \text{ A more general result is available in \cite{9}. These results are based on the probabilistic reformulation of the FM method and the connection to known results of the thermodynamic integration theory \cite{7}.}
\]

Additionally, the RE and FM methods are asymptotically equivalent, for small discrepancies from the PMF, see section VI in \cite{9}.

The DBI, IBI and IMC methods use the pair correlation function \( g^{(2)}(Q) \) and the assumption that the interactions depend only on the distance \( R \) between particles, that is \( g^{(2)}(Q) =: \bar{g}(R) \). Thus the CG effective interaction is given by

\[
\begin{align*}
\bar{U}_{\text{eff}}(R) &= \frac{1}{\beta} \log \bar{g}(R). \\
\bar{U}_{\text{eff}}^{(i+1)}(R) &= \bar{U}_{\text{eff}}^{(i)}(R) + c k_B T \log \frac{\bar{g}^{(i)}(R)}{\bar{g}^{(\text{ref})}(R)},
\end{align*}
\]

where \( c \) is a constant to ensure stability of the iterative process.
where $P_{[0,T]}$ is the path distribution of the original microscopic process and $\Pi_3^t Q_{[0,T]}^\theta$ is the parametrized path-space coarse-grained distribution back-mapped to the microscopic space.

Then, the variational inference problem (16) is equivalent to the following path-space force matching problem

$$\argmin_{\theta \in \Theta} \mathcal{R} (P_{[0,T]} | \Pi_3^t Q_{[0,T]}^\theta) = \argmin_{\theta \in \Theta} \mathbb{E}_{P_{[0,T]}} \left[ \frac{1}{2} \int_0^T || \Pi b(X_s) - \bar{b}(\Pi X_s; \theta) ||^2_{\Pi_{\Xi}} ds \right].$$

where

$$||z||^2_{\Pi_{\Xi}} = z^\text{tr} \Pi^{\text{tr}} \Xi^{-1} \Pi z, \quad z \in \mathbb{R}^m \text{ and } \Xi = [\sigma^\text{tr}(x) \sigma(x)]^{-1} \sigma^\text{tr}(x).$$

under the assumption that the auxiliary reconstructed diffusion process with path-space distribution $\Pi_3^t Q_{[0,T]}^\theta$ has diffusion coefficient $\sigma(x)$. Here $\Pi^z$ is defined such that $\Pi \Pi^z = \Pi$. If moreover $\{X_t\}_{t \geq 0}$ is stationary with the invariant measure $\mu$, then

$$\mathcal{R} (P_{[0,T]} | \Pi_3^t Q_{[0,T]}^\theta) = T \mathcal{H}(P | \Pi_3^t Q^\theta) + \mathcal{R} (\mu | \Pi_3^t \mu_0),$$

and

$$\argmin_{\theta \in \Theta} \mathcal{H}(P | \Pi_3^t Q^\theta) = \argmin_{\theta \in \Theta} \mathbb{E}_{\mu} \left[ \frac{1}{2} || \Pi b(X) - \bar{b}(\Pi X; \theta) ||^2_{\Pi_{\Xi}} \right].$$

Let us now consider Langevin dynamics for the $N$-particle molecular system, described by the process $\{(q_t, p_t)\}_{t \geq 0}$, with positions $q \in \mathbb{R}^{3N}$ and momenta $p \in \mathbb{R}^{3N}$

$$\begin{cases}
    dq_t = M^{-1} p_t dt,
    
    dp_t = F(q_t) dt - \gamma M^{-1} p_t dt + \sigma dB_t,
\end{cases}$$

a Hamiltonian system coupled with a thermostat, where $F(q)$ is the force field that is not necessarily a gradient. $M = \text{diag}(m_1 I_3, \ldots, m_N I_3) \in \mathbb{R}^{3N \times 3N}$ is the mass matrix, $\gamma \in \mathbb{R}^{3N \times 3N}$ is the friction and $\sigma \in \mathbb{R}^{3N \times 3N}$ the diffusion coefficients respectively, and $B_t$ is the $3N$-dimensional Brownian motion. The diffusion and friction coefficients satisfy the fluctuation-dissipation relation $\sigma \sigma^\text{tr} = 2 \beta^{-1} \gamma$.

The proposed dynamics for the coarse variables $\bar{X} = (Q, P) \in \mathbb{R}^{3M}$ are given by the Langevin system

$$\begin{cases}
    dQ_t = M^{-1} P_t \, dt,
    
    dP_t = \bar{F}(Q_t; \theta) dt - \gamma M^{-1} P_t dt + \bar{\sigma} dB_t,
\end{cases}$$

where $\bar{B}_t$ is a $3M$-dimensional Brownian motion. The diffusion coefficient $\bar{\sigma}$ is defined by

$$\bar{\sigma} = \Pi \sigma \sigma^\text{tr} \Pi^p.$$

The optimal parameter set for which the process $\{(Q_t, P_t)\}_{t \geq 0}$ best approximates $\{(q_t, p_t)\}_{t \geq 0}$ at the time interval $[0, T]$ is given by, if $\gamma \Pi = \Pi p \gamma$,

$$\theta^*(T) = \argmin_{\theta} \mathbb{E}_{F_{\bar{X}}} \left[ \frac{1}{2} \int_0^T || \Pi F(q_s) - \bar{F}(Q_s; \theta) ||^2_{\Pi_{\Xi}} ds \right].$$

Note that the norm $||\cdot||_{\Pi_{\Xi}}$ is a weighted Euclidean norm with weights $\Pi_{\Xi} = \Pi^z [\sigma^\text{tr}(x) \sigma(x)]^{-1} \sigma^\text{tr}(x)$. This retrieves the relation of the path-space force matching approach to the diffusion coefficient of the atomistic dynamics. Moreover, if we assume equilibrium dynamics and constant diffusion $\sigma(x) = \sigma$ the optimization problem is reduced to the known FM method, see also Remark 6.3 in [5].

$$\theta^* = \argmin_\theta \mathbb{E}_{\mu} \left[ \frac{1}{2} || \Pi F(q) - \bar{F}(Q; \theta) ||^2 \right] PERIOD$$
4 Models and Simulations

We test the above described methods on different molecular systems at equilibrium:

(a) Methane. Methane liquid was simulated at constant temperature (NVT conditions) at \( T = 100 \text{ K} \) for several ns. 512 CH\(_4\) molecules were modeled, whereas the density was calculated after equilibrating the system in the NPT ensemble for 5 ns \( (\rho = 0.38 \text{ g/cm}^3) \). The time step was 0.5 fs and a cut-off distance of 10 Å was used.

For the coarse-grained representation of CH\(_4\), we have used a one-site representation with a pair potential.

(b) Water. One of the most well-studied liquids both through atomistic and coarse-grained models in the literature is water [17]. Here we have simulated all-atom water, using one of the most typical atomistic force fields, the SPC/E [18]. The model system consists of 1192 molecules at ambient conditions \( (T = 300 \text{ K}, P = 1 \text{ atm}) \). The time step was 1 fs. A cut-off distance of 10 Å was used, while electrostatic interactions were calculated using PME. We first equilibrate the system under NPT conditions for about 50 ns. Then, NVT simulations, in the average density, were performed for 20 ns. All-atom configurations were recorded every 10 ps.

For the coarse-grained representation of H\(_2\)O, we have also used a one-site representation with a pair potential. In the CG representation of water electrostatic interactions were not required to be introduced.

5 Results

5.1 Methane

In the following we present results from the atomistic and CG simulations of the bulk methane liquid. We approximate the many-body PMF between the CG particles through the different approaches discussed above.

First, we apply the IBI method for this system using the all-atom data. IBI converges for this system (tolerance is \( 10^{-4} \)) after 14 iterations. Data for the CG pair correlation function, \( g(R) \), and the resulting potential for various iterations are presented elsewhere [10].

Next, we examine the FM method for the CH\(_4\) fluid, by analyzing the reference data from the all-atom simulations. In order to solve the minimization problem we have tested different basis function sets: linear splines, cubic splines, LJ and Morse. Linear splines, cubic splines and a Morse type basis give the same results, within the numerical accuracy. Only the results using the LJ basis slightly deviate. Results concerning the CG effective interaction were also found to be very close to the data obtained from the two isolated CH\(_4\) in the vacuum [10].

We have also examined the application of the relative entropy minimization problem for the CH\(_4\) liquid. Here we have used a typical Newton-Raphson scheme. Convergence is achieved after about 20 iterations [10].

Data about the pair PMF, that is an approximation of the many-body PMF, for the bulk methane fluid derived from the different approaches (IBI, FM and RE) are shown in Figure[2]. It is clear that different methods give slightly different approximations of the PMF. However, the differences between the various sets of data are rather small, less than 5% in overall.

5.2 Water

The next example considered here is water. First, we apply the IBI method for this system using the all-atom data. Convergence of IBI for water is more sensitive than for the methane fluid discussed before. Indeed more than 100 iterations are required for the CG radial distribution
Function in order to match the atomistic data.

Then, we apply the RE and the FM methods for water. Numerical implementations for these methods are very sensitive to poor sampling. In FM the matrix the canonical system becomes singular, while in IBI and RE the iterative procedure fails. Thus very careful extrapolation methods must be considered at small distances as well as smoothing approaches to reduce the inherent noise of the statistical sampling. Specifically, the speed of convergence for the Newton-Raphson iterative scheme is based on the $\chi$ parameter, whereas its stability primarily depends on the condition number of the Hessian. The latter depends on several parameters: the trajectory length (at what extent does the sample size evenly cover the chosen basis function); the chosen basis set; and the correlation of the above to the model parameters, such as the number of atoms and the complexity of the coarse graining mapping [10].

Another important issue is poor sampling towards the minimum distance of the pair potential $R_{min}$. If this is the case, in the Newton-Raphson method for the RE minimization the Jacobian may involve negative values while the Hessian matrix becomes singular and the iterative scheme either stops or produces enormous fluctuations. A way to overcome this issues is the enrichment of nodes towards $R_{min}$, together with extrapolation of the potential on the first couple of nodes. Another good practice is smoothing out the potential after every iteration to reduce the noise in the updated forces.

In Figure 3 results for the effective CG potential from the IBI and the FM method (using cubic spline basis set) are presented. Note, that results from RE are similar to those of IBI. Although both IBI and FM potentials have a very similar structure with two minima, the actual values of the potential are considerably different, in contrast to the CH$_4$ fluid discussed in the previous subsection. Possible reasons for these discrepancies are related to the fact that FM and RE are only asymptotically equivalent, meaning that finite size basis sets effects might be important during the numerical optimization procedure. Clearly more work is required in order to clarify such differences [19, 9, 10].

Finally, in Figure 4 we show the CG $\bar{g}(R)$ obtained from RE minimization problem together with the reference curve, obtained from the analysis of the all-atom data. The curves are very
close to each other; however there are small differences, in particular in small distances, close
to the first maximum. Note that theoretically it is expected that the RE outcome, at the level of
g(R), should agree with the IBI one [15]. We should report here that we have calculated the CG
potential derivatives appearing in the Jacobian and Hessian in the Newton-Raphson scheme by
direct sampling during the corresponding CG run.

6 CONCLUSIONS

The development of systematic high fidelity coarse-grained models for molecular systems
is a very challenging research area. In this aspect, finding the optimum effective interaction
potential between CG particles for a given model involves crucial theoretical and numerical
issues related to numerical parametrization of the (many-body) potential of mean force.

Here we have discussed different parametrization methods: (a) structural/correlation-based
methods (direct Boltzmann inversion and iterative Boltzmann inversion), (b) Force matching,
and (c) Relative entropy methods. We further examine the implementation of these approaches
to molecular systems at equilibrium and non-equilibrium conditions in various molecular sys-
tems. In more detail, we have seen that CG methods based on relative entropy and force matching
are in principle asymptotically equivalent. If we consider that RE methods are expected to
give the same solution as the IBI methods, for a given CG mapping and a specific basis set, then
we see the direct theoretical relation of all methods discussed here.

Furthermore, we applied the above numerical schemes (IBI, RE and FM) on the same atom-
istic systems. Despite the fact that all the above methods are approximations of the same (many-
body) potential of mean force, it is not clear that their numerical implementation will converge
into the same solution, since there are differences in the derived numerical schemes. We first
consider a simple liquid (methane fluid) the CG effective potentials derived from the different
methods are very similar. Slight differences of the order of 5-10% are found that are within the
numerical accuracy. CG simulations with the derived force field also show structural properties
in very good agreement with the reference (all-atom) data for all models. Different is the case of
the dynamic properties; friction in the CG models is clearly more sensitive to slight differences in the CG potential used. Second, we apply the methods on bulk water, one of the most well studied systems in the literature. Larger, compared to the methane, differences in the derived CG potential from the various methods are observed for water.

A general comment valid for all methods, is related to the actual numerical problems for parts of the phase space, where the energy is very high; i.e. areas with rare sampling. Practically, numerical extrapolation schemes using proper functional forms, as well as smoothing approaches, for the CG potentials should be used, in order to prevent sampling from such regimes.

Application of all above methods requires a very good sampling of the reference all-atom system. Such a sampling could be even more problematic for complex (e.g. polymeric) molecules. On the contrary, the DBI correlation-based approach, that is based on the decomposition of the CG potential in bonded and non-bonded components, can be a computationally efficient alternative. Such a methodology neglects many body terms; however, for several systems can provide an accurate prediction of the structural and thermodynamic properties [20, 21, 22, 23].

Several issues remain to be examined in order to systematically compare different numerical parametrization schemes for realistic molecular systems. For example, all systems studied here concern pair non-bonded CG effective potential; the use of many-body, or density dependent, CG potentials would expect to be important, in particular in systems of high density. In addition, non-linear CG maps could be also relevant especially when free energy differences, such as in thermodynamic integration, are to be computed. Parametrization of the dynamics of CG models is also one of the most challenging issues, in particular for non-equilibrium molecular systems [3, 24, 5, 25].

REFERENCES


MUTLISCALE SURROGATE MODELING AND UNCERTAINTY QUANTIFICATION FOR PERIODIC COMPOSITE STRUCTURES

Charilaos Mylonas¹, Bemetz Valentin¹, Eleni Chatzi¹

¹ Department of Civil, Environmental and Geomatic Engineering, ETH Zürich
Stefano-Franscini-Platz 5, 8093 Zürich
e-mail: mylonasc@ibk.baug.ethz.ch

Keywords: Polynomial Chaos Expansions, Asymptotic Expansion Homogenization, Composite Analysis, Principal Component Analysis, Random Microstructure

Abstract. Computational modeling of the structural behavior of continuous fiber composite materials often takes into account the periodicity of the underlying micro-structure. A well established method dealing with the structural behavior of periodic micro-structures is the so-called Asymptotic Expansion Homogenization (AEH). By considering a periodic perturbation of the material displacement, scale bridging functions, also referred to as elastic correctors, can be derived in order to connect the strains at the level of the macro-structure with micro-structural strains. For complicated inhomogeneous micro-structures, the derivation of such functions is usually performed by the numerical solution of a PDE problem - typically with the Finite Element Method. Moreover, when dealing with uncertain micro-structural geometry and material parameters, there is considerable uncertainty introduced in the actual stresses experienced by the materials. Due to the high computational cost of computing the elastic correctors, the choice of a pure Monte-Carlo approach for dealing with the inevitable material and geometric uncertainties is clearly computationally intractable. This problem is even more pronounced when the effect of damage in the micro-scale is considered, where re-evaluation of the micro-structural representative volume element is necessary for every occurring damage. The novelty in this paper is that a non-intrusive surrogate modeling approach is employed with the purpose of directly bridging the macro-scale behavior of the structure with the material behavior in the micro-scale, therefore reducing the number of costly evaluations of corrector functions, allowing for future developments on the incorporation of fatigue or static damage in the analysis of composite structural components.


1 Introduction

Continuous fiber reinforced polymer composites are light, stiff materials of significantly improved static strength and fatigue resistance. For the engineering analysis of such materials, a direct discretization of the fine spatial variation of the composite material would render the problem computationally intractable. Therefore, composite engineering analysis seeks to deliver a consistent calculation of the effective macroscopic properties by considering the material and geometrical properties of the micro-structure (homogenization) and adequately approximating the stresses in the micro-structure (localization). A mathematically rigorous approach to the problem of homogenization and localization, that further applies to the problem of elasticity in the context of composite materials, was proposed in [1].

The aforementioned technique is often termed Asymptotic Expansion Homogenization (AEH). The application of the method relies on the assumption that displacement appears into well separated spatial scales. The method yields effective elastic properties on the macro-scale without any assumptions on the distribution of strains or stresses in the micro-scale, but only with the assumption of periodicity in the displacements among different representative volume elements. For a more detailed, engineering oriented derivation of the AEH method for elasticity the reader is referred to [2] and [3]. Extending the AEH method to damaged composites has also attracted research interest [4, 5]. In the context of another homogenization framework, it has been shown that efficient hysteretic multi-scale damage models can be derived [6].

The complete determination of material properties and microstructure geometry is, in general, not possible. Therefore, the prediction of the material response in the micro-scale should account for uncertainty. A direct Monte-Carlo approach for the purpose of representing the effect of all the uncertain parameters would quickly become intractable.

In the present work we investigate the potential of non-intrusive probabilistic uncertainty propagation techniques, namely the Polynomial Chaos Expansion (PCE)[7], for the purpose of constructing surrogate models. Efficient surrogate modelling techniques are expected to yield further reductions in the computational cost of multi-scale finite element analysis. An intrusive PCE for the same problem was proposed in [8].

Finally, a dimensionality reduction technique, namely Principal Component Analysis (PCA) was found to be highly efficient on decomposing the stiffness tensor without strong assumptions on the geometry induced symmetries of the homogenized stiffness tensor.

The uncertainty quantification toolbox UQLab was used for deriving the PCE of the homogenized stiffness tensor [9].

2 Computational Methodology

In the following the basic components of the Asymptotic Expansion Homogenization for analyzing elastic periodic structures and the Polynomial Chaos Expansion surrogate modelling technique, used in the present study are going to be briefly presented.

2.1 Asymptotic Expansion Homogenization

This section serves for establishing notation notation and introducing an intuitive understanding of the quantities related to the problem of homogenization and localization for periodic media. We denote $\mathbf{x} = \{x_1, x_2, x_3\}$ as the coordinate system of a composite structure, and further introduce a coordinate system local to every representative volume element (microstructure).
ture) \( y = \{y_1, y_2, y_3\} \). Quantities marked with \( \cdot \) denote the high resolution quantities in the macro-scale. All indices attain values in \( \{1, 2, 3\} \). Einstein summation is implied for repeating indices. We seek to solve the elasticity boundary value problem,

\[
\frac{\partial \sigma_{ij}^c}{\partial x_j^c} + f_i = 0 \quad \text{in } \Omega \tag{1}
\]

\[
u_i^c = 0 \quad \text{on } \partial_1 \Omega \tag{2}
\]

\[
\sigma_{ij}^c n_j = F_i \quad \text{on } \partial_2 \Omega \tag{3}
\]

\[
\epsilon_{ij}(u^c) = \frac{1}{2} \left( \frac{\partial u^c_i}{\partial x_j^c} + \frac{\partial u^c_j}{\partial x_i^c} \right) \tag{4}
\]

\[
(5)
\]

where \( \partial_1 \Omega \) and \( \partial_2 \Omega \) denote different boundaries, \( u_i^c = u^c(x) \) is the displacement of the macro-structure, \( F_i \) a traction force, and \( f_i \) the body force. The constitutive relation simply reads

\[
\sigma_{ij}^c = D^c_{ijkl}\epsilon_{kl} \tag{6}
\]

Due to the geometry of the continuous fiber reinforced composites, \( D^c_{ijkl} \) is varying periodically in the material, in a scale much finer than the scale of the structure. It is convenient to define the so-called scale parameter \( \epsilon << 1 \), which represents the ratio between the microscopic and macroscopic scale. Considering the coordinates of the micro-scale and the macroscale, one may write \( y_i = x_i/\epsilon \). By the chain rule we have

\[
\frac{\partial \cdot}{\partial x_i^c} = \frac{\partial \cdot}{\partial x_i} + \frac{1}{\epsilon} \frac{\partial \cdot}{\partial y_i} \tag{7}
\]

The displacements are represented with the following expansion in, \( \epsilon \), as

\[
u_i^c(x) = u_i^{(0)}(x) + \epsilon u_i^{(1)}(x) + \epsilon^2 u_i^{(2)}(x) + \cdots \tag{8}
\]

It has been rigorously established [1], that by plugging Equation 8 into the problem of elasticity, and by passing to the limit \( \epsilon \to 0 \), the elasticity problem boils down to a hierarchical set of partial differential equations. It is assumed that the displacements in the representative volume elements are connected to the gradients of the displacement in the macro-scale \( \frac{\partial u_i^{(0)}}{\partial x_k}(x) \) by a certain vector valued function \( \chi_i^{kl}(y) \). This approximation reads

\[
u_i^{(1)}(x, y) = -\chi_i^{kl}(y) \frac{\partial u_k^{(0)}}{\partial x_i}(x) + \bar{u}_i^{(1)}(x), \tag{9}
\]

where \( \bar{u}_i^{(1)}(x) \) denotes the average displacement of the representative unit cell in the macro-scale coordinate system. Function \( \chi_i^{mn} \) is often termed the elastic corrector. Note that every pair of components \( mn \) correspond to a different spatial gradient. The accuracy of this approximation relies on the existence of the gradients \( \frac{\partial u_k^{(0)}}{\partial x_i}(x) \) and assumes a slow variation in the macroscopic scale.

For continuous fiber composites, without stress concentrations this is a reasonable assumption. A stress concentration may be due to localized damage, i.e., due to a macroscopic crack or very close to the boundaries of the composite structure\(^2\).

\(^2\)On the other hand, the effect of diffuse slowly spatially varying damage may be well approximated without the presented framework to break down.
For a first order (first order perturbation) approximation of the perturbed displacement field, assuming $\chi_{mn}^{i}$ smooth in $\Omega$ and smooth and periodic with zero mean in the RVE or $\in V_{\text{per}}$ on $\Omega_Y$, the variational problem

$$
\int_{\Omega_Y} D_{ijkl} \frac{\partial \chi_{mn}^{kl}}{\partial y_l} \frac{\partial \nu_i}{\partial y_j} d\mathbf{y} = \int_{\Omega_Y} D_{ijmn} \frac{\partial \nu_j}{\partial y_i} d\mathbf{y} \quad (10)
$$

holds. We seek solutions for $\chi_{mn}^{i}$ so that Equation 10 holds for all $\nu_i \in V_{\text{per}}$.

Due to the symmetries of the stiffness tensor, we have $D_{ijmn} = D_{ijnm} = D_{jimn}$. Therefore, we only need to consider $mn = \{11, 22, 33, 23, 13, 12\}$ for the full computation of the elastic corrector. In practice, Equation 10 results in 6 variational problems for the computation of the corrector, one for every different value of $mn$.

The variational problem allows for a finite element approximation of the corrector function. By considering the RVE averaged strains and stresses, an approximation of the stiffness tensor $D_{ijkl}^{c} \approx D_{ijkl}^{h}$ in the macro-scale is possible. Namely,

$$
D_{ijkl}^{h} = \frac{1}{|Y|} \int_{\Omega_Y} D_{ijkl}(\mathbf{y}) \left[ \delta_{kl} \delta_{ln} - \frac{\partial \chi_{mn}^{kl}}{\partial y_l} \right] d\mathbf{y} \quad (11)
$$

In practice, even for the case of homogeneous materials described by Lamé parameters in the micro-scale, the homogenized stiffness tensor turns out anisotropic. Some symmetries may be induced by the geometry, such as orthotropy and transverse isotropy, but the framework presented in the present work is concerned with the case of the fully anisotropic material.

It is apparent that since the corrector connects the displacements of the macro-structure to the displacements of the micro-structure, strains and stresses can be straight-forwardly computed for the micro-structure. Namely the micro-stresses are computed with

$$
\sigma_{ij}^{(1)}(\mathbf{x}) = D_{ijkl}(\delta_{mk} \delta_{nl} - \frac{\partial \chi_{mn}^{kl}}{\partial y_l}) \frac{\partial u^{(0)}}{\partial x_n}. \quad (12)
$$

Therefore by storing the solution of the corrector we may directly compute stresses in the micro-scale without making any strong assumptions on the distribution of stresses or strains on the boundaries of the RVE. The only assumption required for this framework is the periodicity of displacements in the boundaries of the RVE.

For the actual solution of the finite element discretization of Equation 10, periodic boundary conditions have to be enforced. In addition, one arbitrary point must be constrained to zero in all components of $\chi_{mn}^{i}$ since the weak form has a unique solution up to an additive constant. Due to the periodicity of the corrector, and the fact that homogenization and localization problems are concerned only with derivatives of the corrector, the boundary conditions are essentially equivalent to the zero-mean requirement for the corrector function.

### 2.2 Polynomial Chaos Expansions

Polynomial chaos expansions (PCE) were first introduced in [10] for Gaussian input variables and generalized in [7] for classical probability distribution functions. Consider a set of random inputs $\mathbf{X} = \{x_1, x_2, \cdots, x_n\}$ to a deterministic model $Y = \mathcal{M}(\mathbf{X})$. The method relies in the construction of a tensor product basis of univariate polynomials $\Phi_{i}(x_n)$, orthogonal with respect to inner products weighted by probability distribution functions $f_{X}(x_n)$. The orthogonality relation reads,

$$
\langle \Phi_{i}^{(m)}, \Phi_{j}^{(m)} \rangle_{f_{X}} = \delta_{ij} \quad (13)
$$
Charilaos Mylonas, Valentin Bemetz and Eleni Chatzi

where

\[ \langle f, g \rangle_{f_X} = \int f(x)g(x)f_X(x)dx \] (14)

and \( \delta_{ij} \) is the Kronecker delta. The tensor product basis set reads,

\[ \Psi(X) = \otimes_{m=1}^{n} \Phi^{(m)} \] (15)

where \( \Phi^{(m)} = \{ \Phi_1^{(m)}, \Phi_2^{(m)}, \ldots \} \) with superscript denoting the input dimension and subscript denoting the order of the orthogonal polynomial.

A PCE model, is a linear combination of the elements of Equation 15,

\[ \mathcal{M}(X) = \sum_{a \in \mathcal{A}} c_a \Psi_a(X) \] (16)

indexed by \( a = \{ a_1, a_2, \ldots, a_m \} \), which is a multi-index that denotes the degree of the univariate polynomials of each of the input variables, and \( \mathcal{A} \) the set of multi-indices. For example,

\[ \Psi_a = \Phi_{a_1}^{(1)}(x_1)\Phi_{a_2}^{(2)}(x_2)\cdots \Phi_{a_n}^{(n)}(x_n) \] (17)

where \( a_i \) denotes the degree of orthogonal polynomials along dimension \( i \). In the presented case the number of random input dimensions is \( n = 6 \). In practice, the set of multi-indices, is truncated for numerical implementation. Also the PCE is considered up to a certain degree of univariate polynomials in each dimension to render the problem numerically tractable. According to the Cameron-Martin theorem \([11, 7]\), such an expansion converges in the \( L_2 \) sense, when \( \mathcal{M}(X) \) has finite variance.

There are several approaches for the purpose of determining the coefficients \( c_a \). The most versatile method, that also deals automatically with adaptively selecting basis elements, is the Least Angle Regression (LAR)\([12]\) approach. LAR is the method of choice for the present work. See \([13]\) and \([14]\) for a discussion of the benefits of LAR.

3 Example application on continuous fibre reinforced composites

A typical composite structure, composed of transversely isotropic glass fibers embedded in a polymer matrix with stacking sequence \([0, -\phi, +\phi] \), was analyzed as a proof of concept. The material properties adopted herein, are given in Table 1. In this study, only geometric variation of the micro-structure was considered. For the purpose of demonstrating the effectiveness of the PCE surrogate model of the homogenization process, relatively large variations on the geometrical parameters of the micro-structure were chosen. The ranges of the parameters chosen for the present work are given in Table 2.

\( V_{f_1} \) corresponds to the volume fraction of the 0\(^\circ\) fibers and \( V_{f_2} \) the volume fraction of each of the layers of the \( \pm 45^\circ \) fibers. Correspondingly, \( a_1, a_2 \) are the major radii of the elliptical cross section of the fibers and \( b_1, b_2 \) the minor radii (Figure 1). A uniform distribution is considered for the aforementioned parameters, in the ranges presented in Table 2. In the current study, 200 model runs were used, with input vectors randomly sampled with Latin Hypercube Sampling (LHS) in order to explore the parameter space as well as possible with the limited budget of model runs. A visual account of the solution for the corrector function for a particular set of parameters is given in Figure 2.
Charilaos Mylonas, Valentin Bemetz and Eleni Chatzi

<table>
<thead>
<tr>
<th>parameter</th>
<th>Fiber</th>
<th>Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_1$ [GPa]</td>
<td>31</td>
<td>2.79</td>
</tr>
<tr>
<td>$E_2$ [GPa]</td>
<td>7.59</td>
<td>2.76</td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>$G_{12}$</td>
<td>3.52</td>
<td>1.1</td>
</tr>
<tr>
<td>$G_{23}$</td>
<td>2.69</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Table 1: Material properties of the micro-structure.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{f_1}$</td>
<td>0.600</td>
<td>0.74</td>
</tr>
<tr>
<td>$V_{f_1} \cdot V_{f_2}$</td>
<td>0.600</td>
<td>1.00</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.450</td>
<td>0.55</td>
</tr>
<tr>
<td>$a_1 \cdot b_1$</td>
<td>0.167</td>
<td>0.250</td>
</tr>
<tr>
<td>$a_2 \cdot b_2$</td>
<td>0.167</td>
<td>0.250</td>
</tr>
<tr>
<td>$\phi$</td>
<td>15</td>
<td>75</td>
</tr>
</tbody>
</table>

Table 2: Assumed micro-structure geometry parameter variations.

Figure 1: Random geometric parameters of the micro-structure. The volume fractions affect the intra-fiber spacing.

3.1 Dimensionality reduction with PCA for the homogenized stiffness tensor

It is natural to expect that the components of the homogenized stiffness tensor co-vary. In general, for arbitrary micro-structure geometries it is not trivial to assess intuitively the effect of geometric variation on the stiffness tensor directly. In the present study, Principal Component Analysis (PCA) is implemented for the reduction of the $6 \times 6$ homogenized random stiffness tensor. In order to apply PCA on the homogenized tensors, the components of every random tensor are first flattened to a row vector as indicated in Equation 18.

A set of $N_{PCA}$ principal components $D_{PCA}^{(m)}$ (corresponding to tensor components) is sought, that satisfy Equation 19, where $\mu_{D_{PCA}}$ is the empirical mean of the homogenized stiffness, $X_n$ is the $n^{th}$ realization of the random input vector and $\lambda^{(m)}(X_n)$ denotes a random coefficient that depends on the $n^{th}$ realization of the random input data.

$$D_{ijkd}(X_i) = \begin{bmatrix} D_{1111} & D_{1122} & D_{1133} & 0 & 0 & 0 \\ D_{2222} & D_{2233} & 0 & 0 & 0 \\ D_{3333} & 0 & 0 & 0 & 0 \\ D_{2323} & 0 & 0 & D_{1313} & 0 \\ Sym & 0 & 0 & 0 & D_{1212} \end{bmatrix} \rightarrow \begin{bmatrix} D_{1111} \\ D_{2222} \\ D_{3333} \\ D_{2323} \\ D_{1313} \\ D_{1212} \end{bmatrix}^T$$ (18)
Figure 2: A visual account of the corrector function for a composite with $\phi = 45^\circ$. The correctors are plotted only on the surface of the fibers and the top layer of $0^\circ$ fibers are hidden. The color corresponds to $(\chi_{11}^{mn})^2 + (\chi_{22}^{mn})^2 + (\chi_{33}^{mn})^2$. Although not easily visible due to 3D plotting, the solution for the corrector is periodic.

For the present study, the tensor is symmetric and it is expected to correspond to an orthotropic elastic material. This results in 9 non-zero components. For a general anisotropic elastic material, up to 21 components would be expected. Polynomial surrogates and sensitivity analysis for generally anisotropic materials described by probabilistically modelled random materials and random geometry may be treated via the same framework in a straightforward manner without placing any assumptions on the form of the stiffness tensor. The 4 first principal components were employed herein. Their contributions to the variance of the data are summarized in table Table 3. Considering the variance explained, it is concluded that 4 components are sufficient.
to capture the main variations on the homogenization data. The variance due to the remaining 5 principal components is considered insignificant, and attributed to the slight inaccuracies of the FE solution. For illustrative purposes, the two first principal components are presented in Table 4.

<table>
<thead>
<tr>
<th>Component</th>
<th>$D_{PCA}^{(1)}$</th>
<th>$D_{PCA}^{(2)}$</th>
<th>$D_{PCA}^{(3)}$</th>
<th>$D_{PCA}^{(4)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explained Variance</td>
<td>50.01%</td>
<td>35.45%</td>
<td>14.10%</td>
<td>0.39%</td>
</tr>
</tbody>
</table>

Table 3: Variance explained by the first 4 principal components.

In what follows, polynomial chaos expansions and Sobol’ sensitivity analysis are implemented on the coefficients of the 4 principal components. Polynomial chaos expansion is constructed from a tensor product basis of polynomials orthogonal with respect to the probability distribution of the random inputs of our problem. In the present problem, since the distribution of all input random variables is uniform, a basis composed of multivariate tensor products of Legendre polynomials in each of the input variables of Table 2 is employed. Namely, polynomial chaos expansion is sought in the form

$$
\hat{D}_{PCA}^h(X) = \sum_{m=1}^{N_{PCA}} \sum_{a \in A} c_a(m) \Psi_a(X) D_{PCA}^{(m)}
$$

with $X = \{x_1, \ldots, x_6\}$ denoting the random parameters of the micro-structure, $D_{PCA}^{(m)}$ and $m = \{1, 2, \ldots, N_{PCA}\}$ denoting the principal components. In our case $N_{PCA} = 4$.

The linear PCA approach adopted herein straightforwardly allows for the approximate reconstruction of the original stiffness tensors. The same approach was assessed in the context of health monitoring in [15].

The polynomial chaos expansion is computed by means of Least Angle Regression (LAR) [13]. The quality of the PCE least angle regression fit is measured with the generalized LOO error [16]. In Table 5 various parameters indicative of the quality of the PCE regression fit are summarized, separately for different principal components of the homogenized tensor. A visual account of the quality of the fit for all reconstructed stiffness tensor components, is demonstrated by plotting the reconstructed components against the original simulation data in Figure 3.

The performance of the fit is considered as satisfactory. In the next section the effect of the variability of the input variables to the homogenized stiffness is to be quantified by means of Sobol’ sensitivity indices. A set of histograms for the stiffness matrix component coefficients is given in Figure 4. These histograms were computed by sampling from the polynomial chaos surrogate with $10^4$ samples.

### 3.2 Sobol’ Sensitivity Analysis

As demonstrated in [17] it is possible to efficiently compute the Sobol’ global sensitivity indices through the coefficients of a polynomial chaos surrogate model. Sensitivity analysis is performed separately for each on of the 4 principal components of the PCA. The results are presented in Figure 5. It should be noted that the results of the sensitivity analysis on the $\lambda^{(m)}$ have a meaning that is not decoupled from the values of the principal components $D_{PCA}^{(m)}$ themselves (Table 4). In a setting where the principal components had an interpretable meaning such an analysis would have been more beneficial.

413
\[
D_{\text{PCA}}^{(1)} = \begin{bmatrix}
0.60 & 0.11 & 0.05 & 0 & 0 & 0 \\
0.11 & 0.46 & 0.05 & 0 & 0 & 0 \\
0.05 & 0.05 & 0.16 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.25 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.27 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.50 \\
\end{bmatrix}
\]

\[
D_{\text{PCA}}^{(2)} = \begin{bmatrix}
0.74 & -0.12 & -0.01 & 0 & 0 & 0 \\
-0.12 & -0.14 & 0.02 & 0 & 0 & 0 \\
-0.01 & -0.02 & -0.04 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -0.10 & 0 \\
0 & 0 & 0 & 0 & -0.04 & 0 \\
0 & 0 & 0 & 0 & 0 & -0.64 \\
\end{bmatrix}
\]

Table 4: First two principal components of the random homogenized stiffness tensor.

<table>
<thead>
<tr>
<th>Component</th>
<th>PCE-LOO Error</th>
<th>Normalized MSE</th>
<th>PCE Maximum Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda^{(1)})</td>
<td>(1.90e-3)</td>
<td>(1.6e-3)</td>
<td>6</td>
</tr>
<tr>
<td>(\lambda^{(2)})</td>
<td>(2.05e-3)</td>
<td>(1.7e-3)</td>
<td>5</td>
</tr>
<tr>
<td>(\lambda^{(3)})</td>
<td>(1.81e-3)</td>
<td>(1.3e-3)</td>
<td>5</td>
</tr>
<tr>
<td>(\lambda^{(4)})</td>
<td>(7.17e-2)</td>
<td>(5.0e-2)</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 5: PCE least-squares regression fit quality measures and maximum degree of expansion for the principal components with LAR.
Charilaos Mylonas, Valentin Bemetz and Eleni Chatzi

Figure 3: Quality of fit for the stiffness tensor. The tensor components are retrieved by using the PCE approximated PCA component coefficients with the principal component vectors.

Nevertheless, in our setting, it is clear that the angle of the $\pm \phi^\circ$ fibers is a significant factor, along with $V_{f_2}$ and the ratio of the volume fractions of $\pm \phi^\circ$ and $0^\circ$ fibers. It is interesting to observe, that the shape of the fibers, represented by $a_1, a_2, b_1, b_2$ has an almost negligible effect on the homogenization problem, at least for the range of variations considered in the present study. The high sensitivity index in the $4^{th}$ principal component is considered negligible, in light of the small contribution to the variance in the context of PCA of $\lambda^{(4)}$.

3.3 Conclusion

A framework for the construction of efficient and accurate polynomial surrogate models is presented for the problem of homogenization of parametrized, probabilistically modelled random microstructures. A limited budget of random Monte-Carlo runs is employed together with a non-intrusive surrogate modelling approach. Linear Principal Component Analysis was found sufficient for the data-driven dimensionality reduction of the random realizations of the stiffness tensor. Efficient PCE-based global sensitivity analysis was performed, yielding quantitative results on the effect of different random input parameters on the composite macro-scale response.

The utility of PCE models for the homogenization and localization problems is not limited to the gaining of a deeper insight on the effect of uncertainty of input parameters on homogenization through sensitivity analysis, as demonstrated in the present study. Although in the present work homogenization surrogates are exclusively presented, a rather simple extension in the same framework would pertain to the construction of surrogate models for the problem of micro-strain computation under uncertainty. This will form part of future investigations. The efficient solution of the stress localization problem efficiently is an important stepping stone towards the goal of highly efficient multi-scale damage prediction for composites of an arbitrary micro-structure.

Acknowledgement: The authors would like to gratefully acknowledge the support of the European Research Council via the ERC Starting Grant WINDMIL (ERC-2015-StG #679843) on the topic of Smart Monitoring, Inspection and Life-Cycle Assessment of Wind Turbines.
Figure 4: Histogram of coefficients of the homogenized stiffness matrix for the selected variation of parameters.

Figure 5: Sobol’ sensitivity indices for the different components of the PCA of the homogenized tensor data.
REFERENCES


CONSTRUCTION OF STATISTICALLY SIMILAR RVES FOR THE QUANTIFICATION OF UNCERTAINTY ASSOCIATED WITH THE MATERIAL’S MICROSTRUCTURE MORPHOLOGY

Daniel Balzani\textsuperscript{1,2}, Niklas Miska\textsuperscript{1}, and Stefan Prüger\textsuperscript{1}

\textsuperscript{1}Institute of Mechanics and Shell Structures
Technische Universität Dresden
August-Bebel-Str. 30, 01219 Dresden
e-mail: \{daniel.balzani, niklas.miska, stefan.prueger\}@tu-dresden.de

\textsuperscript{2}Dresden Center for Computational Materials Science
01062 Dresden

Keywords: Uncertainty Quantification, Statistically Similar Representative Volume Elements, Morphology, Microstructure, Dual Phase Steel.

Abstract. Various engineering applications require the use of modern materials. In particular with view to automotive applications, where increased safety standards at reduced weight are important, multiphase steels are advantageous. These steels make use of a pronounced microstructure in order to achieve a high ductility with high strength. The morphology of these microstructures varies over the location in the macroscopic part and over different specimen. As the macroscopic response of the steel is governed by the microstructure morphology, the randomness of the microscopic morphology implies uncertainties regarding the macroscopic material response. We propose to create statistically similar representative volume elements (SSRVE) to enable access to the incorporation of these uncertainties in numerical computations. The SSRVEs are obtained by minimizing a least-square functional consisting of higher order statistical measures, which describe the morphology of the microstructure. The resulting geometries are significantly less complex than the real microstructure and exhibit an advantage regarding meshing and computing the problem. Aside from these advantages, the method also provides a basis to construct various SSRVEs which are within predefined bounds regarding the microstructure statistics. These bounds may be obtained from measurements performed by analyzing the microstructures at different locations in one material. Based on this variety of applicable SSRVEs multiple Finite Element (FE) simulations can be performed to obtain the homogenized response and thus, to quantify statistics regarding macroscopic material parameters. In order to automatize these numerical simulations the Finite Cell Method (FCM) can be applied such that a conforming FE mesh for each of the SSRVEs is not needed to be constructed.
1 Introduction

Many modern materials make use of a pronounced microstructure to achieve advantageous properties. Advanced high strength steels (AHSS) for instance combine high strength with high ductility and are therefore used e.g. to create light-weight car bodies whilst enhancing the crash-safety. As the macroscopic behavior depends on the morphology of the microstructure, the variation of the morphology of the microstructure over different specimens or different locations in one specimen leads to an uncertainty of the material’s response on the macroscale. Thus, it may be necessary to consider this uncertainty in computations of AHSS (among other uncertainties, e.g. implied by variation of the material parameters). Nowadays most approaches to quantify the uncertainty regarding the microstructure use stochastic approaches, e.g. [10, 4], where Monte Carlo simulations are connected with extended finite elements (XFEM) to quantify uncertainties in effective properties or [6], in which random microstructures are combined with model reduction methods. For such analysis a large number of microstructures all representing possible real microstructures is required. The question in this context is how to obtain these microstructures, in particular with view to the fact that they will in most cases not be experimentally accessible. So far, most approaches make use of a rather random generation of artificial and idealized microstructures without taking into account the statistics of the real microstructure morphology. Therefore, here we propose a method based on such statistics which are experimentally measured to enable a suitable and representative set of microstructures. For this purpose we exploit the idea of statistically similar RVEs (SSRVEs) [1], which are obtained by minimizing a least-square functional defined in terms of differences of statistical measures evaluated for the real microstructure and the SSRVE. This contribution is organized as follows: Section 2 briefly introduces the concept of direct micro-macro-computations, before Section 3 describes the method to create the set of microstructures. In Section 4 a simple computational example is provided, Section 5 concludes this paper.

2 Micro-Macro Modeling of DP-Steel

In this contribution we focus on DP steel, whose microstructure consists of a ferrite matrix in which martensite inclusions are embedded. The ferrite matrix with its relatively low yield stress is responsible for the high ductility respectively good formability whilst the martensite inclusions with a higher yield stress act as a reinforcement and increase the strength of the DP steel. Since the different elasto-plastic properties of the microscopic constituents and their interaction govern the macroscopic response, it is advantageous to include the micro-heterogeneity in the modeling. A suitable direct micro-macro approach is the FE2 scheme, cf. e.g. [9]. In this scheme, the evaluation of the material law at every integration point of the macroscopic problem is replaced by homogenizing the microscopic stress distribution, which is in turn obtained by numerically solving a microscopic boundary value problem based on a representative volume element (RVE) in terms of Finite Elements. When suitable boundary conditions are applied to the RVE, the stresses and tangent moduli can be homogenized by

\[
\bar{\mathbf{P}} = \frac{1}{V} \int_B \mathbf{P} dV \quad \text{and} \quad \bar{\mathbf{A}} = \frac{1}{V} \int_B \mathbf{A} dV - \frac{1}{V} \mathbf{L}^T \mathbf{K}^{-1} \mathbf{L}, \quad \mathbf{L} = \int_B \mathbf{B}^T \mathbf{A} dV, \tag{1}
\]

cf. [12, 13] for further details. Herein, the microscopic 1st Piola-Kirchhoff stresses are denoted by \( \mathbf{P} \) and the associated moduli are \( \bar{\mathbf{A}} = \partial_{\mathbf{F}} \mathbf{P} \) with the deformation gradient \( \mathbf{F} \). The macroscopic counterparts are denoted by an overline. \( \mathbf{K}^{-1} \) is the inverse of the global stiffness matrix of the microscopic boundary value problem, \( \mathbf{B} \) is the classical Finite Element B-matrix.
Although the FE\(^2\) scheme replaces a constitutive material law for the composite material, constitutive material laws for the individual microscopic phases are still required to solve the microscopic boundary value problem. Here, we consider a classical finite strain elasto-plasticity formulation with isotropic exponential hardening [14], which is implemented in the material setting following [5]. This model is adjusted to experimental data obtained from mechanical tests, which were performed for the individual phases.

3 Method for the Quantification of Microstructure-Related Uncertainty

This section is devoted to the proposed method for the quantification of microstructure-based uncertainties of the DP steel properties. The method makes use of statistically similar representative volume elements such that variations regarding the morphology of microstructures can be taken into account.

The direct incorporation of the microstructure in FE simulations described in section 2 creates an additional problem: the complex geometry of the microstructure requires many elements to create a sufficient mesh, so that every microscopic boundary value problem ends in an expensive numerical computation with large demands not only with respect to the computing time but also with respect to the required memory. This is particularly important for any problem where a large number of microscopic boundary value problems has to be solved, i.e., for FE\(^2\)-simulations, but also for Monte-Carlo uncertainty quantification. A possible way out is proposed in [1, 11], where an artificial microstructure, the SSRVE, is constructed which is as similar as possible in a statistical sense to the real microstructure. This SSRVE is characterized by a less complex morphology, which therefore requires fewer degrees of freedom for the Finite Element discretization. The SSRVEs are obtained by optimizing a least square functional \(\mathcal{L}\), which is given by

\[
\mathcal{L}(\gamma) := \sum_{L} \omega_L \left( \mathcal{P}^{\text{real}}_L - \mathcal{P}^{\text{SSRVE}}_L(\gamma) \right)^2.
\]  

(2)

In Equation (2), the parameters characterizing the inclusions are collected in \(\gamma\) whereas \(\mathcal{P}^{\text{real}}\) and \(\mathcal{P}^{\text{SSRVE}}\) denotes a statistical descriptor evaluated for the real microstructure resp. the SSRVE. \(\omega_L\) represents a weighting factor for the individual statistical descriptors. For the evaluation of the statistical measures for the real microstructure, the required data can be obtained from segmenting three dimensional electron backscatter diffraction measurements, cf. [2]. For the representation of DP steel, in [1] it turns out that a parameterization with 3 ellipsoidal inclusions results in an acceptable mechanical error. Therefore, we also consider SSRVEs with 3 inclusions and incorporate the same statistical measures as in [1], namely the volume fraction \(\mathcal{P}(V)\), the spectral density \(\mathcal{P}(SD)\) and the lineal path function \(\mathcal{P}(LP)\).

In principle, if unlimited data regarding the 3D microstructure was available, the uncertainties resulting from variations of the microstructure morphology can be quantified by a Monte-Carlo calculation. Then, a rather large number of 3D microstructures from either different specimens or from different locations within one specimen is simulated and the discrete probability distribution of the resulting homogenized response is obtained. However, this unlimited data is typically not available due to the high costs associated with the measurement of 3D microstructures. Therefore, a method to construct artificial microstructures, which are still in a defined interval of acceptable statistical variations, is unavoidable.

Therefore, we propose an approach based on the least square functional used for the construc-
tion of SSRVEs. We assume, that every set of parameters $\gamma$, that yields a value $\mathcal{L}$ in an interval

$$\mathcal{L} \in \left[ \min[\mathcal{L}], \mathcal{L}_B \right],$$

represents an RVE with an admissible microstructure variation provided that $\mathcal{L}_B$ is suitably defined for the considered real microstructure. The acceptable range is limited by a boundary value $\mathcal{L}_B$, so that a set of statistically admissible artificial RVEs can be computed as

$$\{\gamma^{\text{SSVE}}\} = \arg[\mathcal{L}(\gamma) < \mathcal{L}_B].$$

The abbreviation SSRVE stands by definition for the best solution of the least square functional and thus for the RVE best matching the statistics of the complete real microstructure. All other considered parameters lead to artificial microstructures which are thus referred to as statistically similar volume elements (SSVE). All SSVEs in this set can now be used to quantify the uncertainty of macroscopic properties by performing a Monte-Carlo calculation of the homogenized properties. In detail this means that each SSVE is discretized in terms of Finite Elements and a suitable boundary value problem representing a mechanical test is solved to obtain the homogenized macroscopic response. From this, selected macroscopic properties can be computed.

The bound $\mathcal{L}_B$ may be obtained by incorporating some additional information regarding the real 3D microstructures, e.g. a larger measurement space. For this larger microstructure first a suitable RVE defined as sub-section of the real microstructure is found by choosing a reasonable RVE size and minimizing (2), where $\gamma$ solely represents the position of the sub-section. Then, $\mathcal{L}$ is computed for all possible sub-sections in order to identify $\mathcal{L}_B$, which is now the largest possible deviation from $\min \mathcal{L}$. If we relate one of these sub-sections to index $k$, the bound is obtained by

$$\mathcal{L}_B := \max |\mathcal{L}(\gamma_k)|.$$  

If sufficient microstructure information is available such that even the probability distribution could be computed based on the real microstructure, the computational evaluation of this large set of microstructures would still be challenging. Then, the method proposed here could be used instead to enable a calculation of significantly reduced computing time. Furthermore, the probability distribution of $\mathcal{L}$ in the real microstructure could be computed and an according set of SSVEs could be constructed using the Metropolis-Hastings algorithm.

4 Example

As a simplified example we consider the real microstructure of DP steel from [1] and compute the uncertainties on the macroscopic yield stress and hardening modulus, which are considered to result from microstructure variations. Since only limited information regarding the real 3D microstructure is available, the bound is set to $\mathcal{L}_B = 0.0075$, whilst the SSRVE is obtained for $\mathcal{L} = 0.005347$. While optimizing the least-square functional (2) to obtain the SSRVE using the framework Mystic and the differential evolution algorithm, cf. [8, 7], 11 evaluations were performed in which (4) was satisfied. For simplicity, we consider these as SSVEs for the representation of some microstructure variation. Figure 1 shows the SSRVE and one example of an SSVE. As it can be seen, the morphology of the microstructure is similar.

The associated parameters are transferred to Gambit, which is used to create the geometry and the mesh of the SSVEs, which are then transferred to FEAP 8.2, which is used to solve
the boundary value problem using periodic boundary conditions. All meshes are in the range of 25,000 to 30,000 elements and can thus be calculated rather efficiently. Here, only a macroscopic tension test in two different spatial directions is considered. The computations are strain driven and carried out until 10 percent elongation. Figure 2 shows the VON MISES stress plotted against the deformation gradient component in tensile direction. The curves splay in the hardening region due to the different morphologies. There is no spreading in the elastic regime, as the material model for the individual phases shares the same elastic properties and differs only in the hardening parameters. Additionally, the expected tiny degree of anisotropy as typically observed in DP steel is obtained. On the basis of the resulting stress strain curves effective material properties on the macroscale are computed. First, we analyze the effective yield stress \( R_{p,0.2} \), whose histograms are displayed for both tensile tests in Figure 3. Note that due to the intuitively chosen bound \( \mathcal{L}_B \) the obtained uncertainty is rather small, i.e. in a range of only a few MPa. Apparently this bound is not realistic for DP steels, however, it shows the principle
procedure proposed here. With the consideration of additional real microstructures and a larger bound a greater variance in the material parameters is expected. The same effect is noticeable for the hardening modulus at the macroscopic strain \( \bar{\varepsilon} = 1.1 \), for which the histogram is shown in Figure 4. A slight uncertainty is also found here due to the chosen bound.

5 Conclusion

In this paper a method was proposed, that is capable to quantify uncertainties of the material’s response due to variations in the microstructure morphology. It makes use of artificial RVEs, so that the demand of additional scans of the microstructure is reduced to a minimum. Here, a simplified example was considered in the sense that only few SSVEs were taken into account in the uncertainty quantification process. For future development it is planned to implement an automatic simulation procedure based on an extended Finite Cell Method [3]. Furthermore, the statistical variation bound \( L_B \) was chosen rather arbitrarily and thus, the quantitative results do not represent realistic variations in macroscopic properties. More microstructure information shall also be included, so that the bound can be extended to more realistic values. However, the example was able to demonstrate the principle procedure of the proposed approach which enables the construction of a statistically matching set of artificial microstructures to be used in the uncertainty quantification.

Acknowledgements

The authors greatly appreciate financial funding by the German Science Foundation (Deutsche Forschungsgemeinschaft, DFG) as part of the priority program “Polymorphic uncertainty modelling for the numerical design of structures” (SPP 1886), project BA2823/12-1, and as part of the Institutional Strategy “The Synergetic University” at TU Dresden.

REFERENCES


Stochastic Galerkin Projection and Numerical Integration for Stochastic Systems of Equations

Markus Wahlsten\textsuperscript{a}, Jan Nordström\textsuperscript{b}

\textsuperscript{a}Department of Mathematics, Computational Mathematics, Linköping University, SE-581 83 Linköping, Sweden (markus.wahlsten@liu.se).

\textsuperscript{b}Department of Mathematics, Computational Mathematics, Linköping University, SE-581 83 Linköping, Sweden (jan.nordstrom@liu.se).

Abstract

An incompletely parabolic system in three space dimensions with stochastic boundary and initial data is studied. The intrusive approach where we combine polynomial chaos with stochastic Galerkin projection is compared to the non-intrusive approach, where quadrature rules in combination with probability density functions of the prescribed uncertainties are used. The two methods are compared when calculating statistics for the compressible Navier–Stokes equations. As a measure of comparison, variance size, computational efficiency and accuracy are used.

Keywords: uncertainty quantification, stochastic data, polynomial chaos, stochastic Galerkin, intrusive methods, non-intrusive methods, Navier–Stokes equations.

1. Introduction

There are essentially two different methods for uncertainty quantification related to initial boundary value problems. Non-intrusive methods solve the original deterministic problem using a particular stochastic input. Standard quadrature techniques, often combined with sparse grid techniques can be used to obtain the statistics of interest. Intrusive methods are based on polynomial chaos (PC) expansions leading to a system of equations for the expansion coefficients. In contrast to the non-intrusive case, a new non-deterministic code must be developed.

The focus in this paper will be on the comparison in performance between numerical integration (NI) and PC with stochastic Galerkin (SG) approach.
The comparison is performed using the variance reducing boundary conditions derived in [1] and [2] for the compressible Navier–Stokes equations.

The rest of the paper proceeds as follows. In Section 2 a general continuous incompletely parabolic system of equations is introduced. Next, in Section 3, the PC-SG procedure is presented and the system of equations for the expansion coefficients is derived. The non-intrusive approach using NI is described in Section 4. Section 5 introduces a stable and accurate semi-discrete finite difference formulation of the two continuous problems. Further, the technique is applied to the compressible Navier–Stokes equations in Section 6 and a comparison of NI and PC-SG is done. Finally, conclusions are drawn in Section 7.

2. The continuous problem

Consider the following system of equations with stochastic boundary and initial data,

\[
\begin{align*}
    u_t + Au_x + Bu_y + Cu_z &= F_x(u) + G_y(u) + H_z(u), \quad \vec{x} \in \Omega, \quad t > 0, \\
    Lu &= g(\vec{x}, t, \vec{\xi}), \quad \vec{x} \in \partial \Omega, \quad t > 0, \\
    u &= f(\vec{x}, \vec{\xi}), \quad \vec{x} \in \Omega, \quad t = 0,
\end{align*}
\]

(1)

where,

\[
\begin{align*}
    F(u) &= D_{11}u_x + D_{12}u_y + D_{13}u_z, \\
    G(u) &= D_{21}u_x + D_{22}u_y + D_{23}u_z, \\
    H(u) &= D_{31}u_x + D_{32}u_y + D_{33}u_z.
\end{align*}
\]

(2)

We denote the solution by the vector \( u = u(\vec{x}, t, \vec{\xi}) = [u^{(1)}, \ldots, u^{(N)}] \), where, \( \vec{x} = (x, y, z) \), and \( \vec{\xi} = (\xi_1, \xi_2, \ldots, \xi_L) \) is the vector of variables representing the uncertainty in the solution. The \( N \times N \) matrices \( A, B, C, D_{ij} \) are constant and symmetric. The matrices \( D_{ij}, i, j = 1, 2, 3 \) are singular, leading to an incompletely parabolic problem. \( L \) is the boundary operator defined on \( \partial \Omega \), while \( f(\vec{x}, \vec{\xi}) \) and \( g(\vec{x}, t, \vec{\xi}) \) are the stochastic initial and boundary data.

The boundary operator can be written on the general form

\[
L = L^- - RL^+
\]

(3)

with a certain restriction on the matrix \( R \). With a slight abuse of notation we denote \( L^+u \) and \( L^-u \), the outgoing and ingoing generalized characteristic variables respectively. See [2, 3] for a complete derivation of \( L^+ \) and \( L^- \) and conditions on the matrix \( R \) leading to well-posedness.
3. Polynomial chaos expansion with stochastic Galerkin projection

The polynomial chaos framework considered in this paper is based on expansions of polynomials introduced by Ghanem and Spanos [4] and later generalized by Xiu and Karniadakis [5].

3.1. Polynomial chaos

Let \((\Omega_\xi, \mathcal{A}, \mathcal{P})\) be the probability space, where \(\Omega_\xi\) is the event space, \(\mathcal{A}\) the \(\sigma\)-field of subsets of \(\Omega_\xi\), and \(\mathcal{P}\) the probability measure [6]. Consider a general orthogonal chaos basis \(\{\psi_i(\xi)\}_{i=0}^{\infty}\), satisfying

\[
\langle \psi_i, \psi_j \rangle = \int_{\Omega_\xi} \psi_i(\xi)\psi_j(\xi)\,d\mathcal{P}(\xi) = \delta_{ij}.
\]  

(4)

A second order random field \(u(x, t, \xi)\) satisfying \(\int_{\Omega_\xi} u(x, t, \xi)^2\,d\mathcal{P}(\xi) < \infty\), can be written as

\[
u(x, t, \xi) = \sum_{k=0}^{\infty} u_k(x, t)\psi_k(\xi)
\]

(5)

where the coefficients \(\{u_k(x, t)\}_{k=0}^{\infty}\) are given by the projections

\[
u_k(x, t) = \langle u(x, t, \xi), \psi_k(\xi) \rangle, \quad k = 0, 1, \ldots
\]

(6)

The mean and variance can be expressed as

\[
\mathbb{E}[u] = \int_{\Omega_\xi} u\,d\mathcal{P}(\xi) = u_0,
\]

(7)

and

\[
\text{Var}[u] = \mathbb{E}[u^2] - \mathbb{E}[u]^2 = \int_{\Omega_\xi} \sum_{k=0}^{\infty} u_k^2\psi_k^2\,d\mathcal{P}(\xi) - u_0^2 = \sum_{k=1}^{\infty} u_k^2
\]

(8)

which follows from the orthogonality property of the basis functions \(\{\psi_i\}_{i=0}^{\infty}\). For more details on the polynomial chaos technique, see [7].

3.2. The stochastic Galerkin projection

Next, in order to approximately compute the various statistics of the solution to the problem, we insert the truncated expansion

\[
u(\vec{x}, t, \vec{\xi}) = \sum_{k=0}^{M} u_k(\vec{x}, t)\psi_k(\vec{\xi}),
\]

(9)
into (1), to get
\[
\sum_{k=0}^{M} (u_k)_t \psi_k + \sum_{k=0}^{M} A(u_k)_x \psi_k + \sum_{k=0}^{M} B(u_k)_y \psi_k + \sum_{k=0}^{M} C(u_k)_z \psi_k \\
= \sum_{k=0}^{M} F_x(u_k \psi_k) + \sum_{k=0}^{M} G_y(u_k \psi_k) + \sum_{k=0}^{M} H_z(u_k \psi_k),
\]
(10)

\[
\sum_{k=0}^{M} L u_k \psi_k = g(\vec{x}, t, \vec{\xi}),
\]
\[
\sum_{k=0}^{M} u_k \psi_k = f(\vec{x}, \vec{\xi}).
\]

Next, we perform a Galerkin projection, that is, we multiply (10) by \( \psi_l \), for \( l = 0, 1, \ldots, M \) and integrate over the stochastic domain \( \Omega \), to obtain
\[
\int_{\Omega} \sum_{k=0}^{M} (u_k)_t \psi_k \psi_l d\mathcal{P} + \int_{\Omega} \sum_{k=0}^{M} [A(u_k)_x \psi_k \psi_l + B(u_k)_y \psi_k \psi_l + C(u_k)_z \psi_k \psi_l] d\mathcal{P} \\
= \int_{\Omega} \sum_{k=0}^{M} F_x(u_k \psi_k \psi_l) + G_y(u_k \psi_k \psi_l) + H_z(u_k \psi_k \psi_l) d\mathcal{P},
\]
(11)
\[
\int_{\Omega} \sum_{k=0}^{M} L u_k \psi_k \psi_l d\mathcal{P} = \int_{\Omega} g(\vec{x}, t, \vec{\xi}) \psi_l d\mathcal{P},
\]
\[
\int_{\Omega} \sum_{k=0}^{M} u_k \psi_k \psi_l d\mathcal{P} = \int_{\Omega} f(\vec{x}, \vec{\xi}) \psi_l d\mathcal{P}, \text{ for } l = 0, 1, \ldots, M,
\]
which in compact notation can be written

\[ \sum_{k=0}^{M} (u_k)^t \langle \psi_k, \psi_l \rangle + \sum_{k=0}^{M} [A(u_k)_x + B(u_k)_y + C(u_k)_z] \langle \psi_k, \psi_l \rangle = \sum_{k=0}^{M} [F_x(u_k) + G_y(u_k) + H_z(u_k)] \langle \psi_k, \psi_l \rangle, \]  

(12)

\[ \sum_{k=0}^{M} L u_k \langle \psi_k, \psi_l \rangle = \langle g(\vec{x}, t, \vec{\xi}), \psi_l \rangle, \]

\[ \sum_{k=0}^{M} L u_k \langle \psi_k, \psi_l \rangle = \langle f(\vec{x}, \vec{\xi}), \psi_l \rangle, \quad \text{for } l = 0, 1, \ldots, M. \]

From the orthogonality property (4), (12) reduces to

\begin{align*}
(u_l)^t + A(u_l)_x + B(u_l)_y + C(u_l)_z &= F_x(u_l) + G_y(u_l) + H_z(u_l) \\
L u_l &= \langle g(\vec{x}, t, \vec{\xi}), \psi_l \rangle \\
u_l &= \langle f(\vec{x}, \vec{\xi}), \psi_l \rangle
\end{align*}

(13)

for \( l = 0, 1, \ldots, M. \) Hence, a deterministic system of dimension \( M + 1 \) times the original system is obtained. From (13), the deterministic coefficients \( u_0(\vec{x}, t), u_1(\vec{x}, t), \ldots, u_M(\vec{x}, t) \) are computed.

4. Numerical integration

The most commonly used non-intrusive technique, the Monte Carlo method \[8\], will not be considered here. The method is advantageous for large number of stochastic dimensions, but not competitive for low to moderate number of uncertainties. In this paper we consider numerical integration based on quadrature techniques.

Integration and quadrature relations in one dimension can be written

\[ \int_{a}^{b} f(q) \rho(q) dq \approx \sum_{m=1}^{M} f(q^m) w^m \]  

(14)

where \( f \) is the function we want to integrate and \( \rho \) is the corresponding density function. We denote \( M \) as the number of grid points, \( q^m \) and \( w^m \) as the quadrature points in probabilistic space and the corresponding weights respectively. The choice of quadrature points and weights determines the
accuracy of the method. For simplicity, we will use Simpson’s rule [9] as the integration technique in this paper.

The NI method can easily be extended to several dimensions, that is

$$\int_{\Gamma} f(q) \rho(q) \, dq \approx \sum_{m_1=1}^{M_1} \cdots \sum_{m_p=1}^{M_p} f(q_{m_1}^1, \ldots, q_{m_p}^p) w^{m_1} \cdots w^{m_p},$$

where $\Gamma$ is the $p$-dimensional domain and $\tilde{q} = (q_1, \ldots, q_p)$. The $q^m$’s and $w^r$’s correspond to the points and weights of their respective quadrature rule.

5. The semi-discrete formulation

Despite the fact that uncertainties are present in our model, the final problem formulation that arises from the stochastic Galerkin projection is strictly deterministic. We will solve (13) using a semi-discrete finite difference formulation based on the SBP–SAT technique [10, 11, 12, 13]. The reader is referred to [2, 3] for complete technical details.

A stable and accurate semi-discrete formulation of (13) on SBP–SAT form can be expressed as,

$$v_t + (I_M \otimes D_x \otimes I_y \otimes I_z \otimes A)v + (I_M \otimes I_x \otimes D_y \otimes I_z \otimes B)v + (I_M \otimes I_x \otimes I_y \otimes D_z \otimes C)v = (I_M \otimes I_x \otimes I_y \otimes I_N)F + (I_M \otimes I_x \otimes D_y \otimes I_z \otimes I_N)G + (I_M \otimes I_x \otimes I_y \otimes D_z \otimes I_N)H + (I_M \otimes P_x^{-1} E_{N_x} \otimes I_y \otimes I_z \otimes I_N)(\hat{\Sigma} - \hat{R}\hat{L}^+)v - \tilde{g} \otimes e_{N_x})$$

$$v(0) = \tilde{f}.$$

In (16), $v$ represent $u$ numerically, $\tilde{g}$ and $\tilde{f}$ are the numerical approximations of $< g(\vec{x}, t, \vec{\xi}), \psi_l >$, and $< f(\vec{x}, t, \vec{\xi}), \psi_l >$ for $l = 0, 1, \ldots, M$ respectively. To ease the notation, we only consider the boundary at $x = 1$. The treatment of the remaining boundaries is completely analogous. Note that the complete system (16) computes all coefficients and evaluations in $\xi$ for PC-SG and NI respectively.

**Remark 1.** The semi-discrete formulation is similar when computing the solution using NI. The differences will be pointed out below by formal remarks.
The numerical solution is arranged in the following way

\[ v = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_m \\ \vdots \\ v_M \end{bmatrix}, \quad [v_m] = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{N_x} \\ \vdots \\ v_{M} \end{bmatrix}, \quad [v_i]_m = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{N_y} \end{bmatrix}_{mi}, \]

\[ [v_{j}m] = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_k \\ \vdots \\ v_{N_x} \end{bmatrix}_{mij}, \quad [v_{n}m]_{ij} = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{N_y} \end{bmatrix}_{mijk} \]

where \( v_{mijn} \) approximates the polynomial chaos coefficient \( u_{mijn}^{(n)}(x, y, z, t) \).

**Remark 2.** The vector \( v_{mijn} \) approximates \( u^{(n)}(x, y, z, t, \xi_m) \) when computing the solution using NI.

The first derivative in the \( x, y \) and \( z \) direction is approximated by \( D_{x,y,z} = P_{x,y,z}^{-1}Q_{x,y,z} \), respectively. The matrices \( P_{x,y,z} \) are positive definite diagonal matrices and \( Q_{x,y,z} \) are almost skew-symmetric matrices satisfying \( Q_{x,y,z}^T = E_{N_x+N_y} - E_{0_{x,y,z}} = \mathcal{B} = \text{diag}[-1, 0, \ldots, 0, 1] \). Further, we denote \( I_x, I_y, I_z, I_N \) and \( I_M \) as identity matrices of dimension \( N_x + 1, N_y + 1, N_z + 1, N \) and \( M + 1 \). The matrices \( E_{0_x} \) and \( E_{N_x} \) are zero except for the element \((1, 1)\) and \((N_x + 1, N_x + 1)\) respectively, which is 1. Moreover, the vector \( e_{N_x} \) is a zero vector with the exception of the last element which is 1.

The numerical fluxes in (16) are given by

\[ F = (I_M \otimes \tilde{I} \otimes D_{11})v_x + (I_M \otimes \tilde{I} \otimes D_{12})v_y + (I_M \otimes \tilde{I} \otimes D_{13})v_z, \]

\[ G = (I_M \otimes \tilde{I} \otimes D_{21})v_x + (I_M \otimes \tilde{I} \otimes D_{22})v_y + (I_M \otimes \tilde{I} \otimes D_{23})v_z, \quad (17) \]

\[ H = (I_M \otimes \tilde{I} \otimes D_{31})v_x + (I_M \otimes \tilde{I} \otimes D_{32})v_y + (I_M \otimes \tilde{I} \otimes D_{33})v_z, \]

together with the notation \( \tilde{I} = (I_x \otimes I_y \otimes I_z) \) and the abbreviations

\[ v_x = (I_M \otimes D_x \otimes I_y \otimes I_z \otimes I_N)v, \]

\[ v_y = (I_M \otimes I_x \otimes D_y \otimes I_z \otimes I_N)v, \]

\[ v_z = (I_M \otimes I_x \otimes I_y \otimes D_z \otimes I_N)v. \quad (18) \]
The boundary and initial data \( \tilde{g} \) and \( \tilde{f} \), are the projections

\[
\tilde{g} = \begin{bmatrix}
< \tilde{g}(\xi), \psi_0(\xi) > \\
< \tilde{g}(\xi), \psi_1(\xi) > \\
\vdots \\
< \tilde{g}(\xi), \psi_M(\xi) >
\end{bmatrix}, \quad \tilde{f} = \begin{bmatrix}
< \tilde{f}(\xi), \psi_0(\xi) > \\
< \tilde{f}(\xi), \psi_1(\xi) > \\
\vdots \\
< \tilde{f}(\xi), \psi_M(\xi) >
\end{bmatrix}.
\tag{19}
\]

In (19), we denote \( \tilde{g}(\xi) \) and \( \tilde{f}(\xi) \) as the original boundary and initial data vector as a function of \( \xi \) injected on all grid points at the plane \( x = 1 \) and \( t = 0 \) respectively. The inner products \( < \tilde{g}(\xi), \psi_m(\xi) > \) and \( < \tilde{f}(\xi), \psi_m(\xi) > \) are computed numerically using NI.

**Remark 3.** When computing the solution using NI, the vectors \( \tilde{g} \) and \( \tilde{f} \) instead denote \( \tilde{g} = [\tilde{g}(\xi_0), \tilde{g}(\xi_1), \ldots, \tilde{g}(\xi_M)]^T \) and \( \tilde{f} = [\tilde{f}(\xi_0), \tilde{f}(\xi_1), \ldots, \tilde{f}(\xi_M)]^T \). Hence, the boundary and initial data \( \tilde{g}(\xi) \) and \( \tilde{f}(\xi) \) are discretized in \( \xi \).

The discrete boundary operators \( \tilde{L}^+ \) and \( \tilde{L}^- \) are decomposed as

\[
\tilde{L}^+ = \left( I_M \otimes I_x \otimes I_y \otimes I_z \otimes L_0^+ \right) + \left( I_M \otimes D_x \otimes I_y \otimes L_0^+ \right) + \left( I_M \otimes D_y \otimes I_z \otimes L_0^+ \right) + \left( I_M \otimes D_z \otimes L_0^+ \right)
\]

\[
\tilde{L}^- = \left( I_M \otimes I_x \otimes I_y \otimes I_z \otimes L_0^- \right) + \left( I_M \otimes D_x \otimes I_y \otimes I_z \otimes L_0^- \right) + \left( I_M \otimes D_y \otimes I_z \otimes L_0^- \right) + \left( I_M \otimes D_z \otimes L_0^- \right) \tag{20}
\]

In (20), the matrices \( L_0^+, L_0^- \), \( L_0^{+x} \), \( L_0^{-x} \), \( L_0^{+y} \), \( L_0^{-y} \), \( L_0^{+z} \), and \( L_0^{-z} \) correspond to the continuous boundary operators, see [2, 3] for a complete description. In a similar fashion, the matrix \( \tilde{R} \) is defined as

\[
\tilde{R} = (\tilde{I} \otimes R \otimes I_M). \tag{21}
\]

where \( R \) is the matrix corresponding to the continuous boundary conditions (3). The penalty matrix \( \tilde{\Sigma} \) is chosen such that stability is achieved. Again, for a complete derivation and proof of stability for the semi-discrete formulation, the reader is referred to [2, 3]

6. **An example**

To exemplify the difference between NI and PC-SG, we consider the linearized symmetrized Navier–Stokes equations in one dimension [14],

\[
U_t + AU_x = \epsilon BU_{xx}, \quad 0 < x < 1, \quad t > 0, \tag{22}
\]
where

\[
B = \text{diag} \left( 0, \frac{\lambda + 2\mu}{\bar{\rho}}, \gamma \mu \right). \tag{23}
\]

We also use

\[
A = X \Lambda X^T,
\]

where

\[
\Lambda = \begin{bmatrix}
\bar{u} & 0 & 0 \\
0 & \bar{u} + \bar{c} & 0 \\
0 & 0 & \bar{u} - \bar{c}
\end{bmatrix}, \quad
X = \begin{bmatrix}
-\sqrt{\frac{\gamma - 1}{\gamma}} & \frac{1}{\sqrt{2\gamma}} & \frac{1}{\sqrt{2\gamma}} \\
0 & \frac{1}{\sqrt{2\gamma}} & -\frac{1}{\sqrt{2\gamma}} \\
\frac{1}{\sqrt{\gamma}} & \sqrt{\frac{\gamma - 1}{2\gamma}} & \sqrt{\frac{\gamma - 1}{2\gamma}}
\end{bmatrix}. \tag{24}
\]

The parameters in (22), (23) and (24) are \(\mu, \lambda, \text{Pr}, \epsilon\) and represent the dynamic and second viscosity, the Prandtl number and the inverse Reynolds number. In the calculations we use the following parameters

\[
\bar{\rho} = 1, \quad \bar{u} = 1, \quad \bar{c} = 2, \quad \bar{\rho} = 1, \quad \lambda = -2/3, \quad 
\gamma = 1.4, \quad \epsilon = 0.01, \quad \text{Pr} = 0.7, \quad \mu = 1. \tag{25}
\]

The boundary conditions are of the form (3), where the matrices \(R_0\) and \(R_1\) are of sizes 3 \(\times\) 2 and 2 \(\times\) 3 defined on the boundaries \(x = 0\) and \(x = 1\) respectively. For simplicity, we use zero matrices for \(R_0\) and \(R_1\).

Randomness is imposed in the initial and boundary data given by the manufactured solutions

\[
\rho_1 = u_1 = p_1 = \sin(2\pi(x - t)) + \xi \sin(\xi/2 - t) \cos(x - \xi),
\]

\[
\rho_2 = u_2 = p_2 = \sin(2\pi(x - t)) + \sin(5(\xi/2 - t)) \cos(2(x - 5\xi)),
\]

with the additional forcing function

\[
F = W_t + AW_x - \epsilon BW_{xx}, \quad W = [\rho_{1,2}, u_{1,2}, p_{1,2}]^T. \tag{26}
\]

We have chosen the manufactured solutions with subscript 1 and 2 in (26) to illustrate the effects of smoothness. The solution with subscript 1 is considered to be smooth, while 2 is considered to be less smooth. The resulting system is then

\[
U_t + AU_x = \epsilon BU_{xx} + F, \quad 0 < x < 1, \quad t > 0,
\]

\[
L_0 U = L_0 W \quad \text{for} \quad x = 0, \quad t > 0
\]

\[
L_1 U = L_1 W \quad \text{for} \quad x = 1, \quad t > 0
\]

\[
U = W \quad 0 < x < 1, \quad t = 0
\]

\[
(28)
\]
The normed error of the variance that is used for comparison is given by

\[ \| \text{Var}[e] \| = \int_0^T \| \text{Var}[U] - \text{Var}[W] \|_2 \, dt. \]  

(29)

In the calculations, we use 3rd-order SBP-operators with 40 grid points in space together with the 4th-order Runge-Kutta scheme as time integrator. To minimize the effects of the deterministic errors we use a PC-SG computation with 30 basis functions instead of an exact solution as reference solution. The uncertainty \( \xi \) is uniformly distributed between \(-1 \) and \(1\).

Figure 1 and 2 show the normed error of the variance as a function of number of evaluations \( (M) \) for PC-SG and NI using different manufactured solutions. Note that NI requires at least three points, hence it does not overlap with PC-SG. As can be seen, the rate of convergence is significantly higher for the PC-SG method than for NI. The rate of convergence for the NI (Simpson’s rule) is in line with theory, that is 4th order. Also, we note that the PC-SG requires more evaluations to perform better than NI.

Figure 3 and 4 illustrate how the total CPU time is affected by the number of evaluations for both methods. Finally, Figure 5 and 6 present the normed error of the variance as a function of CPU time. For a small number of evaluations, the results indicate a better performance for NI, however for moderate and large number of evaluations PC-SG performs significantly better.

7. Conclusions

A comparison between the intrusive PC-SG and non-intrusive NI have been presented. The study has been carried out on a general incompletely parabolic system of equations. The PC-SG procedure was applied to the continuous problem and a provably stable numerical formulation based on SBP-SAT was constructed. Due to linearity, the numerical scheme for the PC-SG and NI differs only for the initial and boundary data treatments.

The linearized Navier–Stokes equations using generalized characteristic boundary conditions are considered when comparing the two methods. As comparison, the normed error of the variance was used. The numerical results indicate that the PC-SG outperforms NI for the problems considered. Although the cost is higher for the PC-SG, the convergence is faster. The effects are significant for larger number of realizations. We conclude that
Figure 1: The normed error of the variance as a function of $M$ for a uniformly distributed $\xi$ for the methods NI and PC-SG, using a smooth manufactured solution.

Figure 2: The normed error of the variance as a function of $M$ for a uniformly distributed $\xi$ for the methods NI and PC-SG, using a less smooth manufactured solution.
Figure 3: The total computation time as a function of $M$ for a uniformly distributed $\xi$ for the methods NI and PC-SG, using a smooth manufactured solution.

Figure 4: The total computation time as a function of $M$ for a uniformly distributed $\xi$ for the methods NI and PC-SG, using a less smooth manufactured solution.
Figure 5: The normed error of the variance as a function of computation time for a uniformly distributed $\xi$ for the methods NI and PC-SG, using a smooth manufactured solution.

Figure 6: The normed error of the variance as a function of computation time for a uniformly distributed $\xi$ for the methods NI and PC-SG, using a less smooth manufactured solution.
the reduced performance for less smooth problems is more pronounced for PC-SG compared to NI.

References


APPROACH TO PROVE THE EFFICIENCY OF THE MONTE CARLO METHOD COMBINED WITH THE ELEMENTARY EFFECT METHOD TO QUANTIFY UNCERTAINTY OF A BEAM STRUCTURE WITH PIEZO-ELASTIC SUPPORTS

Sushan Li¹, Benedict Götz¹, Maximilian Schaeffner¹, and Roland Platz²

¹ Research group System Reliability, Adaptive Structures, and Machine Acoustics SAM
TU Darmstadt
Magdalenenstr. 4, Darmstadt, Germany
e-mail: {sLi, goetz, schaeffner}@sam.tu-darmstadt.de

² Fraunhofer Institute for Structural Durability and System Reliability LBF
Bartningstr. 47, Darmstadt, Germany
e-mail: roland.platz@lbf.fraunhofer.de

Keywords: Uncertainty quantification, Monte Carlo method, Elementary Effect method, beam, piezo-elastic support

Abstract. In this paper, a new approach is presented to prove the efficiency of the direct Monte Carlo method combined with the Elementary Effect method to quantify structural data uncertainty under uncertain input parameters of a beam structure. Normally, the application of the direct Monte Carlo method requires high computational cost when all input parameters are taken into account. It is proposed to use a combination of the direct Monte Carlo method and the Elementary Effect method for the variance-based sensitivity analysis, named the combined Monte Carlo method. By the application of the Elementary Effect method as a screening method, the truly influential input parameters are identified. Then, the parametric uncertainty is analyzed only under these influential input parameters’ uncertainty by the use of the Monte Carlo method. Through a combination of these two methods, the number of simulations can be significantly reduced due to the reduction of the number of analyzed input parameters.

The novelty of this paper is to investigate the accuracy and the efficiency of this combined approach by the use of a beam structure with piezo-elastic supports for buckling and vibration control as a reference structure. The uncertain structural input parameters are the geometric, material, and stiffness parameters of the piezo-elastic supports. The output variable is the first lateral resonance frequency of the beam structure. Its uncertainty will be analyzed by the application of the combined Monte Carlo method applied for only a few but influential input parameters and will also be analyzed by the application of the direct Monte Carlo method for all input parameters. The results by the two methods will be compared based on the analysis accuracy to estimate the sensitivity of the input parameters on the first lateral resonance frequency and the minimal required number of the simulations.
1 INTRODUCTION

Global sensitivity analysis shows how the uncertainty of a structure is apportioned to the uncertainty of its input parameters by quantifying the relative importance of each input parameter [1]. The underlying goal of the sensitivity analysis is to identify the influential input parameters and then, through controlling and reducing the uncertainty of the influential input parameters, to reduce the uncertainty of the structure [2], e.g., to reduce the uncertainty of the structural vibration amplitude.

The methods that are used for global sensitivity analysis can be divided into two categories, the qualitative analysis methods and the quantitative analysis methods [2]. The qualitative analysis methods, also named screening methods, are feasible to identify if the input parameters have relevant influence on the structure or not [3]. Commonly, they identify the relevant influence based on a small number of simulations. In most cases, “a small number” implies that the simulations do not result in convergence or trustworthy outcomes. Similarly, “a relatively large number” implies that the simulations do converge and trustworthy outcomes are reached, but usually with high costs in simulation time. Therefore, a small number of simulations in the qualitative analysis is commonly not sufficient for ranking the relevant influence of the input parameters on the structure. In contrast, the quantitative analysis methods aim at precisely quantifying and ranking the influence of the input parameters on the structure, based on a relatively large number of simulations [3]. Hence, an approach by combination of the qualitative and the quantitative analysis method can be an efficient approach in sensitivity analysis.

The global sensitivity analysis methods are classified in terms of required number of simulations and the structure’s complexity in [2]. Based on this classification it can be concluded that the Elementary Effect method, which was proposed by Morris in 1991 [4], is a simple but effective screening method [5]. The studies in [6, 7] also support this statement. In the category of the quantitative analysis methods, the Monte Carlo method for variance-based sensitivity analysis is more feasible than other methods, e.g., the Fourier Amplitude Sensitivity Test (FAST), as it requires no assumption and considers not only the linear relation between the input parameters and output variable but also the nonlinear relation [8]. The studies in [9, 10] also support this statement. In this study, the Elementary Effect method is applied at first for screening the sensitivity of each input parameter to identify the input parameters having dynamic influence on the structure. Based on the screening results, the parametric uncertainty is analyzed only under the uncertainty of the influential input parameters by the Monte Carlo method for variance-based sensitivity analysis. The idea of the combination of these two methods is to reduce the number of simulations without losing the accuracy of the uncertainty analysis. This approach was recommended by Saltelli in [5]. The efficiency of this approach will be tested on a real beam structure subject to buckling and vibration control under axial loading with piezo-elastic supports.

The beam structure is described in section 2 and modeled mathematically in section 3. The Elementary Effect method, the direct Monte Carlo method, and the combined Monte Carlo method are introduced in section 4. The Elementary Effect method is applied to the beam structure and the screening results are described in section 5. The sensitivity analysis results of the beam structure based on the direct Monte Carlo method and the combined Monte Carlo method are also described and compared in section 5 with respect to their analysis accuracy and computational costs. The goal of this paper is to investigate if the efficiency of the quantification of uncertainty is improved through the combination of the sensitivity analysis methods without losing the accuracy of the uncertainty analysis.
2 STRUCTURE DESCRIPTION

The reference structure for the application of the global sensitivity analysis is a beam with piezo-elastic supports that is used for two different applications of passive and active structural control: active buckling control [11] and passive lateral vibration attenuation [12]. Figure 1 shows the beam structure with piezo-elastic supports and the sectional view of the piezo-elastic support. It consists of a membrane-like spring element, an axial extension of the beam, and two piezoelectric stack transducers, which are mechanically prestressed by a stack of disc springs [13].

![Figure 1: (a) CAD model of the beam with piezo-elastic supports, (b) Sectional view of piezo-elastic support B](image)

The various components of the piezo-elastic supports may each influence the dynamic behavior of the beam structure. The input parameters of the piezo-elastic supports that have significant influence on the dynamic behavior of the beam structure will be identified. The first lateral resonance frequency $f_1$ is an important measure for active buckling control and passive lateral vibration attenuation. Therefore, it is chosen as the output variable in this study for the global sensitivity analysis.

The mechanical sketch of the beam structure with piezo-elastic supports is shown in Figure 2. The beam is made of aluminum alloy EN AW-7075 with Young’s modulus $E_b$, density $\rho_b$, length $l_b$, and circular solid cross-section of radius $r_b$. The circumferential lateral stiffness is homogeneous and has no preferred direction of lateral deflection, so the beam may deflect in any plane lateral to the longitudinal $x$-axis.

Two piezo-elastic supports A and B are located at $x = 0$ mm and $x = l_b$. The elastic membrane-like spring elements made of spring steel 1.1248 in both supports A and B are represented by lateral stiffness $k_{y,A} = k_{z,A} = k_{l,A}$ and $k_{y,B} = k_{z,B} = k_{l,B}$ in $y$- and $z$-direction and...
rotational stiffness $k_{\varphi,y,A} = k_{\varphi,z,A} = k_{t,A}$ and $k_{\varphi,y,B} = k_{\varphi,z,B} = k_{t,B}$ around the $y$- and $z$-axes. The spring stiffness values are assumed to be equal in $y$- and $z$-direction for this study since both directions are modeled symmetric and independent of each other. The spring element stiffness values are obtained from experimental static load measurements.

In each piezo-elastic support A and B at $x = -l_{\text{ext}}$ and $x = l_b + l_{\text{ext}}$, two piezoelectric stack transducers $P_1$ and $P_2$ as well as $P_3$ and $P_4$ are arranged in the support housing at an angle of 90° to each other orthogonal to the beam’s $x$-axis, Figure 2. All transducers are mechanically prestressed by a stack of disc springs with stiffness $k_{\text{pre}}$. The value of $k_{\text{pre}}$ is smaller than 5% of the stiffness of the piezoelectric transducers. Therefore, its influence is neglected in this study. The transducers in supports A and B are connected to the beam via relatively stiff axial extensions made of hardened steel 1.2312 with Young’s moduli $E_{\text{ext},A}$ and $E_{\text{ext},B}$, densities $\rho_{\text{ext},A}$ and $\rho_{\text{ext},B}$, lengths $l_{\text{ext},A}$ and $l_{\text{ext},B}$, as well as edge lengths $t_{\text{ext},A}$ and $t_{\text{ext},B}$. With that, lateral beam deflections in $y$- and $z$-direction are transformed into the stack transducer’s axial deformation in $y$- and $z$-direction and vice versa. The piezoelectric transducers $P_1$ and $P_2$ in support A and $P_3$ and $P_4$ in support B are P-885.51 stack transducers by PI Ceramic with mechanical stiffness $k_{p,A}$ and $k_{p,B}$ [12].

The sensitivity analysis identifies the input parameters of the piezo-elastic supports that have significant influence on the first lateral natural frequency of the beam structure. For that, only the geometric and material properties of the beam, $E_b$, $\rho_b$, $l_b$, $r_b$, are constant, Table I. The
geometric, material, and stiffness parameters of the components in the piezo-elastic supports are assumed uncertain parameters $p_k$ with $k = 1, 2, \ldots, K = 14$, Table 1. Their assumed normal or uniform distribution functions $\mathcal{N}$ and $\mathcal{U}$ are also given in Table 1. For the normal distribution function $\mathcal{N}(\mu; \sigma)$, the mean $\mu$ and standard deviation $\sigma$ are given and for the uniform distribution $\mathcal{U}(\text{min}; \text{max})$, the minimum and maximum values are given.

<table>
<thead>
<tr>
<th>input parameters</th>
<th>deterministic value</th>
<th>distribution function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_b$</td>
<td>beam Young’s modulus in GPa</td>
<td>73.9</td>
</tr>
<tr>
<td>$\rho_b$</td>
<td>beam density in kg/m$^3$</td>
<td>2 775</td>
</tr>
<tr>
<td>$l_b$</td>
<td>beam length in mm</td>
<td>400</td>
</tr>
<tr>
<td>$r_b$</td>
<td>beam radius in mm</td>
<td>5</td>
</tr>
<tr>
<td>$E_{\text{ext,A}}$</td>
<td>axial extension Young’s modulus in GPa</td>
<td>210</td>
</tr>
<tr>
<td>$\rho_{\text{ext,A}}$</td>
<td>extension density in kg/m$^3$</td>
<td>7800</td>
</tr>
<tr>
<td>$l_{\text{ext,A}}$</td>
<td>extension length in mm</td>
<td>7.8</td>
</tr>
<tr>
<td>$r_{\text{ext,A}}$</td>
<td>extension radius in mm</td>
<td>6</td>
</tr>
<tr>
<td>$E_{\text{ext,B}}$</td>
<td>axial extension Young’s modulus in GPa</td>
<td>210</td>
</tr>
<tr>
<td>$\rho_{\text{ext,B}}$</td>
<td>extension density in kg/m$^3$</td>
<td>7800</td>
</tr>
<tr>
<td>$l_{\text{ext,B}}$</td>
<td>extension length in mm</td>
<td>7.8</td>
</tr>
<tr>
<td>$r_{\text{ext,B}}$</td>
<td>extension radius in mm</td>
<td>6</td>
</tr>
<tr>
<td>$k_{1,A}$</td>
<td>spring lateral stiffness in N/m</td>
<td>22 · 10$^6$</td>
</tr>
<tr>
<td>$k_{1,B}$</td>
<td>spring rotational stiffness in N · m/rad</td>
<td>449.7</td>
</tr>
<tr>
<td>$k_{3,B}$</td>
<td>spring lateral stiffness in N/m</td>
<td>22 · 10$^6$</td>
</tr>
<tr>
<td>$k_{3,B}$</td>
<td>spring rotational stiffness in N · m/rad</td>
<td>449.7</td>
</tr>
<tr>
<td>$k_{p,A}$</td>
<td>piezo transducer A lateral stiffness in N/m</td>
<td>49 · 10$^6$</td>
</tr>
<tr>
<td>$k_{p,B}$</td>
<td>piezo transducer B lateral stiffness in N/m</td>
<td>49 · 10$^6$</td>
</tr>
</tbody>
</table>

Table 1 summarizes the constant and varied geometric, material, and stiffness properties for the following study. The input parameters relating to the axial extensions are $p_1$–$p_8$. Due to the manufacturing tolerances the geometric properties $p_3$, $p_4$, $p_7$, as well as $p_8$ are assumed uniformly distributed. The variation of the material properties $p_1$, $p_2$, $p_5$, and $p_6$ are assumed normally distributed. Their distribution functions are defined according to the tolerance’s specifications and material data found in literature. The input parameters describing the stiffness of the membrane-like spring elements are $p_9$–$p_{12}$. The spring elements are produced by an actively controlled forming process. Thus, the variation of the spring stiffness properties are assumed to be normally distributed. Their distribution functions are defined according to the authors’ static stiffness measurements. Finally, the lateral stiffnesses of the piezoelectric stack transducers are $p_{13}$ and $p_{14}$. As they depend on the material properties, they are assumed normally distributed and their distribution functions are defined according to manufacturer data by PI Ceramic [14].

3 MATHEMATICAL MODEL

To calculate the first lateral resonance frequency $f_1$ of the beam, a finite element (FE) model was derived in [11]. The free vibration of the beam with piezo-elastic supports in Figure 2 is modeled by the homogeneous FE equation of motion

$$M \ddot{r} + K r = 0.$$
Damping in equation (1) is neglected because it is insignificantly small according to the authors’ experimental measurements. In equation (1), \( \mathbf{r} \) is the \([4I \times 1]\) FE displacement vector for a beam discretized by \((I - 1)\) elements with \(4I\) degrees of freedom, two translational and two rotational displacements in and around \(y\)- and \(z\)-direction. \( \mathbf{M} \) is the \([4I \times 4I]\) mass matrix and \( \mathbf{K} \) is the \([4I \times 4I]\) stiffness matrix. The discrete stiffness parameters of the elastic membrane springs, disk springs, and piezoelectric stack actuators are included in stiffness matrix \( \mathbf{K} \).

The first lateral resonance frequency of the beam \( f_1 \) is calculated by the solution of the characteristic equation

\[
\det [\mathbf{K} - (2\pi f_1)^2 \mathbf{M}] = 0.
\]

(2)

The influence of the \( K \) varied input parameters, which are contained in the \([1 \times K]\) vector \( \mathbf{p} = [p_1, p_2, \cdots, p_k, \cdots, p_K] \) with \( k = 1, 2, \cdots, K = 14 \), on the output parameter, which is the first lateral resonance frequency \( f_1 \) in equation (2), will be identified by the sensitivity analysis.

4 METHODS FOR SENSITIVITY ANALYSIS

As already mentioned, the goal of this study is to prove the accuracy and the efficiency of the combination of the Elementary Effect method and the Monte Carlo method for the variance-based sensitivity analysis according to [5] for adequate sensitivity analysis. The Elementary Effect method, the direct Monte Carlo method for the variance-based sensitivity analysis, and the combined Monte Carlo method for the variance-based sensitivity analysis are briefly introduced in this section.

4.1 Screening method: the Elementary Effect method

The Elementary Effect method is a commonly used screening method as it can identify the influential input parameters of most structures [5]. After sampling and numerical simulations, the elementary effect of each input parameter is calculated. For analyzing a structure with \( K \) independent input parameters \( p_k \), \( k = 1, 2, \cdots, K \), which spans a \( K \)-dimensional input space \( \Omega^K \), the Elementary Effect method identifies the input parameters having an influence on the output variable by comparing the mean and the standard deviation of the elementary effects of each input parameter.

For a structure with \( K \) input parameters, the sampling process for the Elementary Effect method begins by randomly choosing simulation points \( \mathbf{p}_{r,0} = [p_{r,0,1}, p_{r,0,2}, \cdots, p_{r,0,K}] \) in \( \Omega^K \) with \( r = 0, 1, 2, \cdots, R \). These simulation points are used as base points and \( R \) represents the number of the base points used in the screening. Based on the base point \( \mathbf{p}_{r,0} \), simulation points \( \mathbf{p}_{r,k} = [p_{r,k,1}, p_{r,k,2}, \cdots, p_{r,k,K}] \) are generated by randomly varying the \( k \)-th input parameter for \( k = 1, 2, \cdots, K \) inside \( \Omega^K \). As the Elementary Effect method is used for parameter screening, it is suggested that the number of the base points \( R = 10 \) is sufficient to produce valuable results [5].

The sampling process is illustrated by using a structure with only two input parameters \( K = 2 \) as an example, Figure 3.

- For \( r = 1 \),
  - Step 1.0: The first base point \( \mathbf{p}_{1,0} \) is randomly chosen in \( \Omega^K \),
  - Step 1.1: based on \( \mathbf{p}_{1,0} \), the input parameter \( p_1 \) is varied with a random length inside \( \Omega^K \) for the simulation point \( \mathbf{p}_{r,k} = \mathbf{p}_{1,1} \).
Step 1.2: based on $p_{1,0}$, the input parameter $p_2$ is varied with a random length inside $\Omega^K$ for the simulation point $p_{r,k} = p_{1,2}$;

- For $r = 2$,
  Step 2.0: The second base point $p_{2,0}$ is randomly chosen in $\Omega^K$,
  Step 2.1: based on $p_{2,0}$, the input parameter $p_1$ is varied with a random length inside $\Omega^K$ for the simulation point $p_{r,k} = p_{2,1}$.
  Step 2.2: based on $p_{2,0}$, the input parameter $p_2$ is varied with a random length inside $\Omega^K$ for the simulation point $p_{r,k} = p_{2,2}$;

- For $r = 3$,
  Step 3.0: The third base point $p_{3,0}$ is randomly chosen in $\Omega^K$,
  Step 3.1 and step 3.2: repeat the processes as in steps 1.1 and 1.2 but based on the third base point $p_{3,0}$;

... ;

- For $r = R = 10$,
  Step 10.0: The tenth base point $p_{10,0}$ is randomly chosen in $\Omega^K$, also the last base point,
  Step 10.1 and step 10.2: repeat the processes as in steps 1.1 ... 1.2 but based on the tenth base point $p_{10,0}$.

For a structure with $K$ input parameters, based on each base point $p_{r,0}$, $K$ simulation points $p_{r,1}, p_{r,2}, \ldots, p_{r,K}$ are sampled. Hence, the number of simulations are $N_{EE} = R \cdot (1 + K)$. The function $f(p)$ describes the relation between the input parameters $p = [p_1, p_2, \ldots, p_K]$ and the output variable of the structure. Therefore, the results at the base points $p_{r,0}$ are $f(p_{r,0})$ and the results at the simulation points $p_{r,k}$ are $f(p_{r,k})$. In this study the function $f(p)$ is evaluated according to equations (2) with $f(p) = f_1$. The simulations provide $R \cdot K$ elementary effects $EE_r^k$; one elementary effect per input parameter $p_k$ based on each base point. The elementary effects $EE_r^k$ of each input parameter are

$$EE_r^k = \frac{f(p_{r,k}) - f(p_{r,0})}{\Delta}, \quad (3)$$

$$\Delta = \frac{p_{r,k,k} - p_{r,0,k}}{p_k_{\max} - p_k_{\min}}. \quad (4)$$
4.2 Quantitative method: the direct Monte Carlo simulation for variance-based sensitivity analysis

Variance-based sensitivity analysis is a widely used approach to quantify the influence of the $K$ input parameters $p = [p_1, p_2, \ldots, p_k, \ldots, p_K]$ on the output variable $f(p)$. The direct Monte Carlo method is used for the variance-based sensitivity analysis to generate a large number of random samples to obtain numerical results without prior knowledge of the investigated structure. To reduce the number of samples, the direct Quasi Monte Carlo method is used as a sub-method of the direct Monte Carlo method [8]. It uses a Quasi random number generator, e.g., the Latin-Hypercube generator or the Sobol’ generator. The Quasi random number generators are designed to generate random samples for a faster convergence. The drawback of the Latin-Hypercube generator is that its algorithm for the randomness of the input parameters is dependent on the sample size. The sample size cannot be changed without starting the simulation again from the beginning. In contrast, the Sobol’ random number generator can freely

\[ \hat{\mu}_k = \frac{1}{R} \sum_{r=1}^{R} |EE_k^r|, \tag{5} \]

\[ \hat{\sigma}_k = \sqrt{ \frac{1}{R-1} \sum_{r=1}^{R} (EE_k^r - \hat{\mu}_k)^2}. \tag{6} \]

The mean $\hat{\mu}_k$ of the elementary effect for the $k$-th input parameter is a sensitivity measure proposed to assess the overall influence of the $k$-th input parameter on the output variable [6]. A relatively high value of $\hat{\mu}_k$ indicates that the $k$-th input parameter has a significant influence on the result of the calculated output variable. In contrast, a relatively low value of $\hat{\mu}_k$ implies that the $k$-th input parameter has only minor influence on the result of the output variable.

Another measure is the standard deviation $\hat{\sigma}_k$ of the elementary effect for the $k$-th input parameter, which estimates whether the effect of the $k$-th input parameter is linear or nonlinear or whether this input parameter has a interacted functional relation with other input parameters to calculate the output variable or not [4, 6]. In case the $k$-th input parameter’s functional relation is linear, the elementary effects would be identical everywhere in the input space $\Omega^K$. Therefore, $\hat{\sigma}_k$ is close to 0. In contrast, if the input parameter’s functional relation is nonlinear and/or interacts with other input parameters, the elementary effects would vary in the input space $\Omega^K$. As a result, $\hat{\sigma}_k$ is larger than 0. The scale ratio $\hat{\sigma}_k/\hat{\mu}_k$ is used to decide if the $\hat{\sigma}_k$ is close to 0 or larger than 0 by comparing $\hat{\sigma}_k$ to $\hat{\mu}_k$.

Through the screening based on the Elementary Effect method, a number of $M < K$ input parameters is identified as the relevant influential input parameters. Then the structure sensitivity is analyzed only under the uncertainty of these $M$ influential input parameters by the Monte Carlo method for variance-based sensitivity analysis.

4.2 Quantitative method: the direct Monte Carlo simulation for variance-based sensitivity analysis

The direct Monte Carlo method for variance-based sensitivity analysis again from the beginning. In contrast, the Sobol’ random number generator can freely
extend the sample size even after finishing the simulation [8]. Therefore, the Sobol’ random number generator is used in this study to generate random samples of the input parameters according to their defined distribution functions $\mathcal{N}$ and $\mathcal{U}$ in Table 1.

The associated sensitivity indices for variance-based sensitivity analysis are the main effect index $S_k$ and the total effect index $S_{Tk}$ [5, 8]. The main effect index $S_k$ is a quantitative measure of the direct influence of the $k$-th input parameter on the output variable. The total effect index $S_{Tk}$ sums the main effect $S_k$ and the interaction effects of the $k$-th input parameter with other input parameters. The analytical calculation of the main effect $S_k$ and total effects $S_{Tk}$ of the input parameters is not possible for complex structures such as the described beam structure. They can only be approximated by application of the Monte Carlo method.

The main effect $S_k$ of each input parameter $p_k$ is approximated according to the Sobol’ Estimator $\hat{S}_k = g(N, A, B, A_B^{(k)})$ [15] as a function $g$ of $N$ numbers of the sampling trials and the simulation results of sampling matrices $A$, $B$, and $A_B^{(k)}$. The total effect $S_{Tk}$ of each input parameter $p_k$ is approximated according to the Jansen Estimator $\hat{S}_{Tk} = h(N, A, B, A_B^{(k)})$ [1, 16] as a function $h$ of $N$, $A$, $B$, and $A_B^{(k)}$. These two estimators are proven to work well for analyzing many different kinds of complex structures [1, 9]. The sampling steps to generate the sampling matrices $A$, $B$, and $A_B^{(k)}$ for calculating the estimators are described in detail in [8].

For the estimation of $\hat{S}_k$ and $\hat{S}_{Tk}$ of a structure with $K$ varied input parameters, $2N$ simulations are carried out for $A$ and $B$ and $N \cdot K$ simulations are carried out for $A_B^{(k)}$. Totally, the number of the direct Monte Carlo simulations for variance-based sensitivity analysis is $N_{DMC} = N \cdot (2 + K)$ [8].

### 4.3 Quantitative method: the combined Monte Carlo method for variance-based sensitivity analysis

The number of the direct Monte Carlo simulations for calculating the sensitivity indices depends on the number of sampling trials $N$ and the number of the varied input parameters $K$. If either $N$ or $K$ can be reduced, the sensitivity analysis can be more efficient. In case of not changing the random number generator, to ensure the accuracy of the sensitivity analysis, the number of sampling trials $N$ should not be reduced. However, based on the Elementary Effect method, the varied input parameters are sorted as influential and non-influential parameters. Therefore, by combining the Elementary Effect method and the Monte Carlo method for variance-based sensitivity analysis, only the $M$ influential input parameters are analyzed. The $(K - M)$ non-influential input parameters are fixed at their deterministic values. In this way, the number of the combined Monte Carlo simulations is reduced to $N_C = N_{EE} + N_{CMC} = R \cdot (1 + K) + N \cdot (2 + M)$, with $M < K$, which is the sum of the number of the simulations for the Elementary Effect method $N_{EE}$ and the number of the simulations for the direct Monte Carlo method $N_{CMC}$ for $M < K$ influential input parameters.

### 5 RESULTS AND COMPARISON OF RESULTS BASED ON THE DIRECT AND THE COMBINED MONTE CARLO METHOD

The simulations of the beam structure based on the direct Monte Carlo method and the combination of the Elementary Effect method and the Monte Carlo method are carried out according to the description in section 4. The results are discussed and compared in this section.
5.1 The screening results based on the Elementary Effect method

In this case study, \( R = 10 \) base points are randomly chosen for screening the influence of the \( K = 14 \) input parameters in Table 1 of the beam structure on the first lateral resonance frequency \( f_1 \) from equation (2). The number of the simulations based on the Elementary Effect method is \( N_{\text{EE}} = R \cdot (1 + K) = 150 \). After the simulations, the mean \( \hat{\mu}_k \) and the standard deviation \( \hat{\sigma}_k \) of each input parameter’s elementary effect are estimated according to equations (5) and (6) and illustrated in Figure 4.

By looking at Figure 4, it can be concluded that the lateral and rotational stiffness properties of the springs \( k_{1,A}, k_{1,B}, k_{1,B} \), as well as \( k_{r,B} \) and the stiffness of the piezo transducers \( k_{p,A} \) and \( k_{p,B} \) have influence on \( f_1 \) when they are varied in the range shown in Table 1. The length of the axial extension on both beam ends \( l_{\text{ext},A} \) and \( l_{\text{ext},B} \) also affects \( f_1 \). This is seen by the high value of the mean \( \hat{\mu}_k \). The input parameters located in the lower left corner in Figure 4 are \( E_{\text{ext},A}, \rho_{\text{ext},A}, l_{\text{ext},A}, E_{\text{ext},B}, \rho_{\text{ext},B}, \) and \( l_{\text{ext},B} \). As their \( \hat{\mu}_k \) is close to 0, they do not show relevant influence on \( f_1 \). For the eight influential input parameters \( k_{1,A}, k_{r,A}, k_{1,B}, k_{r,B}, k_{p,A}, k_{p,B}, l_{\text{ext},A}, \) as well as \( l_{\text{ext},B} \), their scale ratios \( \hat{\sigma}_k/\hat{\mu}_k \) are between 5\% for \( l_{\text{ext},A} \) and 16\% for \( k_{p,B} \). That means the standard deviation \( \hat{\sigma}_k \) of these eight input parameters is fairly small in comparison to their mean \( \hat{\mu}_k \) and implies that these eight influential input parameters’ functional relations to \( f_1 \) are practically linear.

According to the screening result, the \( M = 8 \) input parameters \( k_{1,A}, k_{1,B}, k_{r,A}, k_{r,B}, k_{p,A}, k_{p,B}, l_{\text{ext},A}, \) as well as \( l_{\text{ext},B} \) are concluded as influential input parameters and their influences on \( f_1 \) will be quantified by application of the Monte Carlo method.

5.2 The convergence of the sensitivity analysis results based on the direct Monte Carlo method

In this case study, the number of sampling trials is defined as \( N = 50\,000 \). Therefore, the number of the direct Monte Carlo simulations is \( N_{\text{DMC}} = N \cdot (2 + K) = 800\,000 \). The convergence of the sensitivity indices \( \hat{S}_k \) and \( \hat{S}_{T_k} \) according to the Sobol’ estimator [15] and the Jansen estimator [11,16] of each input parameter is analyzed to examine if the sample size \( N = 50\,000 \) is high enough to adequately estimate the sensitivity indices \( \hat{S}_k \) and \( \hat{S}_{T_k} \). Both sensitivity indices of each input parameter are calculated by every increase of 250 samples. In this study, all
sensitivity indices $\hat{S}_k$ and $\hat{S}_{Tk}$ are assumed to have converged when they stay within a variation range $\pm 0.05$ of the value that is approximated with the sample size $N = 50\,000$, see the horizontal dashed lines in Figure 5. The estimated sensitivity indices of the spring rotation stiffness of the support $A \; p_{10} = k_{r,A}$ by increasing the sample size $N$ are illustrated in Figure 5 as an example.

For the beam structure, the sensitivity index $\hat{S}_{10}$ of $p_{10} = k_{r,A}$ is converged when the estimation is based on more than $N = 16\,250$ samples, the sensitivity index $\hat{S}_{T10}$ of $p_{10} = k_{r,A}$ is converged when the estimation is based on more than $N = 31\,250$ samples. These are the highest required numbers for the convergence among all $K = 14$ input parameters in Table 1. Therefore, the sample size $N = 50\,000$ is sufficient for the sensitivity analysis in this case study.

The values of $\hat{S}_k$ and $\hat{S}_{Tk}$ of each input parameter on the output variable $f_1$ are illustrated as solid bars in Figure 7. The results are discussed and compared with the results based on the combination of the Elementary Effect method and the Monte Carlo method in section 5.4.

5.3 The convergence of the sensitivity analysis results based on the combined Monte Carlo method

The number of sampling trials is also defined as $N = 50\,000$. Therefore, the number of the combined Monte Carlo simulations is $N_{CMD} = N \cdot (2 + M) = 500\,000$. The same convergence analysis as for the direct Monte Carlo simulation is carried out to examine if the sample size $N = 50\,000$ is high enough to adequately estimate the sensitivity indices $\hat{S}_k$ and $\hat{S}_{Tk}$. The convergence criterion is the same as that for the direct Monte Carlo Simulation in Figure 5. Sensitivity indices are assumed to have converged when they stay within the variation range $\pm 0.05$ of the value that is, again, approximated with the sample size $N = 50\,000$, see the horizontal dashed lines in Figure 6. The estimated sensitivity indices $\hat{S}_{10}$ and $\hat{S}_{T10}$ of the spring rotation stiffness of the support $A \; p_{10} = k_{r,A}$ are illustrated in Figure 6 as an example.

For the beam structure, the sensitivity index $\hat{S}_{10}$ is converged when the estimation is based on more than $N = 9\,500$ samples, and the sensitivity index $\hat{S}_{T10}$ is converged when the estimation is based on more than $N = 26\,500$ samples. These are the highest numbers for the convergence among the $M = 8$ input parameters. In comparison to the direct Monte Carlo simulation, the combined Monte Carlo simulation requires a smaller number of sampling trials.

The values of $\hat{S}_k$ and $\hat{S}_{Tk}$ of the eight influential input parameters with $k = 3, 7, 9, \ldots, 14$
in Table 1 are illustrated as non-solid bars in Figure 7. The results are discussed and compared with the results based on the direct Monte Carlo method in section 5.4.

5.4 Comparison of the results based on the direct Monte Carlo method and the combined Monte Carlo method

The sensitivity indices of the main effect $\hat{S}_k$ and the total effect $\hat{S}_{Tk}$ of the $K = 14$ input parameters based on the direct Monte Carlo method and the combined Monte Carlo method based on the $M = 8$ influential input parameters are illustrated in Figure 7.

The black solid bars in Figure 7 represent the main effect $\hat{S}_{k,DMC}$ and the blue solid bars in Figure 7 represent the total effect $\hat{S}_{Tk,DMC}$ of the $K = 14$ input parameters based on the direct Monte Carlo simulation. By comparing $\hat{S}_{k,DMC}$ with $\hat{S}_{Tk,DMC}$, it can be observed that the value of $\hat{S}_{k,DMC}$ is almost equal to the value of $\hat{S}_{Tk,DMC}$ for each input parameter. It indicates that the functional relations between the input parameters and the output variable are nearly linear. This
conclusion agrees with the result based on the Elementary Effect method in section 5.1.

The estimated main and total effects \( \hat{S}_{k,DMC} \) and \( \hat{S}_{k,DMC} \) in Figure 7 based on the direct Monte Carlo method show that the varied input parameters \( p_3 = l_{ext,A} \) and \( p_7 = l_{ext,B} \) have more influence on \( f_1 \) than the other varied input parameters with highest values of \( \hat{S}_{k,DMC} \approx \hat{S}_{k,DMC} \approx 0.25 \). Second highest are \( p_{10} = k_{r,A} \) and \( p_{12} = k_{r,B} \) with \( \hat{S}_{k,DMC} \approx \hat{S}_{k,DMC} \approx 0.15 \). Third highest are \( p_6 = k_{r,A} \), \( p_{11} = k_{r,B} \), \( p_{13} = k_{p,A} \), and \( p_{14} = k_{p,B} \) with small values of \( 0.05 \leq \hat{S}_{k,DMC} \). Moreover, by comparing the sensitivity indices of the input parameters of the support A and those of the support B it can be found that they are almost equal, \( \hat{S}_{k,DMC} \approx \hat{S}_{k,DMC} \approx \hat{S}_{k,DMC} \approx \hat{S}_{k,DMC} \approx \hat{S}_{k,DMC} \approx \hat{S}_{k,DMC} \). This is reasonable since the beam structure is symmetric and the two supports are built based on the same design.

The rest of the varied input parameters, \( p_1 = E_{ext,A} \), \( p_2 = \rho_{ext,A} \), \( p_4 = l_{ext,A} \), \( p_5 = E_{ext,B} \), \( p_6 = \rho_{ext,B} \), and \( p_8 = l_{ext,B} \), show no relevant influence on \( f_1 \) with \( \hat{S}_{k,DMC} \approx \hat{S}_{k,DMC} \approx 0 \). These six input parameters are concluded as non-influential input parameters by screening based on the Elementary Effect method. Therefore, they are kept constant at their deterministic values for variance-based sensitivity analysis are compared according to the minimal required number of simulations \( N_{DMC,\min} \) for the direct Monte Carlo method and \( N_{C,\min} \) for the combination of the Elementary Effect method and the Monte Carlo method, Table 2. It is known from the description in section 5.2 and 5.3 that the combined Monte Carlo simulation requires a lower sampling size in comparison to the direct Monte Carlo method. In this study, the sum of the minimal required number of simulations based on the direct Monte Carlo method is \( N_{DMC,\min} = 500\,000 \) and the sum of the minimal required number of simulations based on the combination of the Elementary Effect method and the Monte Carlo method is only \( N_{C,\min} = N_{EE} + N_{CMC} = 265\,150 \). The numerical cost of the combined Monte Carlo method is only half the cost of the direct Monte Carlo method.

6 CONCLUSION

The combined Monte Carlo method that includes a screening of the relevant influential parameters by the Elementary Effect method was proposed as an efficient application of the Monte Carlo method to quantify structure uncertainty under input parameter uncertainty. The accuracy and the efficiency of the combined Monte Carlo method was investigated in this study by using a beam structure with piezo-elastic supports for buckling and vibration control as a reference structure. Its uncertain structural input parameters are geometric, material, and stiffness pa-
Table 2: The minimal required number of simulations based on the direct Monte Carlo method and the combined Monte Carlo method

<table>
<thead>
<tr>
<th></th>
<th>the direct Monte Carlo method</th>
<th>the combined Monte Carlo method</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of simulations for screening</td>
<td>0</td>
<td>( N_{\text{EE}} = R(1 + K) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( = 10 \cdot (1 + 14) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( = 150 )</td>
</tr>
<tr>
<td>min. sample size for ( \hat{S}_k )</td>
<td>16 250</td>
<td>9 500</td>
</tr>
<tr>
<td>min. sample size for ( \hat{S}_{Tk} )</td>
<td>31 250</td>
<td>26 500</td>
</tr>
<tr>
<td>min. sample size ( N_{\text{min}} )</td>
<td>( N_{\text{DMC, min}} = 31 250 )</td>
<td>( N_{\text{CMC, min}} = 26 500 )</td>
</tr>
<tr>
<td>min. number of simulations for ( \hat{S}<em>k ) and ( \hat{S}</em>{Tk} )</td>
<td>( N_{\text{DMC}} = N_{\text{DMC, min}} \cdot (2 + K) )</td>
<td>( N_{\text{CMC}} = N_{\text{CMC, min}} \cdot (2 + M) )</td>
</tr>
<tr>
<td></td>
<td>( = 31 250 \cdot (2 + 14) )</td>
<td>( = 26 500 \cdot (2 + 8) )</td>
</tr>
<tr>
<td></td>
<td>( = 500 000 )</td>
<td>( = 265 000 )</td>
</tr>
<tr>
<td>sum of minimal number of simulations</td>
<td>( N_{\text{DMC, min}} = 500 000 )</td>
<td>( N_{\text{C, min}} = N_{\text{EE}} + N_{\text{CMC}} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( = 150 + 265 000 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( = 265 150 )</td>
</tr>
</tbody>
</table>

Parameters of the piezoelectric supports. The first lateral natural frequency of the beam structure is subject to be estimated as the output variable. The influence of each varied input parameter on the first lateral resonance frequency was quantified by estimated sensitivity indices. The proposed combination of the Elementary Effect method and the Monte Carlo method was compared with the direct Monte Carlo method. According to the convergence analysis, the combined Monte Carlo method required only half the number of simulations as for the direct Monte Carlo simulation in analyzing the uncertainty of the first lateral resonance frequency of the beam structure. The comparison showed that the result based on the combined Monte Carlo method is similar to the result based on the direct Monte Carlo simulation. Consequently, the combined Monte Carlo method is proven to be an efficient method for uncertainty quantification in comparison to the direct Monte Carlo simulation in the structural design phase.

Acknowledgement

The authors would like to thank the German Research Foundation (DFG) for funding this research within the Collaborative Research Centre (SFB) 805 “Control of Uncertainties in Load-Carrying Structures in Mechanical Engineering”.

REFERENCES


A CONTRIBUTION TO THE EVALUATION OF IMPRECISE AVAILABILITY OF COMPLEX SYSTEMS USING MARKOV MODELS

Joanna M. Akrouche ¹, Mohamed Sallak¹, Eric Châtelet², Fahed A. Abdallah³, and Hiba Z. Haj Chhadé³

¹ Sorbonne universités, Université de technologie de Compiègne, CNRS,
Heudiasyc UMR 7253, CS 60 319, 60 203 Compiègne cedex
e-mail: {joanna.akrouche, mohamed.sallak}@hds.utc.fr

² Université de technologie de Troyes, CNRS, Institut Charles Delaunay/LM2S
12 rue Marie Curie, CS 42060, 10004 Troyes cedex, France
eric.chatelet@utt.fr

³ Université Libanaise
Beirut, Hadath
{fahed.abdallah76,hiba.hajchhade}@gmail.com

Keywords: Dependability, imprecise availability, imprecise Markov chain, interval probabilities, contraction techniques.

Abstract. When using classical methods for the availability assessment of a multi-state system, the precise values of state’ probabilities are required. But, in many cases the available data does not describe the system's components, defining the system's state, precisely. To cope with this problem, the imprecision can be incorporated into the method in terms of imprecise rates [1] (failure and repair rates) by using imprecise probability theory [2]. Markov chain models are known for their simplicity and their great ability to model reparable systems, thus, they are well adapted for modeling stochastic failure and repair processes, where conditional probability distribution of future states depends only on the present state, and then computing the system's availability. To our best of knowledge, only a few works were developed in the context of imprecise continuous Markov chain [3]. The idea in this paper is to replace precise initial distributions and transition matrices by imprecise ones where imprecise rates are expressed in terms of intervals which are supposed to contain the true unknown initial probability and transition matrix. The contribution of this work is twofold: first, applying interval analysis techniques on existing algorithms for availability assessment of multi-state systems, and second, studying the stationarity, convergence and ergodicity properties related to the new proposed technique.
1 INTRODUCTION

The exponential evolution of technology has increased the complexity of systems and has further reduced their design and manufacturing costs. Correspondingly, manufacturers rely on the criterion of quality to be distinguished in the market. To achieve this goal, they must master various tools that will enable them to keep a competitive position and take actions for improvement at all levels. All these reasons make operating reliability the undeniable means that must be mastered when designing any system. Dependability is defined as the property that enables system users to place a justified confidence in the service it delivers to them and it groups two key concepts: reliability and availability. The reliability $R(t)$ is the ability of a system to remain constantly operational in a given duration, in other words, the probability that an item survives the time interval $[0, t]$ and still functions at time $t$. The availability $A(t)$ is the ability of a system to be operational at a specific moment [5].

The principle is to choose amongst several components with different performance, the best components of the system, and also find the optimal configuration, that is the connection between the different components entities that constitute the system. And this must be done while respecting that a certain level of availability is ensured by the complete system.

Dependability domain is extremely large and complex. In this paper we are particularly interested in the optimization of the availability of a product taking into account the multi-state case which is commonly encountered in practical life. We also consider that systems and their components can operate in different performance levels between working and failure states. Availability analysis helps to calculate the ability of a system to provide a required level of performance depending on the level of degradation.

Several methods are employed to calculate the availability, amongst them, there are: Universal Generating Function method (UGF) [6], Inclusion-Exclusion technique [7], and Markov Chain approaches [3], etc. There are many probabilistic techniques that can evaluate this criterion, but these techniques are only effective for very specific cases, for example the case of binary systems. The transition to multi-state [7,8] systems drastically restricts the application of most of these methods. Indeed, the evaluation of the reliability of a multi-state system [9] is more difficult than in the binary case because we have to take into account the different combinations of the component failure modes. In addition to the multi-state aspect, there is the structural aspect of the system which illustrates the type of connection connecting the components and of course the existence of uncertainties. In general, most methods can be applied efficiently only to systems with a simple structure: series, parallel, bridge or mixed. Nevertheless, there is virtually no practical and effective method that relates to systems with complex structures.

In this work, we demonstrate how Markov chains can be used to study the availability of a system by treating uncertainties like intervals and using interval analysis techniques. The structure of the paper is as follows. Section 2 explains the Markovian approach with imprecise data for multi-state systems with several components. Section 3 introduces contractors and how we use them to obtain the availability. Section 4 presents some applications and the last section concludes the paper.

2 MARKOVIAN APPROACH

Traditional binary–state reliability models allow only two possible states for a system and its components: perfect functioning (up) and complete failure (down). However, a system can have a finite number of performance rates. Also, many real–world systems are composed of various components which in turn can have different performance levels and for which one cannot formulate an "all or nothing" type of failure criterion. Failures of some system elements lead, in these cases, only to performance degradation. Such systems are called
multi-state systems (MSS). Traditional reliability theory, which is based on a binary approach, has recently been extended by allowing components and systems to have an arbitrary finite number of states. Many methods have been proposed to study the availability of a MSS, in our approach we use the Markovian approach.

2.1 General description of the Markov Model

A stochastic process describes the evolution of a system over time using a set of probabilities of the system's states (or a subset of states) at time instants \( t \{X(t), t \geq 0\} \); where \( X(t) \) is a random variable that denotes the state of the process at time \( t \). A Markov model is a class of stochastic processes where the future state depends only on the present state. When time is discrete, we have a discrete-time Markov chain; when time is continuous, we have a continuous-time Markov chain which is also called a Markov process. Formally, a discrete-time Markov chain is characterized by a (discrete) set of states \( S \) and the transition probabilities \( p_{ij} \), where \( p_{ij} \) is the probability that the Markov chain moves at the next time instant to state \( j \), given that it is at the present time point at state \( i \). The matrix \( P \) grouping elements \( p_{ij} \) is called the transition probability matrix of the Markov chain. Note that the definition of the \( p_{ij} \) implies that the row sums of \( P \) are equal to 1. A continuous-time Markov process is also described by a discrete set of states \( S \), in this case, however, transition probabilities \( p_{ij} \) are replaced by transition rates \( q_{ij} \) which will be grouped into a transition rate matrix denoted hereafter as \( Q \). The matrix \( Q \) is an array of numbers describing the rate a continuous time Markov chain moves between states, it is expressed in terms of failure rates \( \lambda \) and repair rates \( \mu \). Diagonal elements \( q_{ii} \) are defined such that: \( q_{ii} = - \sum_{i \neq j} q_{ij} \), thus the row sums of \( Q \) are equal to 0. A Markov process is a stochastic process in which the future state does not depend on the past trajectory. Markov models are frequently used in RMS work where events, such as the failure or repair of a module, can occur at any point in time. The Markov model evaluates the probability of jumping from one known state into the next logical state (e.g., from “everything is working” state to "first item failure" state, then from “first item failure” state to “second item’s failure”, and so on,) until the system reaches the final or totally failed state (which depends upon the configuration of the system being considered). The basic assumption of a Markov process is that the behavior of a system in each state is memoryless. A memoryless system is characterized by the fact that the future state of the system depends only on its present state.

The state equations of a system \( S \) are defined in a discrete state space \( E \) by taking into account the transition rates \( q_{ij}(t) \) existing between consecutive system states. By considering the probabilities that the system stays in its current state or moves to any of the possible states in \( E \) for an elementary time interval \( [t, t + dt] \), we obtain a system of differential equations, called equations of Chapman-Kolmogorov. Thus, for each state \( e_i \) [10]

\[
P_i(t + dt) = P(S \text{ in state } e_i \text{ at } t \text{ and in } [t, t + dt]) + \sum_{e_j \in E - e_i} P(S \text{ in state } e_j \text{ at } t \text{ and in } e_i \text{ at } [t, t + dt])
\]  

(1)

The behavior of a Markov chain is fully probabilistically described if the initial and transition probabilities are given, that is, if the following probabilities are known:

\[
P(X_0 = x_i) = p_i \quad \text{and} \quad P(X_{n+1} = x_j | X_n = x_i) = q_{ij}
\]

(2)

From (1) we obtain a system of Chapman-Kolmogorov equations:

\[
\dot{P}(t) = P(t)Q
\]

(3)

with \( P \) is a vector representing the probabilities of being in a state at a certain time \( t \) with
\[
\sum_{j=1}^{n} P_j(t) = 1 \tag{4}
\]

\(P\) is the solution of the Chapman-Kolmogorov and \(Q\) is the transition matrix between the states of the system. The solution of (2) is expressed in exponential form as:
\[
P(t) = \exp(Qt) \cdot P(0) \tag{5}
\]

Where \(\exp(Qt)\) is an \(n \times n\) matrix and \(P(0)\) is the initial probability vector describing the initial state of the system.

2.2 Stationarity

For each system of components, the availability of being in some state, either working or not, will change at the end [3]. So in our study, we focus on finding the availability after an infinite time which is equivalent to calculating the stationarity of the system. The Chapman-Kolmogorov equations at time \(t \to \infty\) lead to:
\[
\Pi \cdot Q = 0 \tag{6}
\]

where \(\Pi\) is the stationarity vector representing the probabilities that the system will be in a certain state:
\[
\Pi = [\pi_1 \pi_2 \ldots \pi_i \ldots \pi_n ] \tag{7}
\]

An element \(\pi_i\) of \(\Pi\) is hence the probability that the system is at state \(e_i \in E\), and \(\Pi\) verifies:
\[
\sum_{i=1}^{n} \pi_i = 1 \tag{8}
\]

2.3 Imprecise Markov Chain

The Markov assumption stating that \(X_{t+dt}\) is conditionally independent of \(X_{s}\) for \(s < t\), knowing \(X_t\) may not be realistic, especially for repair, also the transition rates may not be constant in time, but are usually affected by a variety of factors, and the estimation of the rates themselves may be difficult due to the lack of data. Particularly, under constant transition rates, repair times are exponentially distributed and are independent of the history of the system, but repairs will often follow a binomial distribution rather than an exponential distribution [10]; the same applies for failure rate. A full modeling of these details requires a lot of data and expert knowledge. Instead of ignoring this problem, a better way to cope with it is to incorporate the imprecision into the models. This becomes possible with the development of models of imprecise probabilities, such as the interval probability model, it seems therefore convenient to consider our transition rates as not being fixed, but instead being bounded by an interval.

Imprecision may exist on the initial probabilities or the transition matrix, and sometimes even on both. To model this imprecision, probabilities in (2) will be replaced by intervals. Thus we have the following imprecise Markov model [11]:
\[
P(X_0 = x_i) \in [\underline{p}_i, \overline{p}_i] \text{ and } P(X_{n+1} = x_j | X_n = x_i) \in [\underline{q}_{ij}, \overline{q}_{ij}] \tag{9}
\]

According to the above model, any probability vector that satisfies \(p_i \in [\underline{p}_i, \overline{p}_i]\) for each state with \(i = 1, \ldots, n\), can be considered as an initial distribution, and similarly any transition matrix \(Q \in [\underline{Q}, \overline{Q}]\) can be the transition matrix at time \(t\).

As stated before, the purpose is to find the availability at \(t \to \infty\). Under the assumptions in (9), the stationarity of the system is hence determined in form of a vector of intervals \(\Pi = \{\pi_i\}, i = 1, \ldots, n\), where \(\pi_i \in [\underline{\pi}_i, \overline{\pi}_i]\) is the probability of being in a state \(e_i\). Finding the interval...
of stationarity for each state is not as simple as it might seem. In the case of precise data, we have a constant transition matrix so we can find the answer by solving a system of equations, but in the case of imprecise data, the bounds of the intervals of stationarity cannot be obtained just by taking into account the two bounds of the transition matrix as some previous work suggests [3].

Several methods have been proposed to solve this problem by finding the solution of equation (6) in the presence of imprecise data. The exact method is a technique where we find all the possible transition matrices of a system and then for each matrix we solve (6) and we find a vector of stationarity, to form at the end a vector of intervals which contains all of the obtained vectors, this method gives an exact result but its complexity increases drastically with the system's size since it computes all the possible state values. BUGF (Believe Universal Generated Function) [12] and IUGF (Interval Universal Generated Function) [11] are two methods based on the UGF (Universal Generated Function) but they are applied on intervals and therefore are used in the case of interval-modeled imprecision, these two methods are efficient and give good results, the IUGF is also noted as more efficient than the BUGF in [12]. In our approach we propose to determine the availability of the system by using a new technique applied on intervals, that is the technique of contractors [4] which we introduce in section 3.

3 THE TECHNIQUE OF CONTRACTORS

3.1 Definition of contractors

Consider \( n_x \) variables \( x_i \in \mathbb{R}, i = 1, \ldots, n_x \), linked by \( n_f \) relations (or constraints) [4] of the form

\[
f_j(x_1, \ldots, x_{n_x}) = 0 \quad , \quad j \in 1, \ldots, n_f
\]

(10)

Each variable \( x_i \) is known to belong to a domain \( X_i \). For simplicity, these domains will be intervals, denoted by \( [x_i] \). Define the vector \( x \) as:

\[
x = (x_1, \ldots, x_{n_x})^T
\]

(11)

and the prior domain for \( x \) is a box as:

\[
[x] = [x_1] \times \ldots \times [x_{n_x}]
\]

(12)

Let \( f \) be the function whose coordinate functions are the \( f_j \)'s. Equation (10) can be written in vector notation as:

\[
f(x) = 0
\]

(13)

This corresponds to a constraint satisfaction problem (CSP) \( H \), which can be formulated as:

\[
H: (f(x) = 0 , x \in [x])
\]

(14)

The solution set of \( H \) is defined as:

\[
S = \{ x \in [x]|f(x) = 0 \}
\]

(15)

Contracting \( H \) means replacing \( [x] \) by a smaller domain \( [x'] \) such that the solution set remains unchanged, i.e. \( S \subset [x'] \subset [x] \). There exists an optimal contraction of \( H \), which corresponds to replacing \( [x] \) by the smallest box that contains \( S \). A contractor for \( H \) is any operator that can be used to contract it.

A contractor \( C \) is defined as an operator used to contract the initial domain of the CSP, and thus to provide a new box [11].
These relations mean that contractions gives a sub-domain of the input domain \([x]\), and the resulting subdomain \(C([x])\) contains all the feasible points with respect to the constraints. No solution is “lost”.

### 3.2 Types of contractors

Several contractors exist, each works in a different manner and is efficient only for specific CSPs and for certain cases \([13,14,15]\). “Intervalization with Gauss elimination” \([4]\) is an important class of CSPs for which intervalization of finite subsolvers can be employed only if the system is formed of linear interval equations and if all elements on the diagonal of the matrix are different than zero. “Gauss-Seidel” contractor is an efficient way to contract intervals of a linear system but the matrix in this case must be reversible \([4]\). “Krawczyk and Newton contractors” are two contractors which are not always applicable and are much complicated than the other contractors \([4]\). Finally, one popular contraction technique, which will be used in our approach, is the “Forward-backward propagation (FBP) contractor”. This technique is known for its simplicity and ease, it is also more general than the others since it works on all type of systems \([4]\). It also gives guaranteed results which means that during the contraction we always get an interval belonging to the initial interval. As compared to the “Gauss-Seidel” technique, the FBP offers comparable accuracy. The “Gauss-Seidel” approach is however less general and becomes time-consuming when the system's size increases. For all these reasons we chose to use the “Forward-backward” propagation technique to determine the stationarity as given in equation (6) for the case of imprecise data.

### 3.3 Forward-backward propagation technique

Forward-backward (FBP) contractor \(C_{\downarrow \uparrow}\) (also known as HC4Revise \([16]\)) is a classical algorithm in constraint programming for contracting. This contractor makes it possible to contract the domains of the CSP \(H\) by taking into account each one of the \(n_f\) constraints apart. The algorithm works in two steps \([4]\). The forward step applies interval arithmetic to each operator of the function \(y = f(x)\), from the variable's domain \((x)\) up to the function's domain \((y)\), this step considers the direct forms of the equations. The backward step sets the interval associated to the new function's domain \([y]\) to \([0, 0]\) (imposes constraint satisfaction, since we are solving \(f(x)=0\)) and, then, applies backward arithmetic from the function's domain to the variable's domain, which means using the inverse of the functions that appear in the equations \(f(x)\). The following example explains the procedure of the FBP technique.

**Example:** Consider the constraint \(y = -5x_1 + 2x_2 = 0\) and the initial box-domain \([x] = [1, 4] \times [-3, 7]\). This constraint can be decomposed as shown in (18) into three primitive constraints (i.e. constraints associated with a unique elementary function: multiplication or addition) by introducing two intermediate variables \(a_1\) and \(a_2\) defined as: \(a_1 = -5x_1\) and \(a_2 = 2x_2\).

Initial domains for these variables are determined as follows:

\[
\begin{align*}
a_1 &= -5x_1 = -5 \times [1,4] = [-20, -5] \\
a_2 &= 2x_2 = 2 \times [-3,7] = [-6,14] \\
y &= a_1 + a_2 = [-20,5] + [-6,14] = [-26,9]
\end{align*}
\]

and this step is called the "forward propagation". A method for contracting \(H\) with respect to the constraint \(f(x) = 5x_1 + 2x_2 = 0\) is to contract each of the primitive constraints in (18) until the contractors become inefficient. For this example:

Since \(f(x) = 0\), the domain for \(y\) should be taken equal to \([0]\), so we can add the step:
\[ y := [y] \cap \{0\} \quad (19) \]

If \([y]\) as computed in (18) turns out to be empty, then the CSP has no solution. Else, \([y]\) is replaced by 0, which is the case in this example. After, a backward propagation is performed, updating the domains associated with all the variables to get:

\[
\begin{align*}
[a_1] &:= ([y] - [a_2]) \cap [a_1] = [-14, -5] \\
[a_2] &:= ([y] - [a_1]) \cap [a_2] = [5, 14] \\
[x_1] &:= ([a_1]) / -5) \cap [x_1] = [1, 14/5] \\
[x_2] &:= ( [a_2]) / 2) \cap [x_2] = [5/2, 7]
\end{align*}
\]

Thus, we obtain the new box:

\[ [x](1) = [1, 14/15] \times [5/2, 7] \quad (21) \]

which is the result of the first FBP contraction. Iterating this procedure, the resulting sequence of boxes \([x](k)\) converges towards the smallest possible domain, after which the domains no longer change following another iteration of FBP.

4 APPLICATION

![Figure 1: Flow transmission system](image)

To illustrate the technique of contraction with the Forward-Backward propagation, we use the example presented in [11]. In this example, we evaluate the availability of a flow transmission system design presented in Fig 1 and made of three pipes. The flow is transmitted from left to right, and the performances of the pipe are measured by their transmission capacity (tons of per minute). It is supposed that components 1 and 2 have three states: a state of total failure corresponding to a capacity of 0, a state of full capacity, and a state of partial failure. The component 3 only has two states: a state of total failure, and a state of full capacity. All state performances of the components are precise. We want to calculate the availability of the system by using Markov chain and we will compare to the IUGF proposed in [11] and the BUGF proposed in [12]. For every component \(G_i\):

\[
\begin{align*}
p_i^j & : \text{The probability of being at state } g_i^j \\
\lambda_{i,k}^j & : \text{The transition or degrading rate from state } g_i^j \text{ to } g_{i-k}^j \\
\mu_{i,k}^j & : \text{The repair rate from state } g_i^j \text{ to } g_{i+k}^j
\end{align*}
\]

In our framework, both transition and repair rates are given by intervals. Here are the transition and repair rates for each of the three components:

- For the first component \(G_1\), there are three possible states:
  1. State 1 \(g_1^1 = 1.5\) represents completely successful operation.
  2. State 2 \(g_1^2 = 1\) represents degraded successful operation.
  3. State 3 \(g_1^3 = 0\) represents total failure.
Let the possible transition rates (in terms of hours) be:

\[ \lambda_{3,1}^1 = [10^{-5}; 3 \times 10^{-4}] \ h^{-1} \]
\[ \lambda_{3,1}^2 = [4 \times 10^{-5}; 5 \times 10^{-4}] \ h^{-1} \]
\[ \mu_{2,1}^1 = [2 \times 10^{-2}; 5 \times 10^{-2}] \ h^{-1} \]
\[ \mu_{1,1}^1 = [4 \times 10^{-2}; 8 \times 10^{-2}] \ h^{-1} \]

- For the second component \( G_2 \), there are three possible states:
  1. State 1 \( g_2^1 = 2 \) represents completely successful operation.
  2. State 2 \( g_2^2 = 1.5 \) represents degraded successful operation.
  3. State 3 \( g_2^3 = 0 \) represents total failure.

Let the possible transition rates be:

\[ \lambda_{3,1}^2 = [2 \times 10^{-5}; 6 \times 10^{-4}] \ h^{-1} \]
\[ \lambda_{3,1}^3 = [3 \times 10^{-5}; 4 \times 10^{-4}] \ h^{-1} \]
\[ \mu_{2,1}^2 = [3 \times 10^{-2}; 6 \times 10^{-2}] \ h^{-1} \]
\[ \mu_{1,1}^2 = [3 \times 10^{-2}; 7 \times 10^{-2}] \ h^{-1} \]

- For the third component \( G_3 \), there is two possible states:
  1. State 1 \( g_3^1 = 4 \) represents completely successful operation
  2. State 2 \( g_3^2 = 0 \) represents total failure

Let the possible transition rates be:

\[ \lambda_{3,1}^3 = [10^{-5}; 4 \times 10^{-4}] \ h^{-1} \]
\[ \mu_{1,1}^3 = [5 \times 10^{-2}; 9 \times 10^{-2}] \ h^{-1} \]

Figure 2: Markov chain of the system.

In this example, each of the components 1 and 2 have three possible states, while component 3 has only two states, hence the whole system has at most eighteen (3×3×2) possible states degrading from the total functional state to the failure state as shown in Fig. 2. In this figure, \( 0_j \) means that component \( j \) is in the completely working state, \( 1_j \) means that component \( j \) is in the partial working state and \( 2_j \) means that component \( j \) is in the completely failure state. In each state, the performance level \( g \) is calculated by taking into account the performance level of each component. For example, state 1 denoted by \( 0_10_20_3 \) refers to the state where all three components are completely working.
To obtain the performance level of the state: Components 1 and 2 are placed in parallel so the performance level resulting of them is the summation of the two performances which means 2 + 1.5 = 3.5. Component 3 is placed in series, so the total resulting performance level g is the minimum, i.e. \( g = \min(3.5, 4) = 3.5 \).

In this example, we are studying the availability of the system for a demand level \( w = 1.5 \), we will consider the working states when the total performance level is greater than \( w \), in Fig. 2 the colored states are the working states.

First, we calculate the corresponding availability of the system using IUGF and BUGF, results are grouped in table 1. After, we use the Forward-backward propagation contracting technique to solve the system of equations ensuring that the product of the stationarity vector times the transition matrix is equal to zero (18 equations) plus one last equation that is the summation of each stationarity is equal to one.

Our obtained results are shown in table 1. The result of the exact method is also given in this table.

<table>
<thead>
<tr>
<th>Contraction technique</th>
<th>BUGF</th>
<th>IUGF</th>
<th>Exact method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[0.95; 1.03]</td>
<td>[0.913; 1.09]</td>
<td>[0.952; 0.99]</td>
</tr>
</tbody>
</table>

Table 1: The availability at \( t \to \infty \) for each method.

Table 1 shows that the accuracy offered by the FBP contraction technique is close to that obtained when using the exact method; the interval availability of the contraction method is also more conservative than the other two methods (BUGF and IUGF). Therefore, the FBP contraction method turns out as an efficient technique for availability assessment since it offers accurate results and is more simple especially when handling complex systems.

5 CONCLUSION

We have applied the forward-backward contraction method on the equations obtained by the transition matrix of a Markov chain handling MSSs. Since multi-state systems are much more complicated than traditional two-state systems, an efficient method is desirable which can deal with system's size and complexity. Markov models for our best of knowledge are the most suitable methods, especially when the given data is imprecise. To be able to calculate the availability of a system by applying Markov model in the presence of imprecise data we need an accurate method for calculation. In our work, we proposed to use interval Contraction method. The main goal of the method is to reduce an initial big interval to its most possible minimum size. IUGF and BUGF are also good methods to calculate the imprecise availability of a MSS. These two methods are a pair of variants of the UGF methods, the first allows getting the availability by using imprecise probabilities, and the second calculates \( bel \) and \( pl \) by using mass functions. After testing these three approaches on different examples and comparing the results, we can clearly see that the FBP Contraction method is the most accurate amongst them, its results giving the smallest intervals containing those obtained by the exact method.

As future work, we can extend the proposed interval-based approach to more complicated scenarios and study the availability of a more complex system with multi-state components.
ACKNOWLEDGEMENT

This work was carried out in the framework of the regional project ODISYM.

REFERENCES


FORMULATION OF POTENTIAL FOR DYNAMICAL PARTICLE SYSTEM APPLIED TO MONTE CARLO SAMPLING

Jan Mašek\textsuperscript{1} and Miroslav Vořechovský\textsuperscript{2}

\textsuperscript{1} Brno University of Technology
Veveří 331/95, 602 00 Brno, Czech Republic
e-mail: jan.masek1@vut.cz

\textsuperscript{2} Brno University of Technology
Veveří 331/95, 602 00 Brno, Czech Republic
e-mail: vorechovsky.m@vut.cz

Keywords: Particle dynamical system, statistical sampling optimization, periodic space, particle interaction laws, short- and long-range interactions, self-similar designs.

Abstract. The presented paper investigates the effect of formulation of energy potential of a dynamical particle system as used for optimization of statistical point sampling. The dynamical particle system, originally developed as a physical analogy of the Audze-Eglajs (AE) optimization criterion and its periodical modification (PAE), effectively demonstrated that the originally proposed potential performs well only in poorly applicable scenarios of design spaces of low dimension filled with rather high number of design points. A remedy lying in a refined formulation of energy potential as well as its derivation and reasoning are presented.
1 INTRODUCTION

Numerical integration of Monte-Carlo type requires sampling of integration points that are uniformly distributed in the design domain. The layout of design points crucially affects the performance of such a numerical integration. The problem of using an ideally distributed set of integration points is also of interest in many other engineering and research fields. While sampling from a random vector or integrating an unknown function, using a uniform layout of integration points is the only possible way for minimization of the lower bound of the resulting error.

However, the "uniformity" itself is not a recognized property. Many criteria have been proposed during the recent years for evaluation of the uniformity of point layouts containing $N_{\text{sim}}$ points within a design space of the dimension of $N_{\text{var}}$. Typically, these criteria investigate point layouts with a tendency to prefer designs with points distributed equally distant from each other. Certain criteria are derived from analogies with physical problems. 

Namely, the Audze-Eglājs (AE) criterion [1] may be considered as an elegant instance of these. The objective of the AE criterion is in minimization of potential energy of a system of mutually repelling particles. With their positions, these particles represent positions of sampling points within a unit hypercube design domain.

During the recent years, it has been proposed that the original Audze-Eglājs criterion does suffer from the existence of boundaries of the design space [2, 3]. A remedy of this behavior was also proposed [2, 3] assuming a periodically extended design hypercube in case of which the boundaries naturally disappear. Building on such a refined Periodic Audze-Eglājs criterion (PAE), it has been proved that usage of the PAE criterion leads to statistically uniform designs (from design to design) and to well distributed set of points for every single point layout.

2 AUDZE-EGLĀJS AND PHI CRITERIA

The original formulation of the AE criterion [1] (see also [4, 5, 6, 7]) considers the analogy between the sampling plan and a system of charged particles with repulsive forces. The potential energy of the system is a sum of energies $1/L_{ij}^2$ accumulated by each pair of points $i$ and $j$. Instead of the sum of energies, one can alternatively calculate the average potential energy, i.e. divide the total energy by the number of pairs of points $N_p = \left(\frac{N_{\text{sim}}}{2}\right)$. The AE criterion then reads:

$$E_{n}^{\text{AE}} = \frac{1}{N_p} \sum_{i=1}^{N_{\text{sim}}} \sum_{j=i+1}^{N_{\text{sim}}} \frac{1}{L_{ij}^2}$$

where $L_{ij}$ is the inter-site Euclidean distance between points $i$ and $j$, depending on the Cartesian coordinates of said points, $x_i = \{x_{i,v}\}, i = 1, \ldots, N_{\text{sim}}, v = 1, \ldots, N_{\text{var}}$:

$$L_{ij} = \sqrt{\sum_{v=1}^{N_{\text{var}}} (\Delta_{ij,v})^2}$$

where:

$$\Delta_{ij,v} = |x_{i,v} - x_{j,v}|$$

is the projection of the distance $L_{ij}$ onto the axis $v$.

A generalization of the AE criterion is the $\phi$ criterion [8]:

$$\phi(d, \lambda) = \frac{1}{N_p} \left(\sum_{i=1}^{N_{\text{sim}}} \sum_{j=i+1}^{N_{\text{sim}}} \frac{1}{d^\lambda(x_i, x_j)}\right)^{\frac{1}{\lambda}}$$
This criterion considers a general power, \( \lambda \), of the distance (metric) \( d \). The combination of \( \lambda = 2 \) and the Euclidean intersite distance \( d \) leads \( \phi(d, \lambda) \) identical to the AE criterion.

As \( \lambda \to \infty \), the criterion increasingly prioritizes designs where the minimal distances are maximized which is, in limit, the Maximin distance criterion \([9]\). Using the analogy with the system of charged particles, one can say that with increasing power \( \lambda \) a greater portion of energy is stored in the short-range interactions.

The power \( 1/\lambda \) upon the entire sum is, in fact, a monotonous transformation (the difference between designs is not distorted) and can be dropped as well. Therefore, a simplified version of the criterion with parameters of a used metric \( d(x_i, x_j) \) and exponent value \( \lambda \) can be considered:

\[
\phi(d, \lambda) = \frac{1}{N_p} \sum_{i=1}^{N_{sim}-1} \sum_{j=i+1}^{N_{sim}} \frac{1}{d^\lambda(x_i, x_j)}
\]  

The standard way of measuring the inter-point distances is the Euclidean length defined above in Eq. [2] \( d(x_i, x_j) = L_{ij} \). It has been shown that due to the presence of boundaries of the design domain, this metric used in the criteria Eqs. [1][4][5] leads to non-uniform point distribution \([2][10][11]\). Moreover, the authors of \([2][10][11]\) have shown that if the metric \( d(x_i, x_j) \) is modified such that it considers distances measured in the periodically repeated design domain (a periodic metric), the criterion becomes invariant with respect to arbitrary shifts along individual dimensions. A simplified version of the periodic space considers only the shortest distances, i.e. the distance is taken as \( d(x_i, x_j) = L_{ij} \), where \( L_{ij} \) is the Euclidean distance between the \( i \)-th point and the closest image of the \( j \)-th point within the periodic space:

\[
L_{ij} = \sqrt{\sum_{v=1}^{N_{var}} \left( \Delta_{ij,v} \right)^2}
\]  

where:

\[
\Delta_{ij,v} = \min\{(\Delta_{ij,v}, 1 - \Delta_{ij,v})
\]

is the shortest projection of the distance between point \( i \) and the nearest image of point \( j \) onto the axis \( v \).

Using this definition of distance, Eq. [5] reads:

\[
\phi(L, \lambda) = \frac{1}{N_p} \sum_{i=1}^{N_{sim}-1} \sum_{j=i+1}^{N_{sim}} \frac{1}{L_{ij}^\lambda}
\]  

Utilizing the above-described nomenclature, the AE criterion as written in the can be denoted as \( \phi(L, 2) \). The periodic version of the AE criterion (the PAE criterion \([2]\)) uses a combination of exponent \( \lambda = 2 \) and the shortest (periodic) metric \( L_{ij} \) and therefore it can be denoted as \( \phi(L, 2) \).

The authors of \([2]\) argue that already for exponent \( \lambda = 2 \), the consideration of the shortest distance \( L_{ij} \) suffices to deliver invariance with respect to random shifts along individual dimension and thus prioritizes designs leading to statistical uniformity of coverage. Moreover, it is argued that the shortest distance is the one associated with the highest contribution to the criterion and therefore the PAE criterion captures the important features of the full periodic repetition of the design space. The next section generalizes the criterion in Eq. [8] by considering a higher number of copies of the design domain.
3 PERIODIC EXTENSION OF THE DESIGN DOMAIN

In this section we consider a generalized model in which a certain number of periodic repetitions of the original design domain are considered. Using the nearest image of point \( j \) with respect to point \( i \), as considered in Eq. 8, does not cover a true periodic repetition of the design domain. We argue that the above presented approach is a simplification that can be shown to yield identical results to the fully repeated system in case of sufficient point count \( N_{\text{sim}} \). If the number of points in the original domain is too small to carry enough information about the pattern of a periodically repeated system, making a periodic extension to a sufficient level is desirable. In a true periodic domain, infinite number of images of point \( j \) would interact with point \( i \). When a finite number of copies of the design domain is considered, not only the real particle \( j \), but also all periodically repeated images of the particle \( j \) will contribute to the potential:

\[
\phi(L, \lambda, k_{\text{max}}) = \sum_{i=1}^{N_{\text{sim}}-1} \sum_{j=i+1}^{N_{\text{sim}}} \left( \frac{1}{N_p} \sum_{k=1}^{k_{\text{max}}} \sum_{c=1}^{c_{\text{max}}} \frac{1}{L_{ij}^\lambda (x_i, x_j)} + \frac{1}{N_p} \right)
\]

where \( k_{\text{max}} \) introduced as an additional parameter is the number of added periodical extensions (envelopes) of the design space. In the fully repeated system \( k_{\text{max}} = \infty \) and analogically, for a non-extended system \( k_{\text{max}} = 0 \). Therefore \( \phi(L, \lambda) = \phi(L, \lambda, 0) \), compare Eqs. 8 and 9.

When a certain number of envelopes \( k_{\text{max}} \) is considered, the number of copies of the design domain is denoted as \( c_{\text{max}} = 0 \). The vector \( s_c \) is the vector needed for shifting the original point \( j \) to the particular periodically repeated version indexed by \( c \). Let us denote that the distances to the periodically repeated images of the point \( j \) must be measured as the standard Euclidean distances.

A single level of periodic extension adds another envelope of periodically repeated images of all other particles around each point, see Fig. 1 and 2e. Such an extension does provide additional information about the point layout within the domain.

The level of the periodic extension is quantified by a positive integer \( k_{\text{max}} \). Within an extended periodic domain of finite value of \( k_{\text{max}} \), the particle \( i \) interacts not only with the actual particle \( j \), but with all of \( c_{\text{max}} = \left( \frac{2k_{\text{max}} + 1}{N_{\text{var}}} - 1 \right) \) images of the particle \( j \) as well, see Fig. 1 for \( N_{\text{var}} = 1 \). The envelopes are considered to be centered around the shortest distance with \( L_{ij} \).

\[
\begin{align*}
&L_{ij} - k_{\text{max}} & L_{ij} + k_{\text{max}} \\
&\Delta_{ij} - k & \Delta_{ij} + k \\
&\Delta_{ij} - k_{\text{max}} & \Delta_{ij} + k_{\text{max}}
\end{align*}
\]

Figure 1: 1D example of the periodic extension of level \( k_{\text{max}} = 2 \)

4 THE EXPONENT \( \lambda \)

This section focuses on the exponent \( \lambda \) in the periodic \( \phi \) criterion [10], see Eq. 8. In the original AE criterion and also in the periodic version (PAE), the potential energy between each
pair of points is not dependent on the dimension, $N_{\text{var}}$. It has been found \cite{12} that the character of the criterion is different for various $N_{\text{var}}$ and also for various numbers of points, $N_{\text{sim}}$. In 1D situation, the energy tends to infinity linearly with increasing $N_{\text{sim}}$. In a 2D, the energy tends to infinity as $\ln(N_{\text{sim}})$ which is not a power law. For dimensions $N_{\text{var}} \geq 3$, the energy tends to a constant for increasing $N_{\text{sim}}$.

This behavior can be explained by the fact that for a given $N_{\text{var}} > 1$, various numbers of points yield to different proportions between energy due to the long-range and short-range interactions. The higher is the number of points, the higher the proportion of energy stored in long interactions is. This may not be desirable behavior as the criterion in high dimensions and also for high number of points becomes insensitive to local clusters of points: it becomes dominated by long-range interactions.

The power is suggested to be at least $\lambda \geq N_{\text{var}} + 1$. With this power, the interaction is dominated by short-range interactions. With such a sufficient exponent $\lambda$, the convergence of the potential energy $\phi(L, \lambda)$ or better $\phi(L, \lambda)$ towards infinity for a uniform distribution of points is a power law. Such a convergence signalizes self-similarity of the problem or absence of a length scale. In other words, a zoom into sufficiently dense uniform design with a window greater than a certain size (see below) carries all features of the full design and the energy value can be easily scaled from the value corresponding to the smaller zoom.

This can be shown by studying the behavior of the radial part of the integral of the potential over the volume $V$ of $N_{\text{var}}$-dimensional domain. The potential energy for a uniform design reads:

$$I = \int \frac{1}{L^\lambda} d^{N_{\text{var}}} V$$

where $L$ is used to denote one-dimensional distance between points (the symbol $d$ is not used to avoid confusion with the symbol $d$ for the differential). Transforming this into polar co-ordinates gives:

$$I = \int \phi d^{N_{\text{var}}} V |J| \frac{1}{L^\lambda} dL$$

where $|J|$ is the Jacobian. The volume element is thereby given as:

$$d^{N_{\text{var}}} V = L^{N_{\text{var}} - 1} dL \cdot d\phi \prod_{i=1}^{N_{\text{var}} - 2} \sin^{N_{\text{var}} - 1 - i}(\phi_i)$$

Therefore, the integral is performed over the product $L^{N_{\text{var}} - 1 - i}$. Performing just the radial integration leads to:

$$I_r = \int \frac{L^{N_{\text{var}} - 1}}{L^\lambda} dL = \int L^{N_{\text{var}} - 1 - \lambda} dL$$

For $\lambda = 2$ as used in the AE criterion, we get the behavior described above. Using $\lambda = N_{\text{var}}$ leads to:

$$I_r = \int L^{-1} dL = \ln(L)$$

which diverges logarithmically and the interaction is still long-ranged. Using $\lambda = N_{\text{var}} + 1$ yields

$$I_r = \int L^{-2} dL = \frac{1}{L}$$
which has the desired asymptotic behavior dominated by short-ranged interactions. Using powers \( \lambda > N_{\text{var}} + 1 \) only increases the (asymptotically constant) ratio between short-range and long-range interactions.

Fig. 2 shows the convergence of the normalized potential energy \( \phi(L, \lambda) \) with rise of the number of particles, \( N_{\text{sim}} \). Instead of presenting the results for the point count, \( N_{\text{sim}} \), we introduce a variable \( l_{\text{char}} \), the characteristic length that involves also the dimension of the space. The characteristic length is defined as:

\[
\phi(L, \lambda) \approx \pi \ln(N_{\text{sim}}) + \frac{1}{\sqrt{N_{\text{sim}}}} - \frac{1}{N_{\text{sim}}} \tag{16}
\]

In higher dimensions \( N_{\text{var}} \geq 3 \), the exponent \( \lambda = 2 \) further leads to convergence of the potential energy to a constant [12].

Using the above proposed exponent \( \lambda = N_{\text{var}} + 1 \), the potential energy value tends to a power law as \( N_{\text{sim}} \to \infty \):

\[
\phi(L, N_{\text{var}} + 1) \approx \frac{1}{l_{\text{char}}} \tag{17}
\]

Such a behavior is desired as the designs for a given dimension \( N_{\text{var}} \) tend to have a universal self-similar pattern and the dependence on sample size disappears (no length scale is present). Thus the character of the criterion is kept independent of \( N_{\text{sim}} \) and the proportion between short-range interactions and long-range interactions is constant. This stabilization is obtained for a sufficient number of points within the design domain (a kind of tile). The self-similarity manifested is by the power law dependence (a straight line in Fig. 2). When the exponent is taken even higher \( \lambda > N_{\text{var}} + 1 \), the self-similar regime is achieved for even smaller number of points (greater \( l_{\text{char}} \)).

Graphs in Fig. 2 suggest that there must be link between (a) the exponent (responsible for the proportion between long- and short-range interactions) and, (b) the number of “dummy” copies of the design domain that also modify the proportions. This aspect is discussed in the next section.

5 ON THE SEEMINGLY SIMILAR EFFECT OF RISING THE EXPONENT \( \lambda \) AND INCREASING THE LEVEL OF PERIODIC EXTENSION \( k_{\text{max}} \)

Let us consider a few-body \( (N_{\text{sim}} = 3) \) particle system in design space of \( N_{\text{var}} = 2 \) while using a potential with the exponent \( \lambda = 2 \). When considering the point layout via the \( \phi(L, \lambda) \) interaction, for each point, there exist only two mutual distances to other points, i.e. two forces acting upon each particle.

With an exponent of such insufficient magnitude, these forces do not differ significantly enough to represent correctly which particle shall be considered to be close (short-range interaction) and which to be far (long-range interaction). A rise of the exponent above certain threshold (discussed above) does lead to the needed qualitative change of the ratio between the acting forces: the closer particles start to act as short-range and the farther particles as long-range. In fact, the higher the exponent, the larger portion of potential energy will be stored in the short-range interactions.
Figure 2: Convergence of the normalized potential energy $\phi(L, N_{\text{var}})$ depending on the exponent $\lambda$. Coloring that designate $N_{\text{var}}$ is identical for the two bundles of curves. Solid circles are accompanied by the sample count corresponding to the $l_{\text{char}}$ and $N_{\text{var}}$.

The analogy between the effect of increasing the number of envelopes in and rising the exponent in is evident. While using the original (low) value of the exponent $\lambda$, majority of the potential energy is stored in the long-range interactions. However, there is not a sufficient number of particles for the criterion to distinguish between short and long range. All particles seem to be at similar distance as the design is not filled enough.

The $\phi(L, \lambda, k_{\text{max}})$ interaction adds one or multiple additional envelopes of neighboring images of actual particles. These images, naturally, will act as long-range. Even longer-range than the real particles previously acting as long-range. This leads to a qualitatively more accurate distribution of forces acting upon the real particles. Hence the identical behavior of the $\phi(L, \lambda, k_{\text{max}})$ and the $\phi(L, \lambda)$ interaction with a correct exponent $\lambda$:

- the $\phi(L, \lambda, k_{\text{max}})$ interaction does add long-range points for the actual particles to seem closer,
- the $\phi(L, \lambda)$ interaction with corrected exponent changes the ratio between forces for the close particles to seem closer and the distant particles to seem farther.

It can be therefore shown that while simulating a few-body problem with the $\phi(L, \lambda)$ interaction, it is advised to rise the exponent $\lambda$ even above the lower bound of $N_{\text{var}} + 1$ to force the desired self-similarity for various $N_{\text{var}}$.

When using the $\phi(L, \lambda, k_{\text{max}})$ interaction, especially for few-body problems, greater context of the pattern is carried within the interaction for there is considered $3N_{\text{var}}$ images of each particle. Effectively, a system mimicking $N_{\text{sim}} \cdot 3N_{\text{var}}$ particles is being simulated and the identical pattern should be obtained, see Fig. 3b and 3f.

On a side note, the simulation of a greater (extended) system might be the slower option compared with the correction of the exponent in the energy potential as the number of additional...
points within the envelopes rises steeply.

6 COMPUTER IMPLEMENTATION

A simulation of a particle system is typically a computationally demanding task. However, it is possible and beneficiary to conduct the implementation of solution as parallel as possible. The degree up to which the parallelism can be reached depends dominantly on the nature of the problem at hand and also on the possibilities of the used hardware. The implementation of solution of the proposed dynamical particle system has been conducted using the nVidia CUDA platform.

The equations of motion of the dynamical particle system contain independent accelerations on the left-hand side. This means that accelerations of particles can be solved separately, without solving a system of equations. Furthermore, computation of the mutual distances as well as the numerical integration of equations of motion using the semi-implicit Euler method can be conducted in parallel.

The derivation of equations of motion, the nature of implementation and speed-up are not trivial and were already covered in detail in concurrent publications [3, 13].

7 RESULTS OF NUMERICAL SIMULATIONS

The following section presents examples of numerical simulations of the dynamical particle system, mainly focusing on the effect of the value of the exponent $\lambda$ in the energy potential as well as on the effect of periodical extension.

For start, let us show how the ability of a self-similar design (power-law quality) disappears while the number of particles $N_{\text{sim}}$ decreases. In a design space of dimension $N_{\text{var}} = 2$, the original exponent value yields ideal distribution only if the design space is filled enough. For $N_{\text{var}} > 2$, the original exponent leads to primitive patterns of even incomplete ortho-grids.

Studying the design space of dimension $N_{\text{var}} = 2$, Fig. 3a, 3b and 3c show that an ideal pattern of a triangular grid (if possible) is reached as long as the information needed to creating of such a self-similar pattern is provided (the number of points is sufficient). When the number of particles becomes insufficient, the interaction is not approximated well enough (in a pursuit of a perfect triangular grid) and primitive (even incomplete) ortho-grids are produced, see Fig. 3d.

A remedy, as described above, can be conducted by two seemingly dissimilar approaches. First, see Fig. 3e for an example of a layout obtained with a sufficiently high value of the exponent $\lambda$. Second, a result of a simulation using a periodic extension of a single envelope ($k_{\text{max}} = 1$) is provided. Such a simulation effectively considers $4 \cdot 3^2 = 36$ points (4 of which are the actual particles), see Fig. 3f. Hence, the entire extended space is filled equivalently to the non-extended ($k_{\text{max}} = 0$) simulation of 36 particles, see Fig. 1b. An identical point layout of 36 points is therefore achieved and, in a way, the resulting layout for the system of $N_{\text{sim}} = 4$ is cut out of a greater scenario.

It is worth noting that, when simulating such an extended system (36 particles instead of 4 particles), the power-law quality is easier to achieve as the higher theoretical $N_{\text{sim}}$ (lower $I_{\text{char}}$ value) occurs closer to the power-law asymptote, see Fig 4.

Another portion of simulation examples concerns the results of simulations within a three-dimensional design space ($N_{\text{var}} = 3$). For this and higher dimensions, the original formulation of the (P)AE potential is already malfunctioning and yields poor designs based dominantly on ortho-grids, see Fig. 4a. The result of the remedy of rising the value of the exponent $\lambda$ is
Figure 3: Disappearing of the quality of a self-similar design as yielded by the original PAE formulation and the effect of remedies proposed.

provided, see Fig. 4b.

Such a refined formulation of the energy potential, we believe, will lead to perfect designs not only within the design space of the complete dimension, but also in all sub-spaces of lower dimension, see Fig. 4.

8 RESULTS OF NUMERICAL SIMULATIONS

The paper investigates the formulation of the energetic potential of the Audze-Eglājs optimization criterion and its periodical modification PAE. Remedies of an incorrect behavior of systems of low numbers of particles and high dimensions are proposed.

An indirect solution leading for appropriate layouts of few-body systems lies in simulating a greater scenario: considering additional envelopes of the periodically repeated system. That way, a richer information about the pattern is provided and the optimal point layout can be obtained. However, such a scenario mimics simulation of a system of higher number of particles and does not lead to a correction of the malfunctioning energy potential.

An overarching remedy was therefore pursued, leading to a potential which takes into account the dimension of the problem, $N_{\text{var}}$. First, a generalization of the potential based on the $\phi$ criterion is provided so the crucial parameters of the potential, the metric $d(\cdot, \cdot)$ and the exponent $\lambda$, can become subject of a refinement.

Further derivation was based on the desire for an ability of creating self-similar patterns of point layouts for various point counts. With the proposed value of the exponent $\lambda$, convergence of the potential energy of the criterion towards infinity for a uniform distribution of points is a power law. Such a convergence signalizes self-similarity of the problem or absence of a length scale. Using such a refined interaction, optimal (self-similar) designs are produced even for scenarios of arbitrary dimension, $N_{\text{var}}$, or few-body systems, as was shown.

In this way, the role of the exponent featured in the $\phi$ criterion is explained using the analogy with a system of mutually repelling particles.
Figure 4: 3D designs: a) the original PAE (\( \lambda = 2 \)), b) the corrected potential exponent (\( \lambda = N_{\text{var}} + 1 \)).

ACKNOWLEDGEMENT

This work has been supported by the Grant agency of the Czech Republic under projects Nos. GA16-22230S and GA15-07730S. This support is gratefully acknowledged.

The authors would like to thank Dr. Godehard Sutmann for valuable discussions about physics of interacting particles.
REFERENCES


VORONOÏ WEIGHTING OF SAMPLES IN MONTE CARLO INTEGRATION

Miroslav Vořechovský¹, Václav Sadílek² and Jan Eliáš³

¹ Brno University of Technology
Veveří 331/95, 602 00 Brno, Czech Republic
e-mail: vorechovsky.m@vut.cz, {sadilek.v, elias.j}@fce.vutbr.cz

Keywords: Voronoi tessellation, Reweighting, Monte Carlo integration, Weighted statistical estimators, Sampling plan.

Abstract. The standard way to numerically calculate integrals such as the ones featured in estimation of statistical moments of functions of random variables using Monte Carlo procedure is to: (i) perform selection of samples from the random vector, (ii) approximate the integrals using averages of the functions evaluated at the sampling points. If the $N\text{\textsubscript{sim}}$ points are selected with an equal probability (with respect to the joint distribution function) such as in Monte Carlo sampling, the averages use equal weights $1/N\text{\textsubscript{sim}}$. The problem with Monte Carlo sampling is that the estimated values exhibit a large variance due to the fact that the sampling points are usually not spread uniformly over the domain of sampling probabilities. One way to improve the accuracy would be to perform a more advanced sampling.

The paper explores another way to improve the Monte Carlo integration approach: by considering unequal weights. These weights are obtained by transforming the sampling points into sampling probabilities (points within a unit hypercube), and subsequently by associating the sampling points with weights obtained as volumes of regions/cells around the sampling points within a unit hypercube. These cells are constructed by the Voronoi tessellation around each point. Supposedly, this approach could have been considered superior over the naive one because it can suppress inaccuracies stemming from clusters of sampling points.

The paper also explores utilization of the Voronoi diagram for identification of optimal locations for sample size extension.
1 INTRODUCTION

Monte Carlo estimation of statistical integrals is encountered in numerous applications. A typical example is the computer exploration of functions that feature random variables. These random variables form an $N_{\text{var}}$-dimensional vector, where $N_{\text{var}}$ is the number of random variables considered. In computer experiments the first step is a selection of optimal sample set, i.e. selection of $N_{\text{sim}}$ points from the $N_{\text{var}}$ dimensional space. These points then form the sampling plan which is an $N_{\text{sim}} \times N_{\text{var}}$ matrix. The methods used for formulating the plan of experimental points are collectively known as Design of Experiments (DoE). The purpose of DoE is to provide a set of points lying inside a chosen design domain that are optimally distributed; the optimality of the sample depends on the nature of the problem. Various authors have suggested intuitive goals for good designs, including “good coverage”, the ability to fit complex models, many levels for each factor/variable, and good projection properties. At the same time, a number of different mathematical criteria have been put forth for comparing designs.

The design of experiments is typically performed in a hyper-cubical domain of $N_{\text{var}}$ dimensions, where each dimension/variable, $U_v$, ranges between zero and one ($v = 1, \ldots, N_{\text{var}}$). This design domain is to be covered by $N_{\text{sim}}$ points as evenly as possible as the points within the design domain represent sampling probabilities. The probability that the $i$-th experimental point will be located inside some chosen subset of the domain must be equal to $V_S/V_D$, with $V_S$ being the subset volume and $V_D$ the volume of the whole domain (for unconstrained design $V_D = 1$). Whenever this is valid, the design criterion will be called statistically uniform. Moreover, each separate sampling plan should have the points spread evenly over the design domain. Even though such uniformity is conceptually simple and intuitive on a qualitative level, it is somewhat complicated to describe and characterize it mathematically. Though some problems do not require this uniformity, it is the crucial assumption in Monte-Carlo integration and its violation may lead to significant errors [5, 11].

There exist many other criteria of optimality of the sampling plan: e.g. the Audze-Eglajs (AE) criterion [1] later generalized into the so-called $\phi$ criterion, the Euclidean MaxiMin and MiniMax distance between points, various measures of discrepancy, criteria based on correlation (orthogonality), designs maximizing entropy and many others. It should also be noted that an experimental design can be also obtained via so-called “quasi-random” low-discrepancy sequences (deterministic versions of MC analysis) that can often achieve reasonably uniform sample placement in hypercubes (Niederreiter, Halton, Sobol’, Hammersley, etc.).

As mentioned above, the selection of the sampling points is a crucial step when evaluating approximations to integrals as the ones performed in Monte Carlo simulations (numerical integration). In such applications, equal sampling probabilities inside the design domain are required.

In this article, it is assumed that the sampling points have already been selected and they are not spread optimaly over the design domain. A typical example may be a sample selected using crude Monte Carlo sampling. The article considers the possibility to improve quality of Monte Carlo estimation with such a given sample. The only possibility to improve the estimations of the integrals is to vary the weights associated with individual sampling points. Motivated by the MiniMax criterion of optimality [6], we explore the possibility to improve the quality of statistical estimations using Voronoi tessellation, i.e. a particular form of partitioning of the design domain around given sampling points. The design domain to be partitioned is the unit hypercube described above and therefore the volumes around individual sampling points represent weights (probabilities) to be used in the weighted averages that estimate the integrals.
2 STATISTICAL MOMENT ESTIMATION USING MONTE CARLO SAMPLING

As mentioned in the introduction, one of the frequent uses of DoE is statistical sampling for Monte Carlo integration. We present the application of statistical sampling to the problem of estimating statistical moments of a function of random variables. In particular, a deterministic function, $Z = g(\mathbf{X})$, is considered, which can be a computational model or a physical experiment. $Z$ is the uncertain response variable (or generally a vector of the outputs). The vector $\mathbf{X} \in \mathbb{R}^{N_{\text{var}}}$ is considered to be a random vector of $N_{\text{var}}$ continuous marginals (input random variables describing uncertainties/randomness) with a given joint probability density function (PDF).

Estimation of the statistical moments of variable $Z = g(\mathbf{X})$ is, in fact, an estimation of integrals over domains of random variables weighted by a given joint PDF of the input random vector, $f_\mathbf{X}(\mathbf{x})$. We seek the statistical parameters of $Z = g(\mathbf{X})$ in the form of the following integral:

$$E[S[g(\mathbf{X})]] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} S[g(\mathbf{x})] \, dF_\mathbf{X}(\mathbf{x})$$  \hspace{1cm} (1)

where $dF_\mathbf{X}(\mathbf{x}) = f_\mathbf{X}(\mathbf{x}) \cdot dx_1 \, dx_2 \cdots dx_{N_{\text{var}}}$ is the infinitesimal probability ($F_\mathbf{X}$ denotes the joint cumulative density function) and where the particular form of the function $S[g(\cdot)]$ depends on the statistical parameter of interest. For example, to gain the mean value of $g(\cdot)$, $S[g(\cdot)] = g(\cdot)$; higher statistical moments of $Z$ can be obtained by integrating polynomials of $g(\cdot)$. The probability of an event (an event defined as $g(\cdot) < 0$) is obtained in a similar manner: $S[\cdot]$ is replaced by the Heaviside function (or indicator function) $H[-g(\mathbf{X})]$, which equals one for a failure event ($g < 0$) and zero otherwise. In this way, the domain of integration of the PDF is limited to the failure domain.

In Monte Carlo sampling, which is the most prevalent statistical sampling technique, the above integrals are numerically estimated using the following procedure: (i) draw $N_{\text{sim}}$ realizations of $\mathbf{X}$ that share the same probability of occurrence $1/N_{\text{sim}}$ by using its joint distribution $f_\mathbf{X}(\mathbf{x})$; (ii) compute the same number of output realizations of $S[g(\cdot)]$; and (iii) estimate the desired parameters as arithmetical averages. We now limit ourselves to independent random variables in vector $\mathbf{X}$. The aspect of the correct representation of the target joint PDF of the inputs mentioned in item (i) is absolutely crucial. Practically, this can be achieved by reproducing a uniform distribution in the design space (unit hypercube) that represents the space of sampling probabilities.

Assume now a random vector $\mathbf{U}$ that is selected from a multivariate uniform distribution in such a way that its independent marginal variables $U_v, v = 1, \ldots, N_{\text{var}}$, are uniform over intervals $(0;1)$. A vector with such a multivariate distribution is said to have an “independence copula” $\newline$

$$C(u_1, \ldots, u_{N_{\text{var}}}) = P(U_1 \leq u_1, \ldots, U_{N_{\text{var}}} \leq u_{N_{\text{var}}}) = \prod_{v=1}^{N_{\text{var}}} u_v$$  \hspace{1cm} (2)

These uniform variables can be seen as sampling probabilities: $F_{X_v} = U_v$. The joint cumulative distribution function then reads $F_\mathbf{X}(\mathbf{x}) = \prod_v F_{X_v} = \prod_v U_v$, and $dF_\mathbf{X}(\mathbf{x}) = \prod_v dU_v$. The individual random variables can be obtained by inverse transformations

$$\{X_1, \ldots, X_{N_{\text{var}}]\} = \{F_1^{-1}(U_1), \ldots, F_{N_{\text{var}}}^{-1}(U_{N_{\text{var}}})\}$$  \hspace{1cm} (3)

and similarly the realizations of the original random variables are obtained by the component-wise inverse distribution function of a point $\mathbf{u}$ (a realization of $\mathbf{U}$) representing a sampling
probability
\[ x = \{ x_1, \ldots, x_{N_{\text{var}}} \} = \{ F_1^{-1}(u_1), \ldots, F_{N_{\text{var}}}^{-1}(u_{N_{\text{var}}}) \} \]  

(4)

With the help of this transformation from the original to the uniform joint PDF, the above integral in Eq. (1) can be rewritten as

\[
E[S[g(X)]] = \int_0^1 \ldots \int_0^1 S[g(x)] \ dC(u_1, \ldots, u_{N_{\text{var}}})
\]

\[
= \int_{[0,1]^{N_{\text{var}}}} S[g(x)] \prod_{v=1}^{N_{\text{var}}} dU_v
\]

(5)

so that the integration is performed over a unit hypercube with uniform unit density.

We now assume an estimate of this integral by the following statistic (the average computed using \( N_{\text{sim}} \) realizations of \( U \), namely the sampling points \( u_j \) \(( j = 1, \ldots, N_{\text{sim}} )\))

\[
E[S[g(X)]] \approx \frac{1}{N_{\text{sim}}} \sum_{i=1}^{N_{\text{sim}}} S[g(x_i)]
\]

(6)

where the sampling points \( x_i = \{ x_{i,1}, \ldots, x_{i,v}, \ldots, x_{i,N_{\text{var}}} \} \) are selected using the transformation in Eq. (4), i.e. \( x_{i,v} = F_v^{-1}(u_{i,v}) \), in which we assume that each of the \( N_{\text{sim}} \) sampling points \( u_i \) \(( i = 1, \ldots, N_{\text{sim}} )\) were selected with the same probability of \( 1/N_{\text{sim}} \). Violation of the uniformity of the distribution of points \( u_j \) in the unit hypercube may lead to erroneous estimations of the integrals.

If the sampling points were not selected with respect to equal probabilities in the design domain, the possibility to improve the accuracy in Eq. (6) is to use weights different from \( 1/N_{\text{sim}} \). These weights reflect the probability content of the cells around individual sampling points

\[
E[S[g(X)]] \approx \frac{1}{W} \sum_{i=1}^{N_{\text{sim}}} S[g(x_i)] \cdot w_i
\]

(7)

where \( W = \sum_{i=1}^{N_{\text{sim}}} w_i \) is the sum of weights for \( N_{\text{sim}} \) points (normalization). The proposed approach aims at finding appropriate weights that are calculated considering the spatial distribution of the points. In Monte Carlo sampling, for example, the sampling distribution may correspond to the joint distribution function (CDF) of the random vector, but the sampling strategy is so inefficient that the sample of \( N_{\text{sim}} \) points does not reproduce the CDF well. Reweighting of samples based on a true distribution of points seems to be a way to improve the accuracy. Obviously, unvisited regions of the design domain can not be explored by a nonuniform design.

In any case, partitioning the space into cells around the given sampling points may help to (i) reduce probabilities associated with points that are participating in clusters of points and, at the same time, (ii) identification of unexplored regions may help in adjusting the weights of the existing samples with respect to the volumes of regions they occupy, (iii) the identified unexplored regions can be used for sample size extension by new points in which the function can be additionally evaluated, if possible.

Voronoi tessellation has been selected for partitioning of the design space into volumes that are used as the weights \( w_i, i = 1, \ldots, N_{\text{sim}} \). The following section describes the Voronoi tessellation procedures.
3 WEIGHTS OBTAINED AS VOLUMES OF VORONOI REGIONS

The weights associated with the design points are considered as volumes of Voronoi regions [2] computed on the sampling points. The Voronoi tessellation in \( N_{\text{var}} \)-dimensional space results in \( N_{\text{sim}} \) convex polyhedrons \( V_i \) that enclose all the points that are closer to \( i \)-th sampling point than any other. Defining the distance of point \( u \) from sampling point \( u_i \) as \( d_i(u) \), the Voronoi region associated with \( i \)-th sampling point can be formally defined as

\[
V_i = \left\{ u \in \mathbb{R}^{N_{\text{var}}} \mid \forall j \neq i : d_i(u) \leq d_j(u) \right\}
\]

(8)

Two alternatives of Voronoi tessellation that differ in the boundary regions are investigated:

- **clipped** Voronoi tessellation that is limited to the unit hypercube only

\[
V_i = \left\{ u \in (0, 1)^{N_{\text{var}}} \mid \forall j \neq i : d_i(u) \leq d_j(u) \right\}
\]

(9)

- **periodic** Voronoi tessellation which assumes that every sampling point is periodically repeated in the space along all the dimensions.

These two different concepts are demonstrated in Fig. 1. The reason for studying the periodic tessellation is that the authors have shown recently [5, 11] that the presence of boundaries in the hypercubical design domain cause problems. Briefly, one may think of a problem of packing (hyper)balls into a (hyper)cube. It is clear that the boundary is responsible for a kind of wall-effect. It has been shown [5, 11] that this problem can be removed by considering periodic extension of the design domain. The balls then permeate through the boundaries without interacting with them, see Fig. 1 right.

The clipped Voronoi diagrams [4, 14] are used mostly for construction of meshes and therefore available software to compute such tessellation is limited to two and three dimensional space. A similar situation exists for periodic Voronoi tessellation [13, 10]. In the field of design of experiments more than three variables (factors) can be present and therefore the tessellation must be performed in higher dimensions. In this contribution, Qhull software [3] is utilized for both clipped and periodic tessellations because it can compute Voronoi tessellation for arbitrary dimension. On the other hand, it cannot work directly with neither clipped nor periodic boundary condition and therefore simple tricks are used.

These tricks consist in manipulations of the design domain (together with the sampling points contained) by adding new design domains around it. In order to obtain the clipped structure,
the design domain is extended by reflecting the original design domain along each dimension. There are two reflections of the original unit interval along each dimension to obtain intervals \([-1, 0)\) and \((1, 2]\). Therefore, the tessellation is performed on \(N_{\text{sim}}(1 + 2^{N_{\text{var}}})\) points. The use of reflection automatically provides edges between cells that coincide with the boundary of the original design domain and therefore the volumes outside the design domain can be ignored. The use of reflection to obtain clipped tessellation was proposed in [9].

The periodic structure is obtained by periodic extension (replication) of the original design domain along each direction and additionally the replication must be performed to obtain all the “corner” domains to fill a hypercube \([-1, 0)^{N_{\text{var}}}\). Therefore, \(N_{\text{sim}} \cdot 3^{N_{\text{var}}}\) points in total are used for the periodic tessellation.

The computational times needed for the both tessellation types can be substantially reduced if it involves only reflected or periodically repeated points that are close to the original hypercube, because only these points affects the tessellation inside the hypercube. Unfortunately, no effective algorithm has been developed yet to identify such points and therefore the full set of points must be involved for certainty.

In both alternatives, the weights for individual sampling points are the volumes of regions surrounding points. There are three algorithms available for the volume computation: (i) direct integration, (ii) Monte-Carlo integration and (iii) division into simplexes for which analytical formula is available. The first two algorithms are nicely elucidated in [8]. Here, we perform the third algorithm. Each Voronoi region is (with a help of the Qhull) divided into simplexes. Each simplex has \(N_{\text{var}} + 1\) vertices denoted \(v_j\). The total volume of the region is simply the sum of simplex volumes, that are calculated based on the determinant of coordinate matrix.

\[
V_{\text{simplex}} = \frac{1}{N_{\text{var}}!} \left| \begin{array}{c}
    v_1 - v_0 \\
    v_2 - v_0 \\
    \vdots \\
    v_{N_{\text{var}}} - v_0
\end{array} \right|
\]

These volumes are used directly as weights of sampling points enclosed within these cells.

4 FREQUENCY ANALYSIS OF WEIGHTS

It turns out to be important to see (i) whether the weights are very scattered compared to \(1/N_{\text{sim}}\) and, (ii) whether their magnitude tend to depend on the position inside the domain. This is achieved by studying \(N_{\text{run}} = 1000\) realizations of samples, each having \(N_{\text{sim}}\) points within an \(N_{\text{var}}\)-dimensional hypercube. For each sample, both types of Voronoi tessellation is constructed and the weights are statistically processed.

The results will be presented for two sampling schemes: the classical (crude) Monte Carlo sampling without any optimization (MC-RAND) and LHS (Latin Hypercube Sampling) optimized using the periodic criterion (LHS-PAE). PAE stands for an enhanced version of the Audze-Eglais criterion, see [5, 11].

Figure 2 shows one sample \((N_{\text{sim}}=16)\) of a bivariate random vector \(U_v\) for both sampling schemes. For the two sampling schemes, both types of Voronoi diagrams \((clipped\ and\ periodic)\) are constructed and visualized with colors depending on the area. The LHS-PAE sampling plans show more uniform distribution of points because the PAE-optimized LH-sampling better avoids clustering and limit the occurrence of empty regions. Therefore, the cells in LHS-PAE have similar volumes and the sampling points are closer to the centers of Voronoi regions. The small differences among weights in LHS-PAE with periodic tessellation suggest that weight-
Figure 2: Voronoi weights for MC-RAND and LHS-PAE sampling plans ($N_{var} = 2, N_{sim} = 16$). Comparison of the clipped and periodic tessellations.

Integration will not make much difference in comparison with integrals evaluated using equal weights $1/N_{sim}$. The MC-RAND sampling plans suffer from point clustering and therefore, high variability in volumes of the Voronoi cells is observed. It should be noticed that the choice of tessellation (clipped vs. periodic) affects only the boundary regions while the central part of the hypercube is identical.

Figure 3: Bivariate histograms of the mean value and the standard deviation of cell volumes for both sampling plans and both tessellation alternatives ($N_{var} = 2, N_{sim} = 16$ and $N_{run} = 1000$ realizations).

In order to judge about the spatial distribution of weights within the design domain, the above-mentioned $N_{run}=1000$ realizations of samples accompanied by Voronoi tessellations were prepared and for each spatial location, the mean value and standard deviation of weights occurring at that location have been calculated. The weights (volumes of Voronoi regions $V_{simplex}$ in a hypercube) depend on the type of tessellation but they are independent of the sampling method (MC vs. LHS). The bivariate histograms in Fig. 3 document the dependency of the mean value and the standard deviation of Voronoi region volumes on the position of the sampling point in a square. In the case of clipped tessellation, both the mean value and the standard deviation of weights are not uniform in the hypercube. Three zones can be distinguished: (a) the boundary region where the mean value (shown in blue) of weights is underestimated. The boundary strip is followed/balanced by (b) zone parallel to the boundary where the weights are overestimated.
(see the yellow to red color) and finally, (c) the bulk zone sufficiently far from the boundary, where the weights (volumes) are constant on average. The width of the two boundary zones is decreasing with increasing sample size $N_{\text{sim}}$.

Such a biased representation of different regions in the hypercube partitioned by the clipped tessellation has consequences in Monte Carlo integration. If the points are sampled uniformly, and that is indeed the case of both MC-RAND and LHS-PAE, errors are introduced due to introduction of nonuniform weighting. If the functions are sensitive to inaccuracies in representation of the boundary regions, their weighted MC integration may yield biased results. Therefore we conclude that the clipped tessellation generally should not be used for weighting in MC integration.

The periodic tessellation provides more promising bivariate histograms: no bias around the boundaries is visible for both MC-RAND and LHS-PAE sampling schemes. The statistics of the weights do not depend systematically on the position in the hypercube.

Numerical simulations focused on MC integration presented in Fig. 4 of our recent paper [12] revealed that reweighting the samples according to the volumes of the Voronoi cells in periodic space does not significantly improve the accuracy estimators compared to the standard estimations using equal weights $1/N_{\text{sim}}$. The conclusion was that the gain in accuracy was not worth the effort. Anyway, the periodicity of the Voronoi tessellation was found an important ingredient that guaranties that the Voronoi reweighting does not systematically bias the results.

Fig. 2 shows that good designs such as those obtained by LHS-PAE do not need any reweighting at all as the points have quite regular distribution in the the design domain. MC-RAND designs, on the other hand, can lead to Voronoi cells of quite different volumes. Unfortunately, the tessellation proposed so far does not deal with neither clusters of points nor large empty spaces.

5 ADAPTIVE REFINEMENT BY POINT CLUSTERING & IDENTIFICATION OF EMPTY REGIONS

In this section we present an enhanced weighting algorithm that is able to improve tessellation of a given point layout by an adaptive sequence of two different kinds of steps: (a) grouping of a pair of points that form a cluster (occupy relatively small region) and, (b) insertion of dummy points within regions that are not occupied by the original sampling points. The dummy points help the tessellation to identify volumes of regions that are not represented by any point and thus the volumes must be subtracted form the total unit volume.

The process is performed by individual steps of either point grouping or point insertion and the decision of which step is taken is driven by given rules. The sequence of steps may evolve differently in various designs depending on the particular point layout. At every stage, the key number in the decision of what step to take is the characteristic length that we define as:

$$l_{\text{char}} = \frac{1}{N_{\text{var}}/N_{p}}$$

where $N_{p}$ is the current number of points (including the dummy points and without the redundant points in clusters). The formula makes $l_{\text{char}}$ the average distance to the nearest neighbor in and “ideal” design. The rationale behind this definition of $l_{\text{char}}$ is a regular orthogonal grid of $N_{p} = N^{N_{\text{var}}}$ points in the unit hypercube, where the distance to the nearest neighbor is measured along any dimension and reads $1/N$. The number of points $N_{p}$ is initially equal to the number of points in the design, $N_{\text{sim}}$, and after insertion of a new dummy point is increased by one and
after *grouping* of two points is decreased by one. After any of these steps is taken, the list of points and the value of $l_{\text{char}}$ must be updated.

The characteristic length serves for comparison with two distances of the current design stage, see Fig. [4].

- the smallest inter-point distance, $l_{\text{min}}$, (measured in the periodic space), and
- the diameter of the largest empty $N_{\text{var}}$-dimensional hypersphere, $l_{\text{max}}$, (again in the periodic design domain).

These lengths are also calculated before every decision whether to perform the point *grouping* step or *insertion* step. With these lengths at hand, two parameters are calculated, see Fig. [4]:

$$d_{\text{min}} = |l_{\text{min}} - l_{\text{char}}|, \quad d_{\text{max}} = |l_{\text{max}} - 1.25l_{\text{char}}|$$ \hspace{1cm} (12)

If $d_{\text{min}} > d_{\text{max}}$ the algorithm prefers *grouping* of the closest pair of points with the distance $l_{\text{min}}$ by placing the new point in the centroid of the cluster. Otherwise ($d_{\text{min}} \leq d_{\text{max}}$), the algorithm prefers *insertion* of one dummy point into the center of the largest empty hypersphere of diameter $l_{\text{max}}$.

In any of the two cases, an additional condition have to be met prior to performing any of the two steps. For *point grouping* to be performed, the condition $l_{\text{min}} < 0.5l_{\text{char}}$ must be fulfilled, otherwise the algorithm terminates the adaptive process. Similarly, *point insertion* is made only if the sphere diameter satisfies: $l_{\text{max}} > 1.75l_{\text{char}}$, otherwise the algorithm terminates.

Regarding the grouping step, we mention that any of the two grouped points may already represent a cluster of points. Therefore the points have weights equal to the number of point they represent already (initially they have unit weights). This weighting guarantees that if a cluster of more than two original points occurs, the inserted point after grouping is in the centroid of the whole cluster. The additional conditions serve for halting the process and the constants are tuned manually such that the algorithm leads to quite uniform distribution of Voronoi cells with points relatively close to the centroids of the Voronoi cells (the points are good repentants of their cells). Three kinds of points are available at the end of the adaptive process:

- the original sampling points that have not been clustered (see the black solid circles in Fig. [5]). They are associated with weights according to the volumes of Voronoi cells they occupy.
- the centroids of clusters (see the green solid circles in Fig. [5]). The original points within a single cluster have share weights equal to the volume of the corresponding cell divided by the number points inside it.
• the dummy points (see the red solid circles in Fig. 5). They have zero weights as they do not represent any original point. These points will be considered as good candidates for sample size extension.

Fig. 5 shows examples of Voronoi diagrams obtained with the adaptive algorithm, applied to initial designs obtained by four different techniques. In the same figure, they are compared with the basic periodic Voronoi diagrams as described in Sec. 3, see the top row. The weights of points are represented by the darkness of blue color. Green points represent clusters of the original (white) points.

**Figure 5**: Examples of Voronoi tessellations of various designs with the initial sample size of $N_{\text{sim}} = 32$ points. Top row: periodic tessellation. Bottom row: adaptive tessellation with identified clusters and the “dummy” points.

In the basic setting of the adaptive algorithm, the red points are dummy points and the associated red areas are excluded. In other words, the existence of the red areas signalize that the sum of weights (volumes), $W$, is less than one. The adaptive algorithm basically redistribute the total unit volume among the cells (including the red ones) such that the volumes are almost identical and the points tend to be close to the centroids of the cells. The fact that the blue regions have similar volumes means that the weights among samples are similar, approximately $1/N_p$.

One of the outcomes of the proposed adaptive algorithm is that the points inside the red areas can be used as candidate points for sample size extension. They are placed approximately in the centers of the largest empty hypersphere’s and therefore they are points for the design refinement. The next section studies two aspects of the proposed adaptive Voronoi weighting, namely

(i) whether the reweighting of existing $N_{\text{sim}}$ points leads to improvement in Monte Carlo integration (blue regions in Fig. 5 bottom are considered), and

(ii) how much gain in accuracy and variance reduction is achieved by extending the sample with the proposed “dummy points”, evaluating the studied function and considering them in addition to the existing $N_{\text{sim}}$ points with the adaptive Voronoi weights (both the blue and the red regions in Fig. 5 bottom are considered).
6 NUMERICAL EXAMPLES OF MC INTEGRATION & DISCUSSION

This section studies whether weighting in MC integrals based on the Voronoi tessellation improves the quality of the estimates. Three basic transformations \( g(\mathbf{X}) \) of standard independent Gaussian random variables \( X_v, v = 1, \ldots, N_{\text{var}} \) have been selected for the numerical study. The following equation array presents formulas of the three functions (first column), the analytical solutions for the mean values (second column) and the standard deviations (third column):

\[
\begin{align*}
Z_{\text{sum}} &= g_{\text{sum}}(\mathbf{X}) = \sum_{v=1}^{N_{\text{var}}} X_v & \mu_{\text{sum}} = 0 & \sigma_{\text{sum}} = \sqrt{2} \\
Z_{\text{exp}} &= g_{\text{exp}}(\mathbf{X}) = \sum_{v=1}^{N_{\text{var}}} \exp(-X_v^2) & \mu_{\text{exp}} = \frac{\sqrt{3}}{3} N_{\text{var}} & \sigma_{\text{exp}} = \sqrt{N_{\text{var}}} \sqrt{\frac{\sqrt{5}}{5} - \frac{1}{3}} \\
Z_{\text{prod}} &= g_{\text{prod}}(\mathbf{X}) = \prod_{v=1}^{N_{\text{var}}} X_v & \mu_{\text{prod}} = 0 & \sigma_{\text{prod}} = 1
\end{align*}
\]

Two unoptimized sampling schemes, MC-RAND and LHS-RAND, have been used to prepare \( N_{\text{run}} = 1000 \) sampling plans for various sample sizes \( N_{\text{sim}} \) ranging from 2 to 512. Only bivariate cases are studied: \( N_{\text{var}} = 2 \).

The performance of the approaches will be demonstrated by showing their ability to estimate the mean value and standard deviation of the transformed variable \( Z = g(\mathbf{X}) \). The estimated mean value and standard deviation are denoted as \( \bar{\mu}_Z \) and \( \bar{\sigma}_Z \), respectively, and can be estimated in the spirit of Eq (7) as:

\[
\bar{\mu}_Z = \frac{1}{W} \sum_{i=1}^{N_{\text{sim}}} g(\mathbf{x}_i) \cdot w_i \quad (16)
\]

\[
(\bar{\sigma}_Z)^2 = \frac{1}{W} \sum_{i=1}^{N_{\text{sim}}} [g(\mathbf{x}_i) - \bar{\mu}_Z]^2 \cdot w_i \quad (17)
\]

where \( W = \sum_{i=1}^{N_{\text{sim}}} w_i \) is the sum of weights for \( N_{\text{sim}} \) points.

Three approaches to the weighting in Monte-Carlo type numerical integration were studied:

1. uniform weights that assign each design point a constant weight \( w_i = 1/N_{\text{sim}} \).
2. Voronoi weights from periodic tessellation. Weights are equal to the “volumes” of cells obtained by the adaptive algorithm \( W \leq 1 \).
3. Voronoi weights from periodic tessellation. Weights are equal to the “volumes” of cells obtained by the adaptive algorithm \( W = 1 \) and considering the additional samples (the red regions).

We only used the periodic tessellations as the clipped tessellations provide systematically wrong results [12].

The results of numerical study are presented in Fig. 6 for all three functions, two estimated statistical moments, two sampling schemes and two alternatives of weighting. The second alternative is not presented because the adaptive algorithm without the additional points leads to approximately equal weights \( 1/N_{\text{p}} \) for all \( N_{\text{sim}} \) samples and thus has no effect in Eq. (17). The results are almost identical with those obtained with weight equal to \( 1/N_{\text{sim}} \) (approach 1).

The third technique improves the accuracy of average estimates and it also significantly decreases the variance of the estimator compared to alternative standard equal weighting (alternative 1). This improvement comes at a price of additional effort associated with the new
Figure 6: Convergence of the estimated mean values and standard deviations of the three transformed variables $Z_{\text{sum}}$, $Z_{\text{exp}}$ and $Z_{\text{prod}}$, computed for $N_{\text{ref}} = 2$. Two weighting alternatives are compared for two initial sampling schemes: MC-RAND and LHS-RAND. LHS-PAE is used as a reference technique.

samples. We have found that the number of additional samples is such that for small designs is around 20% of $N_{\text{sim}}$ and this percentage increases to approximately 33% for large sample sizes. In order to show how much gain for the additional sample size is obtained, we have compared the results with those obtained by employing LHS-PAE designs.

Each alternative is represented by a solid line showing the average estimations ± one sample standard deviation (a scatter-band shown by the shaded area); both computed using the $N_{\text{run}} = 1000$ realizations.

The mean values obtained by LHS-schemes for $Z_{\text{sum}}$ and $Z_{\text{exp}}$ has no scatter at all. This is due to the fact that for these additive functions, the LH-samples are transformed to the exact
mean value of each separate variable as these variables are sampled regularly with respect to probabilities. The pairing does not matter. Therefore, the estimations using LH-designs can not improved by adding new points.

In most cases, however, the standard equal weights \(1/N_{\text{sim}}\) provide the largest variance of the estimators; employing LHS instead of MC helps to reduce the variance; and the best results are obtained with LHS-PAE. The accuracy of existing LHS and MC samples is always improved by using the extended sample size with weighting obtained using the proposed adaptive algorithm.

7 CONCLUSIONS

Various alternatives of Voronoi tessellation was studied in an attempt to improve the accuracy of Monte-Carlo integration schemes for small \(N_{\text{sim}}\) by weighting individual sampling points. The weights were obtained as volumes of the Voronoi cells – the regions surrounding the sampling points in the design domain (unit hypercube).

Weighting using the clipped Voronoi tessellation (a tessellation limited to the design domain) was found inapplicable due to problems related to the presence of boundaries of the unit hypercube. The tessellation results in systematic appearance of underestimated regions near the boundaries followed by regions with over-weighted regions. The periodic tessellation removes the systematic bias.

Pure reweighting the sample using Voronoi weights does not help much to increase the accuracy. The proposed adaptive algorithm that identify and remove unvisited regions does not help much despite its ability to find clusters of points and decrease their influence accordingly.

The proposed adaptive algorithm, however, seems to drastically improve the results by adding up to 33% of additional point to the design. Results obtained initially with unoptimized designs seems to improve significantly at the price of the additional evaluations at the new points. The performance of a such an extended sample is almost as good as performance obtained with designs that were optimized in advance. Therefore, the proposed technique can be viewed as a sophisticated sample size extension technique.

ACKNOWLEDGEMENT

This work has been supported by the Grant agency of the Czech Republic under projects Nos. GA16-22230S and GA15-07730S. This support is gratefully acknowledged.

REFERENCES


BAYESIAN FORMULATIONS FOR FORCE RECONSTRUCTION PROBLEMS

Mathieu Aucejo¹ and Olivier De Smet¹

¹Structural Mechanics and Coupled Systems Laboratory - Conservatoire National des Arts et Mîiers
2 rue Conté, 75003 Paris
e-mail: {mathieu.aucejo, olivier.desmet}@lecnam.net

Keywords: Force reconstruction, Bayesian inference, Generalized Gaussian priors, MCMC.

Abstract. To identify mechanical sources acting on a structure, Tikhonov-like regularizations are generally used. These approaches, however, only provide point estimates, meaning that the uncertainty about the regularized solution is not quantified. In practice, such information is essential to guarantee the quality of reconstructed sources. In this contribution, three possible Bayesian formulations of the source identification problem are presented and their limitations discussed. To assess the posterior uncertainty on the parameters appearing in each formulation given a simulated vibration field and a mechanical model, a Gibbs sampler is implemented. The proposed numerical validations highlight the practical interest of these formulations in terms of parameters estimations and posterior uncertainty quantification.
1 INTRODUCTION

In structural dynamics, most of the research papers are generally focused on the modeling of the dynamic response of structures subject to perfectly determined excitation sources. However, the latter are only roughly or partially known in practice. As a result, an error in the definition of the excitation vector is propagated to the dynamic response through the model and can have a significant impact on the subsequent mechanical analysis. However, direct measurement of excitation sources can be practically unfeasible. A possible alternative is to perform indirect measurements using a model of the dynamic behavior of the studied structure and accessible quantities such as displacement or acceleration fields. Unfortunately, the reconstruction of mechanical sources from vibration measurements is an ill-posed inverse problem. A classical approach to bypass this difficulty consists in constraining the space of solutions by using prior information on the noise and the sources to reconstruct. A convenient and efficient way to deal with such prior information is the Bayesian framework, because it allows combining both probabilistic and mechanical data. The most widespread and popular approaches deriving from Bayesian statistics are certainly Tikhonov-like regularizations \([1, 2, 3]\). Although widely used and deeply studied, these methods generally provide only point estimates. In other words, there is no information on the uncertainty about the regularized solution given the measured data and the mechanical model. However, such information is of primary interest for industrial applications, in which it is essential to guarantee the quality of obtained results. In the present contribution, three possible Bayesian formulations of the source identification problem, based on the use of Generalized Gaussian distributions, Gamma and truncated Gamma distributions, are proposed. The limitations of each formulation are discussed from the identification of two close point forces acting on a free-free beam. To assess the posterior uncertainty on the parameters appearing in each formulation given a simulated vibration field and a mechanical model, a Gibbs sampler \([4]\), including Hamiltonian Monte Carlo \([5]\) updates, is implemented to perform the inference. The proposed MCMC procedure is detailed in this contribution. The proposed numerical validations highlight the practical interest of the proposed formulations in terms of parameters estimations and posterior uncertainty quantification.

2 BAYESIAN FORMULATIONS OF THE RECONSTRUCTION PROBLEM

This section aims at introducing the three formulations of the Bayesian force reconstruction problem as well as the related parameters. To render this section more didactic, each formulation is applied on an academic test case in order to better highlight its advantages and limitations. The MCMC algorithm used to perform the inferences will be detailed in section \([5]\).

2.1 Description of the benchmark test case

Before presenting the three Bayesian formulations of the reconstruction problem, we propose to introduce the test case that will serve as a benchmark to analyze and compare each of them. The studied structure is a free-free steel beam with dimensions \(1 \times 0.03 \times 0.01 \text{ m}^3\) excited by two point forces of unit amplitude at 350 Hz. The coordinate of the point forces, measured from the left end of the beam, are \(x_1 = 0.6 \text{ m}\) and \(x_2 = 0.7 \text{ m}\) [see Fig. \([\text{I}]\)].

To perform the reconstruction, a model has to be derived to relate the measured vibration field and the excitation field. For this purpose, let us consider the practical situation where the vibration field \(X\), measured over the surface of a structure, is caused by an unknown excitation field \(F\). If the structure is linear and time invariant, its dynamic behavior is completely deter-
mined by the transfer functions matrix $H$, relating the vibration field $X$ to the excitation field $F$, so that:

$$X = HF + N,$$

where $N$ is the measurement noise vector.

A finite element model of the beam made up with 20 plane beam elements has been used to compute the transfer functions matrix $H$ by assuming that only bending motions are measurable. In other words, the transfer functions matrix is dynamically condensed over the measurable dofs, corresponding here to the bending motions. To simulate the measured vibration field $X$, Eq. (1) is applied using a noise vector $N$ corresponding to a Gaussian white noise, which is calculated assuming a signal-to-noise ratio equal to 34 dB. Finally, it is worth mentioning that the model defined in Eq. (1) is the basis of the proposed Bayesian formulations, meaning that the transfer functions matrix $H$ is also used in the inverse problem. In other words, modeling errors are not considered here.

### 2.2 Standard Bayesian formulation

Formally, the Bayesian paradigm considers all the parameters of the problem as random variables. Consequently, the uncertainty on each parameter is modeled by a probability distribution, describing the state of knowledge or the prior on this parameter. From a mathematical standpoint, the Bayesian reconstruction problem relies on the Bayes’ rule:

$$p(F|X) \propto p(X|F) \ p(F),$$

where:

- $p(F|X)$ is the posterior probability distribution, representing the probability of observing $F$ given a vibration field $X$. It defines what it is known about the excitation field $F$ after making vibration measurements;

- $p(X|F)$ is the likelihood function, representing the probability of measuring $X$ given an excitation field $F$. It reflects the uncertainty related to the measurement of the vibration field $X$;
• $p(F)$ is the prior probability distribution, representing our knowledge on the unknown excitation field $F$ before measuring the vibration field $X$.

Generally, the quality of the force reconstruction strongly depends on the choice of the likelihood function and the prior probability distribution. That is why, the choices made in this contribution have to be carefully explained.

### 2.2.1 Choice of the likelihood function

The likelihood function reflects the uncertainty related to vibration measurements. By definition, this uncertainty is mainly related to the measurement noise $N$. Consequently, the likelihood function $p(X|F)$ can be written under the following form:

$$p(X|F) = p(X - HF|N),$$

representing the probability of obtaining $X - HF = 0$ given the measurement noise $N$.

If the noise is supposed to be due to multiple independent causes, then the likelihood function can be represented by a complex multivariate normal distribution with zero mean and precision $\tau_n$:

$$p(X|F, \tau_n) = \left[\frac{\tau_n^2}{\pi}\right] \frac{1}{\sqrt{\pi}} \exp\left[-\tau_n \|X - HF\|^2_{\ell_2}\right],$$

where $N$ is the number of measurement points.

### 2.2.2 Choice of the prior probability distribution

The prior probability distribution reflects the uncertainty related to the unknown excitation field $F$. Actually, it can be seen as a measure of the a priori knowledge of the experimenter on the sources to identify.

For practical reasons, the excitation field $F$ is supposed to be a real random vector, whose components are independent and identically distributed random variables following a Generalized Gaussian distribution. As a result, the prior probability distribution is written:

$$p(F|\tau_s, q) = \left[\frac{q}{2 \Gamma(1/q)}\right]^M \tau_s^{M \frac{M}{q}} \exp\left[-\tau_s \|F\|^q_{\ell_q}\right],$$

where:

- $q$ is the shape parameter of the distribution. Its value is defined in the interval $[0, +\infty[$;
- $\| \cdot \|_{\ell_q}$ is the $\ell_q$–norm or quasi-norm, if $q \geq 1$ and $q < 1$ respectively;
- $\tau_s$ is the scale parameter of the distribution, which can be viewed as a generalized measure of the precision of the distribution;
- $M$ is the number of reconstruction points;
- $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} \, dt$ is the gamma function.
It should be noted that the choice of a multivariate generalized Gaussian distribution allows a great flexibility for describing prior knowledge of the sources to identify. Formally, sparse excitation fields are promoted for \(q \leq 1\), while distributed excitation fields are favored for \(q = 2\) \cite{6}. In practical situations, the possible values of the shape parameter are bounded in the interval \([0, 2]\).

### 2.3 Summary and application

From the explanations given above, the standard Bayesian formulation of the reconstruction problem finally writes:

\[
p(F|X, \tau_n, \tau_s, q) \propto p(X|F, \tau_n) p(F|\tau_s, q).
\] (6)

This formulation is said standard, because it leads to Tikhonov-like regularizations, which correspond to the MAP estimate of Eq. (6). To explore the posterior probability distribution, \(10^5\) samples are drawn from Eq. (6). Fig. 2 and Tables 1 and 2 present the excitation fields and the estimated values of the forces \(F_1\) and \(F_2\) obtained using either with a proper choice of \(q\) (\(q = 0.5\)) [see Fig. 2a] or with a poor choice of \(q\) (\(q = 2\)) [see Fig. 2b]. It should be noted that optimal values of \(\tau_n\) and \(\tau_s\) are strongly related to the value of \(q\) and are computed accordingly as explained in section 3. Here, \((\tau_n, \tau_s) = (9.77 \times 10^{15}, 1.48)\) for \(q = 0.5\) and \((\tau_n, \tau_s) = (2.87 \times 10^{15}, 12.44)\) for \(q = 2\).

![Figure 2: Standard Bayesian formulation – Real part of the reconstructed force vector (a) with a proper choice of \(q\) (\(q = 0.5\)) and (b) a poor choice of \(q\) (\(q = 2\)) – (—) Reference, (−−) Median of the samples and (□) 95% credible interval](image)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Median</th>
<th>Mode</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>(F_1)</td>
<td>0.996</td>
<td>0.997</td>
<td>[0.942, 1.049]</td>
</tr>
<tr>
<td>(F_2)</td>
<td>0.991</td>
<td>0.993</td>
<td>[0.934, 1.048]</td>
</tr>
</tbody>
</table>

Table 1: Standard Bayesian formulation – Summary of the inference result on the model parameters for \(q = 0.5\)

Several interesting conclusions can be drawn at the light of the results presented in Fig. 2 and Tables 1 and 2. First of all, the choice of the supposedly known parameters have a significant
Table 2: Standard Bayesian formulation – Summary of the inference result on the model parameters for $q = 2$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Median</th>
<th>Mode</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>0.542</td>
<td>0.538</td>
<td>[0.253, 0.831]</td>
</tr>
<tr>
<td>$F_2$</td>
<td>0.448</td>
<td>0.498</td>
<td>[0.199, 0.777]</td>
</tr>
</tbody>
</table>

impact on the quality of reconstruction. Indeed, a proper choice of the shape parameter leads to a very good agreement of the reference field with the median of the samples, associated with sharp posterior uncertainty region characterized by the related 95% credible interval. On the contrary, a poor choice of the shape parameter leads to a disappointed result, since the shape of the median of the samples does not allow to discriminate the two point forces. Furthermore, the posterior uncertainty region is quite large, which is closely related to the value of the shape parameter $q$. All things being equal, the larger the value of the shape parameter is, the larger is the area of the posterior uncertainty region.

The standard formulation gives information on the credibility of the identified excitation field given the measured vibration field $X$, the precisions $\tau_n$ and $\tau_s$ and the shape parameter $q$. Actually, this is the main drawback of the standard formulation, because the quality of the inference is conditioned to the knowledge of the precisions and the shape parameter. If their values are poorly chosen, then the resulting inference won’t be representative of the actual distribution. As a consequence, it is compulsory to determine near-optimal values of $q$, $\tau_n$ and $\tau_s$ if one wants to perform a relevant statistical inference.

2.4 Extended Bayesian formulation

To alleviate the limitations of the standard formulation, the approach generally adopted in the literature consists in considering the precisions $\tau_n$ and $\tau_s$ as random variables, while letting fixed the shape parameters $q$. If we further consider the precisions as independent variables, the following extended formulation is obtained:

$$p(F, \tau_n, \tau_s | X, q) \propto p(X|F, \tau_n) p(F|\tau_s, q) p(\tau_n) p(\tau_s),$$

where $p(\tau_n)$ and $p(\tau_s)$ are the prior probability distributions of the precisions $\tau_n$ and $\tau_s$ respectively.

2.4.1 Choice of the prior probability distribution of the precisions

The choice of the priori probability distributions $p(\tau_n)$ and $p(\tau_s)$ is first limited to distribution having a strictly positive support, because the precisions $\tau_n$ and $\tau_s$ are real positive numbers. The common choice, made in the literature, is the Gamma distribution. The reason for this is rather clear, since the conjugate prior for the precision of a generalized Gaussian distribution is a Gamma distribution [7]. Practically, the Gamma distribution is defined by:

$$G(\tau | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} \exp(-\beta \tau) \quad \text{with} \quad \alpha > 0, \ \beta > 0,$$

where $\alpha$ and $\beta$ are respectively the scale parameter and the rate parameter of the distribution.
However, the use of a Gamma distribution is questionable, since it has been chosen for mathematical convenience and does not reflect any real prior information on the precisions, except their positiveness. That is why, the prior distribution on $\tau_n$ and $\tau_s$ should be as minimally informative as possible [8]. To this end, one sets $\alpha_n = \alpha_s = 1$ and $\beta_n = \beta_s \to 0$.

### 2.4.2 Summary and application

Considering the previous choice, the extended Bayesian formulation of the reconstruction problem is given by:

$$p(F, \tau_n, \tau_s | X, q) \propto p(X | F, \tau_n) p(F | \tau_s, q) p(\tau_n | \alpha_n, \beta_n) p(\tau_s | \alpha_s, \beta_s),$$

where $(\alpha_n, \beta_n)$ are the hyperparameters related to the precision $\tau_n$, while $(\alpha_s, \beta_s)$ are the hyperparameters related to the precision $\tau_s$.

This extended formulation has given rise, when $q = 2$, to the augmented Tikhonov regularization [9]. This method provides a point estimate corresponding to a critical point of the opposite of the logarithm of the posterior probability distribution. The main advantage of this approach is to determine the regularized solution and the precision simultaneously using an iterative process.

As previously done, $10^5$ samples have been drawn from the MCMC algorithm presented in section 3 to explore the posterior probability distribution. Fig. 3 presents the excitation fields obtained using either a proper choice of $q$ ($q = 0.5$) [see Fig. 3a] or with a poor choice of $q$ ($q = 2$) [see Fig. 3b], while Tables 3 and 4 summarize the inference results on each parameter of the model for $q = 0.5$ and $q = 2$ respectively.

![Figure 3: Extended Bayesian formulation – Real part of the reconstructed force vector (a) with a proper choice of $q$ ($q = 0.5$) and (b) a poor choice of $q$ ($q = 2$) – (—) Reference, (−−) Median of the samples and (□) 95% credible interval](image)

Obtained results clearly show that the quality of the inference strongly depends on a proper choice of the shape parameter $q$. Consequently, when setting $q = 2$, as classically done in the literature to perform the inference, one takes the risk to draw erroneous conclusions if the structure is actually excited by localized sources, even if the precisions are estimated from Bayesian inference.
Mathieu Aucejo and Olivier De Smet

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Median</th>
<th>Mode</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>0.993</td>
<td>0.993</td>
<td>[0.933, 1.056]</td>
</tr>
<tr>
<td>$F_2$</td>
<td>0.977</td>
<td>0.977</td>
<td>[0.908, 1.037]</td>
</tr>
<tr>
<td>$\tau_n$</td>
<td>$4.77 \times 10^{15}$</td>
<td>$4.51 \times 10^{15}$</td>
<td>[2.81, 7.57] $\times 10^{15}$</td>
</tr>
<tr>
<td>$\tau_s$</td>
<td>4.98</td>
<td>4.96</td>
<td>[3.48, 6.94]</td>
</tr>
</tbody>
</table>

Table 3: Extended Bayesian formulation – Summary of the inference result on the model parameters for $q = 0.5$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Median</th>
<th>Mode</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>0.564</td>
<td>0.547</td>
<td>[0.251, 0.889]</td>
</tr>
<tr>
<td>$F_2$</td>
<td>0.497</td>
<td>0.487</td>
<td>[0.184, 0.819]</td>
</tr>
<tr>
<td>$\tau_n$</td>
<td>$4.29 \times 10^{15}$</td>
<td>$4.04 \times 10^{15}$</td>
<td>[2.17, 7.80] $\times 10^{15}$</td>
</tr>
<tr>
<td>$\tau_s$</td>
<td>10.37</td>
<td>9.81</td>
<td>[5.38, 17.88]</td>
</tr>
</tbody>
</table>

Table 4: Extended Bayesian formulation – Summary of the inference result on the model parameters for $q = 2$

2.5 Complete Bayesian formulation

The applications of standard and extended Bayesian formulations have pointed out the need for properly defining the value of the shape parameter $q$. However, choosing a priori relevant value is far from an easy task for non-experienced user. That is why, it is interesting to infer the shape parameter from a Bayesian analysis. Practically, this is done by considering this parameter as a random variable. In doing so, one obtains the complete Bayesian formulation:

$$p(F, \tau_n, \tau_s, q|X) \propto p(X|F, \tau_n) p(F|r, q) p(\tau_n|\alpha_n, \beta_n) p(\tau_s|\alpha_s, \beta_s) p(q),$$

(10)

where $p(q)$ is the prior probability distribution of the shape parameter $q$.

2.5.1 Choice of the prior probability distribution of the shape parameter

The only available information is that the value of the shape parameter is bounded and positive. In absence of more precise knowledge on this parameter, the probability distribution is not only chosen to reflect the available information but also for its mathematical tractability. A probability distribution that meets these requirements is the truncated Gamma distribution defined by:

$$\mathcal{G}_T(q|\alpha_q, \beta_q, l_b, u_b) = \frac{\Gamma(\alpha)}{\Gamma(\alpha_q, \beta_q l_b) - \Gamma(\alpha_q, \beta_q u_b)} \mathcal{G}(q|\alpha_q, \beta_q) I_{[l_b, u_b]}(q),$$

(11)

where:

- $\mathcal{G}(q|\alpha_q, \beta_q)$ is the Gamma distribution defined in Eq. (8);
- $I_{[l_b, u_b]}(q)$ is the truncation function defined between the lower bound $l_b$ and the upper bound $u_b$. More precisely, this function simply writes:

$$I_{[l_b, u_b]}(q) = \begin{cases} 1 & \text{if } q \in [l_b, u_b] \\ 0 & \text{otherwise} \end{cases},$$

(12)
• $\gamma(s,x) = \int_0^x t^{s-1} \exp(-t) \, dt$ is the lower incomplete Gamma function.

Here, the choice of the truncated Gamma distribution has been made for mathematical convenience, because other continuous truncated distribution could have theoretically been used. To avoid biasing the inference, the shape of the prior distribution needs to be weakly informative. Here, this means that the hyperparameters should be defined such that $\alpha_q = 1$ and $\beta_q \rightarrow 0$. On the other hand, even if the lower and upper bounds $\ell_b$ and $u_b$ can theoretically take any positive value, one knows that the value of $q$ practically lies in the interval $[0, 2]$. For this particular reason, we set $\ell_b = 0.05$ and $u_b = 2.05$.

### 2.5.2 Summary and application

From the above considerations, the complete Bayesian formulation is given by:

$$p(F, \tau_n, \tau_s, q \mid X) \propto p(X \mid F, \tau_n) p(F \mid \tau_s, q) p(\tau_n \mid \alpha_n, \beta_n) p(\tau_s \mid \alpha_s, \beta_s) p(q \mid \alpha_q, \beta_q, \ell_b, u_b).$$

(13)

To explore the posterior probability distribution, $10^5$ samples have been drawn from the MCMC algorithm presented in section 3. Fig. 4 presents the excitation field obtained after performing the inference, while Table 5 summarizes the inference results on the model parameters.

![Figure 4](image-url)

**Figure 4: Complete Bayesian formulation – Real part of the reconstructed force vector – (—) Reference, (−−) Median of the samples and (□) 95% credible interval**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Median</th>
<th>Mode</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>0.993</td>
<td>0.992</td>
<td>[0.932, 1.06]</td>
</tr>
<tr>
<td>$F_2$</td>
<td>0.976</td>
<td>0.975</td>
<td>[0.906, 1.037]</td>
</tr>
<tr>
<td>$\tau_n$</td>
<td>$4.76 \times 10^{15}$</td>
<td>$4.60 \times 10^{15}$</td>
<td>[2.79, 7.61] $\times 10^{15}$</td>
</tr>
<tr>
<td>$\tau_s$</td>
<td>4.71</td>
<td>4.44</td>
<td>[2.98, 7.77]</td>
</tr>
<tr>
<td>$q$</td>
<td>0.48</td>
<td>0.47</td>
<td>[0.36, 0.67]</td>
</tr>
</tbody>
</table>

Table 5: Complete Bayesian formulation – Summary of the inference result on the model parameters

As expected, obtained results clearly show that the inference performed from the complete Bayesian formulation is able to provide parameters estimates as well as a quantification of
the posterior uncertainty on those parameters. In particular, it can be stressed that the median of the reconstructed excitation field agrees very well with the reference one. Furthermore, it could be added that the size of the associated 95% credible interval is small, meaning that the reconstructed excitation field is highly probable given the mechanical model and the measured vibration field.

3 BAYESIAN INFERENCE - MCMC ALGORITHM

In the previous section, the main features of each Bayesian formulation have been studied without describing the algorithm behind the results presented. Actually, a Gibbs sampler have been implemented to explore the posterior probability distribution of each Bayesian formulations. To render the presentation of the proposed sampler more concise, the proposed Gibbs sampler is derived for the complete Bayesian formulation, since the samplers associated to the standard and the extended Bayesian formulations are, in fact, only particular cases of this more general sampler.

The implementation of a Gibbs sampler requires the knowledge of the full conditional probability distributions. For the complete Bayesian formulation, one has for:

- the shape parameter $q$:

$$
p(q | X, F, \tau_n, \tau_s) \propto \frac{\tau_n^q}{\Gamma(1/q)^M} q^{q/M} F(\tau_s q) q^{q/M-1} \exp \left[ -\beta q \| F \|_q^2 - \tau_s \| F \|_q \right] ;
$$

- the precision $\tau_s$:

$$
p(\tau_s | X, F, \tau_n, q) \propto \mathcal{G} \left( \tau_s \mid \alpha_s + \frac{M}{q}, \beta_s + \| F \|_q^2 \right) ;
$$

- the precision $\tau_n$:

$$
p(\tau_n | X, F, \tau_s, q) \propto \mathcal{G} \left( \tau_n \mid \alpha_n + N, \beta_n + \| X - HF \|_2^2 \right) ;
$$

- the force vector $F$:

$$
p(F | X, \tau_n, \tau_s, q) \propto \exp \left[ -\tau_n \| X - HF \|_2^2 - \tau_s \| F \|_q^2 \right] ;
$$

which corresponds to the standard Bayesian formulation.

From the previous full conditional probability distributions, we can derive the following Gibbs sampler:

1. Set $k = 0$ and initialize $q^{(0)}, \tau_s^{(0)}, \tau_n^{(0)}$ and $F^{(0)}$;

2. Draw $N_s$ samples from full conditional distributions

   for $k = 1 : N_s$

   a. draw $q^{(k)} \sim p(q | X, F^{(k-1)}, \tau_n^{(k-1)}, \tau_s^{(k-1)})$
b. draw \( \tau_s^{(k)} \sim p \left( \tau_s | X, F^{(k-1)}, \tau_n^{(k-1)}, q^{(k)} \right) \)

c. draw \( \tau_n^{(k)} \sim p \left( \tau_n | X, F^{(k-1)}, \tau_s^{(k)}, q^{(k)} \right) \)

d. draw \( F^{(k)} \sim p \left( F | X, \tau_n^{(k)}, \tau_s^{(k)}, q^{(k)} \right) \)

end for

3. Monitor the convergence of the Markov chains

In the next of this section, each step of the Gibbs sampler is detailed.

### 3.1 Initialization of the sampler

In the present contribution, the sampler is initialized from a starting point having a reasonably high probability. Here, the initial force vector \( F^{(0)} \) is obtained from the MAP estimate of the standard Bayesian formulation, which is defined by [see Ref. [3] for more details]:

\[
F^{(0)} = \arg\min_F \| X - HF \|^2_2 + \lambda^{(0)} \| F \|_{q^{(0)}}^2, \tag{18}
\]

where \( \lambda^{(0)} = \tau_s^{(0)}/\tau_n^{(0)} \) is the regularization parameter.

To obtain a relevant initial force vector, it is necessary to determine reasonable values of the shape parameters \( q^{(0)} \) and the precisions \( \tau_s^{(0)} \) and \( \tau_n^{(0)} \). Practically, the values of the shape parameter \( q^{(0)} \) can be chosen without any calculation using the indications given in section 2.2.2. However, because \( q^{(0)} \) can take any value in the range \( ]0,2[ \), the solution of the previous optimization problem has generally no closed-form expression and hence the optimal value of the regularization parameter \( \lambda^{(0)} \) can not be directly computed from automatic selection procedures such as the L-curve principle [10].

Practically, the optimization problem is solved using an Iteratively Reweighted Least Squares (IRLS) algorithm [11]. From a Bayesian standpoint, the IRLS procedure can be viewed as a transformation of the Generalized Gaussian prior into a multivariate Gaussian-like prior. More precisely, one has:

\[
p(F|X, \tau_s^{(0)}, q^{(0)}) = p(F|\tau_s^{(0)}, q^{(0)}) \\
\propto p(F|W, \tau_s^{(0)}) \\
\propto \exp \left[ -\tau_s^{(0)} \| W^{1/2} F \|_2^2 \right], \tag{19}
\]

where \( W \) is a diagonal global weighting matrix depending explicitly on \( F \) and \( q^{(0)} \). This explains why Eq. (18) can only be solved using an iterative procedure.

At each iteration \( j \) of the IRLS algorithm, one has to solve:

\[
F^{(0,j)} = \arg\min_F \| X - HF \|^2_2 + \lambda^{(0,j)} \| W^{(j)1/2} F \|_2^2, \tag{20}
\]

until the convergence is reached.
After convergence of the iterative process, the optimal force vector $F^{(0)}$, the global weighting matrix $W$, as well as the optimal value of the regularization parameter $\lambda^{(0)}$, that has been updated at each iteration using automatic selection procedure such as the L-curve principle, are obtained. Consequently, to complete the initialization step it remains to determine the values of $\tau_n^{(0)}$ and $\tau_s^{(0)}$. Here, we follow the approach proposed by Pereira et al. [12] consisting in finding the optimal values of $\tau_n^{(0)}$ and $\tau_s^{(0)}$ given the measured vibration field $X$ only using a Bayesian approach. With the proposed parametrization, one has:

$$\tau_s^{(0)} = \frac{N}{X^H \left( \lambda^{(0)} I + HW^{-1}H^H \right)^{-1} X} \quad \text{and} \quad \tau_n^{(0)} = \frac{\tau_s^{(0)}}{\lambda^{(0)}}. \quad (21)$$

### 3.2 Drawing samples for the shape parameter

The Gibbs sampler is generally the first choice for conditionally conjugate models, where samples can be drawn directly from each conditional probability distribution. Unfortunately, this is not the case for the shape parameters $q$, for which the conditional distribution is not standard. To bypass this difficulty, we propose to update the value of the shape parameter using a Hamiltonian Monte Carlo (HMC) step, because it limit the random walk behavior of the sampler.

In HMC, the Hamiltonian function can be written as follows:

$$H(q, s) = U(q) + K(s), \quad (22)$$

where $U(q)$ is called the potential energy and is defined as:

$$U(q) = -\log \left[ p \left( q | X, F^{(k-1)}, \tau_n^{(k-1)}, \tau_s^{(k-1)} \right) \right], \quad (23)$$

while $K(s)$ is called the kinetic energy and is chosen for mathematical convenience. Usually, it is defined such that:

$$K(s) = -\log \left[ N(s | 0, 1) \right] = \frac{1}{2} s^T s + \text{const.} \quad (24)$$

In this framework, $q$ and $s$ define the state space of the dynamical system and can be viewed as the position and the momentum respectively. The evolution of the state $(q, s)$ over the time $t$ is governed by the following system of equations:

$$\begin{align*}
\frac{dq}{dt} &= \frac{\partial H(q, s)}{\partial s} \\
\frac{ds}{dt} &= -\frac{\partial H(q, s)}{\partial q}
\end{align*} \quad (25)$$

Practically, Eq. (25) is solved using the leapfrog method. Starting from an initial state $(q_0, s_0)$, it allows determining a proposal state $(q^*, s^*)$, while taking into account the constraints on the parameter $q$. Once a proposal state is obtained, one has to decide to accept it as a new state with probability:

$$P = \min \left( 1, \exp \left[ H(q_0, s_0) - H(q^*, s^*) \right] \right), \quad (26)$$

or to reject it with probability $1 - P$. 

3.3 Drawing samples for the precisions

The conditional probability distribution for the precisions $\tau_n$ and $\tau_s$ are actually gamma distributions. Consequently, samples can be easily drawn from standard statistical packages.

3.4 Drawing samples for the force vector

At first sight, drawing samples from $p(F|X, \tau_n^{(k)}, \tau_s^{(k)}, q^{(k)})$ seems a difficult task. However, the IRLS algorithm allows writing when the iterative process has converged:

$$p(F|X, \tau_n^{(k)}, \tau_s^{(k)}, q^{(k)}) \propto \exp\left[\frac{-\tau_n^{(k)}\|X - HF\|^2}{2} - \tau_s^{(k)}\|W^{1/2} F\|_q^{(k)}\right]$$

where $\mathcal{N}_c(z|\mu, \Sigma)$ is the (circularly-symmetric) complex multivariate Gaussian distribution with mean $\mu$ and covariance matrix $\Sigma$. Here, the mean $\mu_F$ and the covariance matrix $\Sigma_F$ are expressed as:

$$\mu_F = \tau_n^{(k)} \Sigma_F H^HX$$
$$\Sigma_F = (\tau_n^{(k)} H^H H + \tau_s^{(k)} W)^{-1}.$$  

Once $\mu_F$ and the covariance matrix $\Sigma_F$ are obtained, samples of the force vector $F^{(k)}$ can be easily drawn. A widely used method for drawing samples from a complex multivariate Gaussian distribution consists in:

1. generating a complex random vector $v \sim \mathcal{N}_c(v|0, I)$ from two real normal random vectors $v_1 \sim \mathcal{N}(v_1|0, I)$ and $v_2 \sim \mathcal{N}(v_2|0, I)$:

$$v = \frac{v_1 + j v_2}{\sqrt{2}},$$

where $j = \sqrt{-1}$ is the imaginary unit;

2. determining any matrix $L$ satisfying the relation $LL^H = \Sigma_F$. This can be typically done using either a Cholesky decomposition or a spectral decomposition;

3. computing the new sample $F^{(k)} = \mu_F + L v$.

3.5 Convergence diagnostics

In the present contribution, all the inferences have been performed from one single chain for each monitored parameter. More precisely, we have monitored the convergence of the precisions $(\tau_n, \tau_s)$ and the shape parameter $q$. The procedure used for convergence monitoring is based on the combination of two complementary diagnostics:

1. the Raftery-Lewis diagnostic [13] that estimates the total run length and the burn-in period;

2. the Geweke diagnostic [14] that tests the null hypothesis that the Markov chain is in the equilibrium distribution and produces z-statistics for each estimated parameter.

Practically, the Raftery-Lewis is first applied to the chains to monitor. Then, the Geweke diagnostic is applied to the resulting chains.
3.6 Samplers for the standard and the extended Bayesian formulations

The MCMC algorithm for the complete Bayesian formulation can be used to derive the MCMC algorithms corresponding to the standard and the extended Bayesian formulations. Indeed, the MCMC algorithm related to the extended Bayesian formulation is obtained by not considering the step 2a of the Gibbs sampler (i.e. by fixing the values of $q$), while the sampler associated to the standard Bayesian formulation is derived from the general Gibbs sampler by computing steps 1 and 2c only (i.e. by considering fixed values for $\tau_n$, $\tau_s$ and $q$).

4 CONCLUSIONS

In structural dynamics, regularization approaches can be used to deal with source identification problems. Although widely used and deeply studied, these methods generally provide only point estimates. In other words, there is no information on the uncertainty about the regularized solution given the measured data and the mechanical model. However, such information is of primary interest for industrial applications, in which it is essential to guarantee the quality of obtained results. In this contribution, three Bayesian formulations of the force reconstruction problem have been proposed to tackle this issue and an original hybrid Gibbs sampler has been implemented to perform the inferences. One of the merits of proposed sampler is to be flexible enough to deal with each formulation and provide estimates and uncertainty quantification of all the parameters of the formulation considered. The proposed numerical example clearly highlight the benefits and the limitations of the proposed formulations.

REFERENCES


MLS BASED SEQUENTIAL SRSM IN SPARSE GRID FRAMEWORK FOR EFFICIENT UNCERTAINTY QUANTIFICATION

Amit Kumar Rathi\textsuperscript{1}, and Arunasis Chakraborty\textsuperscript{1}

\textsuperscript{1}Department of Civil Engineering, Indian Institute of Technology Guwahati
Assam 781039, India
e-mail: \{ak.rathi, arunasis\}@iitg.ernet.in

Keywords: Polynomial Chaos Expansion, Sparse Grid, Moving Least Square, Multiple optima, Uncertainty Quantification.

Abstract. High fidelity models for uncertainty quantification of large structures in finite element framework are computationally exhaustive. Thus, there is a constant demand for efficient algorithm that uses optimal computational cost without compromising with the quality of the end results. With this in view, present study aims to develop a sequentially evolving stochastic response surface using Hermite family of orthogonal polynomials whose support points are generated in sparse grid framework. Using the values of the original model at these support points, unknown coefficients of the stochastic response surface are optimized by moving least square technique. It helps to reduce the number of original function evaluations to determine the coefficients as compared to other deterministic or random sampling techniques. Besides sparse grid scheme for support point generation, they are also populated sequentially as the optimization progresses in every iteration. The uniqueness of the proposed scheme is its adaptability by changing the order of the polynomials and the level of the sparse grid to minimize the overall computational cost. Multiple optima present in the original response can be identified by introducing additional penalty functions whenever they are required. Once the global response surface is ready, Monte-Carlo simulation or its advance version (e.g. Latin Hypercube Sampling) is adopted to generate the probability density functions for the output variables. Numerical studies are presented to prove the efficiency and accuracy of the proposed scheme as compared to other techniques available in the literature.
1 INTRODUCTION

Uncertainty is inevitable in physical systems where it naturally propagates affecting the performance. It can be broadly classified based on qualitative and quantitative sources [1]. These include randomness associated with analysis and formulation (i.e. epistemic uncertainty) and/or system parameters (i.e. aleatory uncertainty) [2]. For designers, quantifying and incorporating the amount of uncertainty based on the information w.r.t. parameters like input variables are vitally important. Hence, a demand for high fidelity approximate models to evaluate the uncertainty has been an active area of research. Especially, uncertainty in large structures modelled by finite element method requires better techniques as it is a computationally exhaustive exercise. Convergence of such approximate models is not always ensured which may lead to considerable error [1, 3]. To eradicate this error polynomial chaos expansion (PCE) and its variants [4, 5, 6] have been proposed using orthogonal polynomials. Thus, making the approximate model convergent in $L^2$ sense [4].

Although these methods yield accurate results with increase in order of the polynomial which often lead to large number of actual function calls. This problem is more critical with increase in number of random variables which is widely named as curse of dimensionality. Present study aims to develop a sequentially evolving stochastic response surface using orthogonal polynomials involving sparse support points.

2 PROBLEM STATEMENT

Apart from population based methods like Monte Carlo simulation (MCS), Latin hypercube sampling (LHS) etc. for uncertainty quantification, stochastic response surface method (SRSM) is widely used. It was proposed by Isukapalli [5] using PCE as a dimension reduction technique to simplify the complex relation of input-output variables which is mathematically expressed as

$$
g(\xi) = \alpha_0 + \sum_{i_1=1}^{n} \alpha_{i_1} \Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{n} \sum_{i_2=1}^{i_1} \alpha_{i_1i_2} \Gamma_2(\xi_{i_1}, \xi_{i_2}) + \ldots \\
+ \sum_{i_1=1}^{n} \sum_{i_2=1}^{i_1} \ldots \sum_{i_o=1}^{i_{o-1}} \alpha_{i_1i_2\ldots i_o} \Gamma_o(\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_o}) + \ldots \tag{1}
$$

where, $\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_o}$ are standard normal random variables. The function $\Gamma_o$ is the orthogonal polynomial bases which can be developed using Hermite scheme [5]

$$
\Gamma_o(\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_o}) = e^{\frac{1}{2} \xi^T \xi} (-1)^o \frac{\partial^o e^{-\frac{1}{2} \xi^T \xi}}{\partial \xi_{i_1} \partial \xi_{i_2} \ldots \partial \xi_{i_o}}. \tag{2}
$$

These functions are dictated by the order $o$ to limit the number of unknown coefficients $b = \{\alpha_0 \alpha_1 \ldots \alpha_{mn\ldots n}\}^T$. Thus, the Eq.(1) can be simplified to

$$
\tilde{g}(\xi) = \sum_{i=0}^{o} \Xi_i(\xi)b_i = \Xi(\xi)b \tag{3}
$$

Gauss quadrature points (a.k.a. collocation points) [5] are generated using roots of the Hermite polynomial bases for evaluating the original $g(\cdot)$ at certain locations (further referred as support points). These points help in evaluating the unknowns in the above equation as

$$
b = (\Xi^T \Xi)^{-1} \Xi^T g \tag{4}
$$
In the above equation, $\mathbf{g}$ is the vector of original $g(\cdot)$ values. To solve it, the equation must not be underdetermined system which means number of support points $n_e$ must be equal to or greater than number of coefficients $\mathbf{b}$ (i.e. $n_b$). For collocation scheme, the number of points rises exponentially (i.e. $n_e = (o+1)^n$) which leads to the curse of dimensionality and sabotages the benefits of this surrogate model. The present study focuses on addressing this issue with an efficient method.

3 PROPOSED METHOD: SEQUENTIAL SRSM

The proposed method uses advanced regression based PCE coupled with sparse grid support points in an iterative framework for sequential development of stochastic response surface to determine uncertainty.

3.1 Moving Least Square Technique

To incorporate both global and local approximation for a better response surface of nonlinear function, the unknown coefficients are determine by minimizing weighted mean square error. This weight function $w$ varies with the Euclidean distance of the concerned point and support point and thus, it is widely called as moving least square (MLS) technique. The weighted squared error can be given as [7]

$$
\delta(\xi) = \sum_{i=1}^{p} w(\xi - \xi_i) \{\tilde{g}(\xi_i) - g(\xi_i)\}^2.
$$

Minimizing the above equation with respect to unknown coefficients gives

$$
\mathbf{b} = (\Xi^T \mathbf{W} \Xi)^{-1} (\Xi^T \mathbf{W} \mathbf{g}).
$$

where, weight matrix $\mathbf{W} = \text{diag}[w(\xi - \xi_1) w(\xi - \xi_2) \ldots w(\xi - \xi_{n_b})]$. The weight function $w$ is defined as regularized weight which is given by [7]

$$
w(\xi - \xi_i) = \begin{cases} 
\frac{\left[1 + \left(\frac{\lambda(\xi - \xi_i)}{\hat{d}}\right)^2\right]^{-\frac{1}{2}}} {1 - (1 + \lambda^2)^{-\frac{1}{2}}} & \text{if } \|\xi - \xi_i\| \leq \hat{d} \\
0 & \text{if } \|\xi - \xi_i\| > \hat{d}
\end{cases}
$$

In the above equation, $\hat{d}$ is the influence radius and $\lambda$ is adopted as $10^{-5}$ for better accuracy, as proposed by Most and Bucher [7].

3.2 Sparse Grid

Computational cost is usually associated with number of function evaluations required by approximate methods. In order to reduce this cost, sparse grid scheme is used to generate small product grids. These smaller grids are derived from the full grid as used in collocation scheme. Smolyak’s algorithm is used for generating such points by forming tensor product of smaller grids as [8]

$$
\mathcal{S}_q = \sum_{|\mathbf{i}| \leq q + n - 1} (\Delta_{i_1} \otimes \Delta_{i_2} \otimes \cdots \otimes \Delta_{i_n}) \mathbf{g}
$$
where, $\Delta_i = V_{m_i} - V_{m_{i-1}}$, for $i \geq 1$ and $i \in \mathbb{N}^n$. One of the widely accepted scheme is Clenshaw-Curtis [9] which generates equidistant nodes. The number of points in each direction are determined by [10]

$$m_i = \begin{cases} \frac{1}{2^{i-1}} + 1 & \text{for } i = 1 \\ \frac{i}{2} & \text{for } i > 1 \end{cases}$$

(9)

and the respective coordinates of these points are evaluated using

$$x^j_i = \begin{cases} \frac{i-1}{m_i-1} & \text{for } j = 1, 2, \ldots, m_i \text{ and } m_i > 1 \\ \frac{1}{2} & \text{for } j = 1 \text{ and } m_i = 1 \end{cases}$$

(10)

Fig. 1 shows the support points generated using limited tensor products [10] where $x^j_i \in [0, 1]$.

Figure 1: Sparse grid generated using Clenshaw-Curtis scheme

It can be noticed that the number of points and its coordinates in each direction is dictated by $m_i$ which is further influenced by level of sparse grid $q$. This leads to hierarchically generation of sparse grid with less points than the full grid collocation scheme. In case of large number of variables $n$ the difference between the schemes significantly increases. However, number of points exponentially grows with the level $q$ which might generate more points than the collocation scheme, thus a judicial use is required.

3.3 Proposed Sequential Strategy

For developing the sequential strategy, maxima and minima are determined using the constructed response surface as

$$\max_{\xi} \tilde{g}(\xi)$$

s. t. $\xi \in \Omega_{\xi}$

(11)
\[
\min_{\xi} \quad \tilde{g}(\xi) \\
\text{s. t.} \quad \xi \in \Omega
\]

respectively. Optimization process often encounters more than one maxima and/or minima over the iterations. To identify these points or regions weight function is included for penalty. It modifies the stochastic response surface by a continuous and differentiable function as

\[
\tilde{g}_m(\xi) = \tilde{g}(\xi) + \sum_{k=1}^{n_m} \tilde{g}_k(\xi).
\]

In the above equation, \(n_m\) is the total number of multiple optimas and \(\tilde{g}_k(\xi)\) represents penalty function which is expressed as [11]

\[
\tilde{g}_k(\xi) = \begin{cases} \\
\phi |\xi^{*,i} - \xi|_2 \left( \gamma \left| \xi^{*,i} - \xi \right|_2^2 - \left| \nabla \tilde{g}(\xi^{*,i}) \right|_2 \right) & \text{if } |\xi - \xi^{*,i}|_2 \leq r_k \\
0 & \text{if } |\xi - \xi^{*,i}|_2 > r_k 
\end{cases}
\]

As suggested by Der Kiureghian and Dakessian [11] the constant parameters \(\phi\) and \(\gamma\) are considered to be 0.75 and 1.10, respectively, and influence radius of penalty function \(r_k = \gamma |\xi^{*,i}|_2\). The convergence criteria for the optimization with penalty function is given by

\[
\cos^{-1} \left[ \frac{\xi^{*,k+1} + \xi^{*,k}}{2 |\xi^{*,k}|_2} \right] \leq \theta_m
\]

\[
|\xi^{*,k+1} - \xi^{*,k}|_2 \leq |\xi^{*,k}|_2.
\]

which is required to identify new optimal point. The convergence angle \(\theta_m\) between foot of the penalty weight function and optimal point is assumed to be 66° for \(\gamma = 1.10\). For more information one can refer to [11].

### 4 NUMERICAL ANALYSIS

For brevity, in this section two applications of the proposed method is discussed. One is of a benchmark problem with non-algebraic terms resulting in multiple maxima-minima points whereas the second problem demonstrates the proposed method for a real structure with correlated non-normal random variables.

#### 4.1 Example 1: Franke’s Function

The first problem is a non-algebraic function which is mathematically expressed as

\[
g(x) = \frac{3}{4} \left\{ -\frac{1}{4}(9x_1-2)^2 - \frac{1}{4}(9x_2-2)^2 \right\} + \frac{3}{4} \left\{ -\frac{1}{4}(9x_1-4)^2 - \frac{1}{4}(9x_2-7)^2 \right\} + \frac{1}{4} \left\{ -\frac{1}{4}(9x_1-7)^2 - \frac{1}{4}(9x_2-3)^2 \right\} - \frac{1}{4} \left\{ -(9x_1-4)^2 - (9x_2-7)^2 \right\}
\]

and is widely used for testing surrogate models. The parameters \(x_1\) and \(x_2\) are independent normal random variables with mean 0.400 and standard deviation 0.100. Complexity in the problem is due to multiple optima points. The proposed method is applied which first determines the global maxima as in Fig. 2 (c) which is further penalized for finding any other maxima. The next maxima is evaluated as shown in Fig. 2 (h). Following the similar procedure and on searching the next maxima converges to the first maxima as shown in Fig. 2 (i). Since all the possible maxiums are located, now the proposed method optimizes the response surface for
minima. Similar strategy is followed which results in the sequential stochastic response surface as presented in Fig. 2 (l). Once the response surface is constructed, MCS are conducted over it. The results of uncertainty quantification are presented in Fig. 3 and Table 1. The proposed method yields results close to direct MCS on the Franke’s function with less number of function evaluations.
Amit Kumar Rathi and Arunasis Chakraborty

Figure 3: Comparison of (a) pdf and (b) CDF of Franke’s function using Sequential SRSM

<table>
<thead>
<tr>
<th>Method</th>
<th>Function Calls</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCS</td>
<td>100000</td>
<td>0.7796</td>
<td>0.2828</td>
<td>1.5007</td>
</tr>
<tr>
<td>HOSRSM</td>
<td>108</td>
<td>0.7783</td>
<td>0.2848</td>
<td>2.6217</td>
</tr>
<tr>
<td>Seq. SRSM</td>
<td>66</td>
<td>0.7881</td>
<td>0.2741</td>
<td>1.4977</td>
</tr>
</tbody>
</table>

Table 1: Results for Franke’s function using Sequential SRSM

evaluations than higher order SRSM (HOSRSM) \[12\]. This demonstrates the efficiency of the proposed method.


<table>
<thead>
<tr>
<th>Element number</th>
<th>Area</th>
<th>Moment of inertia</th>
<th>Young’s modulus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$A_5$</td>
<td>$I_5$</td>
<td>$E_1$</td>
</tr>
<tr>
<td>2</td>
<td>$A_6$</td>
<td>$I_6$</td>
<td>$E_1$</td>
</tr>
<tr>
<td>3</td>
<td>$A_7$</td>
<td>$I_7$</td>
<td>$E_1$</td>
</tr>
<tr>
<td>4</td>
<td>$A_8$</td>
<td>$I_8$</td>
<td>$E_1$</td>
</tr>
<tr>
<td>5</td>
<td>$A_1$</td>
<td>$I_1$</td>
<td>$E_2$</td>
</tr>
<tr>
<td>6</td>
<td>$A_2$</td>
<td>$I_2$</td>
<td>$E_2$</td>
</tr>
<tr>
<td>7</td>
<td>$A_3$</td>
<td>$I_3$</td>
<td>$E_2$</td>
</tr>
<tr>
<td>8</td>
<td>$A_4$</td>
<td>$I_4$</td>
<td>$E_2$</td>
</tr>
</tbody>
</table>

Table 2: Properties of the portal frame elements

<table>
<thead>
<tr>
<th>Random variables</th>
<th>Units</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_1$</td>
<td>kN/m²</td>
<td>$2.174 \times 10^7$</td>
<td>$1.915 \times 10^6$</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$E_2$</td>
<td>kN/m²</td>
<td>$2.380 \times 10^7$</td>
<td>$1.915 \times 10^6$</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$P_1$</td>
<td>kN</td>
<td>71.175</td>
<td>28.470</td>
<td>Gumbel</td>
</tr>
<tr>
<td>$P_2$</td>
<td>kN</td>
<td>88.970</td>
<td>35.590</td>
<td>Gumbel</td>
</tr>
<tr>
<td>$P_3$</td>
<td>kN</td>
<td>133.454</td>
<td>40.040</td>
<td>Gumbel</td>
</tr>
<tr>
<td>$A_1$</td>
<td>m²</td>
<td>0.313</td>
<td>0.056</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$A_2$</td>
<td>m²</td>
<td>0.372</td>
<td>0.074</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$A_3$</td>
<td>m²</td>
<td>0.506</td>
<td>0.093</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$A_4$</td>
<td>m²</td>
<td>0.558</td>
<td>0.112</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$A_5$</td>
<td>m²</td>
<td>0.253</td>
<td>0.093</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$A_6$</td>
<td>m²</td>
<td>0.291</td>
<td>0.102</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$A_7$</td>
<td>m²</td>
<td>0.373</td>
<td>0.112</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$A_8$</td>
<td>m²</td>
<td>0.419</td>
<td>0.139</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$I_1$</td>
<td>m⁴</td>
<td>$8.134 \times 10^{-3}$</td>
<td>$1.038 \times 10^{-3}$</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$I_2$</td>
<td>m⁴</td>
<td>$1.151 \times 10^{-2}$</td>
<td>$1.298 \times 10^{-3}$</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$I_3$</td>
<td>m⁴</td>
<td>$2.137 \times 10^{-2}$</td>
<td>$2.596 \times 10^{-3}$</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$I_4$</td>
<td>m⁴</td>
<td>$2.596 \times 10^{-2}$</td>
<td>$3.029 \times 10^{-3}$</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$I_5$</td>
<td>m⁴</td>
<td>$1.082 \times 10^{-2}$</td>
<td>$2.596 \times 10^{-3}$</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$I_6$</td>
<td>m⁴</td>
<td>$1.410 \times 10^{-2}$</td>
<td>$3.461 \times 10^{-3}$</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$I_7$</td>
<td>m⁴</td>
<td>$2.328 \times 10^{-2}$</td>
<td>$5.625 \times 10^{-3}$</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$I_8$</td>
<td>m⁴</td>
<td>$2.596 \times 10^{-2}$</td>
<td>$6.490 \times 10^{-3}$</td>
<td>Lognormal</td>
</tr>
</tbody>
</table>

Table 3: Properties and distribution of random variables in Example 2

4.2 Example 2: Portal Frame

In this section, uncertainty caused in horizontal displacement $u(x)$ of the top floor of portal frame due to random variables is evaluated. The details of the multi-storey multi-bay portal frame is shown in Fig. [4] and Table [2]. Properties of the random variables is presented in Table
These random variables are correlated as

\[
\rho = \begin{bmatrix}
1 & 0.90 & 0 & 0 & 0.95 & 0.95 & 0.13 & 0.13 & 1 \\
0.90 & 1 & 0 & 0 & 1 & 1 & \\
0 & 0 & 1 & 0 & 0.95 & 0.95 & \\
0 & 0 & 0 & 1 & 1 & \\
0 & 0 & 0 & 0 & 1 & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
0 & 0 & 0 & 0 & 0.13 & 1 \\
\end{bmatrix}
\]

(18)

Figure 5: Comparison of (a) pdf and (b) CDF of multi-storey portal frame using Sequential SRSM

<table>
<thead>
<tr>
<th>Method</th>
<th>Function Calls</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCS</td>
<td>100000</td>
<td>0.0177</td>
<td>0.0073</td>
<td>0.0736</td>
</tr>
<tr>
<td>Seq. SRSM</td>
<td>1269</td>
<td>0.0176</td>
<td>0.0075</td>
<td>0.0748</td>
</tr>
</tbody>
</table>

Table 4: Results for multi-storey portal frame using Sequential SRSM

To perform the static analysis, finite element method is performed using software package like ANSYS 13.0. Similar to previous example, MCS is used as benchmark with sample size $10^5$ and sequential SRSM is solved to determine uncertainty. The results of uncertainty quantification are presented in Fig. 5 and Table 4. The proposed method yields results close to direct MCS on the portal frame with very less number of function evaluations. This proves the efficiency of the proposed method for determining uncertainty quantification in real structures with non-normal correlated random variables.
5 CONCLUSIONS

Present study demonstrates an efficient uncertainty quantification method using sequential SRSM. It employs MLS based PCE with Hermite polynomial basis which is formed sequentially with support points generated by Clenshaw-Curtis sparse grid scheme. It produces equidistant points in each progressive iterations based on the optimization of maxima and minima. The method also proposes flexibility in the choice of sparse grid level and polynomial order in each iteration to minimize computational burden with adequate accuracy. Multiple optimal points are determined in this method using penalty function. Numerical study considering benchmark and real problems elucidates the accuracy and efficiency of the proposed sequential SRSM.

REFERENCES


MULTILEVEL PRECONDITIONING OF POLYNOMIAL CHAOS METHOD FOR QUANTIFYING UNCERTAINTIES IN A BLOOD PUMP

Chen Song$^{1,2}$, Vincent Heuveline$^{1,2}$

$^1$Heidelberg Institute for Theoretical Studies
Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany
e-mail: {chen.song, vincent.heuveline}@h-its.org

$^2$Engineering Mathematics and Computing Lab (EMCL)
Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University
Im Neuenheimer Feld 205, 69120 Heidelberg, Germany
e-mail: {chen.song, vincent.heuveline}@iwr.uni-heidelberg.de

Keywords: Inrusive method, Polynomial Chaos Expansion, blood pump, Computational Fluid Dynamics, Multilevel preconditioner, High Performance Computing, Variational Multiscale method, Stochastic Finite Element Method.

Abstract. Heart failure (HF) is a severe cardiovascular disease, it happens when the heart muscle is so weakened such that it can not provide sufficient blood as body needs. More than 23 million people are suffered by HF worldwide. Despite the modern transplant operation is well established, the lack of heart donations becomes a big restriction on transplantation frequency. With respect to this matter, ventricular assist devices (VADs) can play an important role in supporting patients during waiting period and after the surgery.

Moreover, it has been shown that VADs by means of blood pump have advantages for working under different conditions. While a lot of work has been done on modeling the functionality of the blood pump, but quantifying uncertainties in a numerical model is a challenging task. We consider the Polynomial Chaos (PC) method, which is introduced by Wiener for modeling stochastic process with Gaussian distribution. The Galerkin projection, the intrusive version of the generalized Polynomial Chaos (gPC), has been densely studied and applied for various problems. The intrusive Galerkin approach could represent stochastic process directly at once with Polynomial Chaos series expansions, it would therefore optimize the total computing effort comparing with classical non-intrusive methods. We compared different preconditioning techniques for a steady state simulation of a blood pump configuration in our previous work, the comparison shows that an inexact multilevel preconditioner has a promising performance. In this work, we show an instationary blood flow through a FDA blood pump configuration with Galerkin Projection method, which is implemented in our open source Finite Element library Hiflow3. Three uncertainty sources are considered: inflow boundary condition, rotor angular speed and dynamic viscosity, the numerical results are demonstrated with more than 30 Million degrees of freedom by using supercomputer.
1 INTRODUCTION

Since last two decades, medical instruments have been playing an important role in the field of health care, they sustain patients’ life by providing advanced clinical solutions or help surgeons to obtain better medical diagnosis. The innovation of medical devices expanded significantly recent years owing to the improvement of mechanical design, numerical modeling, computing hardware, etc. Especially, the blood pump device became one of the most popular assist apparatuses in heart failure treatment \[4, 8, 23\]. Ventricular assist devices (VADs) can be applied as long-term implant for patients who are not eligible for heart transplantation, or served as a generation of sufficient blood flow for patients during the surgery. However, the verification and validation (V&V) procedure is still demanding due to the lack of knowledge about input parametric data, specially for numerical modeling. Whereas, these uncertainties should not be ignored, because quantifying the impact of these uncertainties can be one of the key points for patient-specific implantation.

In our previous work \[26\], we investigate the steady state of Navier-Stokes equations for laminar flow in a blood pump geometry. We modeled the rotating impact with the Multiple Reference Frame (MRF) method \[7, 20, 10\], for quantifying uncertainties, we employed the intrusive Polynomial Chaos Expansion (PCE) method in order to take geometrical uncertainty and specific model parameters into account. We placed also our attention on comparing different solving strategies for the intrusive Polynomial Chaos approach for fluid problem, as being capable of solving efficiently this highly coupled system is very crucial in practice.

In this paper, we will analyze an unsteady blood flow in a blood pump device based on the Variational Multi-Scale method (VMS) for the incompressible Navier-Stokes equations. We concentrate again three different uncertainty sources: inflow boundary condition, dynamic viscosity and angular speed of the rotor. The geometry of the blood pump is referred to the U.S. Food and Drug Administration (FDA) "critical path" initiative benchmark problem \[13\], the aim of this project is to utilize advanced computational simulations to predict the biological response. So as to cope with the rotating machinery modelization, we employ a shear layer update approach, which is very similar to The Shear-Slip Mesh Update Method (SSMUM) \[2, 3\], the main difference is to deploy two layers instead of one in order to facilitate the update process once a regeneration of the mesh is needed. By this setting, the local linear algebra structure remains the same, hence, only few values demand update. We employ the Variational Multi-Scale method \[11, 22\], which inherits the consideration of two separated scales from the large eddy simulation (LES) model and the concept of stabilized finite element method \[28\], it provides the feasibility of modeling the blood flow within the high rotation speed instrument.

We here apply the stochastic finite element method (SFEM) \[14\], this conveys that the spatial domain is discretized by the finite element method (FEM), and the stochastic space is expressed by Polynomial Chaos Expansion (PCE) \[29\]. One important advantage of utilizing PCE is when the model output is smooth regarding to the input, a spectral convergence of PCE is achieved \[9\]. PCE exhibits the stochastic solution via the orthogonal multivariate polynomials, and the input random variables are represented in predefined polynomials regarding to their probability distribution. The stochastic Galerkin projection \[14, 21\] is a powerful tool, which provides the possibility to obtain the coefficients of the PCE system at once. However, efficient solving strategies are needed for such complex structure. From our previous contribution \[26\], we compared three different preconditioning techniques for a stationary flow in a simplified blood pump geometry, the inexact Multilevel method outperformed the Mean based preconditioner and the exact Multilevel method. We focus therefore in this paper on the inexact Multilevel
method, and apply it to the instationary flow in the pump geometry.

The rest of this paper is organized as follow. In Section 2, we introduce the mathematical modeling for high Reynolds number flow in a rotating machinery, illustrate our moving mesh technique and the modelization of stochastic system. Section 3 presents our Multilevel preconditioner algorithm. The numerical results are showed in Section 4. We conclude our contribution and provide an outlook for further development in Section 5.

2 MATHEMATICAL MODELING

We consider the unsteady state of the incompressible Navier-Stokes equations in a rotating machinery, it yields:

\[
\begin{align*}
\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} - \mathbf{u}^r) \cdot \nabla \mathbf{u} - \frac{\mu}{\rho} \Delta \mathbf{u} + \frac{1}{\rho} \nabla p &= 0, \quad \text{in } \Omega, \quad (1a) \\
\nabla \cdot \mathbf{u} &= 0, \quad \text{in } \Omega, \quad (1b) \\
\mathbf{u}^r &= d \times \omega, \quad \text{in } \Omega_{\text{rot}}, \quad (1c) \\
\mathbf{u} &= 0, \quad \text{in } \Omega_{\text{stat}}, \quad (1d) \\
\mathbf{u} &= g, \quad \text{on } \Gamma_{\text{in}}, \quad (1e) \\
\left(\frac{\mu}{\rho} \nabla \mathbf{u} - p \mathbf{I}\right) &= 0, \quad \text{on } \Gamma_{\text{out}}, \quad (1f) \\
\mathbf{u}^r &= d \times \omega, \quad \text{on } \Gamma_{\text{rot}}, \quad (1g) \\
\mathbf{u} &= 0, \quad \text{on } \partial \Omega \setminus (\Gamma_{\text{in}} \cup \Gamma_{\text{out}} \cup \Gamma_{\text{rot}}). \quad (1h)
\end{align*}
\]

Where, \( \mathbf{u} \) is the velocity, and \( p \) is the pressure. The flow is described on \( \Omega \subset \mathbb{R}^3, \Omega = \Omega_{\text{rot}} \cup \Omega_{\text{stat}}, \Omega_{\text{stat}} \cap \Omega_{\text{rot}} = \emptyset, \) \( \Omega_{\text{stat}} \) and \( \Omega_{\text{rot}} \) denote static and rotating domain respectively [3]. \( \mathbf{u}^r \) is the rotation speed, which pre-describes the motion of a moving mesh, it is defined by the angular speed \( \omega \) and the distance to the rotating axis \( d \). \( \mu \) is the dynamic viscosity and \( \rho \) is the density. We ignore here the external body force \( \mathbf{f} \) in N-S equations.

We prescribe a “do-nothing” condition on the outflow boundary \( \Gamma_{\text{out}} \), two different Dirichlet boundary conditions are applied on the inflow boundary \( \Gamma_{\text{in}} \) and the rotor’s surface \( \Gamma_{\text{rot}} \) separately. The rest of boundary is covered by “no-slip” condition (Figure 1). The inflow boundary condition is modeled with a Poiseuille profile:

\[
g = \begin{bmatrix}
0 \\
0 \\
-U_{\max}(1 - l^2/L^2)
\end{bmatrix},
\]

\( L \) is the radius of the circular inflow boundary, \( l \) is the distance from a point on \( \Gamma_{\text{in}} \) to the center point. \( U_{\max} \) is the maximum inflow speed, and \( U_{\max} > 0 \), as the inflow direction is in \(-e_3 = [0, 0, -1]^T\).

The angular speed is defined as:

\[
\omega = \begin{bmatrix}
0 \\
0 \\
\omega
\end{bmatrix},
\]

the direction of the axis of rotation for the rotor is showed in Figure 1.
2.1 Shear layer update approach

With the purpose of modeling a rotating machinery, we propose here a shear layer update approach, which is inherited from SSMUM [2, 3], we employ though two layers instead of one to achieve the efficiency of the update process once a regeneration of mesh is required. In general, for the mesh patching technique [27], the computational domain is divided by two, static and rotating domain ($\Omega_{\text{stat}}$ and $\Omega_{\text{rot}}$). The static domain stays unchanged, the rotating domain evolves corresponding the speed of blade.

Figure 2 demonstrates the main procedure of moving mesh for rotating machinery modeling, static and rotating domain own one layer separately and each cell on those two layers has the same shape. The border between these two domains is therefore the interface between the two identical layers as well. At the beginning, the initial position of the mesh looks like in step a, when the rotating part moves in relation to the rotor, we first shear one of these layers in order to obtain more freedom of movement, as we are limited by the mesh size most of the time. The connectivity of mesh cell remains unchanged during the shear process (step b), only spatial location needs to be modified, the adaptation is then uncomplicated. When the shearing attends the same length as the cell (step c). Afterwards, the vertices on the common interface detach, and reconnect to the subsequent points (step d).

If we could ensure that the local mesh only contains either a part of static mesh or rotating mesh (it can be accomplished easily by using group of communicator or similar concept of parallel communication), thus, we would benefit from this setting, as the structure of the local vector does not alter [1]. Therefore, only values on the ghost cells in the local vector need to be updated. For that reason, projecting solution into a new linear algebra structure is avoided, the update process after reconnecting vertices can be accelerated.

2.2 Variational Multiscale method (VMS)

The Reynolds number for a rotating machinery is defined like:

$$Re := \frac{\omega D^2 \rho}{\mu},$$  \hspace{1cm} (4)

due to the high rotation speed $\omega$ of our instrument (2500 RPM\(^1\)), the Reynolds number is approximately 21000 in our blood pump, an additional high Reynolds number flow modeling is required, we choose in this work the Variational Multiscale method (VMS) for this purpose.

\(^1\)RPM: revolutions per minute.
This technique considers a separation of complete solution scale into two different groups: resolvable scale and unresolvable scale. In general, VMS can be considered as a kind of stabilized finite element technique [28], whereas, it can also be meanwhile interpreted as large eddy simulation (LES) for turbulence modeling. Without describing the theory, we illustrate only the projection-based finite element variational multiscale method for Navier-Stokes equations. Further insight about this method can be found in [17, 18, 19, 15].

The variational multiscale formulation can only be seen in a weak form:

\[
\int_{\Omega} \left( \frac{\partial \mathbf{u}}{\partial t} + ((\mathbf{u} - \mathbf{u}^r) \cdot \nabla) \mathbf{u} \right) \cdot \mathbf{v} \, d\Omega + \int_{\Omega} \frac{\mu}{\rho} \nabla \mathbf{u} : \nabla \mathbf{v} \, d\Omega - \int_{\Omega} \frac{1}{\rho} \nabla \cdot \mathbf{v} \, d\Omega \, d\Omega \\
+ \int_{\Omega} \left[ \frac{\partial \mathbf{u}}{\partial t} + ((\mathbf{u} - \mathbf{u}^r) \cdot \nabla) \mathbf{u} - \frac{\mu}{\rho} \Delta \mathbf{u} + \frac{1}{\rho} \nabla p \right] \tau_M \left[ ((\mathbf{u} - \mathbf{u}^r) \cdot \nabla) \mathbf{v} \right] \, d\Omega \\
+ \int_{\Omega} \left( \nabla \cdot \mathbf{u} \right) \cdot \tau_C (\nabla \cdot \mathbf{v}) \, d\Omega = 0 ,
\]

\[
\int_{\Omega} \left( \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} - \frac{\mu}{\rho} \Delta \mathbf{u} + \frac{1}{\rho} \nabla p \right) \tau_M \nabla q \, d\Omega + \int_{\Omega} \tau_M \nabla q \left[ \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} - \frac{\mu}{\rho} \Delta \mathbf{u} + \frac{1}{\rho} \nabla p \right] \, d\Omega = 0 .
\]
\(v\) and \(q\) are the test functions for velocity and pressure. \(\tau_M\) and \(\tau_C\) are the elementwise constant stabilization operators, which project the unsolvable solution on the fine scale into the coarse scale based on the residual of the governing equations.

### 2.3 Uncertainty model

We consider three different parametric uncertainties: the inflow boundary condition \(g\), the angular speed \(\omega\) and the dynamic viscosity \(\mu\). We model each of those uncertain parameters with independent, uniformly distributed random variables \(\xi_i \sim U(-1, 1)\), \(i = 1, 2, 3\), it reads:

\[
g = g_0 + g_1 \xi_1 ,
\]

\[
\omega = \omega_0 + \omega_2 \xi_2 ,
\]

\[
\mu = \mu_0 + \mu_3 \xi_3 .
\]

Where, \(g_1 = \sigma_1 g_0, \omega_2 = \sigma_2 \omega_0\) and \(\mu_3 = \sigma_3 \mu_0\). \(\sigma_i\) are the decay factors respect to the mean value, thus \(0 < \sigma_i < 1\). With reference to PCE, we define a multivariate random variable \(\xi := (\xi_1, \xi_2, \xi_3)\). Accordingly, \(\xi\) allows us directly map the outcomes of an abstract probability space \((\Omega, \mathcal{A}, \mathbb{P})\) to a subset \(T\) of \(\mathbb{R}^3\). Afterwards, we can express our stochastic solution immediately with the aid of \(\xi\).

### 2.4 Stochastic Galerkin projection

The velocity and pressure in Equation (1) are expressed with the Polynomial Chaos Expansion (PCE) method \([21]\) by employing the orthogonal polynomial \(\psi\):

\[
u(x, \xi) = \sum_{i=0}^{\infty} u_i \psi_i(\xi) ,
\]

\[
p(x, \xi) = \sum_{i=0}^{\infty} p_i \psi_i(\xi) .
\]

\(\psi_i\) represents the Chaos Polynomials, more specifically in this paper, their are the normalized Legendre Polynomials. \(u\) and \(p\) follow certain general assumptions, such as square-integrable respect to \(\xi\), more details can be found in \([21]\). The orthogonality of \(\psi_i\) with respect to the probability density function of \(\xi\) can be explicit written as:

\[
\int_{[-1,1]^3} \psi_i(\xi) \psi_h(\xi) \frac{1}{2^3} d\xi = \delta_{ij} ,
\]

\(\delta_{ij}\) here is the Kronecker delta function. In order to be able of computing our stochastic solution numerically, we need to truncate Equation (7) up to certain polynomials order \(N_o\), it yields:

\[
u(x, \xi) \approx \sum_{i=0}^{P} u_i \psi_i(\xi) ,
\]

\[
p(x, \xi) \approx \sum_{i=0}^{P} p_i \psi_i(\xi) .
\]
\( P + 1 = (M + N_o)!/(M!N_o!) \) is the total number of the PC modes, \( M \) is the number of uncertain variables, in our case, \( M = 3 \).

Therefore, we could pursue the generalized Polynomial Chaos Expansion (gPCE) procedure by first inserting Equation (9) into Equation (5), then multiplying an additional polynomial \( \psi_k, k = 0, \ldots, P \) on both side of governing equations, and taking \( L^2 \) inner product on \( L^2(T) \). By considering the orthogonality of the polynomials, the regarding momentum equation and mass conservation can be thus written as:

\[
\frac{\partial u_k}{\partial t} v_k + \sum_{i=0}^{P} \sum_{j=0}^{P} ((u_i - u_i') \cdot \nabla)u_j v_k c_{ijk} + \sum_{i=0}^{P} \sum_{j=0}^{P} \frac{\mu_i}{\rho} \nabla u_i : \nabla v_j c_{ijk} + \frac{1}{\rho} p_i \nabla \cdot v_k + \tau_M((u_k - u_k') \cdot \nabla v_k) \cdot \frac{\partial u_k}{\partial t} + \sum_{i=0}^{P} \sum_{j=0}^{P} ((u_i - u_j') \cdot \nabla)u_j c_{ijk}
\]

\[
- \sum_{i=0}^{P} \sum_{j=0}^{P} \frac{\mu_i}{\rho} \Delta u_j c_{ijk} + \frac{1}{\rho} \nabla p_k + (\nabla \cdot u_k) \nabla (\nabla \cdot v_k) , \text{ in } \Omega ,
\]

\[
q_k \nabla \cdot u_k (10a)
\]

\[
\tau_M q_k \frac{\partial u_k}{\partial t} + \sum_{i=0}^{P} \sum_{j=0}^{P} ((u_i - u_i') \cdot \nabla)u_j c_{ijk}
\]

\[
- \sum_{i=0}^{P} \sum_{j=0}^{P} \frac{\mu_i}{\rho} \Delta u_j c_{ijk} + \frac{1}{\rho} \nabla p_k , \text{ in } \Omega .
\]

for \( k = 0, \ldots, P \), and \( c_{ijk} := \langle \psi_i \psi_j, \psi_k \rangle \).

So fair, we introduce the variational multiscale formulation for the incompressible Navier-Stokes equations in a rotating machinery based on our moving mesh technique, the modeling of input uncertain parameters and the generalized Polynomial Chaos Expansion according to considered problem is also presented. In the following section, we will focus on the numerical methods for solving this specially system.

### 3 NUMERICAL METHODS

The nonlinear terms in Equation (10) are linearized with Newton method, the general scheme can be written as:

\[
J_F(x_n)(x_{n+1} - x_n) = F(x_n) ,
\]

\( F \) is the system equation, \( J_F \) is the Jacobian matrix of \( F \). \( x_n \) is the solution vector at \( n \)-th Newton iteration, \( x_n = [u_n, p_n]^T \). At each Newton iteration, an update \( (x_{n+1} - x_n) \) of the solution vector will be computed, until it meets to certain convergence criterion.

The spatial part of our problem is therefore discretized by the finite element method with equal order P1/P1 element for velocity and pressure respectively, thanks to VMS, no Taylor-Hood element is necessary. Afterwards, we will obtain a linear system of equations for the variational multiscale formulation for the Navier-Stokes equations in rotating system. The stiff-
ness matrix $A(x_n) \in \mathbb{R}^{(P+1)N,(P+1)N}$ can be expressed as:

$$A(x_n) = \sum_{i=0}^{P} K_i \otimes A_i(x_n).$$

(12)

Note that, $N$ is the number of degree of freedom for the finite element discretization, $x_n$ is the linearized point. $A_i(x_n) \in \mathbb{R}^{N,N}$, $i = 0, \ldots, P$ are the Kronecker factors and $K_i \in \mathbb{R}^{P+1,P+1}, i = 0, \ldots, P$ denote the stochastic Galerkin matrices defined by $(K_i)_{j,k} = c_{ijk}$, which will play an important role in the multilevel method.

### 3.1 Multilevel method

Solving a such complex stochastic Galerkin system is very challenging, choosing a good solving strategy is very crucial. The multilevel method has been applied with good convergence behavior [26, 24, 25].

The stochastic Galerkin system possesses a hierarchical structure due to the polynomial construction, thus we make use of this feature. If space $S_l$ is spanned by the Chaos Polynomials:

$$S_l = \text{span}\{\psi_0, \ldots, \psi_{P_l}\},$$

(13)

$l \in \mathbb{N}, l \leq No$ is certain polynomial degree, $P_l = (M + l)!/(M!!)$. $M$ is again the number of uncertain parameters, in this paper, $M = 3$. $P_l$ is the total number of PC mode with respect to the polynomial order $l$. The hierarchical space $S$ can be also expressed as a nested sequence of spaces with the consideration of the total polynomial degree No:

$$S_0 \subseteq S_1 \subseteq \cdots \subseteq S_l \cdots \subseteq S_{No}.$$  

(14)

The discretized Newton step of Equation (10) at iteration $n$ based on Equation (13) can be found like:

$$\sum_{i=0}^{P_{No}} K_i \otimes A_i(x_n) \tilde{x}_n^{No} = -b_n^{No}.$$  

(15)

$\tilde{x}_n^{No}, b_n^{No} \in \mathbb{R}^{N(P+1)}$, and $\tilde{x}_n^{No} = x_n^{No+1} - x_n^{No}$. We bring into play the principle of multigrid method [5, 6, 16], and consider the polynomial degree $l$ in Equation (14) as the "grid level". Hence, the restriction operator $R_{l-1}$ and the prolongation operator $P_l$ are naturally defined as a $L^2$ projection from $S_{l-1}$ to $S_l$ or vice versa.

One of very crucial procedure for Multilevel method (as well as for Multigrid scheme) is to choose an appropriate smoother, we utilize here the Mean based preconditioner as our smoother [26], we apply $\theta$ times smoothing process to a given initial solution $x_0^l$:

$$x_{k+1}^l := B_l x_k^l = x_k^l + (I_d \otimes A_0)^{-1}.$$  

(16)

Note that, the subscript $k$ indicates the iteration number within the smoothing procedure, which is not the same for Newton step in Equation (11) and Equation (12). A pseudo algorithm about our stochastic Multilevel solver/preconditioner can be found in Algorithm 1.
Algorithm 1 PCE Multilevel preconditioner/solver : ML

1: if $l = 0$ then
2:   solve $A_0 x^0 = b^0$
3: else
4:   $x^l = S^l x^l$ ($\nu_1$ times pre-smoothing)
5:   $r^l = b^l - \sum_{i=0}^{N}(K_i \otimes A_i)x^l$ (compute residual)
6:   $r^{l-1} = R_{l-1} r^l$ (restriction)
7:   for $i = 1$ to $\mu$ do
8:     ML($b^{l-1}, x^{l-1}, l$) (V-cycle/W-cycle)
9:   end for
10:   $c^{l} = P_l c^{l-1}$ (prolongation)
11:   $x^l = x^l + c^l$ (update with correction)
12: $x^l = S^l x^l$ ($\nu_2$ times post-smoothing)
13: end if

Figure 3: One cycle of the multilevel method: ML($x^l, b^l, l$) given a vector $x^l$ and right hand side $b^l$ on level $l$. $\mu = 1$ results in a V-cycle, $\mu = 2$ in a W-cycle.

4 NUMERICAL RESULTS

We use the Crank-Nicolson time stepping scheme for the instationary Navier-Stokes equations, and the nonlinear system Equation (10) is solved via the inexact Newton scheme, it implies that only an approximated solution is applied for each Newton step. We applied the strategy "choice 1" of Eisenstat and Walker in [12] with an initial forcing term equals to 0.5.

<table>
<thead>
<tr>
<th>Inflow maximal speed (m/s)</th>
<th>0.5</th>
<th>Inflow speed variation ($\sigma_1$)</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamic viscosity (N·s/m²)</td>
<td>0.0035</td>
<td>Viscosity variation ($\sigma_3$)</td>
<td>10%</td>
</tr>
<tr>
<td>Angular speed (rad/s)</td>
<td>261.8</td>
<td>Angular speed variation ($\sigma_2$)</td>
<td>10%</td>
</tr>
<tr>
<td>RPM</td>
<td>2500</td>
<td>Density (Kg/m³)</td>
<td>1035</td>
</tr>
</tbody>
</table>

Table 1: Model parameter values.

For each Newton step, we apply an iterative Krylov subspace solver, the flexible generalized minimal residual method (FGMRES), with a Multilevel method preconditioner, which is presented in previous section. For the mean block solver, we use the generalized minimal residual method (GMRES), which is preconditioned with Schur Complement preconditioner. We use an inexact Multilevel method here, it indicates that we compute the inverse of $A_0$ only respect to certain lower accuracy criterion. In this case, we solve the linear system with $A_0$ (in Equation (16)) only up to a relative error $1.0e-1$. In contract, the relative accuracy of the Newton step is set to $1e-9$.

Table 1 indicates the information about input parameters, the Reynolds number, according to Equation (4) is about 210000. The geometry discretization contains 2,984,259 unstructured cells, it results thus 2,274,904 degrees of freedom for the deterministic case. As mentioned in Section 2.3, we consider three different input uncertain parameters:inflow boundary condition, dynamic viscosity and angular speed of the rotor. In this work, we consider the situation that the polynomial degree equals to 3, which gives us in total 20 PC modes. Therefore, the total
Figure 4: Scalability test with the first 50 time steps, 3 uncertainties parameters and polynomial degrees equals to 2.

Figure 5: The mean value and standard deviation of the pressure at time step = 500.

Figure 6 shows the mean value and standard deviation of pressure distribution on the rotor, the standard deviation follows slightly the magnitude of the mean value, as at the center it is lower and on the hub of the blade is higher. But the uncertainty distribution arises also at certain locations where the pressure is less important. Figure 6 represents the mean value and the standard deviation of the velocity, the standard deviation becomes higher after the flow at the outlet, it might be due to the acceleration after the nozzle structure. Figure 7 represents the value of pressure integration over the rotor. It shows the mean value (red) and mean ± 3 standard deviation involving along time.
Figure 6: The mean value and standard deviation of the velocity at time step = 500.

Figure 7: Mean value and mean ± 3 standard deviation for the pressure integration over the rotor.

5 CONCLUSIONS

This work is a continuation of our last contribution [26], which considered a simplified geometry and an artificial stabilized laminar flow. Within this work, we introduce first with our moving mesh strategy in order to be able of modeling this rotating machine in unsteady state, the evolution speed is therefore coupled into the governing equation. The variational multiscale method offers the viability of modeling the high Reynolds number flow. Hence, the intrusive Polynomial Chaos Expansion method is build upon this modelization for quantifying the impact of different uncertain parameters simultaneously.

Besides mathematical modeling, we place also our interest on the solver/preconditioner technique as solving such a complex system is challenging in practice. We therefore employ the inexact Multilevel preconditioner based on the comparison in our last paper [26] in a large system, we benefit from the low accuracy smoothing process to save computational effort.

Our following work will be focused on developing more analysis about the device’s performance respect to surgical requirements. Further process of post-processing of our numerical simulation data and drawing more insight information about the uncertainties.
6 ACKNOWLEDGMENT

The results of this work is computed on bwForCluster MLS&WISO (production and development). The authors acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant INST 35/1134-1 FUGG. The authors acknowledge support by the Heidelberg Institute for Theoretical Studies, HITS gGmbH.

REFERENCES


REDUCED ORDER MODEL-BASED UNCERTAINTY MODELING OF STRUCTURES WITH LOCALIZED RESPONSE

Pengchao Song and Marc P. Mignolet

SEMTE Faculties of Mechanical and Aerospace Engineering
Arizona State University, Tempe, AZ 85287, USA
e-mail: pengchao.song@asu.edu, marc.mignolet@asu.edu

Keywords: Uncertainty Modeling, Reduced Order Modeling, Maximum Entropy, Structural Uncertainty, Localized Response.

Abstract. This paper focuses on the introduction of uncertainty in reduced order models of structures exhibiting a localized static response in the neighborhood of the excitation. A straightforward application of the maximum entropy framework is first considered to carry out the stochastic modeling of the uncertainty. Quite consistently with the maximization of the entropy, it is found that this modeling may lead to a “globalization” of the response and thus an extension of the nonparametric stochastic modeling approach is sought. To this end, the eigenvalues and eigenvectors of the stiffness matrix of structures exhibiting this localization property are first studied. It is found that their lowest eigenvalues are closely spaced when the corresponding eigenvectors are extended to the entire structure. On this basis, a novel version of the nonparametric stochastic modeling approach is introduced to randomize the entire stiffness matrix while distorting only slightly the closely spaced eigenvalue structure. The above concepts are demonstrated on a thin annulus clamped at its inner radius and a localization of the uncertain response is indeed observed using the proposed approach.
1 INTRODUCTION

In the last two decades or so, significant progress has been achieved toward the rational consideration of uncertainty in engineering systems to better predict their performance and the variability thereof. Among the techniques developed within this time frame is the maximum entropy-based nonparametric approach which was initially proposed [1] within the context of modal models of structures. It is however applicable much more broadly (e.g., see [2] for a review) especially when projection based reduced order models are employed. Such models are traditionally developed for computational reasons, being much faster than their full order counterparts, e.g., finite element models. However, from the standpoint of uncertainty modeling, they have yet another advantage as they “concentrate” the uncertainty originating from multiple sources into the single set of their parameters which are typically matrices. The reduced order modeling framework thus permits the consideration of uncertainty in well defined parameters of the physical systems, e.g., thickness, density, etc., but also of harder to model uncertainties such as geometry, material constitutive behavior, etc., as long as they do not affect the form of the reduced order model.

Since its initial formulation in [1], the maximum entropy-based nonparametric approach has been extended multiple times to cover new classes of problems, e.g., vibro-acoustics [3,4], rotordynamics [5-7], nonlinear structural dynamics [8,9], nonlinear thermoelastic problems [10], linear viscoelastic structures [11], etc., but also in rigid body dynamics [12,13] and micromechanics and multiscale modeling, see [2]. The focus of the present effort is on yet another extension of this approach, more specifically to problems in which (i) the response of the mean model (i.e., the one without uncertainty) is localized to one part of the spatial domain and (ii) the uncertainty to be modeled does not alter this localization. The static response of some structures does satisfy the above conditions, and one such example described next will be used as application. There are however other non-structural applications in which these conditions are encountered, e.g., in some heat conduction problems (see [14]), and to which the present discussion may also apply.

2 REPRESENTATIVE EXAMPLE

To illustrate the class of problems investigated here, consider the annulus shown in Fig. 1(a) of inner radius 0.8m, outer radius 1m, thickness 0.002m clamped on its inner radius and free on the outer one. The material, aluminum, is assumed to be homogenous and isotropic with Young’s modulus $E = 7.3 \times 10^{10}$ Pa and Poisson’s ratio $\nu = 0.316$. The annulus is subjected to a static uniform unit pressure in the quadrant $\theta \in [180, 270]$ degrees highlighted in yellow in Fig. 1(a). To evaluate the displacement field of the annulus, it was modeled by finite elements within Nastran (CQUAD4 elements) with a mesh of 144 nodes around the periphery and 6 in the radial direction. Then, shown in Fig. 1(b) is the transverse displacement of the periphery which is clearly localized near the excitation, i.e., in the band $\theta \in [150, 300]$ degrees.

A reduced order model of the finite element one can be constructed by representing the nodal responses stacked in the vector $u$ as a linear combination of basis functions $\psi_i$, i.e. as

$$u = \sum_{i=1}^{M} q_i \psi_i$$  \hspace{1cm} (1)

where the variables $q_i$ are referred to as generalized coordinates. The adoption of the representation of Eq. (1) transforms the equations for the nodal displacements

$$\bar{K}_{FE} u = \bar{F}_{FE},$$  \hspace{1cm} (2)

where $\bar{K}_{FE}$ and $\bar{F}_{FE}$ are the finite element stiffness matrix and loading, into
\[
\overline{K} \overline{q} = \overline{F}
\]

where

\[
\overline{K} = \Psi^T \overline{K}_{\text{FE}} \Psi \quad \text{and} \quad \overline{F} = \Psi^T \overline{F}_{\text{FE}}
\]

with \( \Psi \) denoting the operation of matrix transposition and

\[
\Psi = \left[ \phi_1, \phi_2, \phi_3, \ldots, \phi_M \right]
\]

The above reduced order model construction was exemplified by selecting the basis functions \( \phi_i \) as the linear modes of the annulus. Then, shown in Fig. 1(b) is the transverse displacement of the periphery obtained with 55 such modes that, as expected, closely approximates the finite element solution.

Figure 1: (a) The annulus and its finite element model with the loading domain highlighted in yellow. (b) Static transverse displacements at the periphery, full finite element (FEA) and reduced order (ROM) models.

3 STOCHASTIC UNCERTAINTY MODELING

3.1 “Standard” Maximum Entropy Nonparametric Approach

The consideration of uncertainty in the properties and/or geometry of the structural model can be carried out on the finite element model by letting a set of its parameters, e.g., Young’s modulus, Poisson’s ratio, thickness, outer/inner radii, etc. be random variables with known joint probability density function. Then, proceeding with a Monte Carlo simulation of these properties and performing a finite element analysis for each sample will provide a population of responses which can be analyzed to assess the effects of uncertainty. This approach is appropriate, although potentially computationally expensive, to address the uncertainty in parameters of the model such as those stated above. It is however much more difficult to implement when the uncertainty originates from waviness in the periphery, deviations from the isotropic and homogenous material properties, etc.

While the finite element model will vary, the form of the equations for the response, i.e., Eq. (2), will remain unchanged as long as the constitutive model is linear and that the deformations remain small. Similarly, Eq. (3) will also remain valid as long as the basis functions, assumed here to be deterministic, provide an appropriately accurate representation of the random displacement field. This observation suggests the possibility of introducing the uncertainty directly in either Eq. (2) or Eq. (3). In fact, proceeding with Eq. (3) is easier as (i) \( \overline{K} \) is

\[
\overline{K} = \Psi^T \overline{K}_{\text{FE}} \Psi
\]

and (ii) \( \overline{F} = \Psi^T \overline{F}_{\text{FE}} \).

Pengchao Song and Marc P. Mignolet
Pengchao Song and Marc P. Mignolet

a typically much smaller matrix than \( \overline{K}_{FE} \) and (ii) the former does not exhibit a particular topology as opposed to the latter.

Having established the desire to introduce the uncertainty directly in the reduced order model stiffness matrix \( \overline{K} \), it remains to address how to proceed and more specifically how to select the joint probability density function of the elements of this matrix. The framework selected here is the maximum entropy-based nonparametric approach proposed by Soize [1], see [2] for an extensive review. Within this framework, the joint probability density function of the elements of the stiffness matrix \( \overline{K} \) is not selected but rather determined to maximize the corresponding entropy under constraints corresponding to all physical properties it must satisfy.

More specifically, it is known that the finite element stiffness matrix \( \overline{K}_{FE} \) must be symmetric and positive definite (strictly positive definite given the clamp boundary conditions) and those properties are transferred to its reduced order model counterpart \( \overline{K} \) through the transformation of Eq. (4). Then, proceeding as in [1], see also [2], this stiffness matrix of the “mean model”, i.e., the model without uncertainty, is first decomposed as

\[
\overline{K} = \overline{L} \overline{L}^T
\]

(5)
e.g., by Cholesky factorization. Then, random matrices \( K \) are generated as

\[
K = \overline{L} H H^T \overline{L}^T
\]

(6)
where \( H \) is a lower triangular matrix such that (see also Fig. 2)

1. its off-diagonal elements \( H_{il}, i \neq l \), are normally distributed (Gaussian) random variables with standard deviation \( \sigma = 1/\sqrt{2\mu} \), and

2. its diagonal elements \( H_{ii} \) are obtained as \( H_{ii} = \sqrt{Y_{ii}/\mu} \) where \( Y_{ii} \) is Gamma distributed with parameter \( (n+1)/2 \) where

\[
p(i) = n - i + 2\lambda_0 - 1 \quad \text{and} \quad \mu = (n + 2\lambda_0 - 1)/2
\]

(7)

In the above equations, \( n \) is the size of the matrices and the parameter \( \lambda_0 > 0 \) is the free parameter of the statistical distribution of the random matrices \( K \). An alternative parametrization is through the dispersion parameter \( \delta \) defined as

\[
\delta^2 = \frac{n + 1}{n + 2\lambda_0 - 1}.
\]

(8)

zero mean Gaussian, independent of each other with standard dev. \( \sigma = 1/\sqrt{2\mu} \)

square root of Gamma, independent of all others

\[
H_{il} = \sqrt{Y_{ii}/\mu}
\]

Figure 2: Structure of the random \( H \) matrices (figures for \( n=8, i=2, \) and \( \lambda=1 \) and 10).
The above approach was applied as is to the 55x55 stiffness matrix of the reduced order model of the annulus and 300 random matrices $K$ were determined. From each one of those, a set of generalized coordinates $q$ was determined that satisfies

$$K q = F$$  \hspace{1cm} (9)$$

and the resulting set of nodal displacements $u$ were obtained from Eq. (1). Shown in yellow in Fig. 3 is the uncertainty band corresponding to the 5th and 95th percentile of the transverse displacement of the periphery as determined from the 300 samples of the response. Comparing Figs 1(b) and 3, it is observed that the above uncertainty modeling approach has induced some globalization of the response, i.e., the response outside of the region $\theta \in [150,300]$ degrees is not nearly zero for all samples as could have been construed from Fig. 1(b). Note that this effect likely leads to a higher entropy of the response as compared to the localized case given its increased spread of the joint probability density function of this response. This finding seems quite consistent with, although separate from, the maximization of entropy of the matrix $K$ which is guaranteed by the construction of Eqs (6)-(8).

It should be recognized that the above globalization effect is certainly physically possible, e.g., it could take place if the disk on which the annulus is clamped is not rigid as specified in the mean model but rather exhibits some flexibility which induces a long range interaction between nodal responses.

If, however, there is additional knowledge about the uncertainty that indicates that such a globalization does not take place, then the above methodology must be modified to reflect it, e.g., by adding constraints in the optimization of the entropy or modeling differently the random stiffness matrix. One such extension is formulated below.

### 3.2 Localization Inducing Property

Before revising the above standard nonparametric approach, it is necessary to identify the property of the stiffness matrices $\overline{K}$ and $\overline{K}_{FE}$ that induces the existence of a localized response. To this end, recall that the solution $u$ of Eq. (1) can be expanded in terms of the ei-
genectors $\phi_j$ and corresponding eigenvalues $\lambda_j$ of $K_{FE}$ as

$$u = \sum_j \left( \frac{\phi_j^T FE}{\lambda_j} \right) \phi_j$$

(10)

owing to the orthogonality of the eigenvectors implied by the symmetry of $K_{FE}$.

To proceed further in the discussion, assume first that the eigenvectors $\phi_j$ are extended to the entire structure; this is true of the annulus of Fig. 1(a) for which $\phi_j$ are harmonic functions of the angle $\theta$ (as discretized by the finite element modeling). Then:

(i) a localized response as shown in Fig. 1(b) is possible only if the dominant coefficients $\left( \frac{\phi_j^T FE}{\lambda_j} \right)$ have somewhat similar values as to create an appropriate mix of the extended functions $\phi_j$ to produce localization.

(ii) the “modal forces” $\phi_j^T FE$ decrease slowly as the index $j$ is increased given the localization of the excitation or equivalently the values $\phi_j^T FE$ for the dominant modes will be quite similar.

Combining these two arguments, it is thus concluded that localization must require that the lowest eigenvalue $\lambda_j$ of $K_{FE}$ be close together if the eigenvectors of this stiffness matrix are extended. This result is confirmed by the plot of eigenvalues of the annulus’ stiffness matrix shown in Fig. 4: the first few of its eigenvalues are indeed closely spaced.

Figure 4. Eigenvalues of the stiffness matrix of the annulus finite element increasing order.

For structures with localized response and well separated eigenvalues of their stiffness matrix, the change in stiffness matrix induced by the random matrix $H$ (Eq. (6)) produces only small changes in eigenvalues and eigenvectors when the uncertainty is small (see [15] for perturbation analysis) and thus small changes in the response. That is, for these structures, the application of the standard nonparametric approach may maintain the localized behavior at least as long as the change in eigenvalues and eigenvectors induced by the uncertainty remains small enough.
The above comments suggest that what is important for localization is the relative separation of the first few eigenvalues which must be small when the eigenvectors are extended. To confirm this statement, rewrite Eq. (10) in the form
\[
u = \frac{1}{\lambda_1} \sum_j \left( \phi_j^T F FE \right) \phi_j + \frac{1}{\lambda_1} \sum_j \frac{\lambda_1 - \lambda_j}{\lambda_j} \left( \phi_j^T F FE \right) \phi_j.
\] (11)
or
\[
u = \left[ \frac{1}{\lambda_1} \sum_j \phi_j \phi_j^T \right] F FE + \left[ \frac{1}{\lambda_1} \sum_j \frac{\lambda_1 - \lambda_j}{\lambda_j} \phi_j \phi_j^T \right] F FE = \overline{P}_1 F FE + \overline{P}_2 F FE.
\] (12)
where \( \overline{P}_1 \) is proportional to the identity matrix and thus \( \overline{P}_1 F \) is a purely local term, i.e., the corresponding displacement at each node only depends on the force acting at that node. This term will dominate when the terms \( \left( \phi_j^T F \right)(\lambda_1 - \lambda_j) / \lambda_j \) are all small, i.e., when the dominant eigenvectors (those with significant values of \( \phi_j^T F \)) have a small relative difference of eigenvalue with eigenvector 1.

A similar discussion to the above one applies for the mean reduced order stiffness matrix \( \bar{K} \) of eigenvectors \( \phi_j \) and eigenvalues \( \mu_j \) so that
\[
q = \overline{Q} F = \left[ \frac{1}{\mu_1} \sum_j \phi_j \phi_j^T \right] F + \left[ \frac{1}{\mu_1} \sum_j \frac{\mu_1 - \mu_j}{\mu_j} \phi_j \phi_j^T \right] F = \overline{Q}_1 F + \overline{Q}_2 F.
\] (13)
Associated to this decomposition of the flexibility matrix \( \overline{Q} \) into a component inducing local effects (\( \overline{Q}_1 \)) and a more global one (\( \overline{Q}_2 \)) is a similar split of the stiffness matrix \( \bar{K} \) into a local component, \( \bar{K}_L \), and a more global one, \( \bar{K}_G \), defined as
\[
\bar{K}_L = \left[ \overline{Q}_1 \right]^{-1} = \mu_1 \sum_j \phi_j \phi_j^T \quad \text{and} \quad \bar{K}_G = \bar{K} - \bar{K}_L.
\] (14a),(14b)

### 3.3 Maximum Entropy Nonparametric Modeling for Localized Responses

The above discussion has demonstrated that a localization of the response will only occur when the relative spread of the dominant eigenvalues of the stiffness matrix is small when the corresponding eigenvectors are extended. However, this spread is often increased, sometimes very significantly for the first few eigenvalues, when using the standard nonparametric approach as shown in Fig. 5.

Thus, imposing a localization constraint on the stochastic modeling will require controlling the eigenvalues of the dominant modes (those with eigenvalues close to \( \lambda_1 \)) separately from the rest of them. In this regard, note that the dominant eigenvectors are mostly present in \( \bar{K}_L \) while those with eigenvalues far from \( \lambda_1 \) are dominant in \( \bar{K}_G \).

On the basis of the above observations, it is proposed here to model the uncertainty in \( \bar{K}_L \) and \( \bar{K}_G \) separately. Since there is no particular requirement on the latter matrix (which is nevertheless symmetric and positive definite) and its randomization, the standard nonparametric approach will be applied leading to an uncertain matrix \( \bar{K}_G \) defined as
\[
\bar{K}_G = \bar{L}_G \bar{H}_G \bar{H}_G^T \bar{L}_G^T \quad \text{where} \quad \bar{K}_G = \bar{L}_G \bar{L}_G^T.
\] (15)
with $H_G$ a lower triangular random matrix as defined by Eqs (7)-(8) and Fig. 2 for a particular dispersion $\delta_G$.

Maintaining small the relative separation between the eigenvalues of the dominant eigenvectors can be achieved by scaling uniformly all eigenvalues. This observation suggests that a first approach to introduce uncertainty in $K_L$ is to simply multiply it by a random variable $\Lambda$ which consistently with the maximum entropy concepts can be selected as $H_1^2$ where $H_1$ is a 1x1 matrix defined as in Eqs (7)-(8) and Fig. 2 with a specified dispersion $\delta_1$. A small change in the relative separation between eigenvalues can also be induced by splitting $K_L$ as $K_G$ in Eq. (15) with a random lower triangular matrix $H_L$ of dispersion $\delta_L$. Combining the above two operations leads to the proposed model

$$K_L = H_1^2 L_L H_L H_L^T L_L^T \quad \text{where} \quad K_L = L_L L_L^T.$$  
(16a),(16b)

Note finally that since $K_L$ is defined by Eq. (14a), its decomposition in Eq. (16b) is readily achieved by selecting

$$L_L = \sqrt{\mu_1} [\varphi_1 \varphi_2 \varphi_3 \ldots]$$  
(17)

The uncertain reduced order stiffness matrix is then obtained as

$$K = K_L + K_G.$$  
(18)

### 3.4 Application to the Annulus

The three-parameter ($\delta_G, \delta_1, \delta_L$) stochastic model formulated above was applied to the reduced order model of the annulus with the first 8 eigenvectors $\varphi_j$ retained in the summation of Eq. (14a). To highlight the effects of each random component on the response, shown in Figs 6 (a),(c),(e) are the uncertainty bands obtained with each of the parameter set to a non-zero value in turn. Also shown, on Figs 6 (b),(d),(f) are 3 samples of the corresponding responses.
Figure 6: Static transverse displacement at the periphery of the mean annulus (in red) and results from the stochastic reduced order model: (a),(c),(e) 5th-95th percentile uncertainty band (in yellow), (b),(d),(f) 3 samples. $(\delta_G, \delta_1, \delta_L) = (a),(b) (0.1,0,0); (c),(d) (0,0.05,0); (e),(f) (0,0,0.02)$. 

Pengchao Song and Marc P. Mignolet
From Figs 6 (a),(b), corresponding to $\delta_G \neq 0$, it is observed that the randomization of $K_G$ only leads to a small, rather constant, global uncertainty band consistently with the above discussions. Next, consider Figs 6 (c),(d) corresponding to $\delta_1 \neq 0$ and which leads to a very localized large variability of the response confirming that $K_L$ does indeed control the localized behavior. Finally, shown in Figs 6 (e),(f) are the results corresponding to $L_\delta \neq 0$ which are very similar to those shown in Figs 6 (a),(b) suggesting tentatively that it is sufficient to only consider the two parameter model defined by $\delta_G$ and $\delta_1$, i.e., with

$$K_L = H_1^2 \bar{K}_L.$$  \hspace{1cm} (19)

Shown in Fig. 7 are the uncertainty band (Fig. 7(a)) and 3 samples of the response (Fig. 7(b)) corresponding to the combined case of all three parameters nonzero. It is seen that their effects approximately superpose, creating a thin uncertainty band away from the localization region but a much more significant one within it and samples that exhibit the localization as desired.

![Figure 7: Static transverse displacement at the periphery of the mean annulus (in red) and results from the stochastic reduced order model: (a) 5th-95th percentile uncertainty band (in yellow), (b) 3 samples. ($\delta_G, \delta_1, \delta_L$)=(0.1,0.05,0.02).](image)

4 SUMMARY

This investigation focused on the development of a stochastic model of uncertainty in structures exhibiting a localized static response in the neighborhood of the excitation. More specifically, this modeling is carried out on a reduced order model of the response by relying on the maximum entropy framework. It is found that the “standard” nonparametric method leads to a local response of the simulated uncertain structures when the lowest eigenvalues of the mean structure’s stiffness matrix are well separated. However, when these eigenvalues are closely spaced, this method leads to a globalization of the response which seems consistent with the maximization of the entropy. For these situations, a novel model is proposed which is based on a split of the mean structure’s stiffness matrix into a component that promotes the local response and one that induces a more global behavior. Then, separate, maximum entropy based, stochastic models of these two components are carried out before they are recombined to form the stiffness matrix of the uncertain structure. This process provides control over the separation of the eigenvalues of the uncertain stiffness matrices and thus permits the
occurrence of localization. The above findings are confirmed on an annulus clamped at its inner radius and free at the outer one and the localization of its uncertain response obtained with the proposed model is confirmed.

ACKNOWLEDGEMENTS

The financial support of this work by the Air Force Multi University Research Initiative contract FA9550-15-1-0038 with Dr. Jean-Luc Cambier as Technical Monitor is gratefully acknowledged.

REFERENCES


DIFFERENT VIEWS ON ADDITIONAL RANDOM PARAMETERS IN EXPERIMENT DESIGN FOR THERMOPHYSICAL PARAMETERS ESTIMATION (UNCECOMP 2017)

D. Jarůšková and A. Kučerová

Czech Technical University in Prague, Faculty of Civil Engineering
Thakurova 7, CZ 166 29 Praha 6, Czech Republic
e-mail: daniela.jaruskova@cvut.cz

Keywords: nonlinear regression, common random factors, nuisance parameters, robust design of experiment

Abstract. Our contribution deals with a design of experiment problem for estimating thermophysical parameters described by Ruffio et al. (2012). In particular, the partial differential model of the experiment is replaced here by nonlinear regression model where temperature is a dependent variable, time of measurement plays a role of independent variable and thermal conductivities along x and y axes are parameters to be estimated. The goal of statistical inference is to find optimal position(s) of sensor(s) to estimate the parameters of interest as accurately as possible.

A statistical analysis becomes more complicated when some additional input parameters are random. In our contribution we discuss the situation, where sensor(s) are not placed exactly to designed position(s) due to an error modelled by random nuisance parameters or the so-called random factors. It appears that the very same problem can be viewed from various perspectives so that its solutions are completely different.
**1 INTRODUCTION**

Recent developments in the field of uncertainty quantification open more possibilities to simulate the nonlinear systems with uncertain input parameters and moreover to design optimized and robust experiments for calibrating the models of such systems. Nevertheless, the difficulties connected to handling the uncertain parameters consist not only in computational requirements, but also in formulation of the problem with uncertain parameters itself, namely with respect to the considered source of uncertainty. The goal of this contribution is to demonstrate two different results obtained due to different perspectives on random parameters in experiment design problem. In particular, the difference originates from two different attitudes of a statistician analyzing the data from an experiment – whether he considers the randomness of parameters in his inverse analysis or not. Different perspectives of the statistician are reflected within the process of experiment design providing different optimal solutions. Particular scenarios are described for an illustrative example of non-stationary heat conduction in two-dimensional square domain inspired by Ruffio et al. [4], where the goal of experiment design problem is to find optimal position of a thermocouple measuring temperature in a set of time steps so as to reduce the uncertainties in estimated thermal conductivities in the two principal directions.

**2 EXPERIMENT DESCRIPTION**

Our paper was inspired by [4] that presents an analysis of a numerical experiment. In our simplified version the aim of the experiment is to identify two thermal parameters of an orthotropic homogeneous material - more precisely, a constant thermal conductivity $\lambda_x$ along the $x$–axis and a constant thermal conductivity $\lambda_y$ along the $y$–axis.

The experiment can be described as follows. A square sample is exposed to a constant and uniform heat flux $\varphi$ on the left and bottom boundaries, while the right and top edges are insulated. The system obeys the following equation:

$$C \frac{\partial T}{\partial t} = \lambda_x \frac{\partial^2 T}{\partial x^2} + \lambda_y \frac{\partial^2 T}{\partial y^2},$$

$$0 \leq x \leq L_x, \ 0 \leq y \leq L_y, \ 0 \leq t \leq \tau.$$

The boundary and initial conditions are defined by:

$$-\lambda_x \frac{\partial T}{\partial x} (x = 0) = \varphi, \quad -\lambda_y \frac{\partial T}{\partial y} (y = 0) = \varphi,$$

$$-\lambda_x \frac{\partial T}{\partial x} (x = L_x) = 0, \quad -\lambda_y \frac{\partial T}{\partial y} (y = L_y) = 0,$$

$$T(x, y, 0) = 0.$$

One assumes that temperature is measured by one or more sensors at equidistant time points $t = 1, \ldots, 60$ (s) and $L_x = L_y = 0.05$ (m). The number of sensors is fixed. In Ruffio et al. [4] three sensors were considered. Jarušková and Kučerová [2] considered either one or three sensors. The aim of an analysis of the numerical model is to determine position(s) of sensor(s) to identify thermal conductivities $\lambda_x$ and $\lambda_y$ as accurately as possible while the thermal capacity
\( C = 1700000 \text{ J m}^{-3}\text{K}^{-1} \) and the heat flux \( \phi = 25000 \text{ W m}^{-2} \) are known. Ruffio et al. \cite{4} claim that a solution of the model above may be found analytically in the form of a quickly converging series. In our paper, we assume that a final solution may be quite accurately approximated by the first term only, i.e.,

\[
T(t; \lambda_x, \lambda_y; x, y) = \theta_x(t; \lambda_x; C; \varphi; x) + \theta_y(t; \lambda_y; C; \varphi; y),
\]

\[
\theta_x(t; \lambda_x; C; \varphi; x) = \frac{2\varphi}{\sqrt{C\lambda_x}} \sqrt{t} F\left( \frac{\bar{x}}{\sqrt{t}} \right), \quad \theta_y(t; \lambda_y; C; \varphi; y) = \frac{2\varphi}{\sqrt{C\lambda_y}} \sqrt{t} F\left( \frac{\bar{y}}{\sqrt{t}} \right),
\]

\[
F(z) = \exp\left( -\frac{z^2}{4\pi} \right) - z\left( 1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-v^2} dv \right), \quad z \geq 0.
\]

with \( \bar{x} = (x/2)\sqrt{C/\lambda_x}, \bar{y} = (y/2)\sqrt{C/\lambda_y} \). Figure 1 presents description of our experiment in a graphical way.

Because a sensor situated at \((x, y)\) measures temperatures \(\{Y_i\}\) with measurement errors \(\{e_i\}\), we may assume that the behavior of a measured temperature can be modeled with the help of a nonlinear regression:

\[
Y_i = T(t_i; \lambda_x, \lambda_y; x, y) + e_i, \quad i = 1, \ldots, 60,
\]

where \(\{e_i\}\) are i.i.d. with \( E e_i = 0 \) and \( \text{Var} e_i = \sigma_m^2 = 0.1^2 \). (In the case of three sensors we have 60 measurements in all three positions \((x_1, y_1), (x_2, y_2), (x_3, y_3)\), i.e. 180 measurements in total.) The estimates \( \hat{\lambda}_x \) and \( \hat{\lambda}_y \) may be obtained using the least squares method.
A decision where to place sensor(s) is an experiment design problem in the nonlinear regression, see [3]. We consider two criterions: the $D$ criterion minimizes a determinant of the variance-covariance matrix of $(\hat{\lambda}_x, \hat{\lambda}_y)$ and the $F$ criterion minimizes $(\text{Var} \hat{\lambda}_x/\lambda_x^2 + \text{Var} \hat{\lambda}_y/\lambda_y^2)^{1/2}$. Due to independence of measurement errors $\{e_i\}$, the variance of the estimates $\hat{\lambda}_x$ and $\hat{\lambda}_y$ converges to zero with increasing number of measurements. As the number of observations is large (here 60 measurements), we may consider the variance of the estimates small enough to replace the nonlinear model by its linear approximation at their true values and the exact variance-covariance matrix by an asymptotically equivalent matrix:

$$
\sigma^2_m \left( F^* F^* \right)^{-1},
$$

where the matrix $F^*$ is a two-column matrix of partial derivatives of $T$ with respect to $\lambda_x$ and $\lambda_y$ computed at their true values. It is important to notice that when one sensor is used, the matrix $F^* F^*$ is singular when the sensor $(x, y)$ is placed to the line $y = \sqrt{\lambda_y/\lambda_x} x$. Singularity is connected to the fact that we estimate two parameters by only one sensor. In case of three sensors the corresponding asymptotic variance-covariance matrix is singular when all sensors are situated at that line. Positions of sensor(s) yielding a singular approximate variance-covariance matrix used in the $F$ or $D$ criterion will be called inadmissible.

Ruffio et al. [4] as well as Jarušková and Kučerová [2] started an analysis supposing that $\lambda_x = 0.6$ and $\lambda_y = 4.7$. For one sensor there exist two optimal solutions with respect to the $D$ criterion being $x_{\text{op}}^{(1)} = 0.0018$, $y_{\text{op}}^{(1)} = 0$ and $x_{\text{op}}^{(2)} = 0$, $y_{\text{op}}^{(2)} = 0.0050$, with the optimal value of the $D$ criterion $3.1 \cdot 10^{-12}$, while there exists only one optimal position $x_{\text{op}} = 0.0024$, $y_{\text{op}} = 0$ with respect to the $F$ criterion with the optimal value 0.0016, see Case 1 in [2]. Notice that the optimal positions are on the boundaries and relatively close to the point $(0, 0)$. It is clear that temperature is most sensitive to a change in thermal conductivities near the point $(0, 0)$, see Figure 2, but on the other hand a sensor must not be too close to the line $y = \sqrt{\lambda_y/\lambda_x} x$ to distinguish between the effect of $\lambda_x$ and the effect of $\lambda_y$. For illustration Figure 2 shows dependence of the euclidean norm of the vector $((\partial T/\partial \lambda_x)(1; \lambda_x, \lambda_y; x, 0), \ldots, (\partial T/\partial \lambda_x)(60; \lambda_x, \lambda_y; x, 0))$ on the position of the $x$ - coordinate of a sensor situated on the bottom boundary. Figure 3 shows the dependence of temperature on time for one sensor and different values of $\lambda_x$ and $\lambda_y$.

The problem becomes more complicated when some input parameters of the model are random as described in the following section.

3 NONLINEAR REGRESSION WITH RANDOM PARAMETERS

In our contribution we assume that a sensor may not be placed exactly to a designed position. Then, instead of being measured at a point $(x, y)$ temperature is measured at a point $(x+\Delta x, y+\Delta y)$, where $\Delta x$ and $\Delta y$ are uncertain. In what follows we describe the case of one sensor but the case where more sensors are used is analogue.

3.1 Nonlinear regression with common random factors

In the first scenario we assume that a statistician analyzing the data does not know about the position errors and estimates the parameters of interest $\lambda_x$ and $\lambda_y$ minimizing the least squares under the condition that $\Delta x = 0$ and $\Delta y = 0$, i.e.

$$
(\hat{\lambda}_x, \hat{\lambda}_y) = \arg\min \sum_{i=1}^{60} \left( Y_i - T(t_i; \lambda_x, \lambda_y; x, y) \right)^2.
$$
D. Jarušková and A. Kučerová

One experiment consists of 60 temperature measurements and one assumes that we perform many experiments. A position of sensor(s) varies from one experiment to the other according to some prescribed distribution, but in one experiment it does not change and thus it plays a role of the so-called common factor. The least squares estimates \( \hat{\lambda}_x \) and \( \hat{\lambda}_y \) are some functions of \( \Delta x, \Delta y, e_1, \ldots, e_{60} \), which will be different in repeated experiments and thus the estimates \( \hat{\lambda}_x, \hat{\lambda}_y \) will attain different values as well. We are indeed interested in a distribution of \( (\hat{\lambda}_x, \hat{\lambda}_y) \) induced by the given distribution of \( \Delta x, \Delta y, e_1, \ldots, e_{60} \).

We assume that position errors of \{\Delta x\} and \{\Delta y\} in different experiments are (in agreement with [4]) i.i.d. variables distributed according to a zero mean normal distribution with a variance \( \sigma^2 = 0.0005^2 \). Jarušková and Kučerová [2] pointed out that due to a large number of measurements the effect of measurement errors \( \{e_i\} \) on the distribution of \( (\hat{\lambda}_x, \hat{\lambda}_y) \) is practically negligible in comparison with the effect of random displacements \( \Delta x \) and \( \Delta y \) that remain constant in one experiment.

When designing an optimal experiment, we assume that the distribution of measurement errors as well as distribution of position errors is known to the designer who derives the optimality criterion based on the variance-covariance matrix of \( (\hat{\lambda}_x, \hat{\lambda}_y) \). If the regression function \( T \) is an approximately linear function of \( \lambda_x, \lambda_y, \Delta x \) and \( \Delta y \), then \( \hat{\lambda}_x \) and \( \hat{\lambda}_y \) are approximately linear functions of \( \Delta x \) and \( \Delta y \) and \( E (\hat{\lambda}_x - \lambda_x) \approx 0 \) and \( E (\hat{\lambda}_y - \lambda_y) \approx 0 \) and a variance-covariance matrix of \( (\hat{\lambda}_x, \hat{\lambda}_y) \) is approximately equal to:

\[
\sigma^2 \left( F^* T F^* \right)^{-1} F^* Z^* Z^T F^* (F^* T F^*)^{-1} + \sigma^2 H (F^* T F^*)^{-1},
\]

where \( F^* \) is a matrix of partial derivatives of \( T \) with respect to components of \( \lambda_x \) and \( \lambda_y \) and \( Z^* \) is a matrix of partial derivatives of \( T \) with respect to \( \Delta x \) and \( \Delta y \), both computed at true values of \( \lambda_x \) and \( \lambda_y \) and \( \Delta x = 0 \) and \( \Delta y = 0 \).

Ruffio et al. [4] suggest looking for an optimal position by applying either the \( D \) criterion or the \( F \) criterion to the approximate variance-covariance matrix \( (3) \). Then, an optimal position of one sensor with the respect to the \( F \) criterion is \( x_{op} = 0.0165 \), \( y_{op} = 0.0158 \) with the optimal
value $F = 0.04$ and an optimal position with the respect to the $D$ criterion is $x_{op} = 0.0167$ and $y_{op} = 0.0310$ with a value $D = 8.4 \cdot 10^{-5}$. The optimal position(s) of sensor(s) were found numerically on a grid $[0 : 0.0001 : 0.05] \times [0 : 0.0001 : 0.05]$. When positions of three sensors are to be found we have to use a faster method that provides us with suboptimal solutions. In our contribution we applied an evolutionary GRADE algorithm extended by niching strategy CERAF, see [1].

However, it may happen that for some $(x, y)$ the regression function is a strongly nonlinear function and then the only way how to get an approximate distribution of $(\hat{\lambda}_x, \hat{\lambda}_y)$ is by Monte Carlo simulations, see [2].

### 3.1.1 Note on the specific distribution of random parameters

Ruffio et al. [4] assume that random displacements $\Delta x$ and $\Delta y$ are independent, distributed according to a normal distribution $N(0, 0.0005^2)$. Clearly, if the distribution of $\Delta x$ and $\Delta y$ is normal, then for all possible designed positions $(x, y)$ there exist displacements $\Delta x$ and $\Delta y$ such that the sensor is shifted to the line of inadmissible positions. This is not a serious problem because we may, for instance, replace the two-dimensional normal distribution by its trimmed version being zero outside a circle $(\Delta x^2 + \Delta y^2)^{1/2} \leq 3 \cdot 0.0005$. This changes the original distribution very slightly. When temperature is measured by one sensor and we calculate values of the criteria using the variance-covariance matrices obtained by Monte Carlo simulations and by numerical minimization of the least squares, we consider a narrow band around the line to be a set of inadmissible solutions and we look for an optimal solution outside this band.

A more serious problem might occur when a designed position of a sensor is at one of the boundaries or close to it. This often happens when three or more sensors are used, see [4] or [2]. We suggested a way how to replace a normal distribution by a distribution that shifts sensors to an interior of the heated square sample, see [2], but we do not know how much such
a replacement changes an optimality of the solutions.

3.2 Nonlinear regression with nuisance parameters

In the second scenario, it is supposed that the statistician analyzing the data from one experiment knows that real positions of sensors might be different from the designed positions due to random displacements $\Delta x$ and $\Delta y$. He considers them being nuisance parameters and estimates them together with the parameters of interest $(\lambda_x, \lambda_y)$.

The most simple way for estimating all unknown parameters is again the least squares method:

$$\left(\hat{\lambda}_x, \hat{\lambda}_y, \hat{\Delta}x, \hat{\Delta}y\right) = \text{argmin} \sum_{i=1}^{60} \left(Y_i - T(t_i; \lambda_x, \lambda_y; x + \Delta x, y + \Delta y)\right)^2. \quad (4)$$

The estimates are indeed maximum likelihood estimates because measurement errors are normal.

We introduce the four column matrix $G^*$ of the partial derivatives of the function $T$ with respect to $\lambda_x$, $\lambda_y$, $\Delta x$ and $\Delta y$. The matrix

$$\frac{1}{\sigma_m^2} G^* T G^* \quad (5)$$

is an expected Fisher information matrix under the assumption that $\lambda_x$, $\lambda_y$, $\Delta x$ and $\Delta y$ are true values of the parameters. Its inverse $V_{x,y}^\lambda = V_{x,y}(1:2,1:2)$ an approximate variance-covariance matrix of $(\hat{\lambda}_x, \hat{\lambda}_y)$ (for $n$ large). The matrix $V_{x,y}^\lambda = V_{x,y}(\lambda_x, \lambda_y, \Delta x, \Delta y)$ depends on true values of parameters $\lambda_x$ and $\lambda_y$ as well as on true values of displacements $\Delta x$ and $\Delta y$.

For a chosen position of a sensor $(x, y)$ the matrix $V_{x,y}^\lambda(0.6, 4.7, \Delta x, \Delta y)$ expresses variability of the estimates $(\hat{\lambda}_x, \hat{\lambda}_y)$ when a position of a sensor remain in all experiments the same $(x + \Delta x, y + \Delta y)$ (supposing the true values $\lambda_x = 0.6$ and $\lambda_y = 4.7$.) The elements of the matrix $V_{x,y}^\lambda(0.6, 4.7, \Delta x, \Delta y)$ are functions of $(\Delta x, \Delta y)$ so that if position errors differ the approximate variance-covariance matrix may be also different. If the position errors $(\Delta x, \Delta y)$ are random variables so are the elements of the matrix $V_{x,y}^\lambda(0.6, 4.7, \Delta x, \Delta y)$. It corresponds to the situation that many experiments are performed and the position errors vary randomly from one experiment to the other. It is not a big surprise that under an assumption that $\lambda_x = 0.6$, $\lambda_y = 4.7$ and $\Delta x = 0$, $\Delta y = 0$, the optimal positions can be found on the boundaries. The optimal positions with respect to the $D$ criterion are $x_{op}^{(1)} = 0.0044$, $y_{op}^{(1)} = 0$ and $x_{op}^{(2)} = 0$, $y_{op}^{(2)} = 0.0123$ with the $D$ value $2.3 \cdot 10^{-8}$, while there is only one optimal position with a respect to the $F$ criterion $x_{op} = 0.057$, $y_{op} = 0$ with the $F$ value $0.012$.

Table 1 summarizes values of $D$ and $F$ criteria for different situations.

4 Optimality criteria using prior knowledge on parameters

In nonlinear regression an exact as well as an approximate variance-covariance matrices of estimates of parameters of interest depend generally on the very same parameters that are to be estimated. Therefore, we always need some prior information, which commonly takes a form of some expert guess about the estimated parameter values, which can be then used for evaluation of approximate variance-covariance matrices.
Sometimes, we have some prior information in the form of feasible bounded intervals for estimated parameters. For instance, we may know that $0.45 \leq \lambda_x \leq 0.75$ and $3.2 \leq \lambda_y \leq 6.2$. Our desire is to find a position that is the best for the “worst” values of $\lambda_x$ and $\lambda_y$ in a sense that for any other $\lambda_x$ and $\lambda_y$ from the corresponding intervals we get smaller values of the considered criteria. For any admissible position $(x, y)$ we calculate the values of the $F$ or $D$ criterion over a dense grid in $(\lambda_x, \lambda_y) \in [0, 0.45; 0.75] \times [3.2; 6.2]$. The value of the $F_R$ criterion is a maximum of the corresponding $F$ values and similarly the value of the $D_R$ criterion is a maximum of the corresponding $D$ values. The position $(x, y)$ with the smallest $F_R$ value, respectively the smallest $D_R$ value, is optimal. When a position of one sensor is to be designed and the values of the $F$ and $D$ criterion are calculated using the approximate variance-covariance matrix (3), we consider a set $\{[x, y] \in [0; 0.05] \times [0; 0.05]; \sqrt{3.2/0.75} x \leq y \leq \sqrt{6.2/0.45} x\}$ to be inadmissible and we are looking for an optimal position outside this set. Following the first scenario and using (3) the optimal position with the respect of the $F_R$ criterion is $x_{op} = 0.0155$ and $y_{op} = 0.0260$. The optimal value $F_R = 0.102$ is attained for $\lambda_x = 0.45$ and $\lambda_y = 3.2$. The optimal position with the respect of the $D_R$ criterion is $x_{op} = 0.0174$ and $y_{op} = 0.0307$. The optimal value $D_R = 1.94 \cdot 10^{-4}$ is attained for $\lambda_x = 0.45$ and $\lambda_y = 6.2$. If the variance-covariance matrices are obtained by Monte Carlo simulations and numerical minimization of the least squares, we look for an optimal solution outside the set $\{[x, y] \in [0; 0.05] \times [0; 0.05]; \sqrt{3.2/0.75}(x - 0.0016) \leq y \leq \sqrt{6.2/0.45}(x + 0.0016)\}$, see Figure 4.

Notice, that the “worst” case occurs for values of $\lambda_x$ and $\lambda_y$ in vertices of the set $[0, 0.45; 0.75] \times [3.2; 6.2]$. We call the criteria based on the worst case approach the robust criteria. Ruffio et al. [4] suggest to calculate the maximum over the edge centers according to the so-called star design.

Similarly, if in the second scenario we knew that position errors take values from some bounded intervals we might take the worst case approach not only for the parameters of interest but also for the nuisance parameters, i.e. to minimize either the $D$ or $F$ criterion based on $V_{(x,y)}(\lambda_x, \lambda_y, \Delta x, \Delta y)$ with respect to $(x, y)$ for the maximum value with respect to $(\lambda_x, \lambda_y, \Delta x, \Delta y)$ over the corresponding hypercube. We did not take this approach as Ruffio et al. [4] assume that the intervals for feasible values of $\Delta x$ and $\Delta y$ are theoretically unbounded.

### 4.2 Prior information in terms of probability densities

Very often we may formulate our prior information in terms of some given probability densities. For instance, we may suppose that $\lambda_x$, $\lambda_y$, $(\Delta x, \Delta y)$ are independent with respective densities $p_1(\lambda_x)$, $p_2(\lambda_y)$, $f(\Delta x, \Delta y)$.

We may employ this knowledge by considering the mean value of the $D$ or $F$ criterion based
on $V_{(x,y)}^{\lambda}(\lambda_x, \lambda_y, \Delta x, \Delta y)$ with respect to prior distribution of $\lambda_x$, $\lambda_y$, $(\Delta x, \Delta y)$. In the other words we weight the values of the criterion over the prior distribution:

$$\int \left( D(\hat{\lambda}_x, \hat{\lambda}_y) \right)p_1(\lambda_x)p_2(\lambda_y)f(\Delta x, \Delta y)d(\Delta y)d\lambda_x d\lambda_y,$$

(6)

see (23) in [3], or

$$\int \left( F(\hat{\lambda}_x, \hat{\lambda}_y) \right)p_1(\lambda_x)p_2(\lambda_y)f(\Delta x, \Delta y)d(\Delta y)d\lambda_x d\lambda_y,$$

(7)

see (20) in [3]. The functions $p_1(\lambda_x)$ and $p_2(\lambda_y)$ might be for instance the densities of uniform distributions on the corresponding intervals. The first criterion is thus a weighted average of determinants of an approximate variance-covariance matrix of the estimates $(\hat{\lambda}_x, \hat{\lambda}_y)$ where the “weights” correspond to the prior distribution of $\lambda_x$, $\lambda_y$, $\Delta x$, $\Delta y$. Similarly, the second criterion is a weighted average of the sum of standardized approximate variances of $\hat{\lambda}_x$ and $\hat{\lambda}_y$ with the same weights.

For any $\lambda_x \in [0.45, 0.75]$ and any $\lambda_y \in [3.2, 6.2]$ the optimal positions with respect to the $D$ and $F$ criteria are either on the left or bottom boundary. Therefore, it seems reasonable to look for optimal positions with respect to (6) and (7) on these boundaries. However, it is again not clear how to define the distribution of $\Delta x$ and $\Delta y$ on the boundaries. After a discussion we decided to consider a trimmed normal distribution on the interval $[-c\sigma; c\sigma]$ for one coordinate and a trimmed half normal distribution on the interval $[0, c\sigma]$ for the second one. Such a prior distribution ensures that the “incorrectly” placed sensor will be always inside the heated sample. The numerical optimization shows that the new optimal positions are shifted from the points obtained by least square methods presented in the subsection 3.2 to points that are slightly more distant from the origin $(0,0)$. (A size of shift depends on $c$ and a numerical approximation of integrals in (6) and (7).)
5 CONCLUSION

In our paper we deal with a nonlinear regression describing dependence of temperature measured by one or more sensors on time. Additionally to the parameters of interest, here the thermal conductivities, the regression function contains some other input parameters, e.g. displacements of sensors from their design positions \{(\Delta x(j), \Delta y(j))\}. The distribution of these input parameters is assumed to be known. The aim of statistical inference is to design optimal position(s) of sensor(s).

It is shown that in the scope of mathematical statistics the problem may be viewed in two different ways at least. In the first model the thermal conductivities are estimated under an assumption that the positions of the sensors are exact. In reality this assumption is true only in average but not for all performed experiments. On the contrary the positions vary from one experiment to the other according to an assumed distribution. In the second model the statistician knows that the sensors positions may be not exact and estimates their \(x\) and \(y\) coordinates together with the thermal conductivities.

In our first model it is assumed that one experiment is performed many times and the considered criteria express a variability of the estimates \((\hat{\lambda}_x, \hat{\lambda}_y)\) in many repeated experiments. A crucial point in our example is that the number of measurements is large. Therefore, in the first model variability of thermal conductivities estimates is to a large extent determined by a variability of common random factors, i.e. random displacements of sensors. To minimize the estimates variability, the optimal positions of sensors are chosen in a way that their random displacements affect the estimates of the parameters of interest (thermal conductivities) as little as possible but they also have to be sensitive to the estimated parameters.

In the second model we are interested in a variability of the estimates \((\hat{\lambda}_x, \hat{\lambda}_y)\) in one experiment where the parameters \{(\Delta x(j), \Delta y(j))\} are fixed but unknown and they are estimated together with the parameters of interest \((\lambda_x, \lambda_y)\). The variability of \((\hat{\lambda}_x, \hat{\lambda}_y)\) is caused by measurement errors. How large the variability is, depends not only on the true values of \((\lambda_x, \lambda_y)\), but also on the true values of \{(\Delta x(j), \Delta y(j))\}. In case the true values of the input parameters \{(\Delta x(j), \Delta y(j))\} vary from one experiment to the other experiment according to a known distribution, the value of any criterion based on the asymptotic variance-covariance matrix is a random variable. We may choose a characteristic of its distribution to be a design criterion. In the subsection 4.2 we have chosen a mean value.

In practice the true values of the parameters of interest are also unknown but we have some prior knowledge about their values. Therefore, we may consider a criterion being the mean with respect to the distribution of nuisance parameters as well as to the parameters of interest. In our example with one sensor the variance-covariance matrix depends on the true parameters values relatively slightly and the optimal design is not too far from the optimal design for a model with the parameters fixed at their mean values. Indeed, we could also consider a robust version of the criteria with the respect to \((\lambda_x, \lambda_y)\), as it is described in the subsection 4.1.

As in two considered models a different type of “errors” in estimating thermal conductivities are considered, it is not surprising that the optimal designs are completely different.

This work was supported by the Czech Science Foundation under grants 15-09663S and 15-07299S.
REFERENCES


MULTIDIMENSIONAL STOCHASTIC MATERIAL MODELING AT LARGE DEFORMATIONS CONSIDERING PARAMETER CORRELATIONS

Eduard Penner¹, Ismail Caylak¹ and Rolf Mahnken¹

¹Paderborn University, Chair of Engineering Mechanics
Warburger Str. 100, 33098 Paderborn, Germany
e-mail: {penner, caylak, mahnken}@ltm.upb.de

Keywords: Uncertainty Quantification, Stochastic Material Modeling, Multivariate Polynomial Chaos, Correlated Inputs.

Abstract. In many engineering applications the material behavior, e.g. hyper-elasticity and plasticity, is described by appropriate mathematical models. However, these become uncertain due to different types of uncertainty such as variation in the manufacturing process, measurement errors and missing or incomplete information on material properties. This contribution presents a framework for nonlinear elastic stochastic material model at large deformations. As a key idea uncertainty of the material is described by parameters, which are modeled as stochastic variables.

To this end, 150 specimens for three different rubber materials are experimentally investigated in tensile tests. Based on experimental results 1000 artificial data are generated by aid of an ARMA process [1]. The artificial data are used for parameter identification of an Ogden material model for rubber materials. Furthermore, statistical analysis of material parameters including their correlations is studied. The number of material parameters define the dimension of the stochastic space. Usually, the stochastic material parameters are considered as stochastically independent. However, in our work we consider the dependency including the correlation obtained from experimental data.

The hyperelastic stochastic material parameters are expanded with the multivariate PCE. In this context, the stresses depend on stochastic variables. To determine the corresponding PC coefficients for non-independent stochastic material parameters we use a Cholesky decomposition. As a numerical example we consider the static problem for uniaxial tension of the rectangular plate. This structure is investigated in order to represent the experimental setup conditions.
1 INTRODUCTION

Most materials in the engineering science are heterogeneous, where polycrystallines and composites are typical examples. The heterogeneity leads to uncertainty in the material properties due to the manufacturing process. When assessing the reliability of components, different uncertainties are distinguished. In our contribution we concentrate on the aleatoric uncertainty caused by material parameters, which is characterized by random variables. Therefore, the mechanical system must be described by stochastic partial differential equations (SPDEs). The solution of the mathematical problem can be determined by the stochastic numerical material model, where the system response of the model renders a distribution with statistical mean and variance. A ubiquitous strategy for its solution is the Monte Carlo method \[2, 3\]. An alternative to reduce the computational effort, is the spectral method by Ghanem and Spanos \[4\], which is considered in this paper. In this context the Polynomial Chaos Expansion (PCE) is often apply to represent random variables with a series of random Hermite polynomials. A necessary condition for these polynomials is the orthogonality of them, which implies that random variables do not correlate. In order to consider arbitrary polynomials, the Gram Schmidt algorithm is implemented \[5\].

Research areas for the stochastic modeling are: linear elasticity of solids and mechanics \[4\], plasticity of solids and mechanics \[6–7\], large deformations \[8–9\], fluid flow \[10–11, 12\], flow-structure interactions \[13, 14\] and linear convection problems \[15\]. An open task is the application of the modeling of rubber-like materials, such as natural rubber, which is constitutively represented by an Ogden model and is therefore the focus of this paper. Another challenge is to calculate eigenvalues of the right Cauchy-Green tensor using PC arithmetic.

The uncertainty is considered by random material parameters, which are modeled as stochastic variables. From experimental data the distribution of the random material parameters are generated by parameter identification of each experimental result. It is well known that material parameters of the Ogden model correlate with each other, which leads to correlated random variables. However, the PCE and the determination of PC coefficients require uncorrelated random variables. Therefore, in this work, correlations are determined from statistics of material parameters and are employed for the transformation into uncorrelated random variables.

The organization of this paper is as follows: Section 2 describes the basics of the multivariate PCE. In Section 3 constitutive equations of hyperelasticity are summarized. Finally, Section 4 presents a numerical example, where stochastic system results are compared with results from experimental investigations.

2 MULTIVARIATE POLYNOMIAL CHAOS EXPANSION

In this work uncertainty is modeled by stochastic random variables. In this context the probability space is denoted by \((\Omega, \Sigma, \mathbb{P})\), where \(\Omega\) is the set of elementary events, \(\Sigma\) is the \(\sigma\)-algebra and \(\mathbb{P}\) is the probability measure. Let \(\omega\) be an element of \(\Omega\). One possibility is to expand random variables with the multivariate PCE, see \[16\]. This involves a basis of known random functions with deterministic coefficients, which is of course an advantage compared e.g. with the Karhunen-Loeve Expansion. Then, the PCE of an arbitrary random variable \(X(\omega)\) can be formulated as

\[
X(\omega) = \sum_{\alpha} \hat{X}_{\alpha} \Psi_{\alpha}(\theta(\omega)),
\]

(1)
where \( \hat{X}_\alpha \) are multivariate deterministic PC coefficients and \( \alpha \) is a multi-index. The PC basis functions \( \Psi_{\alpha}(\theta(\omega)) \) are described by multivariate polynomials with uncorrelated standard distributed random variables \( \theta \), which can be expressed as a tensor product

\[
\Psi_{\alpha}(\theta) = \prod_{i=1}^{m} \psi_{\alpha_i}(\theta_i).
\]

In Eq. (2) \( \psi_{\alpha_i} \) are univariate polynomials, where \( \alpha_i \in \mathbb{N}_0 \) is the degree of the polynomial and \( m \in \mathbb{N}_0 \) specifies the stochastic dimension and/or the number of uncorrelated variables \( \theta_i \). As an example Table 1 summarizes multi-index polynomials for a stochastic dimension with \( m = 4 \).

<table>
<thead>
<tr>
<th>( \Psi_{[0,0,0,0]} )</th>
<th>( \Psi_{[2,0,0,0]} )</th>
<th>( \Psi_{[0,1,1,0]} )</th>
<th>( \Psi_{[0,0,0,2]} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \psi_0^4 )</td>
<td>( \psi_2(\theta_1) )</td>
<td>( \psi_1(\theta_2) )</td>
<td>( \psi_2(\theta_4) )</td>
</tr>
<tr>
<td>( \psi_1(\theta_1) )</td>
<td>( \psi_1(\theta_1) \psi_1(\theta_2) )</td>
<td>( \psi_1(\theta_1) \psi_1(\theta_3) )</td>
<td>( \psi_1(\theta_1) \psi_1(\theta_4) )</td>
</tr>
<tr>
<td>( \psi_1(\theta_2) )</td>
<td>( \psi_1(\theta_1) \psi_1(\theta_3) )</td>
<td>( \psi_1(\theta_1) \psi_1(\theta_4) )</td>
<td>( \psi_1(\theta_1) \psi_1(\theta_4) )</td>
</tr>
<tr>
<td>( \psi_1(\theta_3) )</td>
<td>( \psi_1(\theta_1) \psi_1(\theta_4) )</td>
<td>( \psi_1(\theta_1) \psi_1(\theta_4) )</td>
<td>( \psi_1(\theta_1) \psi_1(\theta_4) )</td>
</tr>
<tr>
<td>( \psi_1(\theta_4) )</td>
<td>( \psi_2(\theta_2) )</td>
<td>( \psi_2(\theta_3) )</td>
<td>( \psi_2(\theta_4) )</td>
</tr>
</tbody>
</table>

Table 1: Multi-index polynomials for the stochastic dimension \( m = 4 \)

A more practical choice instead of the multivariate PCE in Eq. (1) is the single-index representation

\[
X \approx \sum_{k=0}^{P} \hat{X}_k \Psi_k(\theta_1, \theta_2, \ldots, \theta_m).
\]

The multi-index \( \alpha \) is now reformulated as a single-index \( k \) such that the maximum number of PC terms is calculated as

\[
P + 1 = \frac{(m + p)!}{m!p!},
\]

where \( p = \sum_{i=1}^{m} \alpha_i \) describes the polynomial order. Using Eq. (4) with \( m = 4 \) and \( p = 2 \), a single-index representation of Eq. (2) is given in Table 2 where the PC basis \( \Psi_k, k = 0, \ldots, P \), are described by univariate polynomials.

### 2.1 Correlated input represented by uncorrelated variables

As mentioned above, standard distributed random variables \( \theta_i \) of Eq. (2) must be uncorrelated. Therefore, the aim of this section is the transformation of correlated into uncorrelated random variables, using the Cholesky decomposition for an arbitrary matrix \( \mathbf{A} = \mathbf{L} \mathbf{L}^T \). The following vectors are defined

\[
X(\theta) = [X_1(\theta), X_2(\theta), \ldots, X_m(\theta)]^T, \quad \bar{X} = [E(X_1), E(X_2), \ldots, E(X_m)]^T,
\]

where \( X_i, i = 1, \ldots, m \) are correlated random variables and \( E(X_i) \) are corresponding expected values. The covariance matrix of \( X \) can be calculated with

\[
\Sigma_{X} = \text{cov}(X) = E \left( (X - \bar{X})(X - \bar{X})^T \right) = \mathbf{L} \mathbf{L}^T,
\]
where $\mathbf{L}$ is the lower triangular matrix of the Cholesky decomposition. For standardized uncorrelated random variables, the covariance matrix is trivial and equal to the identity matrix $\mathbf{I}$. Therefore, Eq. (6) is reformulated as

$$
\mathbf{L}^{-1} \sum_{\mathbf{X}} \mathbf{X} \mathbf{L}^{-T} = \mathbb{E} \left( \frac{\mathbf{L}^{-1} (\mathbf{X} - \bar{\mathbf{X}}) (\mathbf{X} - \bar{\mathbf{X}})^T \mathbf{L}^{-T}}{\mathbf{Y}^T} \right) = \Sigma_{\mathbf{Y} \mathbf{Y}},
$$

(7)

where $\mathbf{Y}(\theta) = [Y_1(\theta_1), Y_2(\theta_2), \ldots, Y_m(\theta_m)]^T$ in Eq. (7) in contrast to $\mathbf{X}$ in Eq. (6) contains standardized uncorrelated zero-mean random variables $Y_i(\theta_i)$. From Eq. (7) we conclude the relation

$$
\mathbf{X} - \bar{\mathbf{X}} = \mathbf{L} \mathbf{Y},
$$

(8)

where random variables $\mathbf{X}$ and $\mathbf{Y}$ are expanded with the PCE according to Eq. (3)

$$
\sum_{k=0}^{P} \hat{X}_k \Psi_k(\theta) - \bar{\mathbf{X}} = \sum_{k=1}^{P} \hat{X}_k \Psi_k(\theta) = \sum_{k=1}^{P} L \hat{Y}_k \Psi_k(\theta).
$$

(9)

In Eq. (9) we exploit $\hat{X}_0 = \bar{\mathbf{X}}$ and $\hat{Y}_0 = 0$, due to the fact that $\mathbf{Y}$ contains zero-mean random variables. Comparing coefficients of Eq. (9) renders

$$
\hat{X}_k = L \hat{Y}_k, \quad \forall \ k \in \{1, \ldots, P\},
$$

(10)

where the PC coefficients $\hat{X}_k$ are formulated as PC coefficients $\hat{Y}_k$ of uncorrelated random variables. Finally, the polynomials $\Psi_k$ in Eq. (3) have to be specified. In the following sections two different types of polynomials $\Psi_k$ will be discussed.

<table>
<thead>
<tr>
<th>Order $p$</th>
<th>Single-index $k$</th>
<th>$\Psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>$\Psi_0 = \Psi_{[0,0,0]}$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>$\Psi_1 = \Psi_{[1,0,0]}$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$\Psi_2 = \Psi_{[0,1,0]}$</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$\Psi_3 = \Psi_{[0,0,1]}$</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>$\Psi_4 = \Psi_{[0,0,0]}$</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>$\Psi_5 = \Psi_{[2,0,0]}$</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>$\Psi_6 = \Psi_{[1,1,0]}$</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>$\Psi_7 = \Psi_{[1,0,1]}$</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>$\Psi_8 = \Psi_{[1,0,0]}$</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>$\Psi_9 = \Psi_{[0,2,0]}$</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>$\Psi_{10} = \Psi_{[0,1,1]}$</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>$\Psi_{11} = \Psi_{[0,1,0]}$</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>$\Psi_{12} = \Psi_{[0,0,2]}$</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>$\Psi_{13} = \Psi_{[0,0,1]}$</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>$\Psi_{14} = \Psi_{[0,0,0]}$</td>
</tr>
</tbody>
</table>

Table 2: Single-index representation of the multivariate polynomials for a stochastic dimension $m = 4$ and polynomial order $p = 2$
2.2 Multivariate Hermite polynomials

In general, a normal distribution of experimental data is assumed. Therefore, we use Hermite polynomials, which are based on standardized uncorrelated normally distributed random variables \( \theta_i \). The first five univariate Hermite polynomials have the form

\[
\begin{align*}
\psi_0 &= 1; \\
\psi_1 &= \theta; \\
\psi_2 &= \theta^2 - 1; \\
\psi_3 &= \theta^3 - 3\theta; \\
\psi_4 &= \theta^4 - 6\theta^2 + 3. \\
\end{align*}
\] (11)

Due to the orthogonality property of Hermite polynomials it follows, see [17]

\[
E(\Psi_0) = 1, \quad E(\Psi_{\alpha} \Psi_{\beta}) = \alpha! \delta_{\alpha\beta}, \quad E(\Psi_{\alpha}^2) = \alpha!, \quad E(\Psi_{\alpha}) = 0 \quad \forall \ \alpha \neq 0,
\] (12)

where \( E \) and \( \alpha! \) denote the expected value and the factorial with

\[
\alpha! = \prod_{i=1}^{m} \alpha_i!.
\] (13)

These Hermite polynomials are used to calculate PC coefficients \( \hat{Y}_k \) of Eq. (10).

2.3 Input parameter based polynomials

An alternative to Hermite polynomials are input parameter based (IPB) polynomials. In contrast to Hermite polynomials, these are based on standardized uncorrelated distributed random variables and we set \( \theta_i = Y_i \). According to Eq. (2) this yields

\[
\Psi_{\alpha}(Y) = \prod_{i=1}^{m} \psi_{\alpha_i}(Y_i),
\] (14)

where \( \psi_{\alpha_i} \) depends on \( Y_i \). An orthogonal set of univariate polynomials can be computed with the Gram-Schmidt algorithm [18]

\[
\psi_0 = 1, \quad \psi_l(Y_i) = e_l(Y_i) - \sum_{k=0}^{l-1} \frac{E(e_l(Y_i) \cdot \psi_k(Y_i))}{E(\psi_k(Y_i) \cdot \psi_k(Y_i))}, \quad l = 1, 3, \ldots, p.
\] (15)

In Eq. (15) polynomials \( e_l(Y_i) = (Y_i)^l \) are of degree \( l \). The advantage of these polynomials is that the PC coefficients \( \hat{Y}_{ik} \) of the \( i \)-th entry of \( \hat{Y}_k \) are known as

\[
\hat{Y}_{ik} = \begin{cases} 
\delta_{ik} & \text{if } k = 1, \ldots, m \\
0 & \text{else} \end{cases}
\] (16)

and must not be calculated by aid of Hermite polynomials. Then, Eq. (3) leads to

\[
X_i \approx \sum_{k=0}^{P} \hat{X}_{ik} \Psi_k(Y_1, Y_2, \ldots, Y_m),
\] (17)

where \( X_i \) depends on IPB polynomials \( \Psi_k(Y_1, Y_2, \ldots, Y_m) \). The corresponding PC coefficients \( \hat{X}_{ik} \) can be be formulated using Eq. (10) and Eq. (16)

\[
\hat{X}_{i0} = \bar{X}_i, \quad \hat{X}_{ik} = L_{ik}, \quad \hat{X}_{ik} = 0 \quad \forall \ k \in \{(m+1), \ldots, P\},
\] (18)

where \( L_{ik} \) describes the \((i, k)\) entry of the lower triangular matrix, which is calculated from the Cholesky decomposition of \( \Sigma_x \) in Eq. (6).
3 A stochastic variation of Ogden’s material model

The derivation of the constitutive relations of the stochastic Ogden model is analogous to the deterministic problem and is therefore not presented in detail. For this reason, we will refer to relevant literature [19, 20, 21, 22, 23]. Thus, for the stochastic strain energy function we obtain

\[ U(\tilde{\lambda}_i, J, \omega) = \sum_{p=1}^{2} \sum_{i=1}^{3} \mu_p(\omega) \left( \tilde{\lambda}_i(\omega)\alpha_p(\omega) - 1 \right) + \sum_{p=1}^{N} K \left( J(\omega) - 1 \right)^{2p}, \]

where \( \tilde{\lambda}_i \) are the eigenvalues of the deviatoric right Cauchy-Green tensor \( C(\omega) \), \( J \) is the Jacobi determinant of the deformation gradient \( F \), \( K \) is the bulk modulus and \( \alpha_p, \mu_p, p = 1, 2 \) are additional material parameters.

According to [24] expressions of the second Piola-Kirchhoff stress tensor \( S \) is considered. As mentioned in the introduction, one challenge of the stochastic modeling of Ogden’s material model is the calculation of eigenvalues \( \lambda_i \) of the right Cauchy-Green tensor

\[ C(\omega) = \sum_{i=1}^{3} \lambda_i^2(\omega) N_i(\omega) \otimes N_i(\omega) = \sum_{i=1}^{3} \lambda_i^2(\omega) M_i(\omega), \]

where \( N_i(\omega) \) and \( M_i(\omega) = N_i(\omega) \otimes N_i(\omega) \) are eigenvectors and eigenvector basis, respectively. The eigenvalues and eigenvectors of deterministic problems can be calculated easily by standard numerical methods. For stochastic problems the following eigenvector problem must be solved

\[ C(\omega) N_i(\omega) - \lambda_i^2(\omega) N_i(\omega) = 0, \quad i = 1, 2, 3, \]

\[ N_i^T(\omega) N_i(\omega) = 1, \quad i = 1, 2, 3. \]

For plane stress problems Eq. (23) is simplified as described in [25]. The system of equations can be solved by using PC-arithmetic, such as product and addition in combination with an optimization algorithms. Basic operations of the PC arithmetic can be found in [26].

4 STATISTICAL ANALYSIS OF OGDEN’S PARAMETERS AND THE RESPONSE

Tensile tests are used to identify material parameters of natural rubber. Based on 150 experimental force-displacement curves, 1000 artificial data are generated as described in [1].

4.1 Distribution and Cholesky decomposition of material parameters

For all artificial data sets parameter identifications are performed such that a distribution of each parameter is obtained. In Figure 1 the cumulative density function (CDF) of Ogden’s material parameters \( \mu_1(\omega), \alpha_1(\omega), \mu_2(\omega) \) and \( \alpha_2(\omega) \) are illustrated. The solid line represents
the CDF of the random parameter, where as the dashed line shows the Gaussian distribution. Parameters $\alpha_1$ and $\alpha_2$ are dimensionless quantities and $\mu_1$ and $\mu_1$ are given in MPa. The bulk modulus is fixed to $K = 1000$ MPa for all parameter identifications. All random parameters are stored in the vector

$$\kappa(\omega) = [\alpha_1(\omega), \mu_1(\omega), \alpha_2(\omega), \mu_2(\omega)]^T.$$  \hspace{1cm} (24)

With $\kappa(\omega)$ from Eq. (24), the expected values and the lower triangular matrix for the covariance matrix in Eq. (6) can be calculated as

$$\bar{\kappa} = \begin{bmatrix} \bar{\alpha}_1 \\ \bar{\mu}_1 \\ \bar{\alpha}_2 \\ \bar{\mu}_2 \end{bmatrix} = \begin{bmatrix} -6.197 \\ -0.077 \\ -0.058 \\ -5.891 \end{bmatrix}, \quad L_{\kappa\kappa} = \begin{bmatrix} 2.0813 & 0 & 0 & 0 \\ -0.0046 & 0.0375 & 0 & 0 \\ 0.0091 & 0.0079 & 0.0219 & 0 \\ 0.3635 & 0.6167 & -0.0119 & 0.1695 \end{bmatrix} \cdot 10^{-1}. \hspace{1cm} (25)$$

![Figure 1](image_url)

Figure 1: Cumulative density functions for material parameters $\alpha_1$, $\mu_1$, $\alpha_2$ and $\mu_1$ from 1000 deterministic parameter identifications.

### 4.2 PC coefficients of material parameters

In order to determine the PC coefficients, the correlated inputs $\kappa$ are first transformed into uncorrelated random variables $Y$, as described in Eq. (7). When Hermite polynomials are used, PC coefficients of $Y$ are determined using the collocation method [27]. Eq. (10) can be used to calculate the PC coefficients of $\kappa$. Alternatively, the IPB polynomials can be used as described in Section 2.3. In this case, PC coefficients correspond to the expectation values $\bar{\kappa}$ and the lower
triangular matrix $L_{\kappa \kappa}$ with

$$
\hat{\kappa} = \begin{bmatrix}
\hat{\alpha}_{10} & \hat{\alpha}_{11} & 0 & 0 \\
\hat{\mu}_{10} & \hat{\mu}_{11} & \hat{\mu}_{12} & 0 \\
\hat{\alpha}_{20} & \hat{\alpha}_{21} & \hat{\alpha}_{22} & \hat{\alpha}_{23} \\
\hat{\mu}_{20} & \hat{\mu}_{21} & \hat{\mu}_{22} & \hat{\mu}_{23} & \hat{\mu}_{24}
\end{bmatrix} = \begin{bmatrix}
\kappa \\
L_{\kappa \kappa}
\end{bmatrix}, \text{ only for IPB polynomials, (26)}
$$

see also relations in Eq. (18). PC coefficients for both types of polynomials in Section 2.2 and Section 2.3 are summarized in Table 3.

<table>
<thead>
<tr>
<th>$(p, P)$</th>
<th>$k$</th>
<th>$\Psi_k$</th>
<th>$\hat{\alpha}_{1k}$</th>
<th>$\hat{\mu}_{2k}$</th>
<th>$\hat{\alpha}_{2k}$</th>
<th>$\hat{\mu}_{2k}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP</td>
<td>1</td>
<td>1</td>
<td>-6.19857</td>
<td>-0.07724</td>
<td>-0.05881</td>
<td>-5.89198</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$\theta_1$</td>
<td>0.20225</td>
<td>-0.00031</td>
<td>0.00096</td>
<td>0.03813</td>
</tr>
<tr>
<td>(1, 5)</td>
<td>3</td>
<td>$\theta_2$</td>
<td>-0.00844</td>
<td>0.00368</td>
<td>0.00087</td>
<td>0.05811</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>$\theta_3$</td>
<td>-0.00028</td>
<td>-0.00019</td>
<td>0.00211</td>
<td>-0.00478</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>$\theta_4$</td>
<td>0.01105</td>
<td>0.00023</td>
<td>0.00004</td>
<td>0.01894</td>
</tr>
<tr>
<td>IPB</td>
<td>1</td>
<td>1</td>
<td>-6.19752</td>
<td>-0.07723</td>
<td>-0.05881</td>
<td>-5.89157</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$Y_1$</td>
<td>0.20814</td>
<td>-0.00047</td>
<td>-0.00091</td>
<td>0.03635</td>
</tr>
<tr>
<td>(1, 5)</td>
<td>3</td>
<td>$Y_2$</td>
<td>0</td>
<td>0.00375</td>
<td>-0.00079</td>
<td>0.06167</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>$Y_3$</td>
<td>0</td>
<td>0</td>
<td>0.00219</td>
<td>-0.00119</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>$Y_4$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01696</td>
</tr>
</tbody>
</table>

Table 3: PC coefficients of material parameters for Hermite and IPB polynomials

4.3 Comparison of numerical simulation with experiments

As a numerical example we consider the static problem for uniaxial tension of a rectangular plate, with the same condition as in the experimental investigations. We apply the stochastic material model for both types of polynomials, Hermite and IPB polynomials. The stochastic model can then be performed using the specific PC coefficients. For comparing the numerical and experimental results, a 95% confidence interval of the 1.PK-stretch curves is considered as can be seen in Figure 3. In addition, Figure 4 shows the CDFs of the first Piola-Kirchhoff at displacements of 25 mm and 85 mm.

5 CONCLUSION

In this paper the stochastic modeling at large deformations for the Ogden material model is presented. A major focus is the consideration of correlated material parameters in the PCE. In this context, the basic idea are to represent correlated by uncorrelated random variables. This ensures that each variable can be developed independently and the corresponding PC coefficients can be calculated. Two types of polynomials, Hermite and IPB polynomials, are presented. In contrast to Hermite polynomials, where PC coefficients have to be determined numerically, IPB polynomials are calculated based on input variables, which render exact PC coefficients. In the numerical example, results for both methods are compared with experimental results. We conclude that a good agreement is observed using Hermite and IPB polynomials based on correlated random material parameters.
Figure 2: Density functions for material parameters $\alpha_1$, $\mu_1$, $\alpha_2$ and $\mu_2$ from experiments and modeling with both Hermite and IPB polynomials.

Figure 3: 95% confidence interval of the 1.PK-stretch curves.
REFERENCES


A POSSIBILISTIC APPROACH FOR LINEAR ISOTROPIC ELASTICITY USING THE FUZZY FINITE ELEMENT METHOD

Alex Dridger¹, Ismail Caylak¹, and Rolf Mahnken¹

Chair of Engineering Mechanics (LTM)
University of Paderborn
Paderborn, Germany
{dridger,caylak,mahnken}@ltm.upb.de

Keywords: Fuzzy set, $\alpha$-level discretization, possibility measure, probability-possibility transformation, imprecise probability

Abstract. The uncertainty quantification has become an inalienable factor in many physical and engineering applications for determination of reliable results with the finite element method (FEM). In this work we investigate the uncertainties characterized by imprecise probabilities [12, 13]. To this end, a membership function is interpreted as a possibility distribution and a possibility distribution as a family of probability distributions. In the basic preliminaries we summarize the fundamentals of the fuzzy set theory as well as the possibility theory including a probability-possibility transformation [11]. Moreover, a linear elastic body with two fuzzy input material parameters is studied. The main objectives are twofold: Firstly, determination of two unknown fuzzy parameters and secondly, numerical computation of the system response under consideration of the interaction between fuzzy parameters using the possibilistic evaluation of the fuzzy finite element method (FFEM). The $\alpha$-level discretization technique [2] is applied in order to reduce the fuzzy arithmetic based FEM to an interval arithmetic based FEM. Finally, our method is applied in a numerical example for a plate with a ring hole.
1 INTRODUCTION

In modern engineering sciences the uncertainty quantification as well as its suitable application in simulations with the finite element method (FEM) becomes essential in order to obtain realistic results. In this context, the uncertainty arises from linguistic, informal or statistical properties and may be attached to material parameters, geometry parameters and/or loading parameters. Different causes of the uncertainty require different characterization techniques. Thus, statistical (or stochastic) based uncertainty is described by randomness [2], i.e. random variables and/or random fields. This kind of uncertainty is referred to as the aleatoric uncertainty, whereas the other kinds of uncertainties (especially the one based on incomplete information) are assigned to the epistemic type [4]. To this end, the deterministic finite element method is expanded to the stochastic FEM (SFEM) [5, 14, 15] in order to handle the aleatoric uncertainty. Accordingly, there exist the interval FEM and the fuzzy FEM (FFEM) [2, 16] in order to comprise the epistemic uncertainty in the simulation. Naturally, in the engineering sciences a pure aleatoric model is not able to exist, since the determination of an exact probability distribution can only be achieved in the infinite trial. For this reason, every uncertain model is (partly) of epistemic nature. The combination of both types of uncertainty is referred to as imprecise probability [12, 13]. For this, the probability-box FEM (P-box FEM) [17] and/or fuzzy stochastic FEM (FSFEM) [2, 18] are applied in order to consider imprecise probabilities in the FEM.

Furthermore, the possibility theory first introduced by Zadeh [19] and expanded by Dubois and Prade [6, 13] was formulated in order to deal with imprecise probabilities. To this end, membership functions of the fuzzy set theory [2, 6] may be interpreted as possibility distributions where one of them comprises a family of probability distributions [11]. A main application area of the possibility theory is the field of computer science or the cognitive psychology as can be seen e.g. in [20].

This work describes the possibilistic approach in the finite element method based on incomplete information for material parameters referred to as design variables. The incompleteness is justified in a small number of experimental observations. Thus, the underlying probability distribution cannot be specified in contrast to a family of possible probability distributions. That is, we use the possibilistic interpretation of the fuzzy finite element method in order to describe imprecise probabilities caused by fuzzy material parameters.

An outline of this work is as follows: Section 2 summarizes the basic equations of the fuzzy set theory and the related possibility theory including the possibility distribution and possibility measure, respectively. Furthermore, the continuous variational formulation for a linear elastic body with fuzzy material parameters are introduced. In Section 3 the spatial discretization is provided in order to obtain the linear elastic fuzzy equation system. Finally, Section 4 consolidates the previous explanations in the representative example for a two parameter model of isotropic linear elasticity under consideration of the interaction of fuzzy material parameters.

2 BASIC PRELIMINARIES

The aim of this work is a possibilistic evaluation of the (linear elastic) fuzzy finite element method based on fuzzy material parameters. In this context, the fuzziness arises from a small number of experimental observations. This section recaps the essential definitions.
2.1 Fuzzy sets

Based on \[1, 2, 3\] we define for a design space \(S \subset \mathbb{R}^{n_s}\) with \(n_s\) as the number of input (design) variables \(s = (s_1, s_2, \ldots, s_{n_s})\)

\[
\hat{S} = \{(\hat{s}, \mu_S(\hat{s})) \mid \hat{s} \in \mathbb{R}^{n_s}, \mu_S(\hat{s}) = \min[\mu_{S_i}(s_i)]\},
\]

as the \(n_s\) dimensional fuzzy set \(\hat{S}\) with the \(n_s\) dimensional membership function \(\mu_S\), where

\[
\hat{S}_i = \{(s_i, \mu_{S_i}(s_i)) \mid s_i \in \mathbb{R}, \mu_{S_i} \in [0, 1]\}
\]
is the one dimensional fuzzy set for \(S_i \subset \mathbb{R}\) with the one dimensional membership function \(\mu_{S_i}\).

2.2 Fuzzy arithmetic

In order to perform mathematical operations with fuzzy sets the \(\alpha\)-level discretization technique is used \[2, 11\]. For this, the \(\alpha_k\)-cut of a fuzzy set \(\hat{S}\) is defined as the set

\[
S_{\alpha_k} = \{\hat{s} \mid \mu_S(\hat{s}) \geq \alpha_k, \hat{s} \in \mathbb{R}^{n_s}\} = S_{1,\alpha_k} \times S_{2,\alpha_k} \times \cdots \times S_{n_s,\alpha_k} \subset \mathbb{R}^{n_s},
\]

where \(S_{i,\alpha_k} := [s_{i,\alpha_k}^L, s_{i,\alpha_k}^R]\) for \(k \in \mathbb{N}_0\) and \(i = 1, \ldots, n_s\). This approach allows the usage of an ordinary interval arithmetic technique \[7, 21\] in order to perform mathematical operations with fuzzy sets. The higher the number of cuts, the more accurate is result, however, it becomes numerically more demanding. Figure 1 illustrates the \(\alpha\)-level discretization on a one dimensional trapezoidal fuzzy set \(\hat{S}_i\).

![Figure 1: \(\alpha\)-cuts of a fuzzy trapezoidal number \(\hat{S}_i\)](image)

Note that, the inclusion of the interaction (which is defined as being the mutual dependency of fuzzy variables \[2\]) is important in order to obtain most realistic results with the interval arithmetic. Section 4 illustrates the meaning of interaction in the representative example.

2.3 Possibility vs. probability

There are several interpretations of fuzziness (see e.g. \[22\]). Initially, fuzzy sets were introduced to model vague linguistic knowledge. There, the membership degree represents a degree of similarity of preference, acceptability, suitability or (more general) a degree of truth. A further interpretation is to consider a fuzzy set as a set of probability measure \[6, 13, 20\]. To this end, membership functions are interpreted as possibility distributions and a possibility distribution as a family of probability distributions. Thus, in the engineering sciences, the terminology of possibility theory allows the examination of imprecise probabilities, whereas the ordinary fuzzy set theory is used for the epistemic uncertainty.
2.3.1 Possibility distribution as a family of probability distributions

According to [11] a possibility distribution is a function
\[ \pi : S \rightarrow [0, 1] \subseteq \mathbb{R}, \]  
with \( S \) as the design space of design variables. Following [11] we formulate the equations
\begin{enumerate}
\item \( \mathcal{P}(\pi) = \{ p \in \mathcal{P}, \forall A \subseteq \mathbb{R}^n, N(A) \leq P(A) \leq \Pi(A) \} \), with
\item \( \Pi(A) = \sup_{\pi \in A} \pi(s) \) as the possibility measure and
\item \( N(A) = 1 - \Pi(A) \) as the necessity measure,
\end{enumerate}
where \( \mathcal{P} \) is the family of all possible probability distributions and \( P \) is the probability measure associated with the probability distribution \( p \). From Eq. (4) and Eq. (5) we observe, that the possibility distribution \( \pi \) contains all probability distributions \( p \) that are upper bounded by the possibility measure \( \Pi \), whereas the necessity measure \( N \) may be interpreted as the lower bound. For more information about the possibility measure and the necessity measure see e.g. [6, 23]. In this work, we restrict our examination on the possibility. We use this term as an abbreviation for the often (synonymically) used terms possibility distribution, possibility measure and/or possibility degree.

2.3.2 Probability-possibility transformation for limited experimental data

As already indicated our attempt is the inclusion of sparse experimental data in a simulation as well as the interpretation of the corresponding system response. That is, we have to consider especially non-available information in order to obtain most realistic results based on sparse information. The determination of the exact probability distribution is not possible in case of sparse experimental data. Nevertheless, we may embrace the family of possible probability distributions. The idea is to transform the (possible) marginal probability density functions in two possibility distributions. A subsequent combination to one possibility distribution encoding the family of all possible normal density functions which may arise considering the sparse experimental information.

We describe the procedure considering the schematic illustration in Fig. 2. Let \( f_M \) be a (normal) probability density function of the variable \( s \) (as can be seen in Fig. 2a) with corresponding expected value \( \mu_M \) and standard deviation \( \sigma_M \) calculated from \( n \) sparse experimental data. The corresponding possibility distribution \( \pi \) is shown in Fig. 2c.
data of $s$. Although it is unlikely that the density function $f_M$ describes the real density function of the design variable $s$, nevertheless, it is a possible one. With a specific confidence level (of e.g. 95%) one obtains the confidence intervals $\mu = [\mu_L, \mu_R]$ for possible expected values with $\mu_L \leq \mu_M \leq \mu_R$ and $\sigma = [\sigma_L, \sigma_R]$ for possible standard deviations with $\sigma_L \leq \sigma_M \leq \sigma_R$ considering the $n$ experimental data (see [8, 9] for more information about confidence intervals). Thus, we may establish the marginal (normal) density function $f_L$ with expected value $\mu_L$ and standard deviation $\sigma_R$ and the marginal (normal) density function $f_R$ with expected value $\mu_R$ and standard deviation $\sigma_R$ as illustrated in Fig. 2.a. Note that, both marginal density functions have the largest standard deviation $\sigma_R$. Consider for exemplification $f := f_L$ as the standard normal density function, i.e.

$$f = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} s^2\right),$$

with $\mu_L = 0$ and $\sigma_L = 1$. The probability-possibility transformation

$$T: \left\{ \begin{array}{l}
D \rightarrow P \\
n \rightarrow \pi^n
\end{array} \right. $$

assigns to the density function $f \in D$ the maximum specific possibility distribution $\pi^f$, where $D$ is a set of density functions and $P$ is a set of possibility distributions. In order to obtain the maximum specific possibility distribution with the relation in Eq. (7) one has to determine the confidence intervals [11]

$$I_{\beta_k} = \{s \mid s \in f^{-1}[d], \forall d \in [c_k, +\infty)\},$$

with

$$\beta_k = \int_{\{s \in f(s) \geq c_k\}} f(s)ds,$$

for $k = 1, \ldots, n_\alpha$ and $c_k > 0$ where $n_\alpha$ is the number of alpha cuts. Then, considering the relation (derived in [11])

$$S_{\alpha_k} = [s_L^{\alpha_k}, s_R^{\alpha_k}] := S_{1-\beta_k} = I_{\beta_k}$$

one is able to determine the $\alpha$-cuts of the corresponding maximum specific possibility distribution. Figure 3 illustrates the transformation of the standard normal density function $f$ (shown in Fig. 3.a) to the maximum specific possibility distribution $\pi^f$ (shown in Fig. 3.b).

Figure 3: Probability-possibility transformation: a) Density function of a standard normal distribution, b) corresponding maximum specific possibility distribution
Furthermore, Fig. 2.b. illustrates schematically the transformation of marginal density functions $f_L$ and $f_R$ to the corresponding possibility distributions $\pi^f_L$ and $\pi^f_R$ using Eq. (7) - Eq. (10). Subsequently, we formulate the (common) maximum specific possibility distribution $\pi$ as

$$\pi(s) := \begin{cases} 
\pi^f_L(s) & \text{for } s < (\pi^f_L)^{-1}(1) \\
\pi^f_R(s) & \text{for } s > (\pi^f_R)^{-1}(1) \\
1 & \text{for } s \in [(\pi^f_L)^{-1}(1), (\pi^f_R)^{-1}(1)]
\end{cases} \tag{11}$$

Thus, the possibility distribution $\pi$ in Eq. (11) encodes all possible probability density functions which may arise considering the $n$ sparse experimental data. Eventually, $\pi$ constitutes the input function of the design variable $s$ for the finite element method as can be seen in the representative example in Section 4.

### 2.4 Governing equations for an elastic body

The considered linear elastic material behavior is characterized by fuzzy material parameters. Building on deterministic governing equations (see e.g. [10]) on the region $\Omega$ occupied by the elastic body, the following variational formulation is constituted:

$$\int_{\Omega} \delta \hat{\varepsilon} : \hat{\boldsymbol{C}} : \hat{\varepsilon} \ d\Omega = \int_{\partial \Omega} \hat{\mathbf{t}} \cdot \delta \hat{\mathbf{u}} \ d\partial\Omega. \tag{12}$$

In Eq. (12) we have $\hat{\boldsymbol{C}}$ as the fuzzy elasticity tensor, $\delta \hat{\mathbf{u}}$ as the fuzzy virtual displacement, $\hat{\varepsilon}$ as the fuzzy strain tensor and $\delta \hat{\varepsilon}$ as the fuzzy virtual strain tensor with corresponding Dirichlet and Neumann boundary conditions

$$\mathbf{u} = \mathbf{u} \text{ on } \partial \Omega_u, \quad \mathbf{t} = \mathbf{\sigma} \cdot \mathbf{n} \text{ on } \partial \Omega_\sigma \tag{13}$$

with non-intersecting boundaries $\partial \Omega = \partial \Omega_\sigma \cup \partial \Omega_u$ and $\emptyset = \partial \Omega_\sigma \cap \partial \Omega_u$.

### 3 NUMERICAL IMPLEMENTATION

#### 3.1 Spatial discretization with fuzzy material parameters

We consider a linear elastic isotropic material described by fuzzy input material parameters Young’s modulus $\hat{E}$ and Poisson’s ratio $\hat{\nu}$. Using the displacement and strain approximation

$$\hat{\mathbf{u}} \approx \sum_{i=1}^{n_k} N_i \hat{\mathbf{u}}^i = \mathbf{N} \hat{\mathbf{u}}^* \quad \text{and} \quad \hat{\varepsilon} \approx \mathbf{B} \hat{\mathbf{u}}^* \tag{14}$$

with $\mathbf{B}$ as the associated derivative matrix of shape functions $N_i$, $\hat{\mathbf{u}}^*$ as the nodal displacement vector and $n_k$ as the number of nodes per element, the following fuzzy equation system is derived:

$$\hat{\mathbf{K}} \hat{\mathbf{u}}^* = \mathbf{A} \sum_{e=1}^{n_e} \int_{\Omega^e} \hat{\mathbf{B}}^T \hat{\boldsymbol{C}} \mathbf{B} \ d\Omega^e = \mathbf{f} = \mathbf{A} \sum_{e=1}^{n_e} \int_{\partial \Omega^e} \hat{\mathbf{t}} \cdot \mathbf{n} \ d\partial \Omega^e. \tag{15}$$

In Eq. (15) $\mathbf{A}$ represents the assembly operator, $\partial \Omega^e$ is the domain of one finite element and $n_e$ is the number of elements. Furthermore, the fuzzy elasticity matrix is given as

$$\hat{\boldsymbol{C}} = \hat{\boldsymbol{C}}_{\text{dev}} + \hat{\boldsymbol{C}}_{\text{vol}} = 2\hat{G}I_{\text{dev}} C + \hat{K}_{\text{im}} m^T \tag{16}$$
with
\[ \hat{G} = \frac{\hat{E}}{2(1 + \hat{\nu})} \quad \text{fuzzy shear modulus,} \]
\[ \hat{K} = \frac{\hat{E}}{3(1 - 2\hat{\nu})} \quad \text{fuzzy bulk modulus} \]
and
\[ I_{\text{dev}}^C = I_C - \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}, \quad m = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}. \quad (18) \]

The shear modulus and the bulk modulus in Eq. (17) become fuzzy, since they depend on fuzzy material parameters \( \hat{E} \) and \( \hat{\nu} \). Therefore, the elasticity matrix in Eq. (16) is also fuzzy. For arithmetic operations with fuzzy variables in Eq. (16) and Eq. (17) the interval arithmetic technique is used considering the interaction of fuzzy variables.

### 3.2 Possibility determination of the system response

The possibility of the system response \( \hat{u}^* \) in Eq. (15) is constituted as the possibility of a quantity of interest \( \pi(Q(\hat{u}^*)) \). In this context, the quantity of interest may be the displacement at certain regions or the displacement of a single node for example. For more information about quantity of interests see, e.g. [26, 25, 27]. The possibility is calculated by determining the minimum and the maximum of the quantity of interest \( Q(\hat{u}^*)^\alpha_k \) for \( k = 1, \ldots, n_\alpha \) using the interval arithmetic technique.

### 4 REPRESENTATIVE EXAMPLE

Figure 4 illustrates the geometry, the loading and the FE-discretization of a plate with a ring hole with linear triangular elements under plain strain condition. Our goal is the computation of the possibility for the displacement \( \hat{u}^*_A \) of node A (i.e. \( \pi_A := \pi(Q(\hat{u}^*) = \hat{u}^*_A) \)) illustrated in Fig. 4b.

![Figure 4: Plate with a ring hole: a) Geometry and b) discretization with triangular elements](image-url)
4.1 Possibility establishment for the input parameters

Figure 5 shows a histogram of experimental results for Young’s modulus $E$ and Poisson’s ratio $\nu$ for an adhesive material. Experimental observations of this material are published in [24] for combined tension torsion tests. Considering this sparse information we use the probability-possibility transformation explained in Section 2.3.2 for both material parameters in order to obtain their possibilities as illustrated in Figure 6. The corresponding possibilities for the fuzzy shear modulus and the fuzzy bulk modulus from Eq. (17). The interaction domain with the joint possibility $\pi(\xi = (G, K)) = 0.5$ for $\alpha_6 = 0.5$ is illustrated in Figure 7.

![Figure 5: Experimental results for a) Young’s modulus and b) Poisson’s ratio](image)

![Figure 6: Possibility distribution of a) Young’s modulus and b) Poisson’s ratio](image)

![Figure 7: Interaction domain for the $\alpha_6$-cut of the shear modulus and the bulk modulus](image)
4.2 Possibility of the displacement for node $A$

Considering the interaction of the shear modulus and the bulk modulus in Eq. 17 we use the $\alpha$-level discretization technique in order to calculate $\pi_A$. In this context, Fig. 7 illustrates the $\alpha_6$-cut of the joint possibility distribution of the shear modulus and the bulk modulus. By determining the minimum and the maximum of $\hat{u}_A$ in every $\alpha$-cut the possibility $\pi_A$ is calculated. In this context, due to the monotonicity of $\hat{u}_A(G,K)$ the solutions for the minimum and maximum (in every $\alpha$-cut) are in the corners of the interaction domain of the shear modulus and the bulk modulus. Figure 8 illustrates the possibility distribution $\pi_A$ of the displacement $\hat{u}_A := (u_{A,x}, u_{A,y})$ for node $A$ calculated with 11 $\alpha$-cuts (with equidistant distances of 0.1) under the assumption of independency for the $x$ and $y$ directions. In order to establish this two dimensional representation for the possibility $\pi_A$ of the displacement $\hat{u}_A$, the principle of Eq. (1) is used.

![Figure 8: Possibility of the displacement for node $A$ in Figure 4](image)

5 CONCLUSION

Different kinds of uncertainty require different characterization techniques. This work presents a possibilistic evaluation of the linear elastic fuzzy finite element method with fuzzy input parameters based on sparse experimental data. To this end, possible marginal probability density functions of the fuzzy input parameters are transformed into possibility distributions using a probability-possibility transformation. Subsequently, a collective possibility distribution encoding a family of possible density functions is established. The possibility distributions of the material parameters are implemented as input functions in the fuzzy finite element method. In order to perform mathematical operations with possibility distributions respectively membership functions the $\alpha$-level discretization technique is used. It allows the calculation with an interval arithmetic based FEM. To ensure realistic results it is necessary to incorporate the interaction of fuzzy variables while using interval arithmetic. The possibilistic evaluation allows the examination of the uncertainty characterized by imprecise probabilities and may be interpreted as an upper bound of the probabilities.
REFERENCES


A COUPLED NUMERICAL AND LASER OPTICAL METHOD FOR ON-SITE CALIBRATION OF FLOW METERS

A. Weissenbrunner¹, A. Fiebach¹, M. Juling¹ and P.U. Thamsen²

¹Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2-12, 10587 Berlin, Germany
²Technische Universität Berlin, Department of Fluid System Dynamics, Straße des 17. Juni 135, 10623 Berlin, Germany
andreas.weissenbrunner@ptb.de

Keywords: CFD, polynomial chaos, LDV, optimization, calibration, flow rate reconstruction, uncertainty quantification, pipe flow.

Abstract. In this contribution it is intended to combine laser-optical measurements with numerical simulated flow profiles. This approach allows on-site calibration of installed flow meters and improves the accuracy of flow rate measurements under disturbed flow conditions. Since multiple pipe assemblies consecutively disturb the flow profiles, uncertain inflow conditions are considered in a polynomial chaos approach. Reynolds averaged Navier-Stokes equations (RANS) are solved to predict velocity profiles with inflow conditions from elbows out-of-plane with random distance to each other. Further, an optimization problem is solved in order to fit the simulated velocity profiles to the measurement data. The procedure is applied to different measurements from a test rig, for which the reference flow rate is known. The uncertainty of all measurements can be estimated very well. A reduction of the uncertainty by the procedure can however not always be assured, which is due to a lack of accuracy of the CFD simulations.
1 INTRODUCTION

Elbow alignments are necessary in almost all pipe assemblies in industrial fields, especially in district heating systems. They introduce disturbance to the flow profiles that require a straight pipe length of several tenth diameters to be eliminated. Very often those parts are not sufficiently long to redevelop an ideal profile. Thus, flow rate measurements with ideal flow conditions are not possible. It is well known that many flow meters react sensitively to disturbed flow profiles, as they are usually calibrated under ideal conditions on test rigs. This has been demonstrated in numerous publications, for example see [1, 4, 7, 17, 9, 10, 14, 15]. The induced errors for large heat meters tested in district heating pipelines are typically higher than 3% and could reach more then 20%, see [3].

To determine these errors, an in-situ calibration technique with Laser Doppler Velocimetry (LDV) was developed in [12]. The technique enables to measure point wise the axial fluid velocity on a diametrical path through an inserted contour fitting window. Under the assumption of rotationally symmetric flow profiles the integral over the path gives the flow rate. In disturbed flow conditions the uncertainty is highly increased and depends on the exact shape of the flow profile at the measurement cross section, which is generally unknown. Therefore the idea is to simulate the velocity profile at the measurement cross section and combine the result with LDV measurements.

In cooperation with TU Berlin, ILA GmbH and Optolotion Messtechnik GmbH the project ”EnEff:Wärme: On-site calibration of flow meters in district heating” [2] was initiated. The goal is to develop a method which permits on-site calibration of installed flow meters in non ideal installation conditions.

Pipes of district heating systems are characterized by large diameters and high temperatures, which leads to high Reynolds number flows. Therefore, Reynolds Averaged Navier Stokes (RANS) Models are used to predict the flow profiles at the measurement cross sections. In this paper combinations of elbows out-of-plane with different distance to each other are considered as pipe assemblies. LDV Measurements with 4 different setups were conducted. For the simulations the distance of the elbows is considered as random in a non-intrusive polynomial chaos study. To combine measurement and simulation the uncertainty of a single path measurement is predicted with CFD simulations. To further improve uncertainties a minimization problem is formulated for which the distance and additionally the path angle are varied in order to fit the CFD profiles to the LDV.

In Section 2 the experimental setup is described and the measured velocity distributions are shown. In Section 3 the numerical setup is described, the expectation value and the standard deviation of the velocity profile are compared to the measurements. In Section 4 the CFD results are used to estimate and reduce the uncertainty of the flow rate prediction. The results for all measurements are displayed and discussed. The last Section concludes the paper.

2 EXPERIMENTAL SETUP

The most influencing distortions in pipe flow profiles arise from bended pipes out-of-plane. They cause asymmetric and swirling profiles which decay only slowly. Most studied out-of-plane configurations consider two elbows closely coupled with an ideal inlet profile [8, 11, 9, 6]. In industrial pipe assemblies this is not the case. In an earlier work [15] it was shown that upstream pipe elbows with different distances cause diverse flow profiles. Here a combination of three closely coupled bends is studied with preliminary distortion.
The inner pipe diameter $D$ is $53.6\,\text{mm}$, whereas the curvature radius $R_c$ of the $90^\circ$ bends, according to the pipe center line, denotes $2.35\,D$. Compared to most other studied bends found in literature [11,9,5,8] this is a relatively large curvature. Between each of the three closely coupled bends a straight pipe is mounted with a length of $3\,D$. Two assemblies of identical mirrored triple bends with varying distance to each other are installed before the measurement section, see Figure 1 left. Four experimental setups were conducted with intermediate distances $5, 10, 20$ and $50\,D$. These measurements are referred as Measurement 1, 2, 3 and 4. The distance of the measurement section behind the last bend is fixed to $6\,D$.

To measure the velocity in the cross section a laser Doppler velocimeter (LDV) is used. LDV uses the Doppler shift of laser light, that is scattered by small particles within the fluid, to determine the velocity of the particles and hence the velocity of the fluid within a small measuring volume. The measuring volume is positioned on a Grid consisting of 49 diametrical and 10 angular locations, see Figure 1 right. The measurements were conducted on a test rig at the department of fluid system dynamics of the TU Berlin\textsuperscript{1}. The test rig contains plastic pipes with a very smooth surface. The fluid is water, for which the temperature and velocity were chosen to match a Reynolds number of about $3\cdot10^5$.

Centrifugal forces cause the velocity profile behind a bend to be higher at the outer part. If another bend is closely coupled and rotated by $90$ degrees in space, a swirling flow is generated. This causes the axial velocity distribution in a cross section to form a sickle shape. The three closely coupled elbows, which are considered here, cause similar effects. This can be observed in the measured velocity profiles in Figure 2. For different intermediate distances $\xi$ of the two elbow packages the shape of the sickle as well as the azimuthal position are varying. The broadness of the sickle is decreasing for increasing intermediate distance. The azimuthal position of the sickle is located on the right hand side and varies by about $45^\circ$.

\textsuperscript{1}Measurement data obtained by P. Kretschmer and A. Swienty
Figure 2: Normalized axial Velocity distribution \( \frac{u_z}{u_{vel}} \) 6 \( D \) downstream of the elbow combination. From the left top to right bottom: LDV measurements with intermediate distance 5, 10, 20 and 50 \( D \).
3 NUMERICAL SETUP

The geometry of the experimental set up is reconstructed in a CAD model and spatially discretized with a hexahedral grid consisting of about 3.8 million elements, see Figure 3. To sufficiently resolve the wall layer the dimensionless wall distance $y^+$ is chosen to be close to 1. To ensure mesh independence a second Grid with 11 million elements was generated. The results with both grids exhibit only marginal differences. Hence, the coarse grid is sufficient. The solution of the governing Navier-Stokes equations is numerically expensive and for high Reynolds numbers not feasible. Hence a turbulence model is needed. Here the eddy viscosity Reynolds Averaged Navier-Stokes (RANS) $k-\omega$ model [16] is chosen. Due to the smooth pipe surface of the test rig pipes, no-slip boundary conditions on the walls and a zero-gradient condition at the outlet are chosen. At the inlet, a fully developed profile from an earlier simulation with associated turbulence data is used as a natural inflow boundary condition. The length of the straight pipe downstream of the elbows is $65 \, D$. In a non intrusive polynomial chaos approach the distance of the preliminary group of three elbows is chosen as random. It is considered to be uniformly distributed between 4 and $60 \, D$. Therefore all four measurement setups are included in the probability space. 20 sample positions are chosen for the polynomial chaos approach according to the nodes of the Gauß-Legendre quadrature. The commercial solver ANSYS CFX is used for the CFD calculations. The normalized axial velocity profile resulting from a simulation with ideal inflow conditions is shown in Figure 4 left. The expected axial flow profile, calculated with random intermediate distance and the associated relative standard deviation is shown in Figure 4. Similar to the measurements, shown in Figure 2, the position of the sickle is located on the right-hand side, but is more elongated than in the measurements. The standard deviation exhibits values of up to 10 % relative to the mean velocity $u_{vol}$. The lowest values are located on the lower right-hand side and form sickle shape. The highest values can be observed on the right and left side bordering the sickle shape.

4 COMBINATION OF CFD AND LDV

In the following laser optical and numerical methods are combined in order to firstly estimate and secondly reduce the uncertainty of the flow rate calculation. The volume
flow rate from a single path LDV measurement is calculated by

\[ Q = \pi \int_{-R}^{R} u_z(\varphi_p, r) |r| \, dr. \]  

In case of a rotationally symmetric profile this leads to the exact value. However, for non symmetric profiles this leads to errors. In a first step the error range of the measurements is estimated by using CFD simulations. Note, that for all further calculations the reference flow rate \( Q_{\text{ref}} \) is defined by the mean value of the integrated LDV measurements. Due to reflections and other measurement inherit uncertainties, the values deviate slightly to the reference Meter on the test rig, but are neglected here.

4.1 Uncertainty estimation of the flow rate

The simulation result of the flow downstream the three elbows with fully developed inflow condition from Figure 4 is considered and the angle \( \varphi_p \) is set as random. This has been identified as a good approach in earlier research \[15\]. The error is defined as \((Q_p - Q_{\text{ref}})/Q_{\text{ref}} \times 100\), where \( Q_p \) denotes the flow rate calculated with a diametrical path. When discretization errors are neglected, the resulting expected error must vanish, due to its definition \[1\]. The results are shown in Figure 5. The standard deviation closely behind the bend outlet is up to 4 % and decreases until about 5 \( D \) to about 1.5 % and increases again until 20 \( D \). From there on it slowly decreases and is still 2 % after 50 diameters of straight pipe, compare the red error bars. The blue error bars denote the maximum deviations that occur in the particular cross section. They exhibit a similar behavior with increasing distance to the elbows. The amplitudes of the negative deviations are higher and can lead to more than -8 % directly behind the bend and about -6 % in 20 \( D \) distance. The measurements are located in 6 \( D \), which is in the area where the errors are expected to be the lowest. The evaluation of the standard deviation of the measurements (black) is with 1.2 % slightly lower than those predicted by the simulations. Also, the maximum and minimum deviations (green) are in the predicted scope.

4.2 Reduction of the flow rate uncertainty

In the next step a procedure is derived, that allows to reduce uncertainties of the flow rate calculation shown in Figure 5. Therefore, an optimization procedure is proposed in this section. As a measure for the difference \( \varepsilon = u_s(r) - u_m(r) \) between the measured and
Figure 5: Error estimation of a single diametrical LDV-Measurement with CFD with random angular position of the evaluated path, over the distance to the bend outlet, blue: maximum and minimum deviation, red: standard deviation from the CFD simulation, green and black: maximum/minimum and standard deviations from measurements.

Simulated velocity profile along a diametrical path, the $L^2$-norm in polar coordinates

$$
\| \epsilon \| = \left( \int_{-R}^{R} \int_{0}^{\pi} \epsilon^2 |r| d\varphi dr \right)^{\frac{1}{2}} = \pi \left( \int_{-R}^{R} \epsilon^2 |r| dr \right)^{\frac{1}{2}}
$$

is considered, whereas $R$ denotes the maximum radius $r$ of the pipe cross section. The flow rate is the quantity of interest and therefore unknown. However in the simulations the flow rate must be (implicitly) set in the boundary inlet condition. To conduct the simulations the flow rate $Q_s$ is estimated. Those estimations can result from the integral of a single path measurement or an installed flow meter. However the flow rate in the simulation $Q_s$ is defective. To compensate these errors a constant factor $F$ is introduced, which multiplies the simulated velocity profile to best fit the measured velocity profile, i.e. $Q_c = FQ_s$. Additionally the velocity distribution of the simulation is depending on the inlet condition, e.g. the distance to the preliminary bends $\xi \in \Omega = [4, 64] \ D$. The angle $\varphi_p$ of the diametrical path can also be varied. This results in the non-linear, non convex optimization problem

$$
T = \| F \tilde{u}_s (\xi, \varphi_p, r) - u_m (r) \|, \ \min_{\forall \xi \in \Omega, \varphi \in [0, 2\pi], F \in \mathbb{R}^+} T.
$$

The velocity profile for a certain inlet condition $\tilde{u}_s$ is assembled by a sum over all modes for a certain radius and angle $\tilde{u}_i (\xi, \varphi_p, r)$ multiplied with the orthogonal polynomials $P_i (\xi)$

$$
\tilde{u}_s (\xi, r) = \sum_{i=1}^{N_m} \tilde{u}_i (r) P_i (\xi).
$$
From the optimization criteria \( \frac{\partial T}{\partial F} = 0 \) the factor \( F \) can be determined by

\[
F(\xi, \varphi_p) = -\frac{\int_{-R}^{R} u_m(r) \tilde{u}_s(\xi, \varphi_p, r) |r| \, dr}{\int_{-R}^{R} \tilde{u}_s^2(\xi, \varphi_p, r) |r| \, dr}.
\] (5)

The derivatives of \( T \) for the other variables \( \xi \) and \( \varphi_p \) can not be calculated directly and would lead to a new minimization problem, which is not considered here. In order to simplify the calculation the optimal factor \( F(\xi, \varphi_p) \) is inserted into the target function \( T \). The resulting optimization problem is reduced by one dimension

\[
T = \left\| F(\xi, \varphi_p) \tilde{u}_p(\xi, \varphi_p, r) - u_m(r) \right\|, \min_{\forall \xi \in \Omega, \varphi \in [0, 2\pi]} T.
\] (6)

To solve the minimization problem (6) the downhill simplex algorithm according to Nelder and Mead [13] is used. To distinguish the global minimum the algorithm is started with different initial conditions. All Integrals are numerically solved by a trapezoidal rule.

### 4.3 Results

The procedure is applied to Measurement 1-4 in Figure 2. Different variations, e.g. of the angle \( \varphi_p \) and the distance between the elbows \( \xi \) were conducted. The variation of the path angle shows the best results. The errors for all measurements are shown in Figure 3. Each path is treated as a single LDV measurement. The errors are in percent relative to the mean value of all path integrations according to (1). The errors of the integration of each path is shown in blue circles connected with lines. All measurements show a periodic behavior with amplitudes of about 2 \%. Applied to each particular measurement the proposed procedure behaves differently. For Measurement 3 the error decreases for each path. The mean error of the combined procedure, depicted as red dashed line, is staying around zero. For Measurement 2 and 4 the error for about half of the paths is decreased. For the other half the error slightly increases. The mean value in both cases is increasing to about 1 \%. For Measurement 1, in which the distance between the two elbow conjunctions is 5 \( D \), the combined procedure lowers the error only in one path. In every other path the error is increased, the mean error is about 3 \%. The value of the target function (6) is represented by red error bars. It is proposed to be a measure for the uncertainty for a single path measurement. It can be observed that in almost all cases the error bar covers the zero error line. Nonetheless, the size of the error bars do not always correspond to the present error and seem to be too large. To assess the performance of the combined method the standard deviation of the error over the paths is compared in Figure 7. The standard deviation of the errors that are calculated solely by integration of the LDV profiles is depicted in blue and for the combined algorithm in red. While the standard deviation for Measurement 2, 3, and 4 is drastically reduced, it is slightly enhanced for Measurement 1. The yellow bars represent the value of the expected error for the combined procedure. It is added to the standard deviation as an additional uncertainty. For Measurements 1 and 2 the uncertainty is enhanced by the combined algorithm. For Measurement 3 and 4 a reduction of the uncertainty is still achieved by the combined procedure.
Figure 6: Normalized axial Velocity distribution $u_z/u_b$ 6 $D$ downstream of the triple elbow combination with different distances to the preliminary elbows. From the left top to right bottom: 5, 10, 20 and 50 $D$, corresponding to Measurement 1-4 in Figure 2.
5 CONCLUSIONS

In this paper systematic errors of single path laser Doppler velocimetry (LDV) measurements in asymmetric swirling flows behind several elbows out-of-plane are investigated. To do so, simulations with Reynolds averaged Navier-Stokes equations (RANS) are performed. Firstly it is demonstrated, that the uncertainty of the calculated flow rate due to the integration of a single diametrical LDV path can be well estimated with computational fluid dynamics. Therefore, the angle is set as random and the standard deviation is used as a measure for the uncertainty. Secondly, a method is derived to fit a simulation result to a measurement by solving an optimization problem. Due to the adapted simulation a factor is introduced that estimates the flow rate, while the value of the target function is proposed as a measure for the uncertainty of the coupled numerical and laser optical procedure. Four LDV measurements were studied.

For two cases the uncertainty does not change much. Due to the lack of accuracy of the RANS turbulence models for swirling flows the uncertainty for one case is increased. It is expected that with advanced turbulence modeling the uncertainties will be decreased. This is confirmed for one case for which it leads to a significant reduction of the measurement uncertainty. This shows the potential of the combination of CFD and LDV applied to measurements under disturbed flow conditions.

ACKNOWLEDGMENT

A. Weissenbrunner was supported by the Federal Ministry for Economic Affairs and Energy within the EnEff:Wärme project ”On-site calibration of flow meters in district heating”. A. Fiebach was supported by the European Commission and EURAMET within the EMRP NEW-04 project ”Novel mathematical and statistical approaches to uncertainty evaluation”. The authors would like to thank all members of the EnEff-project for their assistance, especially A. Swienty and P. Kretschmer for providing the measurement data.
REFERENCES


ON THE DEVELOPMENT OF THE 3D EULER EQUATIONS USING INTRUSIVE PCE FOR UNCERTAINTY QUANTIFICATION

Kyriakos-Dimitrios Kantarakias, Michail E. Chatzimanolakis, Varvara G. Asouti and Kyriakos C. Giannakoglou

National Technical University of Athens, School of Mech. Eng.,
Parallel CFD & Optimization Unit, Athens, Greece
e-mail: {kyr.kantar,mikechatzimanolakis}@gmail.com,
vasouti@mail.ntua.gr, kgianna@central.ntua.gr

Keywords: Uncertainty Quantification, Intrusive Polynomial Chaos

Abstract. This paper presents a method for the quantification of uncertainty propagation using intrusive Polynomial Chaos Expansion (iPCE) in CFD. In contrast to commonly implemented non–intrusive methods which take advantage of existing CFD evaluation software in order to quantify the statistical behavior of the flow, an intrusive PCE method is developed and implemented to the 3D Euler equations. Uncertainties are introduced through the flow conditions and their propagation throughout the flow field is quantified. A Probability Density Function (PDF) is assumed for each uncertain flow condition and the generalized PCE inviscid equations for the corresponding coefficient fields of the flow variables are derived. Already known properties of the equations, such as the first order-homogeneity, are found to hold in the new set of the equations. The discretization schemes are adapted to the new set of governing equations while a systematic approach to the corresponding eigenproblem is introduced. The method is applied to 3D inviscid flow cases for which the mean value and the standard deviation of specific flow quantities characterizing the flow are quantified and compared with those computed by the non–intrusive PCE and Monte–Carlo methods.
1 INTRODUCTION

Computational Fluid Dynamics (CFD) methods and software are indispensable tools in the analysis and design in various engineering fields. The common practice is to analyze the flow under consideration using a deterministic computational model. However, such ideal situations are rare in real-world applications and the system’s performance is sensible to varying conditions; for instance, a compressor’s efficiency is affected by variations/uncertainties associated with the inlet and/or outlet conditions etc. In order to take into account these uncertainties, non-deterministic approaches should be used.

During the last decade, polynomial chaos expansion (PCE) methods have been used to model uncertainties in engineering applications [3, 4, 5]. These are based on the idea of homogeneous chaos, [1, 2]. In contrast to sampling techniques such as Monte-Carlo, PCE methods are based on the spectral representation of the uncertain quantities.

PCE methods can be intrusive (iPCE) or non-intrusive (niPCE) depending on whether the governing equations are altered or not, [7, 4]. In the non-intrusive variant the CFD code is used in its standard form as an evaluation software along with an integration formula based on Gauss quadrature and the appropriate weighting functions that correspond to the assumed PDFs. Through this method, the statistical moments of an objective function (such as the lift coefficient of a wing) are computed. On the other hand, the intrusive methodology introduces uncertainties in the mathematical model, the effect of which appears in the flow equations through the PCE of the flow variables. This results in a new set of PDEs governing the PCE coefficient fields of the flow variables to be numerically solved. Contrary to its non-intrusive variant, the iPCE method computes the statistical moments of the flow field and, through post-processing, those of the objective function of interest. In terms of computational cost, both PCE variants are far more efficient than the Monte-Carlo method which requires thousands of CFD evaluations in order to compute the statistical moments of the objective function.

In the present paper, the iPCE of the 3D Euler equations is presented. Emphasis is laid on the derivation of the stochastic flow equations and their discretization. Uncertainties are introduced by the boundary conditions. Other forms, such as uncertainties in the geometry or the properties of the gas are not discussed in this paper, however their effect can be quantified through a similar method.

2 UNCERTAINTY QUANTIFICATION (UQ) USING PCE

Let us assume a set of \( m \) uncertain or stochastic variables \( \xi = (\xi_1, ..., \xi_m) \). Their probability density functions (PDF) \( w_i \) are associated with an orthogonal polynomial bases \( \psi^{(i)} = \{\psi_0^{(i)}, \psi_1^{(i)}, ..., \} \), where by definition

\[
< \psi_j^{(i)}, \psi_k^{(i)} >= \int_{\mathcal{E}_i} \psi_j^{(i)} w_i d\xi_i = \delta_{jk} < \psi_j^{(i)}, \psi_j^{(i)}> \tag{1}
\]

In eq. (1) no summation for the repeated indices is implied, \( \delta_{jk} \) is the Kronecker symbol and \( \mathcal{E}_i \) denotes the domain of \( w_i \).

The PCE of any quantity \( \phi(\tilde{\xi}) \) can be expressed by using a polynomial basis, defined as the tensor product of \( \psi^{(i)} \), \( \Psi = \otimes_{i=1}^{m} \psi^{(i)} = \{\Psi_0, \Psi_1, \ldots \} \), as follows

\[
\phi(\tilde{\xi}) = \sum_{i=0}^{\infty} \phi_i \Psi_i(\tilde{\xi}) \tag{2}
\]
Bases $\Psi$ are also orthogonal, namely

$$< \Psi_i, \Psi_j > = \int_{\mathcal{E}} \Psi_i \Psi_j w \delta i j = \delta_{i j} < \Psi_i, \Psi_i >$$  \hspace{1cm} (3)

In eq. 3, $w(\xi) = \prod_{i=1}^{m} w_i(\xi_i)$ and $\mathcal{E} = \bigcup_{i=1}^{m} \mathcal{E}_i$. Based on the above and after normalizing the $\Psi$ polynomials so that $< \Psi_i, \Psi_i > = 1$, the first two statistical moments, namely the mean value and standard deviation, of any quantity $\phi$ are given by

$$\mu_\phi = \phi_0 , \quad \sigma_\phi = \sqrt{\sum_{i=1}^{\infty} \phi_i^2}$$  \hspace{1cm} (4)

In iPCE methods, one should first apply the PCE to each flow variable, truncated to $q+1$ terms, introduce these expansions into the flow equations and derive a set of new equations and boundary conditions which must be numerically solved to compute the PCE coefficient fields of all flow quantities. The statistical moments of the quantity of interest (such as lift, drag, losses or any other, usually integral, quantity) are computed at a post-processing level, using eqs. 4.

To determine the most appropriate value of $q$, a common practice is to retain all polynomials up to a user-defined degree $C$, which is referred to as the chaos order. For a selected chaos order $C$ and $m$ stochastic variables, $q$ is given by

$$q = \frac{(m + C)!}{m!C!} - 1$$  \hspace{1cm} (5)

The required $q+1$ equations per flow variable can be derived through appropriate Galerkin projections of the governing equations/PDEs. The Galerkin projection of any scalar function $\phi(\xi)$ to $\Psi_i$ is defined as

$$G_i[\phi] = \int_{\mathcal{E}} \Psi_i \phi w d\xi$$  \hspace{1cm} (6)

By making use of the properties of the selected orthogonal polynomial bases, the $q+1$ Galerkin projections to the flow PDEs give rise to a new set of PDEs.

3 DERIVATION AND NUMERICAL SOLUTION OF THE iPCE EQUATIONS

3.1 The iPCE Flow Equations

The Euler equations for compressible fluid flows are written, in the standard vector form, as

$$\frac{\partial \tilde{U}}{\partial t} + \frac{\partial \tilde{f}_i}{\partial x_i} = \tilde{0}$$  \hspace{1cm} (7)

In eq. 7, $\tilde{U} = [\rho \rho \tilde{u} E_i]^T$ are the conservative flow variables and $\tilde{f}_i = [\rho u_i \rho u_i \tilde{u} + p \delta_i u_i (E_i + p)]^T$ are the inviscid fluxes in Cartesian coordinates, $\rho$ is the density, $\tilde{u} = [u_1, u_2, u_3]^T$ is the velocity vector, $E_i$ is the total energy per unit mass, $p$ is the pressure and $\delta_i = [\delta_{i1}, \delta_{i2}]^T$. Eqs. 7 can be rewritten as

$$\frac{\partial \tilde{U}}{\partial t} + A_i \frac{\partial \tilde{U}}{\partial x_i} = \tilde{0}$$  \hspace{1cm} (8)

with $A_i = \frac{\partial \tilde{f}_i}{\partial \tilde{U}}$ denoting the Jacobian matrices.
The Euler equations solver this work is based upon is an in-house time-marching code, based on the finite volume technique. It is fully parallelized, based on the multi-domain technique, using the MPI protocol. It has been also ported to NVIDIA GPUs (using CUDA C), offering significant speedup (about $\times 60$).

A systematic approach to the derivation of the PCE Euler equations requires the following two definitions, which extend the notion of Galerkin projections to vectors and matrices:

**Definition 1** Let $\vec{X} = [X_1, \ldots, X_m]^T$ be a vector whose components are $X_j = \sum_{i=0}^{\infty} x_{j,i} \Psi_i(\vec{\xi})$.

The Galerkin projected vector $\vec{X}$ of order $q$ is defined as

$$G_q[\vec{X}] = [g_0[\vec{X}], g_1[\vec{X}], \ldots, g_q[\vec{X}]]^T$$

where $g_i[\vec{X}] = [G_i[X_1], G_i[X_2], \ldots, G_i[X_m]]^T$.

**Definition 2** Let $A$ be an $m \times m$ matrix whose components $A_{ij}$ are given by $A_{ij} = \sum_{k=0}^{\infty} a_{ij,k} \Psi_k(\vec{\xi})$.

The Galerkin projected matrix $A$ of order $q$ is defined as the block matrix

$$G_q[A] = \begin{bmatrix} B_{0,0} & B_{0,1} & \ldots & B_{0,q} \\
B_{1,0} & B_{1,1} & \ldots & B_{1,q} \\
\vdots & \vdots & \ddots & \vdots \\
B_{q,0} & B_{q,1} & \ldots & B_{q,q} \end{bmatrix}$$

where the elements of each $m \times m$ block are

$$(B_{k_1,k_2})_{ij} = G_{k_1}[A_{ij} \Psi_{k_2}] = \int_{\mathcal{E}} \Psi_{k_1} \Psi_{k_2} A_{ij} w d\vec{\xi} \quad 1 \leq i, j \leq m$$

One can verify that the following relation

$$G_q[A\vec{x}] = G_q[A]G_q[\vec{x}]$$

holds for an any vector $\vec{x}$ and matrix $A$. Applying definitions 1 and 2 to eqs. 7, 8 and taking eq. 9 into consideration the PCE Euler equations are derived (subscript $q$ is omitted hereafter)

$$\frac{\partial G[U]}{\partial t} + \frac{\partial G[\vec{f}_i]}{\partial x_i} = \frac{\partial G[\vec{U}]}{\partial t} + G[A_i] \frac{\partial G[\vec{U}]}{\partial x_i} = 0$$

$$G[\vec{U}] = [\rho_0 (\rho \vec{u})_0 E_{t_0} \, \rho_1 (\rho \vec{u})_1 E_{t_1} \, \ldots \, \rho_q (\rho \vec{u})_q E_{t_q}]^T$$

Eqs. 10 resemble their deterministic counterpart in that they are first–order homogeneous, since $\vec{f}_i = A_i \vec{U}$ implies (based on eq. 9) that $G[\vec{f}_i] = G[A_i]G[\vec{U}]$.

### 3.2 Numerical Solution of the iPCE Equations

For the sake of simplicity, the Galerkin operator will be omitted in what follows; thus, below, $A$ and $\vec{U}$ will denote $G[A]$, and $G[\vec{U}]$ respectively. Upwind schemes can be applied to the iPCE equations, given that the latter are first order–homogeneous and have the same form as the deterministic ones. According to the vertex–centered finite volume method, each control volume $\Omega_P$ is formed around the corresponding grid node $P$, as illustrated in fig. 1 for a hybrid
2D grid (its extension to 3D is evident). By integrating eqs. (10) over $\Omega_P$, in a steady flow analysis, we get

$$\Omega_P \left( \frac{\bar{U}_P^{\kappa+1} - \bar{U}_P^{\kappa}}{\Delta t_P} \right) + \sum_{Q \in \text{nei}(P)} \left[ \Phi_{PQ} \right] \partial \Omega_{PQ} = 0$$ (11)

where $\kappa$ is the pseudo–time step counter, $\Phi_{PQ}$ is the inviscid numerical flux crossing the interface ($\partial \Omega_{PQ}$) between two adjacent finite volumes (pointing from $P$ to $Q$); $\text{nei}(P)$ stands for the set of neighbouring finite volumes of node $P$. The inviscid fluxes are computed using the flux vector splitting technique [9], applied between $P$ and $Q$, as follows

$$\Phi_{PQ} = A_{-PQ} \bar{U}_{PQ}^R + A_{+PQ} \bar{U}_{PQ}^L$$ (12)

where $A_{PQ} = \frac{\partial (\bar{f} \bar{n})}{\partial \bar{u}} = A_{+PQ} + A_{-PQ}$ and $A_{PQ}^+$, $A_{PQ}^-$ are defined using the positive and negative eigenvalues of the Jacobian matrix. For second–order spatial accuracy, $\bar{U}_{PQ}^L$ and $\bar{U}_{PQ}^R$ (where $L$ and $R$ denote the two states on both sides of the interface between $\Omega_P$ and $\Omega_Q$) are computed from $\bar{U}_P$, $\bar{U}_Q$, $\nabla \bar{U}_P$ and $\nabla \bar{U}_Q$ as follows

$$\bar{U}_{PQ}^L = \bar{U}_P + \frac{1}{2} (\bar{PQ}) \cdot \nabla \bar{U}_P, \quad \bar{U}_{PQ}^R = \bar{U}_Q - \frac{1}{2} (\bar{PQ}) \cdot \nabla \bar{U}_Q$$

The so–computed fluxes are limited using the van Leer–van Albada limiting function [6]. Spatial gradients are computed using the Green–Gauss integration formula.

The discretized eqs. (11) are solved at each pseudo–time step using the point–implicit Jacobi which is written as

$$D_P^\kappa \Delta \bar{U}^{\kappa+1,\nu}_P + \sum_{Q \in \text{nei}(P)} Z^\kappa_{Q} \Delta \bar{U}^{\kappa+1,\nu}_Q = -R^\kappa_{P,\nu}$$

$$\bar{U}^{\kappa+1}_P = \Delta \bar{U}^{\kappa+1}_P + U^{\kappa}_P$$ (13)

where $\kappa$ is the pseudo–time counter, $\nu$ the Jacobi internal iteration counter, $D_P$, $Z_Q$ stand for the diagonal and non–diagonal matrices respectively and $R^\kappa_{P,\nu}$ is the residual array. Each Jacobi iteration comprises one iteration to solve the equations corresponding to one of the statistical moments by freezing the other terms.
3.3 Eigen–Decomposition

The implementation of the FVS scheme requires the solution of the corresponding eigenproblem. Starting point is the characteristic equations \( \Lambda n_i \frac{\partial \hat{W}}{\partial x_i} = 0 \) where \( \hat{W} = [w^A w^B w^C w^D w^E]^T \) are the characteristic variables, \( \Lambda = diag(u^{(n)}_1, u^{(n)}_2, u^{(n)}_3, u^{(n)}_4) + c, u^{(n)}_i = u_i n_i \) and \( c \) is the speed of sound. By applying the iPCE followed by Galerkin projections, we get

\[
G[\Lambda]n_i \frac{\partial G[\hat{W}]}{\partial x_i} = \vec{0}
\] (14)

Eqs. [14] can also be written as

\[
\begin{bmatrix}
    d_j G_j [\Psi_0 \Psi_0] & d_j G_j [\Psi_0 \Psi_1] & \ldots & d_j G_j [\Psi_0 \Psi_q] \\
    d_j G_j [\Psi_1 \Psi_0] & d_j G_j [\Psi_1 \Psi_1] & \ldots & d_j G_j [\Psi_1 \Psi_q] \\
    \vdots & \vdots & \ddots & \vdots \\
    d_j G_j [\Psi_q \Psi_0] & d_j G_j [\Psi_q \Psi_1] & \ldots & d_j G_j [\Psi_q \Psi_q] 
\end{bmatrix}
\begin{bmatrix}
    \hat{W}_0 \\
    \hat{W}_1 \\
    \vdots \\
    \hat{W}_q 
\end{bmatrix} = \vec{0}
\] (15)

where

\[
d_j = diag(u^{(n)}_j, u^{(n)}_j, u^{(n)}_j + c, u^{(n)}_j - c)
\]

\[
\hat{W}_i = [w_i^A, w_i^B, w_i^C, w_i^D, w_i^E]^T
\]

and \( u^{(n)}_j, w^A, c \) denote the j–th term of the PCE of \( \vec{u}, \vec{n}, w^A \) and of \( c \), respectively. Note that \( G_k[\Psi_i \Psi_j] \) is equal to \( <\Psi_i, \Psi_j, \Psi_k> = \int_{\xi} \Psi_i \Psi_j \Psi_k w d\xi \).

The spectral components of \( c_j \) can be found through the Galerkin projections of the expression of \( c \), written as a function of the conservative variables, as follows

\[
c_j = G_j[c] = \int_{\xi} \sqrt{\frac{\gamma R}{c_v g_i \Psi_i} \left( E_{t_k} \Psi_k - \frac{(q_i u_i \Psi_i)(q_k u_k \Psi_k)}{2 g_i \Psi_i} \right)} w d\xi^{-1}
\] (16)

In eq. [16] \( \gamma \) is the specific heat ratio, \( R \) the specific gas constant and \( c_v \) the specific heat capacity at constant volume. Note that a similar procedure can be used for the PCE of any quantity expressed in terms of the conservative variables. By re-ordering eqs. [15] we get

\[
diag([Z(u^{(n)}), Z(u^{(n)}), Z(u^{(n)} + c), Z(u^{(n)} - c)]) \hat{w} = \vec{0}
\] (17)

where

\[
\hat{w} = [w_0^A \ldots w_0^A w_0^B \ldots w_p^B w_0^C \ldots w_p^C w_0^D \ldots w_p^D w_0^E \ldots w_p^E]^T
\]

\[
Z(\lambda) =
\begin{bmatrix}
    \lambda_j G_j [\Psi_0 \Psi_0] & \lambda_j G_j [\Psi_0 \Psi_1] & \ldots & \lambda_j G_j [\Psi_0 \Psi_q] \\
    \lambda_j G_j [\Psi_1 \Psi_0] & \lambda_j G_j [\Psi_1 \Psi_1] & \ldots & \lambda_j G_j [\Psi_1 \Psi_q] \\
    \vdots & \vdots & \ddots & \vdots \\
    \lambda_j G_j [\Psi_q \Psi_0] & \lambda_j G_j [\Psi_q \Psi_1] & \ldots & \lambda_j G_j [\Psi_q \Psi_q]
\end{bmatrix}
\] (18)

Thus, the solution of an eigenproblem corresponding to a \((5 \times (q+1)) \times (5 \times (q+1))\) matrix is now reduced to one corresponding to the \((q+1) \times (q+1)\) matrix \( Z \). The diagonalization of \( Z \) yields the desired eigenvalues and eigenvectors of \( G[A_i] \).
4 APPLICATIONS

The iPCE method was programmed and applied to the UQ in an external and an internal aerodynamic case. Comparisons with the niPCE method and/or the (much more expensive) Monte–Carlo technique are shown.

4.1 Flow Around an Aircraft Model

The first problem is considered with the inviscid flow around an aircraft model. The study is carried out around half of the aircraft (due to the symmetric flow conditions) and the unstructured CFD mesh consists of about 45K nodes and 256K tetrahedra. The quantity of interest is the lift coefficient of the aircraft, for which the statistical moments must be computed.

Uncertainties are introduced from the stochastically varying infinite flow angle \(a_\infty\) and/or the infinite Mach number \(M_\infty\). Four cases are studied. Results are presented in Table 1, where \(N(\mu,\sigma)\) denotes the normal distribution with mean value \(\mu\) and standard deviation \(\sigma\), whereas \(\Psi_i\) stand for the Hermite polynomials. Also, \(U(a,b)\) denotes the uniform distribution in the interval \([a,b]\) and, in this case, \(\Psi_i\) are the Legendre polynomials.

<table>
<thead>
<tr>
<th>Flow Conditions</th>
<th>iPCE Chaos order (C=1)</th>
<th>niPCE Chaos order (C=1)</th>
<th>iPCE Chaos order (C=2)</th>
<th>niPCE Chaos order (C=2)</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M_\infty = 0.7) (&lt;a_\infty \sim N(5^\circ, 0.5^\circ))</td>
<td>(\mu_{C_L}^{iPCE}) 0.119174</td>
<td>(\sigma_{C_L}^{iPCE}) 0.0095783</td>
<td>(\mu_{C_L}^{niPCE}) 0.119107</td>
<td>(\sigma_{C_L}^{niPCE}) 0.009657</td>
<td>(\mu_{C_L}^{MC}) 0.119106</td>
</tr>
<tr>
<td>(M_\infty \sim N(0.7, 0.02)) (&lt;a_\infty = 5^\circ)</td>
<td>(\mu_{C_L}^{iPCE}) 0.119448</td>
<td>(\sigma_{C_L}^{iPCE}) 0.001792</td>
<td>(\mu_{C_L}^{niPCE}) 0.119312</td>
<td>(\sigma_{C_L}^{niPCE}) 0.001894</td>
<td>(\mu_{C_L}^{MC}) 0.119312</td>
</tr>
<tr>
<td>(M_\infty \sim N(0.7, 0.02)) (&lt;a_\infty \sim N(5^\circ, 0.5^\circ))</td>
<td>(\mu_{C_L}^{iPCE}) 0.119326</td>
<td>(\sigma_{C_L}^{iPCE}) 0.0093172</td>
<td>(\mu_{C_L}^{niPCE}) 0.119237</td>
<td>(\sigma_{C_L}^{niPCE}) 0.0098453</td>
<td>(\mu_{C_L}^{MC}) 0.11925</td>
</tr>
</tbody>
</table>

Table 1: UQ for the flow around an aircraft model. Statistical moments of the lift coefficient values computed using iPCE, niPCE (with \(C=1\) and \(C=2\)) and the Monte–Carlo method with 2000 replicates in each case.

Figure 2 compares the Mach number’s mean and standard deviation fields respectively, in the case uncertainty is only due to the Mach number \(a_\infty = 5^\circ\); second case in Table 1. It can be seen that the iPCE and niPCE results perfectly match each other all over the aircraft surface. One can also notice the increased variance after the supersonic area of the wing surface which, in the case of niPCE, is extended over a greater area along the wing.

4.2 Flow in the CS10 Compressor Cascade

The second case deals with the transonic flow in a compressor cascade. This is practically a 2D airfoil geometry extruded in the spanwise direction and studied herein as a 3D linear cascade. The compressor profile is that of the Standard Configuration 10 (SC10) which results by superimposing the thickness distribution of a modified NACA 0006 airfoil on a circular–arc camber line. The blade stagger angle is 45\(^\circ\) and the pitch–to–chord ratio is 1.0. The quantity of interest is the static pressure rise \(\Delta p\) and uncertainty is introduced by the inlet flow angle \(a_1\) and the outlet isentropic Mach number \(M_{2,is}\). Results are presented in Table 2.

Results of iPCE and niPCE are compared in figure 3. Once more, we may notice the higher...
variance of the Mach number field on the blade surface after the shock. The convergence of the spectral continuity equations, for the iPCE with \( C=2 \), is plotted in figure 4.

<table>
<thead>
<tr>
<th>Flow Conditions</th>
<th>iPCE Chaos order ( C = 1 )</th>
<th>iPCE Chaos order ( C = 2 )</th>
<th>niPCE</th>
<th>niPCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_{2,ls} = 0.4425 )</td>
<td>( \mu_{\Delta p} ) 1.32930</td>
<td>( \mu_{\Delta p} ) 1.32944</td>
<td>( \sigma_{\Delta p} ) 0.02675</td>
<td>( \sigma_{\Delta p} ) 0.03179</td>
</tr>
<tr>
<td>( a_1 \sim N(58^\circ, 1^\circ) )</td>
<td>( \sigma_{\Delta p} ) 0.02579</td>
<td>( \sigma_{\Delta p} ) 0.02956</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M_{2,ls} \sim N(0.4425, 0.005) )</td>
<td>( \mu_{\Delta p} ) 1.33754</td>
<td>( \mu_{\Delta p} ) 1.33728</td>
<td>( \sigma_{\Delta p} ) 0.009685</td>
<td>( \sigma_{\Delta p} ) 0.009786</td>
</tr>
<tr>
<td>( a_1 = 58^\circ )</td>
<td>( \sigma_{\Delta p} ) 0.009135</td>
<td>( \sigma_{\Delta p} ) 0.009346</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: CS10 Compressor Cascade. Statistical moments of \( \Delta p \) computed using iPCE and niPCE, with \( C = 1 \) and \( C = 2 \).

## 5 CONCLUSIONS

This paper presents the development of the intrusive polynomial chaos expansion of the 3D Euler equations through a systematic procedure that ensures the unimpeded numerical solution of the resulting equations. For a single uncertain quantity the iPCE equations ask for about 1.5× more time than the baseline Euler equations which makes it far more efficient than any Monte–Carlo sampling technique and comparable with the niPCE with chaos order equal to one. In any other case, i.e. for more uncertain variables and/or higher PC order, the iPCE vastly
Figure 3: CS10 Compressor Cascade ($M_{2, is} \sim N(0.4425, 0.005)$, $\alpha_1 = 58^\circ$). Mean (top) and standard deviation (bottom) fields of the Mach number computed using the iPCE (left) and niPCE (right), with $C=2$.

Figure 4: CS10 Compressor Cascade ($M_{2, is} \sim N(0.4425, 0.005)$, $\alpha_1 = 58^\circ$). Convergence of the residual of the spectral continuity equations (iPCE, $C=2$).

outperforms the niPCE.

Though, in the cases studied herein, uncertainties are introduced through the flow conditions, the implementation of any other type of uncertainty, such as uncertainties related to the geometry, is straightforward by means of the presented development. Regarding future work, the extension to viscous/turbulent flows along with the corresponding continuous adjoint method, for use in optimization under uncertainties, is foreseen.

REFERENCES


THE NON-PARAMETRIC APPROACH TO THE QUANTIFICATION OF THE UNCERTAINTY IN THE DESIGN OF EXPERIMENTS MODELLING

Jacek Pietraszek\textsuperscript{1}, Renata Dwornicka\textsuperscript{1}, Mariusz Krawczyk\textsuperscript{2}, and Maciej Kołomycki\textsuperscript{1}

\textsuperscript{1}Department of Software Engineering and Applied Statistics, Cracow University of Technology
37 Jana Pawła II, 31-864 Kraków, Poland
e-mail: pmpietra@mech.pk.edu.pl, dwornick@mech.pk.edu.pl, mkolomycki@gmail.com

\textsuperscript{2}Chair of Information Systems and Computer Modeling, Cracow University of Technology
37 Jana Pawła II, 31-864 Kraków, Poland
e-mail: mariusz.krawczyk@mech.pk.edu.pl

Keywords: Non-Parametric Statistical Methods, Design of Experiments, DoE, Factorial Approach, RSM Approach, Uncertainty Quantification.

Abstract. The classic design of experiments (DoE) typically uses the least-square method for a model identification and requires associated assumption about the normality of a noise factor. It is very convenience because it leads to a relative simple computations and well-known asymptotic statistics based on the normality assumption. However, if that assumption is not satisfied it may fail and obtained results may differ radically from the verification tests. The rationale for the caution may be the comparison of interval plots (based on the normality hypothesis) and box-plots (based on raw data). The useful approach is the bootstrap-based methodology which replaces the requirement of the normality assumption with weaker requirement of the independent and identical distribution (i.i.d.) of the random term. The industrial applications of this approach are still rare because the industry is very conservative and usually utilizes old well-known methods and typical numerical software like e.g. Statistica, Statgraphics or Minitab. This paper presents the bootstrap modeling of the random uncertainty in the two cases: the factorial designed experiment and the response surface experiment.

1 INTRODUCTION

Methods of the approximation and the prediction rapidly evolved in recent years. New approaches came from new branches of the statistics and the artificial intelligence area: non-parametric, data-driven, stochastic etc. Apart from this, the classic approach of the design of experiment methodology (DoE) is still very useful and popular in the industry full-scale production as well as in the R&D laboratories.

The typical procedure of the model identification assumes the normal distribution of the data noise [7] and uses the least square method [8] to identify parameters of the model with \textit{a priori} assumed structure what set this approach in the group of parametric methods. The terms of a model are repeatedly eliminated in the specific backward stepwise regression, while three indicators: (a) the least significance of parameters, (b) the significance of the lack of fit and (c) the conformity of residuals with the normal distribution are simultaneously observed to make a decision to stop or to continue the elimination procedure.

Practically, for small datasets, the conformity with the normal distribution has very weak reliability and it leads to the very uncertain assessment of parameters statistical significance and bounds of their confidence intervals. The bootstrap approach [9] with simulation-based identification of parameters confidence intervals appears to be better solution than theoretically proved but only asymptotically equal \( t \)-distribution [7]. It seems to be specifically more important for the non-parametric models where the model structure is created dynamically during data analysis e.g. artificial neural networks, regression trees etc., while theoretically-based distribution cannot be determined, because their final structures are so non-linear that analytical identification of associated distributions and their confidence intervals is not possible.

The key issue in the bootstrap approach is to make proper identification of the random term inside the model structure because the i.i.d. requirement have to be satisfied by this term.

2 METHODS

DoE methodology splits into two branches: the factorial approach [10, 11] for qualitative input variables (factors) and the response surface methodology (RSM) [11, 12] for quantitative input variables. Obviously, the models for mixed inputs, qualitative and quantitative, may be identified however it is not so easy.

The specific models used in these methodologies and the location of the random term are described in subsequent chapters in details.

2.1 Factorial model

The factorial approach is typical for qualitative input variables. Their settings formally are labels, not numerical values, in contrast to quantitative ones. The model predicts response for assumed settings and its typical structure is the fixed-effects model [7] with additive terms which depends on single factors (main effects, linear effects), two factors (second-order interaction), three factors (third-order interaction) and eventually up to the highest-order interaction which merges all factors contained in the model. So rich model requires symmetrically equivalent rich experimental design: all combinations of all factor’s levels have to be tested experimentally.

Such design is known as \textit{full factorial}. It may be described by implication: if you need all possible interactions, you have to test all possible combinations of levels. It is clear that reverse question has to be ask: could I omit tests of some combinations of levels, if I needn’t all
interactions? The answer is positive and was given first by Fisher [2] in the form of *latin squares* and later by Yates [3] who developed two-level *fractional factorials*.

The sample structure of the fixed-effects model with three factors and terms up to the third-order interaction is defined as following:

\[ y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon \]  

(1)

where:

- \( y_{ijk} \) – the random variable describing the predicted output,
- \( \mu \) – the average output,
- \( \alpha_i \) – the linear (main) effect of the first factor \( \alpha \) at its level \( i \),
- \( \beta_j \) – the linear (main) effect of the second factor \( \beta \) at its level \( j \),
- \( \gamma_k \) – the linear (main) effect of the third factor \( \gamma \) at its level \( k \),
- \( (\alpha\beta)_{ij} \) – the second-order interaction effect of the factors \( \alpha \) and \( \beta \) at their levels \( i \) and \( j \),
- \( (\alpha\gamma)_{ik} \) – the second-order interaction effect of the factors \( \alpha \) and \( \gamma \) at their levels \( i \) and \( k \),
- \( (\beta\gamma)_{jk} \) – the second-order interaction effect of the factors \( \beta \) and \( \gamma \) at their levels \( j \) and \( k \),
- \( (\alpha\beta\gamma)_{ijk} \) – the third-order interaction effect of the factors \( \alpha, \beta \) and \( \gamma \) at their levels \( i, j \) and \( k \),
- \( \epsilon \) – the random term describing impact of all uncontrolled noise factors; typically the normal distribution with an unknown variance \( N(0, \sigma^2) \) is assumed, however it is only a hypothesis.

The equation (1) shows the general form of the stochastic model, but the numerical simulation like e.g. the bootstrap requires replacement of the random variable with its particular observations. It leads to smoothly different relationship:

\[ y_{ijk|r} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + r_{ijk|r} \]  

(2)

where:

- \( y_{ijk|r} \) – the output measured at levels \( i, j, k \) of the respective factors and at \( r \) repetition of the test,
- \( r_{ijk|r} \) – the residuum (difference between the real measurement of output and its prediction) observed at levels \( i, j, k \) of the respective factors and at \( r \) repetition of test; it should be noted that \( r \) symbol has two difference means: the residuum and the index of test repetitions (in the subscript).

Just the set of observed residuums \{ \( r_{ijk|r} \) \} will be the source for the future subsampling in the bootstrap approach. The algorithm’s steps are as following [9]:

a) the dataset of measurements \{ \( y_{ijk|r} \) \} is used to the identification of effects (see Eq.1) typically by the least-squares method,

b) the identified effects are used to predict output \{ \( \overline{y}_{ijk} \) \} for those levels \( i, j, k \) at which measurements were obtained,

c) the dataset of residuums \{ \( r_{ijk|r} \) \} is created based on differences between the observed measurements and the predicted outputs, according to the formula:

\[ r_{ijk|r} = y_{ijk|r} - \overline{y}_{ijk} \]  

(3)
d) the bootstrap subsampling is made by a random draw from the dataset of residuums \( \{ r_{ijkr} \} \),
added subsequently to the predicted outputs, resulting in the dataset of bootstrapped outputs:

\[
\hat{y}_{ijkr} = \overline{y}_{jk} + \text{draw}(r_{ijkr})
\]  

(4)
e) the bootstrapped outputs are used to identification of new estimates of model effects.
The steps (d) and (e) should be repeat many times (at least thousands times) resulting in a dataset of enough size to evaluate necessary statistics.

2.2 Response surface model

The response surface approach is typical for quantitative input variables. Their settings are continuous numerical values. The model predicts response for assumed settings and its typical used structure is a second-order polynomial [7] being a local example of Taylor series. Such model is a specific modification of a classic approximation problem due to a random term describing an impact of all uncontrolled noise factors:

\[
y = f(x_1, \ldots, x_i; b_0, b_1, \ldots, b_j) + \epsilon
\]  

(5)

where:

- \( y \) – the random variable describing the predicted output,
- \( f \) – the assumed function, usually first or second-order polynomial,
- \( x_i \) – \( i \)-th input variable,
- \( b_j \) – \( j \)-th model parameter,
- \( \epsilon \) – the random term describing impact of all uncontrolled noise factors; typically the normal distribution with an unknown variance \( N(0, \sigma^2) \) is assumed, however it is only a hypothesis.

Assumption of \( f \) function as the first or the second-order polynomial leads to very convenience situation: all model parameters are linearly set and may be identified by the least-squares method.

Further procedure has to be split. If a dataset of the output is collected in a designed experiment with strictly controlled input variables then the random term is \( \epsilon \) from stochastic Eq.5 is replaced with particular observed residuum \( r \) and the dataset of residuum will be the source for bootstrap draw. The next steps of the bootstrap approach are exactly the same a previously mentioned for the fixed-effect model.

But if the dataset of output is collected in a passive experiment and input variables are loosely controlled (if any), then a point from multidimensional space of observations \((x_1, \ldots, x_i; y)\) is established as a the random element. The algorithm’s steps are as following [9]:
a) the dataset of input values (formally input, but only observed in passive mode) and observed measurements \( \{(x_1, \ldots, x_i; y)_j\} \) is used to the identification of the basic model parameters (see Eq.5) typically by the least-squares method,
b) the bootstrap subsampling is made by a random draw from the dataset of input values and observed measurement \( \{(x_1, \ldots, x_i; y)_j\} \) resulting in a bootstrapped dataset of the same size \( \{ (\hat{x}_1, \ldots, \hat{x}_i; \hat{y})_j \} \), and a bootstrapped model with parameters \( \hat{b}_k \) is identified:

\[
\sum \left[ \hat{y}_{ij} - f(\hat{x}_{ij}, \ldots, \hat{x}_{ij}; \hat{b}_k) \right]^2 \rightarrow \min
\]  

(6)
c) the bootstrapped model and its output are used to identification of new estimates of model effects.
The steps (b) and (c) should be repeat many times (at least thousands times) resulting in a dataset of enough size to evaluate necessary statistics.

3 CASE STUDIES

Two case studies are presented for a bootstrap-based non-parametric estimation:
– estimation of the parameter’s significance for a fixed-effects model,
– estimation of the parameter’s significance for a response surface model in passive mode experiment.

3.1 Case study for a fixed-effects model

The investigation was conducted on the materials science matter: the analysis of of the ceramic shell mould of the airfoil blade casting [13]. One casting was selected for the microstructural and statistical analysis, and finally cut off. The cross-sections were included and prepared as metallographic samples from nickel-based superalloy. To reveal the microstructure of the investigated material the surfaces of the samples were etched, observed by a scanning electron microscope and obtained images analyzed by a computer-aided image analysis program to estimate quantitatively the main parameters describing the ($\gamma+\gamma'$) eutectic islands that occurred in the investigated superalloy. Details of the bootstrap-based simulation provided on the mentioned data are available in [13, 14] and here a briefly summary of analysis and results is included.

Dataset was created as six groups containing areas of eutectic islands identified and measured for each of six traces. The primary selected tool for the test of the traces homogeneity was classic ANOVA, but unfortunately the diagnostic test rejected the variances homogeneity, what is the basic assumption in ANOVA.

As an alternative method, the bootstrap-based identification of confidence intervals was provided. The one-way general linear model with six levels was identified to construct the residuum dataset, which is necessary to make random draw in the bootstrap. Next, the bootstrapped models were identified in 10 thousand iterations and parameters were collected to construct associated statistics (Figure 1).

![Figure 1: Distribution of the bootstrapped F statistics (source: [13])](image)
The bootstrap simulation experiment revealed that ANOVA results are rather uncertain. The bootstrapped $p$-Value is approximately equal to 0.75 while classic ANOVA $p$-Value obtained from experimental data gave 0.84. It leads to the conclusion that criterion $p$-Value should be treated more as a fuzzy basis for the decision, not the sharp one just on 0.05.

3.2 Case study for a response surface model in a passive mode experiment

The case study [15] used data obtained during investigation on compression vertebral fractures prediction based on computer tomography (CT) and microtomography ($\mu$CT) images. The significant difference in a resolution between these two class of images led to different prediction models. The small sample size (23 compressed and scanned vertebrae) and the high dimensionality of detected properties imposed the necessity of an alternative approach to the analysis, other than classic one derived with a requirement of the normality.

The RSM model was used to construct the prediction model to fit the observed strength of vertebrae with the crushing force. The bootstrap method was used to evaluate confidence intervals for effects. Finally, the existence of zero inside the intervals was inspected. If any of intervals contained zero, the null hypothesis of was rejected i.e. the parameter was treated as statistically insignificant.

The source dataset contained 23 records of three variables selected as predictors (input variables) and measured strength selected as output. The number of draw iterations was set to 10,000 to easy selection of quantiles from the bootstrapped dataset. After the full bootstrap procedure, the descriptive statistics were evaluated for model parameters. The bounds of the confidence intervals were easy identified due to the selected number of bootstrap iterations. They were values found at positions 250 and 9750 in the sorted bootstrapped results. Similarly, the bootstrapped $p$-Value was evaluated as relative position of sign switching inside the sorted bootstrapped results.

The obtained results led to the same decision: all linear coefficients significant and an intercept insignificant, but bootstrap-based results showed stronger significance of the coefficient than classic results.

4 CONCLUSIONS

- The non-parametric analysis based on the bootstrap approach allows to conduct effective analysis in such situations, when requirements of classic assumptions are not met.
- The bootstrap-based analysis requires proper identification of the i.i.d. random term (residuum vs. multidimensional point) but its implementation is easy and does not require additional subtle assumptions.
- The non-parametric bootstrap-based analysis is useful for processing data with a weak assessment of the distribution e.g. small datasets.
- The non-parametric bootstrap-based analysis may be easy automatized for processing huge or even big data, which distributions do not fit into any theoretical distribution.
- Further research will be conducted, especially in full-scale industry environment [16, 17, 18], because of engineers expectations about automatized and easy-to-interpret analysis methods.
REFERENCES


RANDOM AND GRADIENT BASED FIELDS IN DISCRETE PARTICLE MODELS OF HETEROGENEOUS MATERIALS

Jan Podroužek\textsuperscript{1,2}, Jan Vorel\textsuperscript{2}, and Roman Wan-Wendner\textsuperscript{2}

\textsuperscript{1}Faculty of Civil Engineering
Brno University of Technology
Veveří 331/95, Brno
podrouzek.j@fce.vutbr.cz

\textsuperscript{2}Institute of Structural Engineering
BOKU Wien
Peter-Jordan-Straße 82, 1190 Wien
\{jan.vorel, roman.wendner\}@boku.ac.at

Keywords: LDPM, random fields, gradient-based fields, spatial variability, coarse aggregates.

Abstract. The proper characterization of the microstructure and macroscopic properties of random heterogeneous materials may help to interpret the observed scatter in many engineering problems. In this contribution, a lattice discrete particle model is presented in a civil engineering framework with a special emphasis on concrete structures and infrastructure. The implications of spatial variability are investigated with regard to the observed scatter in standard concrete tests. In particular, classical experiments, such as three-point bending tests, may exhibit various levels of scatter which have to be accounted for if material characteristics are to be derived from such experiments and consequently used for the analysis and design of structural systems. Therefore, random and gradient based fields are combined in this paper in an attempt to realistically capture the material properties and associated variability stemming from the microstructure and macroscopic features such as the placement of coarse aggregates. Various mechanical and statistical aspects of simulated test series are investigated, such as macroscopic fracture energy or distribution of load bearing capacity. The presented application example of three-point bend specimens incorporates not only the inherent spatial variability owing to the placement of aggregates, modelled by random fields, but also various production artefacts, such as concrete compacting characterized here by gradient based fields.
1 INTRODUCTION

The estimation of uncertainties in probabilistic structural analysis may improve if the spatial variability concept is properly introduced. The presented paper addresses various discrete particle placement choices and schemes for discrete particle models such that the governing realization(s) of random or gradient-based fields is correlated to a particular structural discretization (i.e. radius and placement of particles) in a discrete framework. This work is partially motivated by the increasing demand in developing procedures for statistical estimates of structural response, including deteriorating structures, many of which are still in use due to socio-economic constraints, and by the increasingly available computational resources [1], [2]. This has prompted research into Monte Carlo (MC) based small-sample simulation methods [3], [4], where, despite the capacity of current computers, and in particular in the context of spatial variability, practical utilization requires the availability of an effective sampling strategy that would dramatically reduce the number of required realizations while maintaining accurate estimates of the response characteristics (low-probability large-consequence events) [5]. Recent attempts addressing the sampling strategy for spatial variability in the MC simulation framework include [6], [7], which is based on the original work of [8], where critical samples of stochastic processes are identified.

In this paper the main focus diverts from the reliability aspects of spatial variability and aims to investigate how particle placement schemes may be used to characterize various microstructural and macroscopic properties of random heterogeneous materials [9], while maintaining the material property fields constant. Any observed scatter can therefore be directly attributed to the chosen particle placement scheme and its parameters. This implies a causal relationship between spatial variability, auto-correlation length of the random fields, type of spectral function and meso/micro-structure of the material which is an open research question.

2 LDPM

The so-called lattice discrete particle model (LDPM) naturally accounts for material heterogeneity by random particle placement and size, which is also constrained by a grading curve [10]. This approach captures most microstructural effects of concrete very well, when compared to the continuum framework, however introducing higher order spatial variability enables to control and interpret the response scatter [6], [9].

As a well-established member of the discrete framework, LDPM has been extensively calibrated and validated and has shown superior capabilities in reproducing and predicting concrete behavior [10]–[14]. It simulates the mesostructure of concrete by a three-dimensional (3D) assemblage of particles that are generated randomly according to a given grain size distribution. Delaunay tetrahedralization and 3D domain tessellation are used here to generate a system of cells interacting through triangular facets. Displacements and rotations of such adjacent particles form the discrete compatibility equations in terms of rigid body kinematics. At each cell facet the mesoscale constitutive law is formulated such that it simulates cohesive fracture, compaction due to pose collapse, frictional slip and rate effect. For each single particle, the equilibrium equations are finally evaluated. An extended version of LDPM is currently being developed and simulates various coupled deterioration mechanisms, such as Alkali-Silica reaction (ASR) [12], [15]. A further development is the age-dependent LDPM framework in which the local material properties are derived by
chemo-mechanical coupling with a chemo-hygro-thermal model [16], [17] which also drives
the creep and shrinkage analysis in a rate type form [18].

3 ABSTRACTION LEVELS FOR LDPM

The proposed particle placement schemes have become part of the spatial variability
package for LDPM and, as will be shown in the next chapter, may influence not only the
scattering and asymptotic properties of the spatially variable models but also shift the mean
value of repeated identical experiments. By introducing the spatial variability package for
LDPM a significant contribution to the general understanding of physics and reliability of
spatial variability may be achieved.

The independent and random particle placement (IRPP) scheme, as implemented in the
original LDPM, may be considered as a first abstraction level. Here, the
random diameters are
generated according to the size distribution curve until the required volume fraction is reached
(figure 1). No conflicting requirements are to be solved. Overlapping or less than minimum
distance particles are resampled. Implications are scarcely populated boundary regions and an
invariant coefficient of variation. The original procedure for generating particles is described
in detail in [10].

Figure 1: Maximum (left) and minimum (right) strength realizations of LDPM using the random and
independent particle generation scheme.
The second abstraction level assumes the original particle placement scheme, i.e. the IRPP, combined with one or more spatially variable fields, which are used to describe local fluctuations of material properties resulting from its inherent variability (random field, figure 2) and construction or transport processes (gradient-based fields). Similarly to the previous case, there are no geometry-related conflicting requirements and overlapping or less than minimum distance particles are resampled. Boundary regions may be normally populated by adopting a simple modification to the re-sampling algorithm. Please note that material characterizations derived from random fields must be verified for inadmissible values, such as negative strength, modulus, etc. This may lead to a conflict if the governing probability distribution, used for the random field generation, is to be maintained. Otherwise, truncated distributions may be used or the realization of random field can be rescaled to fit admissible range [6], [19], [20]. This is a quite popular abstraction level, especially for continuum models, but stipulates full independence between microstructure, i.e. particle placement and material properties. The later assumption is reasonable for continuum models but questionable for lower-scale models such as LDPM.

![Figure 2: Sample realization of a Gaussian random field with a power spectral function $#^2$ as a material property field only.](image)

The third abstraction level assumes that an initial random or gradient-based field of choice (or their arbitrary combination) governs the particle generation process (i.e. the position and size of each particle). If the particle generation is to be governed not only by granulometric distributions, but also by an initial random (figure 3) or gradient-based field (figure 4), the particle generation becomes a complex problem and has to be approached by balancing trade-offs between conflicting goals. Clearly, the global requirement on particular size distribution can lead to a local conflict with the initial random field, the role of which can be further ambiguous if we consider it to affect both the position and size of the particles (clustering of large particles). Details regarding the associated steps/choices for random fields were published in [9]. For higher volume fractions this becomes a computationally expensive procedure, however local conflicts can be resolved in parallel and terminate with the first valid particle.
Alternatively, the initial random or gradient based field may not only govern the particle generation process but also be used to modify the material property fields and thus maintain compatibility between the two domains.

Figure 3: Maximum strength realizations of LDPM experiments based on a governing Gaussian random field with a power spectral parameter ($\nu^{-2.00}$) maximizing the COV of load capacity.
RESULTS

Selected observation regarding the statistical characterization of parallel identical LDPM experiments and various particle placement schemes are discussed in this chapter. Two classical experiments for concrete, the cylinder compression test (D150x300 mm) and notched three point bending test (400x400x100 mm), have been statistically reproduced by LDPM and the spatial variability package. Each of the presented numerical experiments is characterized by 20 realizations. The same sequence of seeds for the random number generation is maintained throughout the experiment. Figure 1 shows the ensemble of particles for the abstraction level 1, i.e. the independent and random particle placement scheme for the maximum and minimum strength realizations. Figure 2 conceptually depicts the abstraction level 2, i.e. particular realization of a random material property field which is combined with independently and randomly generated ensemble of particles. Figure 3 shows the maximum strength realizations of a gradient and a Gaussian random field with a power spectral function \( f(x) = #^{-2.00} \) for the cylindrical compression test and three point bending test (notched). Figure 4 shows the maximum strength realizations of cylindrical compression tests with gravity-based gradient field (the failure is depicted in figure 5) and normal-to-gravity-based gradient field.
field. The consequence of rotating the gravity-based gradient field is evident from figure 6 where the scattered load-displacement (LD) diagrams are shifted upwards for the rotated version (figure 4 right), i.e. the mean capacity is increased by ~2%.

Finally, the normalized statistical characteristics are plotted in figure 7 against various power spectral parameter values. Here, the 20 cylindrical compression tests share the same material property fields (constant) and differ only by the particular realizations of the particle generation, in this case based on the abstraction level 3 using a Gaussian random field with power spectral function $\#^n$. For the reference, the dashed line is added into the plot to mark the zero correlation length, which corresponds to the level 1 IRPP scheme. Please note that the range of COVs for the cylindrical compression test series is 0.42 to 0.66 % and for the notched three point bending test series 2.11 to 3.71 %, considering the various particle placement schemes only.

Figure 5: Visualization of a crack opening displacement (red color ~ 0.7 mm) at peak load at maximum strength realization of cylindrical compression tests with gravity-based gradient field (fig. 2 left).
Figure 6: Scattering and shifted L-D diagrams (zoomed-in detail) of LDPM realizations of compression cylinder tests with two gradient-based fields (gravity angle is orange and normal to gravity angle is blue).

Figure 7: Normalized statistical characteristics vs. power spectral parameter $n$; response of cylindrical compression test series (20 realizations) and Gaussian random field with power spectral function $#^n$. The dashed line represents the zero correlation reference, i.e. the IRPP scheme.
5 CONCLUSIONS

- The inverse U-shaped curve for the coefficient of variation (figure 7) and various power spectral parameters clearly supports the hypotheses on a causal relationship between spatial variability, auto-correlation length of the random fields, type of spectral function and meso/micro-structure of the material. Please note that the material property fields are constant for all presented results.

- Although the relative differences resulting from the presented abstraction level 3 (i.e. correlated particle placement scheme and constant material property fields) may not always be significant, the ability to obtain the functional form for mean values and COV of response series represents a fundamental concept for spatial variability based structural reliability.

- The relative differences in terms of COV for parallel identical test series may increase dramatically if (a) the material property fields are randomized, (b) engineering failure probabilities ($\sim 10^{-4}$) are introduced (by effective sampling strategy) rather than a first passage probability based on 20 realizations, and (c) structural components or structures are analyzed that are sensitive to damage localization leading to, e.g., the statistical size-effect.

- Even if the material property field remains constant, the various particle placement schemes derived from initial random or gradient based field may be used to control not only the scattering of simulated experiments, but also to shift the mean values of the response.

- Several abstraction levels for the lattice discrete particle models of concrete have been presented, together with a particle placement schemes derived from initial random or gradient based fields.

- The proposed spatial variability package for LDPM may dramatically increase the consistency and realism of the LPDM paradigm if a physical reference for the governing spatially variable field is established.

- Reliability of spatially variable LDPM models will require an objective statistical characterization in a small sample Monte Carlo simulation framework.

ACKNOWLEDGEMENT

The financial support by the Austrian Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development is gratefully acknowledged. The computational results presented have been achieved in part using the Vienna Scientific Cluster (VSC). The support of the project No. LO1408 "AdMaS UP - Advanced Materials, Structures and Technologies", supported by Ministry of Education, Youth and Sports under the „National Sustainability Programme I“ is also greatly acknowledged.
REFERENCES


FATIGUE DAMAGE MODEL OF WIND TURBINE COMPOSITE BLADES UNDER UNCERTAIN WIND SPEED

Chi Zhang¹, and Huapeng Chen¹*

¹Department of Engineering Science, University of Greenwich
Chatham Maritime, Kent, ME4 4TB, UK
{c.zhang,h.chen}@gre.ac.uk
*Corresponding author: h.chen@gre.ac.uk

Keywords: Composite Blade, S-N Curve, Goodman Diagram, Fatigue, Wind Speed, Blade Element Momentum.

Abstract. This paper presents a load spectrum and estimates the fatigue life of composite blades of horizontal axis wind turbines. The distributions of aerodynamic loads are analysed by beam element momentum theory, which affects the fatigue life of the blade along with other loads such as gravity and wind shear. The wind speed follows Weibull distribution and wind speed over an hour time is generated based on the offshore wind parameters. Then, the maximum stress caused by the deterministic dynamic loads, such as aerodynamic load, gravitational pressure and wind shear effect, is calculated using finite element model by Ansys. The stress cycle is established by Fast Fourier transform of the load spectrum which converts the time domain to frequency domain. Fatigue damage performance and the fatigue life estimate for blades are used to be predicted by fatigue damage rule for each cycle during the service life of offshore wind turbine based on the Goodman diagram and S-N curve. Finally, safe working life is predicted by applying the Miner’s law for linear fatigue damage accumulation. A numerical example of composite blades of a wind turbine is investigated by proposed method to estimate fatigue life under uncertain wind speed in the offshore environment. From the results, the proposed method can provide an effective tool for evaluating the fatigue damage and assessing structural performance of the wind turbine blades under the uncertain offshore wind.
1 INTRODUCTION

With the increasing energy consumption in the world, renewable and clean energy such as wind resource have been considered as an alternative way to resolve the energy crisis. The wind turbine is one of offshore structure that captures the power of the wind and use it to generate clean energy. Since the rate of offshore wind turbines increases to multi-MW-size, one of the critical elements is the large composite blade due to the huge cost and fatigue damage. To adapt harsh marine environments and to resist fatigue load, glass-fibre reinforced composite materials are typically used to manufacture blades since these materials have better fatigue resistance and lighter weight than traditional materials, e.g., metals. One challenge in fatigue damage estimation is that the fatigue damage of reliable wind turbine blades is under complex load conditions, such as wind load, gravitational load, and wind shear effect [1,2]. Among these loads, the wind speed is significantly affected by the environmental condition with relatively high frequency and amplitude in offshore environments, which could be generated by deterministic distribution [3-5]. It is important to develop a method for analysing the blade fatigue damage before the blades fail catastrophically, which may destroy the entire wind turbine.

Most ongoing research on fatigue phenomena uses load spectra obtained by digital sampling or finite element (FE) model of wind turbine blades, which estimated the numerical solution at a location near the root of the blade [4]. Then, to obtain the cycles, traditional counting methods such as Rainflow Counting is used in [3,6]. However, computationally signal processing techniques (such as Fast Fourier transform) can also be used to calculate the fatigue cycles according to the frequency of the load spectra [7]. Based on Goodman diagram and S-N curve, the equivalent load spectra for each cycle can be investigated [7,8,9]. This stress is utilised to estimate fatigue damage in the blade using Miner’s rule [5,6], and then fatigue life is estimated.

This study presents a method for investigating the fatigue damage of a composite wind turbine blade. The Fast Fourier transform is used here to obtain the load cycles, based on the data of stress spectrum by FE model solution. The Miner’s law is used to predict the fatigue damage during the designed service life of offshore wind turbine by the S-N curve and Goodman diagram. A numerical example of composite blades of a wind turbine is investigated to estimate fatigue damage under uncertain wind speed, following Weibull distribution describing the wind in the offshore environment. Results show that FE model can take the variable wind speed with other blade loads into account and the frequency of stress can be obtained. Then, the stress in each cycle can be converted into fatigue stress. The proposed method can be further used for determining the resistance of the composite blades based on S-N curves. The fatigue damage of composite blades can be predicted by Miner’s law based on combining fatigue stress and life cycles in each cycle. The fatigue life can be calculated before the cumulative damage exceeds unity. From the results, the proposed method can both provide a useful tool for evaluating the fatigue damage and assess structural performance of the wind turbine blades under the uncertain offshore wind.

2 ANALYSIS OF BLADE LOADS

Loads of the wind turbine blades contain both deterministic and stochastic load under different wind speed, such as aerodynamic load, wind shear, gravity load and periodic load [1,2]. However, the loads can be imported into the finite element model in fatigue analysis to investigate the maximum stress in the blade.

2.1 Aerodynamic loads

Blade element momentum (BEM) method is the core theory to analyse wind turbine aerodynamics on the blades. This BEM theory treats the revolving rotor as an actuator disc that
combines both blade element theory and one-dimensional momentum theory in both the rotor axial and tangential direction [2]. Both rotor axial and tangential induction factors that describe the airflow speed change are introduced in the one-dimensional momentum method to calculate the wind thrust and rotor motivation torque. The blade element theory separates the blade into several elements and ignores the mutual influence between two adjacent elements. The aerodynamic loads on each element are dependent on its local airfoil characteristics, i.e. its lift and drag coefficients. The sum of these loads yields the total loads on the blade in X and Y directions [1], respectively, namely

\[ F_{Yp} = \int \frac{B}{2\pi} \rho \int_0^{2\pi} \int_0^R (C_L \cos \alpha + C_D \sin \alpha) c \omega^2 d\theta dr \]  

\[ F_{Xp} = \int \frac{B}{2\pi} \rho \int_0^{2\pi} \int_0^R (C_L \sin \alpha - C_D \cos \alpha) c \omega^2 d\theta dr \]  

where \( B \) is the number of blades; \( \rho \) is the air density; \( c \) is the length of chord; \( \omega \) is the angular velocity of the rotor; \( r \) is the radius of the section; \( \theta \) is the azimuth angle of the rotor; \( C_L \) is the airfoil lift coefficient; \( C_D \) is the drag coefficient; and \( \alpha \) is inflow angle.

Therefore, the flapwise moment and edgewise at the root of the blade are

\[ M_{Xp} = \int_0^R r dF_{Yp}, \quad M_{Yp} = -\int_0^R r dF_{Xp} \]  

where \( p \) represents the pressure centre of the airfoil.

2.2 Wind shear

The wind shear effect depends on the height of the wind turbine, and the description of this relationship is exponential. Since the wind turbine becomes MW recently, the radius of the blade increases significantly. The wind speed will significantly change in different positions of the blade, so the wind shear effect should be considered. According to [10], the wind speed at the height of the sea level can express as

\[ V(y) = V_h (1 + \frac{r \cos \theta}{h})^m \]  

where \( h \) is the height of the hub of the wind turbine, \( V_h \) is wind speed at the hub, \( m \) is wind shear exponent, normally between 0.1 and 0.25.

2.3 Gravity

For large wind turbine blades, the blade is subjected to gravity during service time, which is periodic changing by azimuth angle \( \theta \) of the rotor. The gravity per element of the blade on each axis can be obtained as

\[ F_g = \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix} \begin{bmatrix} 0 \\ M_g \end{bmatrix} \]  

where \( F_g \) is the gravity loads, \( M \) is the mass of the blade, and \( g \) is the gravity acceleration. The gravity load is considered as an axial load, which does not vary with time.

2.4 Distribution of wind speed

According to [3], the Weibull frequency distribution is adopted to describe the wind speed. Therefore, the probabilistic density function \( f(v) \) and cumulative distribution function \( F(v) \) are

\[ f(v) = \frac{K}{c} \left( \frac{v}{c} \right)^{K-1} e^{-\left( \frac{v}{c} \right)^K} \]  

\[ F(v) = 1 - e^{-\left( \frac{v}{c} \right)^K} \]
where \( K \) is the shape parameter and \( C \) is the scale parameter of the distribution.

3 FATIGUE ANALYSIS

Fatigue analysis has been considered as a critical part of the design process of wind turbine blades. According to the results of different wind speed, the loads in wind turbine blades are established. Fatigue life was estimated by using the Fast Fourier transform for the cycle, Goodman diagrams and the load spectrum to get the fatigue stress and the S–N linear damage equation for accumulating the damage [5,9]. By combining these methods, a dynamic structural problem for the fatigue can be converted into a static structural problem.

3.1 Finite element blade model

In this study, the finite element model is introduced in Ansys software for fatigue analysis under various wind speeds to get the maximum stress. A composite blade considered in this study is the same geometry to the 1.5MW Sandia reference wind turbine composite blade [4].

![Figure 1: Structural design of 1.5MW wind turbine blade by NuMAD based on the Sandia database [4]](image)

The aerodynamic properties, e.g., airfoil type, chord length, and twist angle, are the same as those from the design data. A refined blade geometry model (Figure 1) was generated in NuMAD by connecting 24 airfoils, which smooth the transition from section to section and reduce the stress concentration.

![Figure 2: The meshed and geometrical model of wind turbine blades based on 1.5MW Sandia using Ansys 18.0](image)
The NuMAD can export the 3D structural design into Ansys file. The meshed model (Figure 2) generated by Ansys 18.0, which implements layered shell 281 elements for the structure. By applying the blades loads into the model, the solution of maximum stress can be analysed in FE model. In this study, the maximum stress is assumed as the fatigue stress in these loads to calculate the fatigue damage in this certain wind speed.

3.2 Fast Fourier transform

According to the previous section, it generates random wind speed numbers by the Weibull distribution with scale parameter \( C \) and shape parameter \( K \) during the given period. The Fast Fourier transform (FFT) is used here to obtain the fatigue cycles, which will be used to analyse the uncertain stress during the particular time. A dynamic structural problem for the fatigue can be converted into a static structural problem [7].

Fourier transform converts the time domain to frequency domain by integrating over the axis of whole time. The Fast Fourier transform technique is used to derive the wind frequency by converting the time-domain component of wind stress into frequency from a spectrum using the transfer equation

\[
X(u) = \int_{-\infty}^{\infty} f(x) e^{-2\pi iux} dx
\]

where \( X(u) \) is the frequency domain function, \( u \) is the spatial frequency, and \( f(x) \) is the spatial domain function.

By the FFT method, the frequency for the uncertain wind speed during the measuring time is established. The fatigue cycles by wind speed can be therefore used to analyse the damage for the wind turbine blades.

3.3 Goodman diagram

Constant life diagrams are graphical representations of the safe regime of constant amplitude loading for a given specified life, e.g. the endurance limit or infinite life [6]. These diagrams can be drawn in some ways, depending on which parameters are selected to describe the constant amplitude cyclic loading, such as Goodman diagram [9].

![Figure 3: An example of Goodman diagram showing amplitude stress \( \sigma_{eqa} \) vs. mean stress \( \sigma_{eqm} \)](image)

Constant amplitude cyclic loading can be defined by specifying the following parameters, i.e. maximum stress and minimum stress. The Goodman diagram can give the equivalent stress
under different mean stress. A simple Goodman diagram without considering the stress ratio can be expressed as

$$\sigma_{eq} = \frac{\sigma_a}{\sigma_m-\sigma_u} (\sigma_{eqm} - \sigma_u)$$

(9)

where $\sigma_{eq}$ is the equivalent stress in this cycle when the equivalent mean stress $\sigma_{eqm}$ is given, $\sigma_a$ is the amplitude stress, and $\sigma_m$ is the mean stress in this cycle, $\sigma_u$ is the ultimate stress for the material.

Figure 3 gives an illustrative Goodman diagram used in this study. The fatigue stress should use the equal amplitude stress when the average stress is null ($\sigma_m=0$) by Goodman diagram for the wind turbine blade, and this amplitude stress is expressed as $\sigma_e$.

### 3.4 S-N curve

To calculate the damage of composite blades by S-N curve, it is necessary to know the fatigue life cycles under constant equal amplitude loading obtained by Goodman diagram. Fatigue properties of materials are usually expressed in the form of S-N curves according to the experimental data, i.e. stress amplitude versus the number of cycles to failure.

According to [6,8,9] a typical following S–N curve model can be equalled as

$$\sigma_e = A - BlgN, \quad \text{or} \quad r = \frac{\sigma_e}{\sigma_u} = a - blgN$$

(10)

where $r=\sigma_e/\sigma_u$ is the stress ratio, $A$ and $B$ are parameters depending on the material, which also can be expressed as $aa_u$ and $b\sigma_u$. Figure 4 represents the relationship between the stress ratio and the stress cycles to calculate the fatigue damage.

The fatigue life cycles $N$ can be obtained in this graph given an amplitude stress $\sigma_e$. The damage in this cycle is $1/N$ by assuming the total damage is unity.

### 3.5 Miner law

The fatigue damage due to each stress component can be accumulated by Miner’s rule. According to Miner’s rule, the structure is defined as failure when the cumulative damage $D$ exceeds unity [1,5,9], where $D$ is defined as

$$D = \sum_i \frac{n_i}{N_i}$$

(11)
where \( n_i \) is the load cycle number for a stress range \( i \), and \( N_i \) is the allowable load cycle number associated with \( i \).

By the method mentioned above, the damage per hour can be generated and the total hour of fatigue life can be estimated. To achieve the design code of wind turbine concerning the total service life, some maintenance strategies may increase the performance of the wind turbine blade and prevent the structural failure.

4 NUMERICAL EXAMPLE

Based on the method described in the previous section, a numerical example can be performed to demonstrate the effectiveness of the proposed method. In this study, fatigue damage in composite blades of the wind turbine under offshore wind condition is analysed, and then the fatigue life can be estimated.

The measurement of deterioration in this paper is therefore based on the Miner’s law. When the cumulative fatigue damage exceeds unity, the composite blade is considered as it loses its mechanical properties and could not service anymore. Before this situation happens, the blade structure can be treated as nearly failed and a proper maintenance strategy will be taken.

In the following example, the 1.5MW Sandia reference wind turbine composite blade [4] is analysed according to the method discussed above. The aerodynamic properties, e.g., airfoil type, chord length, and twist angle, are the same as the design data. The Weibull distribution of the wind speed of European offshore wind farm for the whole year is estimated, where the shape parameter \( K = 2.0 \) and the scale parameter \( C = 15 \text{ m/s} \). Therefore, the wind field simulation per minute for 1 hour can be randomised in MATLAB, as shown in Figure 5.

![Figure 5: Prediction of wind speed generated by Weibull distribution and results of maximum stress to time spectrum during one hour](image)

From the results, the aerodynamic loading can be separated into lift loading and drag loading along with each blade element by BEM method. The offshore wind turbine is typically located just above the sea level thus the wind shear effect could not be ignored. The height of the wind turbine hub is assumed 50 meters, and wind shear exponent is taken 0.1 in this study. The gravity in this study considered as an axial load in the gravity centre of the blade.

The solution of the FE model generated by NuMAD in Ansys involves the blade loads including aerodynamic load, wind shear, and gravity load. The maximum stress in certain wind speed can be regarded as equivalent stress in load analysis for this 1.5 MW blade. The stress to time diagram for 1-hour given by wind field simulation is established in Figure 5. The spectra
of maximum stress versus wind speeds are shown in Figure 6 according to the existing data from [1] by wind turbine blade of FE model.

![Figure 6: The maximum stress in various wind speed](image)

The maximum stress in wind turbine blade depends on the wind speed shown in Figure 6. As the wind speed increases, the stress also increases. However, after reaching a certain wind speed, the maximum stress reduces and then becomes stable. When the wind speed exceeds the design wind speed for the wind turbine, the attack angle decreases, leading to the reducing wind surface. The wind turbine keeps the stable angular velocity to produce the stable power when the turbulence of wind speed is high.

To get the cycle of the fatigue stress in this measuring time, the FFT is used to obtain the frequency during one hour. The cycles can be used to analyse the fatigue damage for the wind turbine blades. Figure 7 shows the result of the frequency of the stress during the time. The first five peak points in amplitude are used to establish the stress cycle during the time.

![Figure 7: The frequency of amplitude and phase by FFT](image)

The number of the cycles can be obtained by the frequency and the whole period. Each cycle has the maximum stress $\sigma_{max}$ and minimum stress $\sigma_{min}$, and therefore the mean stress $\sigma_m$ and amplitude $\sigma_a$ for this cycle can be calculated. For the glass-fibre material, assuming the ultimate...
stress $\sigma_u$ in Goodman diagram $\sigma_u=100\text{MPa}$, the equivalent stress $\sigma_e$ at $\sigma_{eqm}=0$ can be obtained. According to S-N curve, the life cycles $N_i$ at $\sigma_e$ can be calculated. The $a$ and $b$ for one type of glass fibre material is 1.3 and 0.16, respectively. The Miner’s law accumulates damage by the life cycles, and the failure will occur when the damage accumulates to a certain critical value of unity.

Table 1 shows results of the maximum stress $\sigma_{max}$, minimum stress $\sigma_{min}$, the mean stress $\sigma_m$ and amplitude $\sigma_a$ for every cycle in each peak point. The equivalent stress $\sigma_e$ is obtained by Goodman diagram for glass fibre material. The life cycles $N$ in every cycle is predicted by the proposed S-N curve in this table since the equivalent stress $\sigma_e$ has been obtained in each cycle. Then the fatigue damage $D$ is reciprocal of life cycle $N$ obtained by S-N curve in this cycle. From this table, the total damage is 7.55E-06 by summing all the damage in each cycle in this 1 hour.

<table>
<thead>
<tr>
<th>Peak</th>
<th>$\sigma_{max}$ (MPa)</th>
<th>$\sigma_{min}$ (MPa)</th>
<th>$\sigma_m$ (MPa)</th>
<th>$\sigma_a$ (MPa)</th>
<th>$\sigma_e$ (MPa)</th>
<th>$N$ (cycles)</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49.40</td>
<td>5.87</td>
<td>27.63</td>
<td>21.77</td>
<td>30.08</td>
<td>1.76E+06</td>
<td>5.68E-07</td>
</tr>
<tr>
<td>cycle 2</td>
<td>49.35</td>
<td>20.94</td>
<td>35.14</td>
<td>14.21</td>
<td>21.90</td>
<td>5.70E+06</td>
<td>1.75E-07</td>
</tr>
<tr>
<td>2</td>
<td>49.36</td>
<td>5.87</td>
<td>27.61</td>
<td>21.75</td>
<td>30.04</td>
<td>1.77E+06</td>
<td>5.66E-07</td>
</tr>
<tr>
<td>cycle 4</td>
<td>49.40</td>
<td>5.87</td>
<td>27.63</td>
<td>21.77</td>
<td>30.08</td>
<td>1.76E+06</td>
<td>5.68E-07</td>
</tr>
<tr>
<td>4</td>
<td>49.35</td>
<td>20.94</td>
<td>35.14</td>
<td>14.21</td>
<td>21.90</td>
<td>5.70E+06</td>
<td>1.75E-07</td>
</tr>
<tr>
<td>5</td>
<td>48.84</td>
<td>37.83</td>
<td>43.34</td>
<td>5.51</td>
<td>9.72</td>
<td>3.29E+07</td>
<td>3.04E-08</td>
</tr>
<tr>
<td>cycle 6</td>
<td>49.36</td>
<td>5.87</td>
<td>27.61</td>
<td>21.75</td>
<td>30.04</td>
<td>1.77E+06</td>
<td>5.66E-07</td>
</tr>
<tr>
<td>3</td>
<td>49.40</td>
<td>5.87</td>
<td>27.63</td>
<td>21.77</td>
<td>30.08</td>
<td>1.76E+06</td>
<td>5.68E-07</td>
</tr>
<tr>
<td>cycle 8</td>
<td>49.26</td>
<td>5.87</td>
<td>27.56</td>
<td>21.70</td>
<td>29.95</td>
<td>1.79E+06</td>
<td>5.58E-07</td>
</tr>
<tr>
<td>4</td>
<td>49.35</td>
<td>20.94</td>
<td>35.14</td>
<td>14.21</td>
<td>21.90</td>
<td>5.70E+06</td>
<td>1.75E-07</td>
</tr>
<tr>
<td>6</td>
<td>49.17</td>
<td>37.83</td>
<td>43.50</td>
<td>5.67</td>
<td>10.03</td>
<td>3.15E+07</td>
<td>3.18E-08</td>
</tr>
<tr>
<td>8</td>
<td>48.84</td>
<td>37.92</td>
<td>43.38</td>
<td>5.46</td>
<td>9.65</td>
<td>3.33E+07</td>
<td>3.01E-08</td>
</tr>
<tr>
<td>cycle 9</td>
<td>49.36</td>
<td>5.87</td>
<td>27.61</td>
<td>21.75</td>
<td>30.04</td>
<td>1.77E+06</td>
<td>5.66E-07</td>
</tr>
<tr>
<td>5</td>
<td>49.40</td>
<td>5.87</td>
<td>27.14</td>
<td>21.27</td>
<td>29.20</td>
<td>2.00E+06</td>
<td>5.01E-07</td>
</tr>
<tr>
<td>cycle 10</td>
<td>49.26</td>
<td>5.87</td>
<td>27.56</td>
<td>21.70</td>
<td>29.95</td>
<td>1.79E+06</td>
<td>5.58E-07</td>
</tr>
<tr>
<td>7</td>
<td>49.35</td>
<td>20.94</td>
<td>35.14</td>
<td>14.21</td>
<td>21.90</td>
<td>5.70E+06</td>
<td>1.75E-07</td>
</tr>
<tr>
<td>9</td>
<td>49.17</td>
<td>37.83</td>
<td>43.50</td>
<td>5.67</td>
<td>10.03</td>
<td>3.15E+07</td>
<td>3.18E-08</td>
</tr>
<tr>
<td></td>
<td>48.84</td>
<td>41.99</td>
<td>45.42</td>
<td>3.43</td>
<td>6.28</td>
<td>5.40E+07</td>
<td>1.85E-08</td>
</tr>
<tr>
<td></td>
<td>48.41</td>
<td>5.87</td>
<td>27.14</td>
<td>21.27</td>
<td>29.20</td>
<td>2.00E+06</td>
<td>5.01E-07</td>
</tr>
<tr>
<td></td>
<td>49.26</td>
<td>37.88</td>
<td>43.57</td>
<td>5.69</td>
<td>10.08</td>
<td>3.13E+07</td>
<td>3.20E-08</td>
</tr>
<tr>
<td></td>
<td>49.35</td>
<td>5.87</td>
<td>27.61</td>
<td>21.74</td>
<td>30.03</td>
<td>1.77E+06</td>
<td>5.65E-07</td>
</tr>
<tr>
<td></td>
<td>49.17</td>
<td>30.43</td>
<td>43.01</td>
<td>5.18</td>
<td>9.08</td>
<td>3.61E+07</td>
<td>2.77E-08</td>
</tr>
<tr>
<td></td>
<td>48.84</td>
<td>37.92</td>
<td>43.38</td>
<td>5.46</td>
<td>9.65</td>
<td>3.33E+07</td>
<td>3.01E-08</td>
</tr>
<tr>
<td></td>
<td>48.84</td>
<td>41.99</td>
<td>45.42</td>
<td>3.43</td>
<td>6.02</td>
<td>5.40E+07</td>
<td>1.85E-08</td>
</tr>
</tbody>
</table>

$\sum D$ 7.55E-06

Table 1: The fatigue loads and damage in each cycle during 1-hour measuring period

The estimated fatigue life of this glass fibre reinforced composite blade is
\[ Y = \frac{1}{\sum_{D \times 24 \times 365}} = 15.1 \text{ year} \quad (1) \]

Since the design code expects the service time of offshore wind turbine blades is 25 years, some proper maintenance strategies should be taken during the design process.

5 CONCLUSIONS

In this study, the fatigue life was estimated by linear Miner’s law based on the well-known S–N linear damage equation, the equivalent load spectrum by Goodman diagram with analysing the blade loads in FE model of the wind turbine blade. The calculation of the load spectrum and the fatigue performance of wind turbine blade is discussed by analysing the blade loads in FE model using the deterministic method.

From the sample spectrum data during short period operation, the data were rearranged load cycles, maximum stress, minimum stress, mean stress, the amplitude of the stress, and equivalent stress by FFT and Goodman diagram, and then reordered as life cycle and damage for each cycle by S–N curve. Based on the Miner’s law, the accumulation of fatigue damage can be obtained during the sample measuring time. The total service time of the wind turbine blade is estimated as 15.1 years without maintenance. Further work is needed to determine the optimum repair strategy for the composite blades of an offshore wind turbine according to the obtained results.

REFERENCES


RELIABILITY OF PRESTRESSED CONCRETE BRIDGE GIRDERS USING FIELD INFORMATION AND THE COMBINED APPROACH

Adrián David García-Soto¹, Alejandro Hernández-Martínez², Jesús Gerardo Valdés-Vázquez²

¹Universidad de Guanajuato, Av. Juárez 77, Zona Centro, C. P. 36000, Guanajuato, Gto., México
e-mail: adgarcia@ugto.mx

²Universidad de Guanajuato, Av. Juárez 77, Zona Centro, C. P. 36000, Guanajuato, Gto., México
{alejandro.hernandez, valdes}@ugto.mx

Keywords: Prestressed Concrete, Bridge Girders, Reliability, Combined Method, Field Information.

Abstract. Field information on live loads and concrete strength is used to assess the reliability of a prestressed precast concrete girder of an existing bridge in Guanajuato State (Central Mexico). A recently developed method, termed the combined approach (CA), is used to perform the reliability analysis. The approach combines two well-known techniques: the first order reliability method (FORM) and the point estimate method (PEM). Information on the bridge design, live load from weigh-in-motion data and compressive strength of concrete cores, was gathered for this study through two research projects. The information is employed to probabilistically characterized the demand and the capacity of the bridge prestressed concrete girders. A numerical scheme is used to obtain the girder bending capacity, and the bending capacity of the point estimates for the reliability analysis, by considering common design practices in Mexico. The use of a numerical method can be required, or more practical, if non-common geometries and prestressed steel and/or reinforcement steel arrangements are used (e.g., non-tension-controlled girders). This implies that the limit state function (LSF) could only be implicitly known, and the FORM may not be feasible, while the Monte Carlo simulation technique (MCS) may require extensive samples. The present study shows the results of the reliability levels for the bridge girder for a range of mean live to dead load effects. This reliability is obtained with information directly gather on field for the bridge under study, for a traffic jam scenario and for a single vehicle passage scenario. It is concluded that the CA can be an adequate alternative to perform the reliability analysis, especially when there is no explicit limit state function (as for instance required in the FORM), and without the need of extensive simulations (unlike the MCS). Some codified design implications are discussed from the resulting reliability levels, and future research for several issues is recommended.
1 INTRODUCTION

The reliability analysis of prestressed concrete bridge girders is used to estimate the probability of failure of these elements when subjected to load demands of different kind. The computed probabilities of failure are linked to the reliability index, commonly employed to establish code formats to achieve certain safety consistency. The reliability levels of prestressed precast concrete bridge girders is reported in the literature [1, 2]. However, most of the studies use typical bridge sections, and statistics and probabilistic information from other sources, rather than using existing bridges and deriving probabilistic information from field information and/or real bridge projects expressly selected for reliability studies.

Recent studies on reliability of prestressed concrete beams and bridge girders include sensitivity analysis using updated databases on materials [1], comparison of different standards [2], spatial and time depend reliability under corrosion and/or cracking effects [3, 4, 5], beams exposed to fire [6] or considering creep models [7], and calibration tasks [8], among other issues.

In Mexico, there are also some studies on reliability of prestressed concrete bridges dealing with optimal inspection times considering corrosion [9, 10], which use some local guidelines for designing precast prestressed concrete beams. In this study the AASHTO regulations are adopted [11] to obtain the flexural capacity of the bridge girder, since this is not an uncommon practice among Mexican practitioners, according to one of the authors of this study with experience in the field. Nevertheless, the load factors and live loads are based on previous studies, and weigh-in-motion (WIM) data recorded at highways located at Guanajuato State in Central Mexico.

Other field information that is necessary for reliability studies of existing bridges, is that related to the capacity of the prestressed concrete bridge girders. In the present study, the compressive stress, required in the methods to compute the bending capacity of prestressed members, is derived from concrete cores directly extracted from existing bridges and tested to obtain relevant parameters. The bending capacity of the considered prestressed concrete girder is obtained by considering the equilibrium of the section including, both, the reinforcement steel and the prestressing strands, and a numerical scheme (instead of any formulae, which applicability could be limited to certain cases).

There are several available reliability methods to assess the safety levels of bridge structures [12, 13]. Some of them require the establishment of an explicit limit state function, like the first order reliability method (FORM), or a very extensive number of simulations, like the Monte Carlo simulation techniques (MCS). Reliability indices are obtained by using an approach which can leave aside the use of an explicit LSF, using only a few point estimates. The method is termed combined approach [14], since it combines the use of other known techniques, the FORM and the point estimate method (PEM) [15-18].

The assessment of reliabilities of prestressed concrete bridge girders is important to understand the implicit reliability levels for codified design, the possible impact of loads and resistances for a specific geographic and socioeconomic region, and the design implications.

Therefore, the main objective of this study is to obtain the reliability of a prestressed concrete girder of an existing bridge in Guanajuato, Mexico, subjected to bending, using field information on the demand and the capacity, and the combined approach.

2 FIELD INFORMATION

This section is divided in three parts. First, a general description of selected bridges in Guanajuato for research purposes is briefly reported; then, the recorded and used weigh-in-
motion data is outlined; finally, a summary of the tests from concrete cores extracted from the selected bridges is succinctly described.

2.1 Existing bridges

As a part of two research projects financially supported by two Mexican institutions, one to investigate the concrete capacity sponsored by PROMEP (Project “Nuevos PTC”, UGTO-PTC-429), and the other one to investigate the live load demands sponsored by CONACYT (Project “Problemas Nacionales 2014”, No. 248162), five existing bridges in Guanajuato State in Central Mexico, a key transportation region in the country [19], were selected. Prestressed precast concrete medium span bridges were selected for the research, since these are a very common type of bridges in Guanajuato, and possibly in the whole country [20]; note that bridges of this kind are also very common in other countries, like Mainland China, Hong Kong and the United States [2]. These bridges were also chosen because they are located on (or close to) the highways where the WIM is (or is to be) recorded. Other issues which impacted the selection decision of the bridges were time, distance, available financial support, feasibility to be sampled, among many other aspects, not always as initially planned and too long to be described here.

And attempt was made to get the original project information for each of the selected bridges, but so far only partial information is available. In fact, the bridge selected for the present study, is one of the mentioned five bridges with more complete available information. The structure is a 30 m span bridge, simply supported on abutments, and which superstructure is built with prestressed precast type AASHTO concrete girders, and shown in Figure 1.

Figure 1: Bridge considered in this study.
2.2 Demand

As part of the research project to gather information on the demand due to live loads, traffic surveys of different characteristics are conducted. At the time this study is submitted, although some WIM data is already collected in 2017, it is still not processed and available for its use in the present work. Figure 2 shows WIM recorded during February, 2017; this information will be employed in future research. Therefore, probabilistic information derived from WIM data collected in Guanajuato State and reported in a previous study is used instead [21]. Other surveys (static) were carried out by the Mexican Institute of Transportation (IMT for its acronym in Spanish) and are still not available, but will be also included in future studies.

![Figure 2: WIM station installed during 2017.](image)

2.3 Capacity

For the five selected bridges previously mentioned, information on the capacity was obtained by means of another research project. The information consisted on relevant parameters of the precast concrete of the bridge girders. Concrete cores were extracted on field from the prestressed precast concrete bridge girders, and tested in lab to obtain the compressive strength, young modulus and other parameters. For the bridge considered here, Figure 3 shows part of the extraction process and some of the samples before testing.

![Figure 3: Extraction on field and retrieved concrete cores.](image)
3 NOMINAL CAPACITY FROM CODE

The nominal bending capacity for a girder from the bridge in Figure 1 is computed. The procedure described here to obtain the capacity is used later to derive the mean flexural moment due to dead and live loads for the reliability analysis. The bridge girders are AASHTO type girders with 34 tendons separated 5 cm away centroid to centroid; the lowest layer centroid is 5 cm above the bottom side of the girder. A schematic section of the girder and prestressing strands is depicted in Figure 4, and the material properties and forces employed to compute the capacity are listed in Table 1. The regular reinforcement steel is distributed along the inside perimeter of the beam (except at the extreme fiber of the bottom flange) with a total area, $A_s = 28.5 \text{ cm}^2$; this reinforcement is also included in the computing. The nominal capacity is based on the AASHTO code [11] (including the resistance factor, $\phi$), since such practice is not uncommon among designers in Mexico, as mentioned previously. Rather than using formulae, the capacity is calculated by following the design hypotheses and section equilibrium as specified in Section 5.7.2 in the code [11]; this approach is preferred, since it is not limited to specific cases (e.g., tension-controlled beams), but is adequate for wider applicability. The nominal capacity after considering that 30% of the prestressing is lost is $M_n=4102.62 \text{ kN}\cdot\text{m}$ (which is equal to the resistant moment, since $\phi=1$ for the studied case [11]).

Figure 4: Girder section.
Table 1: Properties for the considered girder.

<table>
<thead>
<tr>
<th>Material</th>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concrete</td>
<td>Compressive stress</td>
<td>350 kgf/cm² (34.32 MPa)</td>
</tr>
<tr>
<td>Steel reinforcement</td>
<td>Yield stress</td>
<td>4200 kgf/cm² (411.88 MPa)</td>
</tr>
<tr>
<td>Prestressed reinforcement</td>
<td>Rupture stress</td>
<td>270 ksi (1861.58 MPa)</td>
</tr>
<tr>
<td></td>
<td>Diameter</td>
<td>½ in (12.70 mm)</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>98.7 mm²</td>
</tr>
<tr>
<td></td>
<td>Pre-stress force</td>
<td>13100 kgf (128.45 kN)</td>
</tr>
</tbody>
</table>

4 RANDOM VARIABLES AND LIMIT STATE FUNCTION

The random variables and their probabilistic characterization are listed in Table 2. The random variables in Table 2 are considered independent. Any other variables are taken as deterministic and equal to the nominal values. The moments and probability density functions (PDFs) for the modeling error, B, the reinforcement steel yielding stress, $f_y$, and the flexural moment due to dead load, $D$, are based on a previous study [20]. Other variables in Table 2 account for the uncertainty related to prestressing steel ultimate strength, $f_{pu}$, the prestressing steel area (for each tendon), $A_{ps}$, and the prestress loss, $P_{Loss}$, (it is considered that 30% of the prestressing force is lost for a return period over $T_r=50$ yr). Since statistics for the prestress are not readily available for the Mexican practice, but the nominal values of Mexican manufactures are very similar to those reported in the literature, they are adopted from (or based on) other works for $f_{pu}$ [6], $A_{ps}$ [1, 9] and $P_{Loss}$ [4, 5].

The statistics for flexural moments due to live load, $L$, are based on a previous study [21]; they are dependent on span length, not only for single vehicle passage, but also for multiple presence (both are considered). For the bridge in this study the span length is $L_{span}=30$ m, and lane load is first considered (not impact included; traffic jam scenario [21]). This leads to select the coefficient of variation shown in Table 2.

The PDF and probabilistic moments of the concrete compressive stress, $f'_c$, is based on the results from testing the concrete cores shown in Figure 3. The mean and coefficient of variation are statistics directly computed from the available samples for the bridge, and are listed in Table 2; since only four samples are available, a Normal PDF is arbitrarily adopted. Other alternative to investigate a possible underlying PDF for the compressive strength, is to use the available data for the five bridges in terms of the mean to nominal values; this strategy may be explored in future studies. Since the bridge is relatively new, no correction for concrete age is carried out; this issue will be also considered in further research.

<table>
<thead>
<tr>
<th>Random variable</th>
<th>mean</th>
<th>Coeff. of variation</th>
<th>Coeff. of Skewness</th>
<th>PDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td>1.1</td>
<td>0.10</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td>$f'_c$ (Mpa)</td>
<td>35.62</td>
<td>0.052</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td>$f_y$ (Mpa)</td>
<td>458.8</td>
<td>0.096</td>
<td>0.301</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$f_{pu}$ (Mpa)</td>
<td>1,910.26</td>
<td>0.025</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td>$A_{ps}$ (mm²)</td>
<td>98.7</td>
<td>0.016</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td>$P_{Loss}$ (%)</td>
<td>30</td>
<td>0.30</td>
<td>0.301</td>
<td>Lognormal</td>
</tr>
<tr>
<td>$D$ (kN·m)</td>
<td>**</td>
<td>0.10</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td>$L$ (kN·m)</td>
<td>**</td>
<td>0.069</td>
<td>1.1395</td>
<td>Gumbel</td>
</tr>
</tbody>
</table>

Table 2: Random variables and their characteristics.
Note that the mean dead load effect (flexure moment), $m_D$, and mean live load effect (flexure moment), $m_L$, are not defined in Table 2. They are derived under the assumption that the prestressed concrete bridge girder just meets the requirement of codified design. This assumption requires the use of a load format from code; the load format used here is based on the Mexican bridge regulations and a live load model proposed for bridge design in Mexico [21, 22]. This leads to the following expression

$$1.3m_D + 1.95m_L = \phi M_{\text{mean}}$$  \hspace{1cm} (1)

The right-hand side of Equation (1) is the bending capacity multiplied by the resistance factor, analogous to the nominal bending capacity computed as per the AASHTO code [11], except that mean instead of nominal values are employed. There is not specific procedure referred in the Mexican regulations to obtain the prestressed concrete girder capacity for bridge design [22]; consequently, the prestressed concrete bridge girder capacity is computed with the AASHTO code [11], as previously explained.

The reliability indices reported in the next section are computed for a range of mean live load effect to mean dead load effect ratios ($m_L/m_D$). Note that by considering a certain value of the $m_L/m_D$ ratio, $m_D$ and $m_L$ can be determined from Equation (1) and used for the reliability analysis.

The LSF can be established as follows

$$g_{CA} = R_{\text{pres}} - D - L$$  \hspace{1cm} (2)

where the bending capacity of the prestressed concrete girder is a function of several variables, $R_{\text{pres}} = f (B, f_c, f_y, f_{pu}, A_{ps}, P_{\text{Loss}})$. Since the capacity for the prestressed concrete girder subjected to flexure moments is a random variable, which in turn is a function of several random variables, as mentioned above, the need of a continuous derivable function is required if performing the FORM is of interest. To do this, a closed-form expression is required to compute the bending capacity for prestressed concrete members; alternatively, another technique can be used to carry out the reliability analysis. Since a numerical scheme is used for computing the bending capacity, such capacity can be obtained for a wider range of design cases, but on the other hand, the use of the FORM cannot be directly implemented. In this case, other options (e.g., the MCS) can be used to estimate the probabilities of failure, and so the reliabilities; however, this can be computationally expensive. Therefore, the CA [14] is used in the next section.

Before proceeding to carry out the reliability analysis, it is noteworthy to mention some considerations. It is pointed out that the assumption referred to formulate Equation (1), implicitly considers that other factors, like the multiple lane factor (i.e., the event of more than one lane is loaded in the bridge at the same time), or the girder distribution factor are properly taken into account; these aspects are also related to random variables, and deserve future research. For the probabilistic characterization of the multiple lane factors, the methodology described somewhere else could be followed [23].

It is also noted that the reliability index in the present work, $\beta$, is an annual reliability index, since the used statistics for the live load effects are linked to such time period [21].
5 RELIABILITY ANALYSIS

The combined approach is used here, since it takes advantage of the computational economy of the PEM (only 2xn realizations are required, where n is the number of random variables [18]), unlike other techniques (e.g., MCS), and it also takes advantage of the efficiency of the FORM (once the LSF has been properly reformulated as per the CA, so that it is explicitly established).

Table 3 shows the values of the variable combinations resulting from the PEM (for the first part of the CA), to compute the girder bending capacity using the numerical scheme referred in Section 3 above (but not load and resistance factors are involved this time).

<table>
<thead>
<tr>
<th>Number of realization</th>
<th>Value of each variable (PEM)</th>
<th>Capacity (kN·m)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>$f'_c$ (Mpa)</td>
</tr>
<tr>
<td>1</td>
<td>1.37</td>
<td>458.8</td>
</tr>
<tr>
<td>2</td>
<td>0.83</td>
<td>458.8</td>
</tr>
<tr>
<td>3</td>
<td>1.1</td>
<td>573.6</td>
</tr>
<tr>
<td>4</td>
<td>1.1</td>
<td>357.4</td>
</tr>
<tr>
<td>5</td>
<td>1.1</td>
<td>458.8</td>
</tr>
<tr>
<td>6</td>
<td>1.1</td>
<td>458.8</td>
</tr>
<tr>
<td>7</td>
<td>1.1</td>
<td>458.8</td>
</tr>
<tr>
<td>8</td>
<td>1.1</td>
<td>458.8</td>
</tr>
<tr>
<td>9</td>
<td>1.1</td>
<td>458.8</td>
</tr>
<tr>
<td>10</td>
<td>1.1</td>
<td>458.8</td>
</tr>
<tr>
<td>11</td>
<td>1.1</td>
<td>458.8</td>
</tr>
<tr>
<td>12</td>
<td>1.1</td>
<td>458.8</td>
</tr>
</tbody>
</table>

Table 3: Realizations from the PEM and corresponding obtained bending capacities.

The 12 (2xn) realizations of the prestressed concrete bridge girder capacity are also included in Table 3 (last column, bending capacity in kN·m); these values are employed to obtain the first two probabilistic moments of the girder capacity ($R_{pres}$ in Equation (2)) [18]; the mean and coefficient of variation resulted in 4726.96 kN·m and 0.1391, respectively (note that this implies a mean to nominal value of 1.1522 for the bending capacity of the prestressed concrete bridge girder). With the previous information, the procedure established in the CA [14] is carried out, and the reliability levels are computed by using the FORM and Equation (2), except that $R_{pres}$ in not a function of many random variables, but a single, completely determined random variable, as established by following the CA (details can be found in other study [14]). Figure 5 shows the reliability levels for the prestressed concrete girder by using the CA for a range of $m_l/m_D$ values.

From Fig. 5 it can be observed that the annual reliability index is between around 3.6 and 4.4 for a range of $m_l/m_D$ values (dotted line). As reference, note that a target reliability index for 1 year ($\beta_t$) equal to 3.75 was employed for calibration purposes in previous studies [21], studies from which the live load effects statistics are adopted for the present work; note further that the same load factors format from the Mexican regulations [22] are used (Equation (1)). The values in Fig. 5 are consistent but moderately higher than those in other studies [20, 21].
21], for different span lengths and for reinforced concrete bridges and steel bridges. In those cases, the annual reliability levels, in average, approximately correspond to an annual target reliability index of 3.75, which in turn is associated to a target reliability index of 3.5 for a service period of 75 years [24]; for the prestressed concrete bridge girder studied here, and for the lane load case, larger reliability levels are obtained, but not significant larger than the $\beta_T = 3.75$. Since the statistics of live load effects are span dependent, future studies are recommended to further inspect the reliability levels of prestressed concrete girders for bridges.

Figure 5 also shows the reliability index when a simple consideration on the dynamic live load effect (impact; single vehicle passage scenario [21]) is taken into account. For this purpose, an assumption followed in other study is adopted [1], and Equation (1) is rewritten as

$$1.3 m_D + (1.95+0.1) m_L = \phi M_{\text{mean}}$$

where an extra 10% of live load effect due to static live load is assumed (i.e., the impact is considered as a fraction of the static live load effect).

The same procedure using the CA is repeated by considering Equation (3), except that the coefficient of variation of $L$, is not the one in Table 2, but equal to 0.12. This is derived, in one hand, by considering a coefficient of variation equal to 0.065 for $L_{\text{span}}=30$ m, for the case of single vehicular passage [21], which is the relevant scenario for dynamic amplification; on the other hand, to include the coefficient of variation for the dynamic effect, a value is adopted from other work and equal to 0.10 [5]; the previous coefficients of variation are used to derived the coefficient of variation of the total (static plus dynamic) live load effects, under
the assumption that the square root of the sum of the squares can be used to compute the total coefficient of variation \([25]\) (like if both effects were normally distributed), and this resulted in the value of 0.12 referred above. It is acknowledged that this is not the most rigorous possible treatment of the dynamic live load effect, and future work is recommended.

The results by including the dynamic effects show that a decrease in the reliability index is exhibit (Figure 5, solid line); this is a consequence of the larger dispersion (in terms of the coefficient of variation) introduced by the dynamic effects in the live load. In fact, values closer, in average, to an annual target reliability level of 3.75 (or 3.5 for a service period of 75 years [24]) are obtained.

A couple more of reliability analyses are performed by using the CA, but considering a \(\phi=0.9\) (as proposed in other study [21]), instead of the AASHTO strength factor; this is carried out for both, the lane load case and the single vehicle passage. The obtained reliability indices are those also indicated in Figure 5 for the static case (dash-dotted line) and for the dynamic one (dashed line). The used of \(\phi=0.9\) results in a shifting upwards (in relation to their counterparts as per the AASHTO [11], \(\phi=1\)), increasing the reliability levels for the whole range of \(m_L/m_D\) ratios. This may indicate that, unlike the case of steel and reinforced concrete bridge girders, \(\phi=1\) could be more suitable for prestressed concrete bridge girders, if the target reliability index mentioned above is considered, and the live load models and load factors proposed in a previous work are used [21]; however, more analyses (e.g., by considering other bridge spans, and/or by exploring load factors format as a function of the \(m_L/m_D\) ratio [14]) may be convenient for calibration tasks.

As a final remark, it is noted that the CA procedure can be implemented for other LSFs (e.g., shear), and it can potentially be used when a numerical method is required to compute the capacity of an element (e.g., a finite element model), or maybe even of a system, but further research is recommended to investigate such issues.

6 CONCLUSIONS

- The reliability of prestressed concrete bridge girders is assessed using field information and the combined approach (CA). It is pointed out that it is not common to find in the literature the use of field information on loads and strengths, expressively gathered to compute reliabilities (like in this study). Some relevant findings of this study are given below.

- The CA is a simple method to obtain reliability levels of prestressed concrete bridge girders; it is a convenient alternative when a continuous derivable limit state function (LSF) is not available (unlike the FORM), or if an extensive number of simulations (like in the MCS), computationally expensive methods, or complex techniques are not desired.

- When the lane load (static live load; traffic jam scenario) is considered, an annual reliability index between around 3.6 and 4.4 (bending) for a range of \(m_L/m_D\) values is obtained for the considered prestressed concrete bridge girder.

- The results by including the dynamic effects show a decrease in the reliability index, compared to the static live load case; this is a consequence of the larger dispersion (in terms of the coefficient of variation) when such effect is considered. For this scenario, values closer to an annual target reliability level of 3.75 (or 3.5 for a service period of 75 years) are obtained for the same prestressed concrete girder.
When a $\phi = 0.9$ instead of $\phi = 1$ is used, the reliability indices (for both, the static and dynamic cases) are shifted upwards; i.e., the reliability levels are increased for the whole range of $m_l/m_D$ ratios. It is concluded that the used of $\phi = 1$, together with a previously proposed live load model and load factors, could be more suitable for computing the capacity of prestressed concrete bridge girders, if an annual target reliability level of 3.75 (or 3.5 for a service period of 75 years) is of interest.

The CA could be a valuable alternative for other LSFs and, in general, when a numerical method is involved in computing the capacity of an element (e.g., finite element models), and maybe even in other cases. Further research is recommended in the body of the manuscript, to investigate these and other issues.

7 ACKNOWLEDGEMENTS

The financial support from CONACYT (Project “Problemas Nacionales 2014”, No. 248162) and PROMEP (Project “Nuevos PTC”, UGTO-PTC-429) is gratefully acknowledge, as well as that from DAA, DAIP, and the Department of Civil Engineering of Universidad de Guanajuato. We are very thankful to SCT (Secretaria de Comunicaciones y Transportes de Guanajuato), to the SOPs (Secretaria de Obra Pública del Estado y del Municipio de Guanajuato), to Alejandra Vizguerra and Agustin Nuñez, and to engineers, technicians, and many other workers from the IMCYC, IMT, Guanajuato Federal Police, Rocher Ingeniería, Conseccionaria of Irapuato-La Piedad Highway, and several departments from Universidad de Guanajuato, to a local contractor who provided power plants and other equipment, and to anyone of the directly or indirectly involved in this study, too long to mention, but who helped to achieve this goal.

REFERENCES


ASSESSING HOW THE DEPENDENCE STRUCTURE AFFECTS THE RELIABILITY PARAMETER OF THE STRENGTH-STRESS MODEL

A. Barbiero

1Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, via Conservatorio 7, 20122 Milan (Italy)
e-mail: alessandro.barbiero@unimi.it

Keywords: Bivariate random variable, Copula, Farlie-Gumbel-Morgenstern distributions, Fréchet bounds, Statistical dependence

Abstract. The stress-strength model is a basic modeling tool in reliability analysis. In simple terms, it considers a component (or a system) with an intrinsic strength Y, which is subjected to a stress X; the component works if and only if Y is greater than X. If stress and strength are regarded as random variables, then the probability that the component works is given by P(X < Y) and is usually called “reliability parameter”. Statistical independence is usually assumed between the two random variables X and Y: for this case, the literature on this topic is particularly rich. This strong assumption, in fact, makes the calculation and estimation of the reliability parameter R more tractable. However, this hypothesis is not always verified in practice, and this translates into an over- or under-estimation of R. To avoid this drawback, statistical dependence can be introduced and modeled between X and Y, for example resorting to copulas. In some recent works, the problem of computing and estimating R is considered when the stress and strength variables, belonging to the same parametric family of distributions, are linked by a specific copula. In this work, we further consider the computational issues related to this copula approach applied to the stress-strength model, when other families of copulas are selected. A sort of sensitivity analysis is performed in order to assess how the value of the reliability parameter is affected by the choice of the copula binding X and Y together and of its parameters. As a limit case, we consider the situation where no information on the dependence structure of the stress and strength margins is available and try to provide lower and upper bounds for R.
1 INTRODUCTION

The stress-strength model is a basic modeling tool in reliability analysis. In simple terms, it considers a component (or a system) with an intrinsic strength $Y$, which is subjected to a stress $X$; the component works if and only if $Y$ is greater than $X$; if stress and strength are regarded as random variables (r.v.s), then the probability that the component works is given by $P(X < Y)$ and is usually called “reliability parameter”. The stress-strength model has many applications, especially in engineering concepts such as structures, deterioration of rocket motors, static fatigue of ceramic components, fatigue failure of aircraft structures, and the aging of concrete pressure vessels [1].

Statistical independence is usually assumed between the two r.v.s $X$ and $Y$: for this case, the literature on this topic is particularly rich. The algebraic form for $R$ has been worked out for the majority of the well-known distributions, including normal, uniform, exponential, gamma, beta, extreme value, Weibull, Laplace, logistic, and Pareto distributions [2]. This strong assumption, in fact, makes the calculation and estimation of the reliability parameter $R$ more tractable. However, this hypothesis is not always verified in practice, and this translates into an over- or under-estimation of $R$. To avoid this drawback, statistical dependence can be introduced and modeled between $X$ and $Y$, for example resorting to copulas. Citing [3], “the use of a joint distribution for stress and strength in estimating $R$ is justified by the practice of using stronger prototypes in worse environments which cause greater stresses. Moreover, the use of two independent random samples– one on $X$ and the other on $Y$–to estimate $R$ lacks support from pragmatic considerations”.

The case when $(X, Y)$ follows a bivariate normal distribution has been investigated by [3, 4, 5]. [1] and [6] considered the problem of computing and estimating $R$ for some bivariate gamma and exponential distributions, respectively. Within the Bayesian framework, reference [7] studied the estimation of the reliability parameter assuming for the stress and strength variables asymmetric distributions obtained by skewing scale mixtures of normals; the margins are linked by the Gaussian copula. In [8] a stress-strength model is investigated with stress and strength marginally distributed as non-identical Dagum r.v.s and their dependence described by a Frank copula. In [9] the problem of estimation of the reliability parameter is considered when the Farlie-Gumbel-Morgenstern copula is used to link stress and strength variables, whose marginal distributions both belong to the Burr system. More recently, in [10], an ampler study on the effect of statistical dependence on the distribution of the functions of r.v.s deals also with the computation of $R$ when several statistical distributions are chosen for both $X$ and $Y$ and for various copula structures.

In this work, we further consider the computational issues related to this copula approach applied to the stress-strength model, when other families of copulas are selected. The rest of the paper is organized as follows. In Section 2, we recall some basic notions about copulas. In Section 3, after recalling the main results about the computation of the reliability parameter for stress-strength models involving copulas, we first derive its explicit expression when both stress and strength belong to the exponential family and other copula families than those already analyzed in the literature model their dependence. Then, in Section 4, we consider the situation where no information on the dependence structure of the stress and strength margins is available and try to provide lower and upper bounds for $R$. We examine the case of normally distributed and Cauchy distributed stress and strength. Section 5 concludes the paper with some possible research perspectives.
2 COPULAS

Copulas allow to separate the study of the marginal distributions of a multivariate model from the study of the dependence structure and then to combine more marginal distributions with a variety of possible dependence structures. Formally, a d-dimensional copula \( C(u_1, \ldots, u_d) \) is simply a joint cumulative distribution function (c.d.f.) in \([0, 1]^d\) with standard uniform marginal c.d.f.s. The importance of copulas in the study of multivariate c.d.f.s is summarized by the variety of possible dependence structures. Formally, a 2 COPULAS

For any copula \( C \), the following inequality holds:

\[
W(u, v) \leq C(u, v) \leq M(u, v).
\]

For continuous r.v.s, the comonotonicity copula is the copula of two perfectly positive-dependent r.v.s: in this case, \((X, Y) \overset{d}{=} (F^{-1}(U), G^{-1}(U))\); the countermonotonicity copula is the copula of two perfectly negative-dependent r.v.s: in this case, \((X, Y) \overset{d}{=} (F^{-1}(1-U), G^{-1}(1-U))\); the independence copula is the copula of two independent r.v.s. Note that the Gaussian copula (3) comprises the comonotonicity, countermonotonicity, and the independence copulas as limiting cases (for \( \rho \to +1, \rho \to -1, \) and \( \rho = 0 \), respectively).
3 RELIABILITY FOR STRESS-STRENGTH MODEL WITH DEPENDENT VARIABLES

If the continuous stress and strength r.v.s $X \sim F$ and $Y \sim G$ are linked by a copula $C$, then the stress-strength parameter $R$ is given by (see [10])

$$P(X - Y < 0) = 1 - \int_0^1 \dot{C}_1(1 - w, G(F^{-1}(1 - w))) dw,$$

being $\dot{C}_1 = \partial C(u, v) / \partial u = P(V \leq v | U = u)$. If additionally $X$ and $Y$ are both positive r.v.s, then the stress-strength parameter $R$ is given also by (see [10] again)

$$P\left(\frac{X}{Y} < 1\right) = 1 - \int_0^1 \dot{C}_1(w, G(F^{-1}(1 - w))) dw.$$

Exploiting formulas (5) or (6), in [10] the expressions of the reliability parameter have been explicitly or numerically found for some possible choices of $F$ and $G$ (namely, uniform and exponential) and the linking copulas (Farlie-Gumbel-Morgenstern, Frank, Clayton and Gumbel).

In the next two subsections, we first examine two other kind of copulas and compute the reliability parameter when stress and strength follow an exponential distribution; then we provide sharp bounds for the reliability parameter when no information is available about the copula linking normal or Cauchy stress and strength distributions.

3.1 Extended Farlie-Gumbel-Morgenstern copula

The Farlie-Gumbel-Morgenstern [13], henceforth FGM, is a well-known family of copulas allowing for a moderate level of dependence:

$$C(u, v) = uv(1 + \theta(1 - u)(1 - v)), -1 \leq \theta \leq +1.$$

It can be easily shown that for this copula $\rho_{uv} = \theta/3$, and then $\rho_{uv}$ is bounded between $-1/3$ and $1/3$. [14] showed that if the margins linked by the FGM copula are absolutely continuous, Pearson’s correlation cannot exceed $1/3$. Several extensions of this family have been proposed, especially for enlarging the range of possible correlation values.

Here we first consider the following extension, which is due to [15] and described also in [16]:

$$C(u, v) = uv + \theta u^b v^b (1 - u)^a (1 - v)^a, a, b \geq 1.$$  \hfill (8)

We set $a = 2$, $b = 1$ in Eq.(8) (then the constraint on $\theta$ is $-1 \leq \theta \leq +3$). The conditional distribution $\dot{C}_1$ can be easily derived:

$$\dot{C}_1 = \frac{\partial C(u, v)}{\partial u} = v + \theta v(1 - v)^2(1 + 3u^2 - 4u).$$  \hfill (9)

If $X$ and $Y$ follow exponential distributions with rate parameters $\lambda_x$ and $\lambda_y$, respectively, linked by the above extended FGM copula, then, since $F^{-1}(w) = -\log(1 - w)/\lambda_x$ and $G(F^{-1}(1 -$
\[ w) = 1 - w^{\frac{\lambda_y}{\lambda_x}}, \]
the reliability parameter provided by Eq.\((\ref{5})\) takes the following form:

\[
R = 1 - \int_0^1 (1 - w^\alpha) + \theta(1 - w^\alpha)w^{2\alpha}[1 + 3(1 - w)^2 - 4(1 - w)]dw
\]

\[
= 1 - \int_0^1 (1 - w^\alpha) + \theta(3w^{2\alpha+2} - 2w^{2\alpha+1} - 3w^{3\alpha+2} + 2w^{3\alpha+1})dw
\]

\[
= 1 - \left[ w - \frac{w^{\alpha+1}}{\alpha + 1} + \theta \left( \frac{3w^{3\alpha+3}}{2\alpha + 3} - \frac{2w^{2\alpha+2}}{2\alpha + 2} - \frac{3w^{3\alpha+3}}{3\alpha + 3} + \frac{2w^{3\alpha+1}}{3\alpha + 2} \right) \right]_0^1
\]

\[
= \frac{1}{\alpha + 1} - \theta \left( \frac{3}{2\alpha + 3} + \frac{2}{3\alpha + 2} - \frac{2}{\alpha + 1} \right)
\]

\[
= \frac{1}{\alpha + 1} - \theta \frac{\alpha(\alpha - 1)}{(\alpha + 1)(2\alpha + 3)(3\alpha + 2)}, \tag{10}
\]

where \(\alpha = \frac{\lambda_y}{\lambda_x}\). For \(\alpha > 1\), i.e., if the expected value of \(Y\) is smaller than the expected value of \(X\), \(R\) is a decreasing linear function of \(\theta\); for \(\alpha < 1\), \(R\) is an increasing linear function of \(\theta\). When \(\alpha = 1\), \(R\) is constant and equal to 0.5. The range allowed to \(R\) is quite narrow; for example, when \(\alpha = 1/3\), \(R\) goes from 0.7348485 to 0.7954545. If \(\theta = 0\), Eq.\((\ref{10})\) boils down to the usual formula for two independent exponential distributions: \(R = \frac{1}{\alpha + 1} = \frac{\lambda_x}{\lambda_x + \lambda_y}\).

We consider now the following alternative extension of the FGM copula (see again \([15]\) and \([16]\)):

\[
C(u, v) = uv + \theta uv(1 - u^p)(1 - v^p), p > 0. \tag{11}
\]

Let us set \(p = 2\) in Eq.\((\ref{11})\) (then the constraint on \(\theta\) is \(-1/4 \leq \theta \leq 1/2\)). The conditional distribution \(C_1\) becomes:

\[
C_1 = v + \theta v(1 - v^2)(1 - 3u^2) \tag{12}
\]

and then from Eq.\((\ref{6})\) we derive:

\[
R = 1 - \int_0^1 (1 - w^\alpha) + \theta(1 - w^\alpha)[1 - (1 - w^\alpha)^2][1 - 3(1 - w)^2]dw
\]

\[
= 1 - \int_0^1 (1 - w^\alpha) + \theta(-2w^{3\alpha} + 6w^{2\alpha} - 4w^\alpha - 3w^{3\alpha+2} + 9w^{2\alpha+2} - 6w^{3\alpha+1} + 18w^{2\alpha+1} + 12w^{3\alpha+1})dw
\]

\[
= 1 - \left[ w - \frac{w^\alpha}{\alpha + 1} + \theta \left( \frac{6w^{2\alpha+1}}{2\alpha + 1} + \frac{6w^{3\alpha+2}}{3\alpha + 2} + \frac{12w^{\alpha+2}}{\alpha + 2} + \frac{9w^{2\alpha+3}}{2\alpha + 3} + \frac{2w^{3\alpha+1}}{3\alpha + 1} - \frac{6w^{\alpha+3}}{\alpha + 3} - \frac{14w^{\alpha+1}}{\alpha + 1} \right) \right]_0^1
\]

\[
= \frac{1}{\alpha + 1} - \theta \left( \frac{6}{2\alpha + 1} + \frac{6}{\alpha + 2} + \frac{12}{2\alpha + 3} + \frac{9}{3\alpha + 1} - \frac{2}{\alpha + 3} - \frac{6}{\alpha + 1} \right)
\]

\[
= \frac{1}{\alpha + 1} - \theta \frac{(\alpha - 1)(\alpha(30\alpha^4 + 251\alpha^3 + 470\alpha^2 + 251\alpha + 30))}{(\alpha + 1)(\alpha + 2)(\alpha + 3)(\alpha + 1)(\alpha + 3)(3\alpha + 1)(3\alpha + 2)} \tag{13}
\]

As for the previous case, for \(\alpha > 1\), \(R\) is a decreasing linear function of \(\theta\); for \(\alpha < 1\), \(R\) is an increasing linear function of \(\theta\). When \(\alpha = 1\), \(R\) is constant and equal to 0.5. The range allowed to \(R\) is quite narrow; for example, when \(\alpha = 1/3\), \(R\) goes from 0.7243506 to 0.8012987. If \(\theta = 0\), Eq.\((\ref{13})\) boils down to the usual formula for two independent exponential distributions.
3.2 Ali-Mikhail-Hak copula

We consider the following family of Archimedean copulas, known as Ali-Mikhail-Hak family [17]:

\[ C(u, v) = \frac{uv}{1 - \theta(1 - u)(1 - v)}, -1 \leq \theta \leq +1. \]  

(14)

For \( \theta = 0 \), we obtain the independence copula, whereas positive (negative) values of \( \theta \) lead to positive (negative) dependence. The conditional distribution \( \dot{C}_1 \) can be easily derived:

\[ \dot{C}_1 = \frac{\partial C(u, v)}{\partial u} = \frac{v(1 - \theta + \theta v)}{(1 - \theta(1 - u)(1 - v))^2}. \]  

(15)

If \( X \) and \( Y \) follow exponential distributions with parameters \( \lambda_x \) and \( \lambda_y \), respectively, linked by a copula belonging to the Ali-Mikhail-Hak family, then the reliability parameter is provided, after some calculations, by Eq.(5) or (6):

\[ R = 1 - \int_0^1 \frac{(1 - w^\alpha)(1 - \theta w^\alpha)}{(1 - \theta w^{\alpha+1})^2} \, dw \]  

(16)

The integral is not (easily) analytically solvable except for integer values of \( \alpha \); in particular, if \( \alpha = 1 \), we get

\[ R = 1 - \left[ 2x\theta - \theta - 1 \right]_0^1 = 1 - \left( -\frac{1}{2\theta} - \frac{-\theta - 1}{2\theta} \right) = \frac{1}{2} \]  

(17)

for any possible value of \( \theta \in [-1, +1] \), as one could expect. However, in general, the integral in (16) can be solved numerically through any common statistical or mathematical software, such as R or Mathematica, and then it is possible, for example, to plot the graph of \( R \) as a function of \( \theta \). Such graphs, for several choices of \( \alpha \), are presented in Figure[1] note that \( R \) is an increasing (decreasing) function of \( \theta \) for \( \alpha \) smaller (greater) than 1. The range allowed to \( R \) looks larger than for the extended FGM copulas seen before, for a fixed value of \( \alpha \); for example, when \( \alpha = 1/3 \), \( R \) is allowed to vary between from 0.717419 to 0.8390486.

4 LOWER AND UPPER BOUNDS FOR \( R \) WHEN NO INFORMATION IS AVAILABLE ABOUT DEPENDENCE

Suppose again that \( X \) and \( Y \) are r.v.s with c.d.f.s \( F \) and \( G \), respectively. Let \( S \) denote the c.d.f. of the sum \( X + Y \), i.e., \( S(z) = P(X + Y < z) \). A classical problem in probability is finding \( S^\wedge(z) = \sup S(z) \) and \( S^\vee(z) = \inf S(z) \), where the supremum and infimum are taken over the Fréchet-Hoeffding class \( H(F, G) \) of all joint c.d.f.s \( H \) with marginals \( F \) and \( G \). This problem leads to copulas naturally, since, as we have seen in Section 2, if \( H \) is the joint distribution function of \( X \) and \( Y \), then \( H(x, y) = C(F(x), G(y)) \) for at least one copula \( C \) (exactly one if \( X \) and \( Y \) are continuous). The following result holds (see [12], p.228):

\[ S^\vee(z) \leq S(z) \leq S^\wedge(z) \]

\[ S^\vee(z) = \sup_{x+y=z} W(F(x), G(y)) \]

\[ S^\wedge(z) = \inf_{x+y=z} \tilde{W}(F(x), G(y)) \]  

(18)

with \( \tilde{W} = u + v - W(u, v) = \min \{u + v, 1\} \).
If instead of the $Y$ r.v. we consider its opposite $-Y$, with c.d.f. $1 - G(-y)$, and let $z = 0$, the problem above turns into finding the lower and upper bound (say, $R_\vee$ and $R_\wedge$) of the reliability parameter of a stress-strength model with stress $X \sim F$ and strength $Y \sim G$ over all joint c.d.f.s $H$ with marginals $F$ and $G$: $R_\vee \leq R = P(X < Y) \leq R_\wedge$. Adjusting the results reported in [12, p.229-231], we can state that the lower and upper bounds $R_\vee$ and $R_\wedge$ are sharp since they are attained by (at least) one copula each. Recalling that if the copula of $(X, Y)$ is $C(u, v)$, then the copula of $(X, -Y)$ is $u - C(u, 1 - v)$, we derive for the lower bound:

$$C_\vee(u, v) = \begin{cases} \min \{u - R_\vee, v\} & [R_\vee, 1] \times [0, 1 - R_\vee] \\ \max \{0, u + v - 1\} & \text{elsewhere}; \end{cases} \tag{19}$$

for the upper bound:

$$C_\wedge(u, v) = \begin{cases} \min \{u, v - (1 - R_\wedge)\} & [0, R_\wedge] \times [1 - R_\wedge, 1] \\ \max \{0, u + v - 1\} & \text{elsewhere}. \end{cases} \tag{20}$$

4.1 Normal stress and strength

Let us consider the case of normal stress and strength: $X \sim N(\mu_x, \sigma_x^2), Y \sim N(\mu_y, \sigma_y^2)$. If $\sigma_x^2 = \sigma_y^2 = \sigma^2$ it can be shown (properly adapting the results presented in [12, p.232-233]) that

$$R_\vee = \begin{cases} 1 & \eta \geq 0 \\ 2\Phi\left(\frac{\eta}{2\sigma}\right) - 1 & \eta \leq 0 \end{cases} \tag{21}$$

$$R_\wedge = \begin{cases} 2\Phi\left(\frac{\eta}{2\sigma}\right) & \eta \leq 0 \\ 1 & \eta \geq 0 \end{cases} \tag{22}$$
where \( \eta = \mu_y - \mu_x \). If the variances are not the same,

\[
R^\lor = \Phi \left( \frac{-\sigma_x \eta - \sigma_y \phi}{\sigma^2_y - \sigma^2_x} \right) + \Phi \left( \frac{\sigma_y \eta - \sigma_x \phi}{\sigma^2_y - \sigma^2_x} \right) - 1 \tag{23}
\]

and

\[
R^\land = \Phi \left( \frac{-\sigma_x \eta + \sigma_y \phi}{\sigma^2_y - \sigma^2_x} \right) + \Phi \left( \frac{\sigma_y \eta + \sigma_x \phi}{\sigma^2_y - \sigma^2_x} \right), \tag{24}
\]

where \( \phi = [\eta^2 + 2(\sigma^2_y - \sigma^2_x) \log(\sigma_x/\sigma_y)]^{1/2} \).

Let us remember that if \( X \) and \( Y \) are assumed to be independent normal r.v.s, then the reliability parameter \( R \) would be equal to

\[
R = \Phi \left( \frac{\eta}{\sqrt{\sigma^2_x + \sigma^2_y}} \right). \tag{25}
\]

If \( X \) and \( Y \) are comonotonic, then

\[
R = \Phi \left( \frac{\eta}{|\sigma_y - \sigma_x|} \right) \tag{26}
\]

in case of unequal variances; in case of equal variances, \( R = 1 \) if \( \eta > 0 \); \( R = 0 \) if \( \eta \leq 0 \). If \( X \) and \( Y \) are countermonotonic, then it can be easily proved that

\[
R = \Phi \left( \frac{\eta}{\sigma_y + \sigma_x} \right). \tag{27}
\]

**Example 1 (normal stress and strength with equal variances).** Let \( \mu_x = 6, \mu_y = 7, \sigma_x = \sigma_y = 2 \). If \( X \) and \( Y \) are assumed to be independent r.v.s, then the reliability parameter \( R \) would be equal to \( R^\lor = 0.1974 \) and \( R^\land = 1 \). From Eq. (27), the value of \( R \) corresponding to the countermonotonicity copula \( W(u,v) \) is \( R = 0.5987 \); for the comonotonicity copula \( M(u,v) \), applying Eq. (26), we have \( R = 1 \).

**Example 2 (normal stress and strength with unequal variances).** Let \( \mu_x = 6, \sigma_x = 1, \mu_y = 7, \sigma_y = 2 \). If \( X \) and \( Y \) are assumed to be independent r.v.s, then, applying Eq. (25), the reliability parameter \( R \) would be equal to \( R^\lor = 0.6726 \). Lower and upper bounds for \( R \) are provided through Eqs. (23) and (24) as \( R^\lor = 0.3451 \) and \( R^\land = 0.9551 \). From Eq. (27), the value of \( R \) corresponding to the countermonotonicity copula is \( R = 0.6306 \); for the comonotonicity copula (applying Eq. (26), we have \( R = 0.8413 \). Copulas leading to \( R^\lor \) and \( R^\land \), given by Eqs. (19) and (20), are graphically displayed in Figure 2.

### 4.2 Cauchy stress and strength

Let us now focus on the Cauchy distribution, whose density function is given by

\[
f(t; \alpha, \beta) = \frac{1}{\pi \beta \left[ 1 + \left( \frac{t - \alpha}{\beta} \right)^2 \right]^2}, \quad x, \alpha \in \mathbb{R}, \beta \in \mathbb{R}^+, \tag{28}
\]

and whose c.d.f. is

\[
F(t; \alpha, \beta) = \frac{1}{2} + \frac{1}{\pi} \arctan \left( \frac{t - \alpha}{\beta} \right). \tag{29}
\]
If we assume that $X$ and $Y$ are r.v.s with Cauchy distribution with location parameters $\alpha_x$ and $\alpha_y$, and scale parameters $\beta_x$ and $\beta_y$, respectively, then we have the following bounds for the reliability parameter $R$, derived rearranging [12, p.234]. If $\beta_x = \beta_y = \beta$,

$$R^\vee = \begin{cases} 0 & \eta \leq 0 \\ \frac{2}{\pi} \arctan \left( \frac{\eta}{2\beta} \right) & \eta \geq 0 \end{cases}$$

$$R^\wedge = \begin{cases} 1 + \frac{2}{\pi} \arctan \left( \frac{\eta}{2\beta} \right) & \eta \leq 0 \\ 1 & \eta \geq 0 \end{cases}$$

where $\eta = \alpha_y - \alpha_x$. If $\beta_x \neq \beta_y$,

$$R^\vee = \frac{1}{\pi} \left[ \arctan \left( \frac{-\eta + \beta_y \phi}{\beta_y - \beta_x} \right) + \arctan \left( \frac{\eta - \beta_x \phi}{\beta_y - \beta_x} \right) \right]$$

$$R^\wedge = 1 + \frac{1}{\pi} \left[ \arctan \left( \frac{-\eta + \beta_y \phi}{\beta_y - \beta_x} \right) + \arctan \left( \frac{\eta + \beta_x \phi}{\beta_y - \beta_x} \right) \right]$$

where $\phi = [(\eta^2 + (\beta_y - \beta_x)^2)/(\beta_x \beta_y)]^{1/2}$.

Note that for independent r.v.s we have $R = \frac{1}{\pi} + \frac{1}{\pi} \arctan \left( \frac{\eta}{\beta_x + \beta_y} \right)$. If the copula linking $X$ and $Y$ is the comonotonicity copula we have $R = \frac{1}{\pi} + \frac{1}{\pi} \arctan \left( \frac{\eta}{|\beta_y - \beta_x|} \right)$, if $\beta_x \neq \beta_y$. Otherwise, if $\beta_x = \beta_y$, $R = 1$ if $\eta > 0$, $R = 0$ if $\eta \leq 0$. If the copula linking $X$ and $Y$ is the countermonotonicity copula we have $R = \frac{1}{\pi} + \frac{1}{\pi} \arctan \left( \frac{\eta}{\beta_x + \beta_y} \right)$, as for the independence copula. Note then how two different copulas linking $X$ and $Y$ provide in this case the same value for the probability $P(X < Y)$. 

---

Figure 2: Reliability parameter for normal stress $X$ and strength $Y$. Representation of the supports of copulas leading to the lower bound ($C^\vee(u, v)$, left) and upper bound ($C^\wedge(u, v)$, right) when $X \sim N(\mu_x = 6, \sigma_x = 1)$ and $Y \sim N(\mu_y = 7, \sigma_y = 2)$. $R_L$ stands for the lower bound $R^\vee$ and $R_U$ for the upper bound $R^\wedge$. 

---

648
Example 3 (Cauchy stress and strength equal scale parameters). Let $\alpha_x=6$, $\alpha_y=7$, $\beta_x=\beta_y=2$. Lower and upper bounds for $R$ are provided as $R^\vee=0.1560$ and $R^\wedge=1$. Note that if $X$ and $Y$ were independent, then $R=0.5780$ and the same result would hold if they were countermonotonic; if $X$ and $Y$ were comonotonic, $R=1$.

Example 4 (Cauchy stress and strength unequal scale parameters). Let $\alpha_x=6$, $\beta_x=1$, $\alpha_y=7$, $\beta_y=2$. Lower and upper bounds for $R$ are provided as $R^\vee=0.25$ and $R^\wedge=0.9548$. Note that if $X$ and $Y$ were independent, then $R=0.6024$, and the same result would hold if they were countermonotonic; if $X$ and $Y$ were comonotonic, $R=0.75$.

5 CONCLUSIONS

In this work, we considered the problem of computing the reliability of the stress-strength model when stress and strength variables are dependent. We first examined the case of exponential variables linked by three different types of copulas, deriving analytical expression for the reliability parameter and examining its relationship with the copula parameter. Then, focusing on normal or Cauchy distributed stress and strength, we derived lower and upper bound for the reliability parameter when no assumption is made on their copula.

As for research perspectives, other distributions for stress and strength and other copula families can be examined and the corresponding value of $R$, or sharp lower and upper bounds for $R$, can be analytically or numerically derived. Generalizations to more complex stress-strength models, where stress and strength are themselves functions of several random components (i.e., $X=\xi(X_1,X_2,\ldots,X_h)$ and $Y=\nu(Y_1,Y_2,\ldots,Y_k)$, for some functions $\xi$ and $\nu$) will be considered. Another aspect to be developed is the estimation of lower and upper bounds for $R$ when only random samples from $X$ and $Y$ are available: this means that besides the uncertainty about the dependence structure (copula) of $X$ and $Y$, we will have to take into account also the uncertainty about (the parameters of) the margins, which have to be estimated from the samples.

REFERENCES


ROBUST ARTIFICIAL NEURAL NETWORK FOR RELIABILITY ANALYSIS

Uchenna Oparaji\textsuperscript{1,2}, Rong-Jiun Sheu\textsuperscript{2}, and Edoardo Patelli\textsuperscript{1}

\textsuperscript{1}Institute for Risk and Uncertainty, University of Liverpool
Chadwick Building, Peach Street, Liverpool L69 7ZF, United Kingdom
e-mail: \{u.oparaji, epatelli\}@liverpool.ac.uk

\textsuperscript{2}Institute of Nuclear Engineering and Science
National Tsing Hua University, Hsinchu, Taiwan
e-mail: \{rjsheu\}@mx.nthu.edu.tw

Keywords: Artificial Neural Network, Uncertainty Quantification, Reliability Analysis.

Abstract. Artificial Neural Networks (ANN) are used in place of expensive models to reduce the computational burden required for reliability analysis. Often, ANNs with selected architecture are trained with the back-propagation algorithm from few data representatives of the input/output relationship of the underlying model of interest. However, different performing ANNs might be obtained from the same training data, leading to an uncertainty in selecting the best performing ANN. On the other hand, using cross-validation to select the best performing ANN based on the highest $R^2$ value can lead to a biasing in terms of the prediction made by the selected ANN. This is due to the fact that the use of $R^2$ cannot determine if the prediction made by ANN is biased. Additionally, $R^2$ does not indicate if a model is adequate, as it is possible to have a low $R^2$ for a good model and a high $R^2$ for a bad model. Hence we propose an approach to improve the prediction robustness of an ANN based on coupling Bayesian framework and model averaging technique into a unified framework. The model uncertainties propagated to the robust prediction is quantified in terms of confidence intervals. Two examples are used to demonstrate the applicability of the approach.
1 INTRODUCTION

Nowadays, numerical models are increasingly used to analyze and predict the performance of complex critical systems. Concurrently, engineering practitioners are concerned with uncertainty, which is inherent to these systems. As a consequence, probabilistic analyses, such as reliability analysis [1], robust design optimization [2], and sensitivity analysis [3], have received much attention in the last decades. However, the computational cost required for performing the aforementioned analyses depends on several factors such as: the numerical model representing the system, the type of analysis, and the treatment of uncertainties (i.e. aleatory and/or epistemic uncertainty). In the context of reliability analysis, the propagation of parameter uncertainties from model inputs to outputs is performed by means of Monte Carlo simulation based approaches. These simulation approaches include: Monte Carlo (MC) [4], and advanced MC such as: Importance Sampling [5], Directional Sampling [6], Line Sampling [7], Subset Simulation [8] etc. Although, advanced MC methods are very efficient, the computational cost required to perform reliability analysis is usually expensive. A popular strategy to reduce the computational costs is to replace the real model with a surrogate model such as an artificial neural network (ANN). ANNs can be constructed based on few data sets from the underlying model of interest. On the other hand, the use of an ANN for this kind of analysis introduces model selection uncertainty in addition to biasing and variance in the estimated quantity of interest. As a matter of fact, an ANN with a specific architecture trained repeatedly with a finite data set $D_{\text{train}}(x,y)$ results to different performing ANNs whose cost functions are being trapped at different local minima of the cost function solution space. This phenomenon occurs as a result of the random initialization of the weights within each ANN. Consequently, it is of common practice to select the best ANN from the uncertain set on the basis of performance on an independent validation set, and to keep only the network with the lowest validation error and discard the rest. However, there are two disadvantages to such approach. Firstly, all of the effort required to train the remaining networks is wasted. Secondly, the generalization performance of the networks on the validation set has a random component due to the noise on the data, hence the network which had the lowest error on the validation set might perform poorly on a new test set. These disadvantages can be overcome by combining the networks together to form a committee that can significantly improve the robustness of the predicted quantity. Hence, in this paper an approach is proposed to improve the robustness of a neural network when used to predict the probability of failure $p_F$. The outline of this paper is as follows: In Section 2 a succinct theory of reliability analysis using simulation approach and neural network modelling is discussed. This is followed by the proposed approach (Section 3). Next, to demonstrate the applicability of the proposed approach, two numerical examples are tested in Section 4. Finally, conclusions are provided in Section 5.

2 RELIABILITY ANALYSIS

The limit-state function can simply be defined as a deterministic mapping from the $z$-dimensional input space to a one-dimensional output space:

$$G : x \in D_x \subset \mathbb{R}^z \rightarrow y = G(x) \in \mathbb{R}$$ (1)

where $x$ is the $z$-dimensional state variables and $y$ the performance variable. $G(x)$ indicates if a realization $x \in D_x$ corresponds to the safe state ($G(x) > 0$) or failed state ($G(x) \leq 0$). In the context of probability theory, the failure probability, $p_F$, is defined as the probability that a
realization $\mathbf{x} \in D_x$ corresponds to a failed state in terms of the limit-state function $G(\mathbf{x})$:

$$p_F = \mathbb{P}(G(\mathbf{x}) \leq 0) = \int_{D_f} f_X(\mathbf{x}) d\mathbf{x} \quad (2)$$

where $D_f = \mathbf{x} \in D_x : G(\mathbf{x}) \leq 0$ is the failure region and $f_X(\mathbf{x})$ is the joint probability density function of the state variables $X$. As Eq. (2) is analytically intractable due its multidimensional nature, Monte Carlo simulation (MCS) (see [4]) allows one to numerically compute the estimate of the failure probability $p_F$, considering a large sample of size $N$:

$$\hat{p}_F = \frac{1}{N} \sum_{i=1}^{N} I_{G(\mathbf{x})\leq0}(x_i) \quad (3)$$

where $I_{G(\mathbf{x})\leq0}$ is the indicator function for failure such that $I = 1$ for $G(\mathbf{x}) \leq 0$ and $I = 0$ otherwise.

2.1 MODELLING ARTIFICIAL NEURAL NETWORK FOR RELIABILITY ANALYSIS

A setback on the use of MCS to compute the estimate of $p_F$ is the large number of model evaluation required for computing a robust estimate. Hence, an ANN can be used in place of the limit state function to reduce the computational cost. The construction an ANN requires a set of real-valued input/output data pairs $D_{train}(\mathbf{x}, y)$ of size $N_{train}$ generated according to a signal plus noise model $y = \mu(\mathbf{x}) + \epsilon$, where $y$ is the observed performance generated from the expensive model, $\mathbf{x}$ is the independent state variables sampled from a joint probability density $\Omega(\mathbf{x})$, $\epsilon$ is independent, identically distributed (iid) noise sampled from a density $\Psi(\epsilon)$ (not necessarily Gaussian) having mean of 0 and variance $\sigma^2$, and $\mu(\mathbf{x})$ the unknown function that is needed to be approximated by finding an approximation $\hat{\mu}(\mathbf{x})$ from $D_{train}(\mathbf{x}, y)$. A priori assumptions can be made about the functional form of $\mu(\mathbf{x})$. However, since a parametric function class is usually unknown, non-parametric regression approach must be resorted to. Using the non-parametric approach, one constructs an estimate $\hat{\mu}(\mathbf{x})$ of $\mu(\mathbf{x})$ from a large class of functions $Y$ known to have good approximation properties. The class of approximation functions usually contains a set of estimators $f(w, \mathbf{x}) \subset Y$ for which the elements of each subclass $f(w, \mathbf{x})$ are continuously parametrized by a set of $p$ weights $w^\alpha; \alpha = 1, 2, ..., p$. The gradient decent algorithm [9] which is used to minimize the cost function $J$ of the neural network defined as:

$$J = \frac{1}{N_{train}} \sum_{i=1}^{N_{train}} (y_i - \hat{y}_i)^2 \quad (4)$$

by finding a set of weights $w^\alpha$ such that for any given input, the cost function is sufficiently small. However, a limitation of the gradient decent algorithm to train an ANN is the possibility of the cost function to be trapped in a local minimum, thereby reducing the predictive capability of the network.

3 THE PROPOSED APPROACH

The proposed approach aims towards improving the robustness of the prediction made by an ANN when used to perform reliability analysis [10]. The underlying idea behind the proposed approach is to construct a set of ANNs with the same architecture and based on the same training data set $D_{train}(\mathbf{x}, y)$. By doing so, a distribution of identical ANNs having their error functions
trapped in different local minima is created. The major highlight of this approach is that the solution space of the error function is exploited as many times as possible with the possibility of locating a global minima on the error surface. Further, Bayes’ theorem is used to evaluate the posterior probability of each of the trained ANN based on their likelihood to predict the training data. This is followed by the use of a model averaging technique (adjustment factor approach see [11]) to combine the total prediction made by all the ANNs in the set to yield a robust prediction that converges to the true value. Finally, the model uncertainty propagated to the predicted quantity is quantified in terms of confidence intervals.

3.1 BAYESIAN MODEL SELECTION FOR ARTIFICIAL NEURAL NETWORK

Given a set of $M$ identical (i.e. the same model structure) competing ANNs $N_k, k = 1, 2, ..., M$ trained with same data set $D_{\text{train}}(x, y)$, Bayes theorem can be used to express the posterior probability of the $k^{th}$ ANN in the set which is defined by:

$$
P(N_k|D_{\text{train}}) = \frac{P(D_{\text{train}}(x, y)|N_k)P(N_k)}{\sum_{q=1}^{M} P(D_{\text{train}}(x, y)|N_q)P(N_q)}$$

(5)

where $P(D_{\text{train}}(x, y)|N_k)$ is the likelihood of training data $D_{\text{train}}(x, y)$ for the $N_k$ ANN, and $P(N_k)$ is the prior probability of $N_k$, which is the ANN probability evaluated before observing training data $D_{\text{train}}(x, y)$. The prior ANN probability $P(N_k)$ can be specified depending on the existing prior knowledge about the credibility of ANN $N_k$, or it can be given as a uniform probability, $P(N_k) = 1/M$, if no additional information is provided. The advantage of assigning uniform prior probability to $P(N_k)$ is that the difficulty of estimating the prior probability numerically is avoided. The likelihood $P(D_{\text{train}}(x, y)|N_k)$ may be thought of as the probability of observing the training data $D_{\text{train}}(x, y)$ under $N_k$ ANN. It supplies a relative measure of how well the $N_k$ ANN is supported by the training data $D_{\text{train}}(x, y)$. Since the denominator in Eq.(5) is common for all the ANNs, the posterior ANN probability is proportional to prior probability and the likelihood. The likelihood of each ANN is evaluated by measuring the degree of agreement between the training data $D_{\text{train}}(y)$ and the response $\hat{y}$ for each ANN. Hence, a probabilistic relationship between training data $D_{\text{train}}(x, y)$ and ANN predictions $\hat{y}$ involving uncertainty can be described. Typically, the bias function and noise are included as parts of the probabilistic relationship to match ANN predictions with training data. The bias function captures the discrepancies between the expensive model responses and predictions made by the ANN. The noise is usually assumed to be independent and identically distributed normal random variable with a mean of zero [12]. Various authors, see e.g. [13, 14, 15] have used the Bayesian statistical methodology to quantify the uncertainty in the bias function modelled as a Gaussian process. In their works, a mathematical formulation that combines bias function associated with the ANN and noise from training data is utilized to describe the probabilistic relationship between the training data $D_{\text{train}}(x, y)$ and ANN predictions $\hat{y}$. The mathematical formulation of this probabilistic relationship is given by the following equation:

$$
D_{\text{train}}(y) = \hat{y} - \epsilon
$$

(6)

where $\epsilon$ is a random variable that covers both bias associated with the ANN prediction $\hat{y}$ and the noise in the response training data $D_{\text{train}}(y)$. $\epsilon$ is assumed to be an independent identically distributed random variable with a mean $\mu$ of zero. The use of $\epsilon$ with zero mean does not shift ANN prediction $\hat{y}$. This reflects the fact that $\hat{y}$ is the most probable prediction value for the ANN. The bias function is not included as a separate term in the probabilistic relationship. This
is due to the fact that introducing a separate bias function results in shifting the prediction \( \hat{y} \) of the ANN from the initially predicted value.

The likelihood \( P(D_{train}(x,y)|N_k) \) of training data \( D_{train}(x,y) \) for ANN \( N_k \) is evaluated by observing where the training data points \( D_{train}(y) \) are located in the distribution of \( \hat{y} \) estimated by \( N_k \). The procedures to estimate the distribution \( P(\hat{y}|N_k) \) of \( N_k \) and the likelihood \( P(D_{train}(x,y)|N_k) \) is given. First, the uncertainty in errors of predictions \( \hat{y} \) made by \( N_k \) is quantified by introducing an assumption that the prediction errors are independent and identically distributed normal random variable with a mean \( \mu \) of zero. The error of the prediction of the \( k^{th} \) network is represented by the following:

\[
\varepsilon_{ki} = D_{train}(y_i) - \hat{y}_i, \varepsilon_{ki} \sim N(0, \sigma^2_k), i = 1, 2, ..., N \tag{7}
\]

where \( D_{train}(y_i) \) is the \( i^{th} \) training response output data, \( \hat{y}_i \) the prediction of the training data made by \( N_k \), \( \sigma^2_k \) is the variance of prediction error \( \varepsilon_{ki} \), and \( N \) the number of samples in the training data. The prediction error \( \varepsilon_{ki} \) measured is considered to be a random sample from a normal distribution with a mean \( \mu \) of zero and variance \( \sigma^2_k \). Using the principle of maximum likelihood estimation (MLE) (see [16]), the variance \( \sigma^2_k \) for \( N_k \) can be estimated as:

\[
\sigma^2_k = \frac{1}{N} \sum_{i=1}^{N} \varepsilon^2_{ki} \tag{8}
\]

Secondly, the predictive distribution \( P(\hat{y}|N_k) \) of response \( \hat{y} \) under model \( N_k \) is created by including the prediction error obtained in the previous step into the prediction of \( \hat{y} \) made by \( N_k \). This predictive distribution is defined by the following equation:

\[
P(\hat{y}|N_k) = D_{train}(y) + \varepsilon_k \tag{9}
\]

Lastly, assuming that the residuals between the training data \( D_{train}(x,y) \) and ANN \( N_k \) output \( \hat{y} \) are normally and independently distributed with a mean of zero and constant variance \( \sigma^2_k \), the likelihood function \( P(D_{train}(x,y)|N_k) \) is approximated by:

\[
P(D_{train}(x,y)|N_k) \approx \frac{1}{\sqrt{2\pi\sigma^2_k}} \frac{1}{N} \exp\left\{ \frac{-[y_i - \hat{y}_i]^2}{2\sigma^2_k} \right\} \tag{10}
\]

### 3.2 ROBUST ARTIFICIAL NEURAL NETWORK PREDICTION

To obtain a robust prediction from an ANN, the estimates made by all the subsequent trained ANNs are combined using model averaging technique. Specifically, the adjustment factor approach (see [11]) which is a model averaging technique is combined with Bayes’ theorem. In this way, the ANN having the highest posterior probability is used in conjunction with other respective ANNs trained to correct the bias estimate predicted by the single ANN. The adjustment factor \( A_f \) is evaluated by assuming the error between the prediction of all the subsequent trained ANNs and the training data are normally distributed. The robust ANN prediction can be obtained from the following equation:

\[
y_{\text{robust}} = \hat{y}^* + A_f \tag{11}
\]

where \( \hat{y}^* \) represents the point estimate of the best ANN in the set characterised by the highest probability, and \( y_{\text{robust}} \) represent the robust prediction which incorporates the model uncertainty.
Since the adjustment factor \( A_f \) is assumed to be a normal distribution, the expected value and variance of the adjustment factor \( A_f \) is given by the following relationships:

\[
E(A_f) = \sum_{k=1}^{M} P(N_k|D_{\text{train}})(\hat{y}_k - \hat{y}^*)
\]

\[
\text{Var}(A_f) = \sum_{k=1}^{M} P(N_k|D_{\text{train}})(\hat{y}_k - E(y_{\text{robust}}))^2
\]

Similarly, the expected value and variance of the robust prediction \( y_{\text{adj}} \) can be estimated from the following relationships:

\[
E(y_{\text{robust}}) = \hat{y}^* + E(A_f)
\]

\[
\text{Var}(y_{\text{robust}}) = \text{Var}(A_f)
\]

where \( E(A_f) \) and \( \text{Var}(A_f) \) represents the expected value and variance of the adjustment factor, and \( E(y_{\text{robust}}) \) and \( \text{Var}(y_{\text{robust}}) \) represents the expected value and variance of the robust estimate.

### 3.3 CONFIDENCE INTERVAL FOR ROBUST ESTIMATE

To quantify the uncertainty in the robust prediction \( y_{\text{robust}} \) due to model uncertainty, confidence intervals are established. In particular, the 5\(^{th}\) and 95\(^{th}\) percentiles of the robust prediction are used quantify the model uncertainty. In theory, this interval is likely to contain the true estimated value. As the model uncertainty is assumed to follow normal distribution, the confidence intervals (see [17]) are calculated from the following equations:

\[
CI_{\text{upper}} = E(y_{\text{robust}}) + z^* \sqrt{\text{Var}(y_{\text{robust}})}
\]

\[
CI_{\text{lower}} = E(y_{\text{robust}}) - z^* \sqrt{\text{Var}(y_{\text{robust}})}
\]

where \( CI_{\text{upper}} \) and \( CI_{\text{lower}} \) represents the upper and lower confidence intervals of the robust estimate and \( z^* \) represents the upper critical value of the Gaussian distribution quantifying the model uncertainty.

### 4 NUMERICAL EXAMPLE

#### 4.1 THE HAT FUNCTION

The hat function is defined by the analytical expression [18]:

\[
G(x) = 20 - (x_1 - x_2)^2 - 8(x_1 + x_2 - 4)^3
\]

where \( x_i, i = 1, 2 \) is defined as Gaussian variables with mean \( \mu_{x_i} = 0.5 \) and standard deviation \( \sigma_{x_i} = 1.0 \). Failure is defined as \( G(x) \leq 0 \) hence \( p_F = P(G(x)) \leq 0 \). The limit state surface plot of the hat function is shown in Fig. [1].
The aim of this example is to verify the proposed approach by replacing the limit state function with an ANN, then compute a robust estimate of $\hat{\rho}_F$, quantify the model uncertainty and, finally, verify the number of identical ANNs that must be trained to attain an optimal robust estimate of $\hat{\rho}_F$.

4.2 ANALYSIS 1

Training samples $D_{\text{train}}(x,y)$ of size $N_{\text{train}} = 2000$ have been generated via Latin hypercube sampling (LHS) algorithm[19] from the hat function. Two sets $Z_1 = N_k, k = 1,2,\ldots M$ and $Z_2 = N_i, i = 1,2,\ldots M$ composed of $M = 1000$ identical ANNs have been trained based on $D_{\text{train}}(x,y)$. Specifically, in the first set ($Z_1$), all the training samples in $D_{\text{train}}(x,y)$ have been used to train the ANNs to maximize their predictive performances. For the second set $Z_2$, 80% of the training samples $D_{\text{train}}$ have been used to train the ANNs and the remaining 20% used for validation. The network architecture chosen for the ANNs in both sets composed of three hidden layers (2,7,1). Next, the posterior probability of the ANNs in set $Z_1$ has been estimated using Bayes’ formula by assigning uniform prior probability $P(N_k) = 1/M$ to each ANN. On the other hand, the coefficient of determination $R^2$ for the ANNs in set $Z_2$ have been estimated based on the 20% validation samples. Table[1] shows a comparison of 10 selected ANNs from $Z_1$ and $Z_2$ based on their posterior probabilities and their error values $R^2$. It should be noted that $i^{th}$ ANN in both set ($Z_1$ and $Z_2$) have been trained inside the same iteration loop, hence the initialization of the weight values within each loop it is assumed to be similar. Therefore, their resultant performances are expected to be similar. As shown in Table[1] although the ANNs $N_i, i = 1,2,\ldots M$ in sets $Z_1$ and $Z_2$ are identical as they have been trained in the same iteration loop, the performance measures in terms of the posterior probability and $R^2$ shows no agreement. For example, the $6^{th}$ and $10^{th}$ ANNs have the highest posterior probability, however their corresponding $R^2$ values don’t show a similar trend. Hence, we can support our claim that the use of $R^2$ value to select the best model is a biased method. Further, to implement the proposed approach, the ANNs in $Z_1$ have been chosen as they have better performance (i.e. due more samples used to train them). To accurately compute a robust estimate of $\hat{\rho}_F$, $10^4$ Monte Carlo simulation runs have been used for each ANN, and the proposed approach presented have been used to average out the prediction made by each ANN model into a robust value that is converges to the true value. Finally, the model uncertainty propagated to robust prediction of $\hat{\rho}_F$ has been quantified in terms of confidence intervals estimated.
Table 1: Artificial Neural Networks Posterior Probability Calculated Compared to Corresponding $R^2$ value.

| ANN | $P(D_{\text{train}}|N_k)$ | $R^2$ |
|-----|--------------------------|-------|
| 1   | 0.103                    | 0.9989|
| 2   | 0.087                    | 0.9998|
| 3   | 0.101                    | 0.9993|
| 4   | 0.117                    | 0.9999|
| 5   | 0.070                    | 0.9998|
| 6   | 0.122                    | 0.9998|
| 7   | 0.093                    | 0.9996|
| 8   | 0.099                    | 0.9999|
| 9   | 0.086                    | 0.9999|
| 10  | 0.122                    | 0.9995|

4.3 ANALYSIS 2

To check the number of ANNs that must be trained in order to obtain a robust value (i.e. close to reference value), the real model has been used to estimate the reference value of $\hat{p}_F$ adopting the same failure criteria (i.e $G(x) \leq 0$) and $N = 10^4$ samples. On the other hand, 3 separate tests adopting our approach utilizing $M = 100$, 1000, 10000 identical ANNs respectively have been carried out. As shown in Fig. 2 the robust estimate of $\hat{p}_F$ obtained from the proposed approach converges to the true value (i.e. blue dashed horizontal line) when $M = 1000$ identical ANNs are used. This means that $M = 1000$ ANNs is sufficient enough to explore the entire solution space of the error function, thus locating a global minima. The importance of this approach is that it can lead to significant improvements in the predictions $\hat{p}_F$, while involving little additional computational effort.

![Graph of Failure Probability vs. ANN](image1.png)

![Probability Density Function](image2.png)

Figure 2: Confidence Intervals and Probability Density Functions Representing Model Uncertainty for $M = 100, 1000, 10000$ Identical Trained Artificial Neural Networks

4.4 CANTILEVER BEAM

A cantilever beam of length $L$ and rectangular cross section of width $b$ and height $h$ is loaded at the end by a concentrated point load $P$. The displacement $w$ at the tip of the beam should be
determined for the case where the point load $P$, the Young’s modulus $E$, the density $\rho$ of the material and the height $h$ are uncertain.

![Figure 3: Cantilever Beam](image)

Uncertainties of the width $b$ and of the length $L$ are assumed to be negligible. The displacement $w$ at the tip of the beam where load $P$ is applied can be expressed mathematically as:

$$w = \frac{gbhL^4}{8EI} + \frac{PL^3}{3EI}$$  \hspace{1cm} (19)

where $g$ denotes gravitational constant, and $I$ is given as:

$$I = \frac{bh^3}{12}$$  \hspace{1cm} (20)

The limit state function for the cantilever beam is defined as:

$$G_{\text{cantilever}}(x) = \beta - w(x)$$  \hspace{1cm} (21)

where $\beta = 0.01$ is the maximum allowable displacement of the beam. In this example, the parameter uncertainties are modelled as random variables characterised by probability density function given in Table 2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distribution</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>SI – unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>Log – Normal</td>
<td>5</td>
<td>0.4</td>
<td>KN</td>
</tr>
<tr>
<td>$h$</td>
<td>Normal</td>
<td>0.24</td>
<td>0.01</td>
<td>m</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Log – Normal</td>
<td>600</td>
<td>140</td>
<td>Kg/m$^3$</td>
</tr>
<tr>
<td>$E$</td>
<td>Log – Normal</td>
<td>10</td>
<td>1.6</td>
<td>GN/m$^2$ \hspace{1cm}</td>
</tr>
</tbody>
</table>

Table 2: Model Input Parameters

The aim of this example is to study how small number of training samples (i.e. $N_{\text{train}} = 50, 100, 150, 200$) affects the robust estimate and the corresponding confidence intervals.

4.5 ANALYSIS 3

In this section, 2 sets (i.e., similar to Section 4.2) of identical ANNs (i.e. $M = 1000$) with hidden layer configuration of $(4,7,1)$ have been constructed based on $D_{\text{train}}(x,y), N_{\text{train}} = 50, 100, 150, 200$ obtained via LHS algorithm [19]. The approach used in Section 4.2 has been adopted here to estimate the posterior probability and $R^2$ value of the ANNs. Further, to accurately compute an estimate of $\hat{p}_F$, $10^4$ Monte Carlo simulation runs have been used for each ANN in the first set, and the proposed approach presented have been used to combine the prediction made by each ANN in the first set into a robust $\hat{p}_F$ estimate. Finally, the model uncertainty has been
quantified in terms of confidence intervals of the robust estimate. The results of these analyses are shown in Fig. 4.

![Figure 4: Robust Confidence Intervals Obtained from Different Number of Training Samples (N_{train} = 50, 100, 150, 200)](image)

Notice that the "true" (i.e., reference) value of the failure probability (i.e., $\hat{p}_F = 0.0738$, shown in blue dashed lines in Fig. 4) has been obtained with a large number samples (i.e., $N = 10^4$) of simulations of the original model to provide a robust term for comparison. Also, from the results in Fig. 4 as the number of training samples $N_{\text{train}}$ increases, the width of the confidence intervals decreases and the expected value of the robust estimate approaches the reference value (i.e., $\hat{p}_F = 0.0738$). On the other hand, in the cases of small training data sets (e.g., $N_{\text{train}} = 50, 100$) the failure probabilities are significantly overestimated by the proposed approach (e.g., the expected values of the robust estimate are far off from the reference value) and the associated model uncertainties are quite large. However, in all the cases for small training data sets, the confidence intervals derived is robust enough to capture the true estimate. Hence, in a situation where training data set is small, this approach can be used as a guide to derive a robust confidence interval that is adequate to capture the true value that is being estimated.

5 CONCLUSIONS

Reliability analysis of complex models using the simulation approach is computationally expensive due to the large number of model evaluations required to compute their robust measures. In this paper, an ANN is being used as substitutes for an expensive model to alleviate the computational restrictions. The use of ANN for this kind of analysis introduces additional biasing and variance (i.e uncertainties) to the predicted quantity. It has been shown that the use of cross-validation technique to select the best ANN out of a set of ANN with identical architecture introduces biasing and reduces the robustness of the predicted quantity. Therefore, a novel approach has been presented to enhance the accuracy of the prediction (i.e. robustness) made by an ANN and quantify the model uncertainties in terms of confidence intervals. The proposed approach combines Bayesian model selection and model averaging technique into a unified framework. The applicability of the proposed approach has been demonstrated on two examples. Although the computational effort required for implementing the proposed approach is expensive, parallelization strategies can be adopted to reduce this effort.
Acknowledgements

This work has been partially supported by the EPSRC Grant EP/M018717/1 (Smart on-line monitoring for nuclear power plants (SMART)).

REFERENCES


A NEW NESTED SPACE-FILLING DESIGNS ON EXPERIMENTS WITH DIFFERENT RANGES OF FACTORS

Jin Xu¹, Xiaojun Duan¹, Zhengming Wang¹ and Liang Yan¹

¹College of Science, National University of Defense Technology, Changsha, Hunan, HN 731, PRC
e-mail: xujin_nudt@163.com; xj_duan@163.com; wzm@nudt.edu.cn; yanliang@nudt.edu.cn

Keywords: computer experiment, Nested designs, Radar homing missile, multifidelity computer model, Latin hypercube designs

Abstract. Experiments with different levels of accuracy or fidelity are very common in statistics, applied mathematics and engineering. Nested space-filling designs are useful approach to deal with those questions. But in some special situations, e.g., the captive flight tests and missile tests in radar homing precision experiments, the ranges of factors are different in some accuracy experiments and sometimes they have overlapping portions. The existing methods deal almost exclusively with experiments with same ranges of factors. In this paper, we propose a new construction of nested space-filling designs named NDF(Nested Designs on Factor with different ranges ) to solve these problems. The proposed construction are useful in the experiments with different ranges of factors and achieves uniformity in low dimensions.
1 INTRODUCTION

Computer experiments, which simulate real world phenomena using some method in computer such as finite element analysis and computational fluid dynamics, have become widely used in statistics, applied mathematics and engineering ([1],[2],[7]). A main goal in many computer experiments is to estimate the expected output of a computer model given a distribution of inputs. To address this issue, McKay, Conover, and Beckman ([8]) introduced Latin hypercube designs. Space-filling designs are important in the design and analysis for computer experiments ([1],[2]). It is commonly believed that design points should be evenly spread in the experimental region. Nested Latin hypercube designs are popular researched in recently years ([3],[4],[5]), which mostly deal with the high-accuracy and low-accuracy experiments, where Sliced Latin hypercube designs are similar to the Nested one ([10],[11],[12]). Meanwhile, a new method to construct SLHD is proposed in [9]. But in some special situations, e.g., the captive flight tests and missile tests in radar homing precision experiments, the ranges of factors are different in the original experiment and substitution experiment, NLHD could not deal with this situations and the existing methods deal almost exclusively with experiments which have same ranges of factors. In this paper, we propose a new construction of space-filling designs which first generates two small Latin hypercube designs and then arrange them together to form a new Latin hypercube design. This construction is similar with Sliced Latin hypercube designs (SLHD), however, the proposed construction is easy to implement, capable of accommodating any number of run size.

Firstly, We introduce a motivating example. Captive fight and missile physics experiments are two important methods to test the property of radar homing missile. But the ranges of factors are different in the two experiments when considering the CEP (Circular Error Probability) as the response. It is known that the error of radar homing is related to factors $R$, $\dot{R}$ and $S/C$ which denote velocity, acceleration and signal-to-clutter ratio separately. The parameter interval for different experimental states is displayed in table 1. In general, the cost $W$ of the four different experiments are not same, their relation is $W_1 < W_2 < W_3 < W_4$. The main difference factors which could be designed between captive fight and missile physics experiments are $R$ and $\dot{R}$.

<table>
<thead>
<tr>
<th></th>
<th>State 1</th>
<th>State 2</th>
<th>State 3</th>
<th>State 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{R}$</td>
<td>300–400</td>
<td>300–400</td>
<td>500–900</td>
<td>500–900</td>
</tr>
<tr>
<td>$\dot{R}$</td>
<td>0–30</td>
<td>0–30</td>
<td>0–300</td>
<td>0–300</td>
</tr>
<tr>
<td>$S/C$</td>
<td>1.5–2.5</td>
<td>3–5</td>
<td>1.5–2.5</td>
<td>3–5</td>
</tr>
</tbody>
</table>

Then, we introduce LHD briefly. An $n \times m$ matrix $A = (a_{ij})$ is a called a Latin hypercube if each column of $A$ is a permutation of $\{1, \cdots, n\}$. There are two methods to generate the design points in the unit cube $[0, 1]^n$ from $A$. The first is through $D_{ij} = (a_{ij} - 0.5)/n$, and the other is through $D_{ij} = (a_{ij} - u_{ij})/n$ where $u_{ij}$ are independent random variables with a common uniform distribution on $[0, 1]$. The difference between the two methods can be seen as follows. When projected onto each of the $m$ variables, both methods have the property that one and only one of the $n$ design points fall within each of the $n$ small intervals defined by $[0, 1/n), [1/n, 2/n), \cdots, [(n - 1)/n, 1]$. The first method gives the mid-points of these intervals while the second gives points are uniformly distributed in their corresponding intervals.
2 A New Construction of Latin hypercube design

In this section, we present a simple method to construct a new nested Latin hypercube design, named NDF, which can be used in the experiments with factors that have different ranges. The proposed Latin hypercube design has different construction from the existing one in [3], where we first generates two small Latin hypercube designs and then arrange them together to form the NDF. And this construction. The proposed LHD and its two parts could achieve good uniformity in low dimensions, then this construction could deal with the above mentioned problems efficiently and directly. Meanwhile, its parameters could be more flexible parameters.

Let $Z_n$ denotes the set $\{1, \cdots, n\}$, now we present a $Z_n$-to-$Z_m$ projection where $n, m$ are two positive integers with $n > m$. For any $x$ in $Z_n$, define

$$\varphi (x) = \lfloor x * m/n \rfloor$$

We now introduce some useful notation. For $a \in R$, $\lfloor a \rfloor$ denotes the smallest integer no less than $a$ and $\lceil a \rceil$ denotes the largest integer no greater than $a$. Similarly define $\lceil D \rceil$ and $\lfloor D \rfloor$ for a real matrix $D$. For a matrix $A$, let $A(:, j)$ be its $j$th column, $A(i, :)$ be its $i$th row, and $A A (i, j)$ be its $(i, j)$th element. For the projection $\varphi : Z_n \rightarrow Z_m$, let $y \in Z_m$, the set $\{\varphi^{-1} (y)\}$ denotes all the integers $x \in Z_n$ which satisfy $\varphi (x) = y$. For a matrix $H$ based on $Z_n$, $\varphi (H)$ denotes the matrix obtained form $H$ after the levels of its entries are collapsed according to $\varphi$. Obviously, we have some results as follows: If $n, m$ are two positive integers with $n = k * m + b$, where $k > 0, 0 \leq b < m$ are integers, for the projection $\varphi$, we have that

Lemma 1 If $n, m$ are two positive integers with $n = km + b$, where $k > 0, 0 \leq b < m$ are integers, for the projection $\varphi$, we have that

1. $\varphi$ is a monotonous increasing function;

2. For any integer $y \in Z_m$, the number of the entries $\{\varphi^{-1} (y)\}$ is $k$ or $k + 1$, and the number of $\{\varphi^{-1} (y)\}$ which have $k$ and $k + 1$ entries is $m - b$ and $b$ respectively.

Proof. Only part $(ii)$ needs a proof. Consider $x \in Z_n, y \in Z_m$ and $x \in \{\varphi^{-1} (y)\}$, obviously, $y - 1 < xn/n \leq y$, so we have $n(y - 1)/m < x \leq yn/m$. Thus,

$$x \in \lfloor n(y - 1)/m \rfloor, \lfloor yn/m \rfloor$$

Note that $n(y - 1)/m = k(y - 1) + b(y - 1)/m$ and $yn/m = ky + by/m$. For $x$ is a integer, which yields that the number of the entries $\{\varphi^{-1} (y)\}$ is $k$ or $k + 1$. For there are $m$ sets $\{\varphi^{-1} (y)\}$ and $n$ entries of $Z_n$, the rest of part $(ii)$ holds.

Lemma 2 For $n_1, n_2, n_3, p$ are positive integers with $n_1 = k_1 p, n_2 = k_2 p, n_3 = k_3 p$, and $k_1 \geq k_2 \geq k_3 > 0$ are integers, let $\varphi_1 : Z_{n_1} \rightarrow Z_{n_2}, \varphi_2 : Z_{n_2} \rightarrow Z_{n_3}$, and $\varphi : Z_{n_1} \rightarrow Z_{n_3}$. Then, for any $x \in Z_{n_1}$, the equation

$$\varphi (x) = \varphi_2 [\varphi_1 (x)]$$

is satisfied.

Proof. For any $x \in Z_{n_1}$,

$$\varphi (x) = \lfloor xn_3/n_1 \rfloor = \lfloor xk_3/k_1 \rfloor$$
and
\[ \varphi_2 [\varphi_1 (x)] = [\lceil x \rceil_{n_1}]_{n_3} = [\lceil x \rceil_{k_1}]_{k_3} / k_2 \]

Let \( q \) be an integer and \( \varphi (x) = q \), then, we must have that \( k_1 (q - 1) / k_3 < x \leq k_1 q / k_3 \). Thus
\[
\varphi_2 [\varphi_1 (x)] \leq \lceil [q k_1 k_2 / (k_1 k_3)] k_3 / k_2 \rceil < (q k_2 / k_3 + 1) \frac{k_3}{k_2} = [q + k_3 / k_2]
\]
\[ = q + 1 \]

Similarly,
\[
\varphi_2 [\varphi_1 (x)] > \lceil [k_1 k_2 (q - 1) / (k_1 k_3)] k_3 / k_2 \rceil \geq \lceil (k_2 (q - 1) / k_3) k_3 / k_2 \rceil = q - 1
\]

For \( \varphi_2 [\varphi_1 (x)] \) is an integer, \( \varphi_2 [\varphi_1 (x)] = q \) holds, which completes the proof.

Let \( A_1, A_2 \) are two LHDs generated independently, where \( A_1 \) contains \( n_1 \) points in \( m \) factors with its factor levels \( 1, \ldots, n_1 \) and \( A_2 \) contains \( n_2 \) points in \( m \) factors with its factor levels \( 1, \ldots, n_2 \), without losing generality, let \( n_1 \leq n_2 \). Stack \( A_1 \) and \( A_2 \) to form an \( (n \times m) \) matrix \( H = A_1 \cup A_2 \), where \( n = n_1 + n_2 \), \( H_1 \) denotes the submatrix of \( H \) corresponding to \( A_1 \) and \( H_2 \) denotes the submatrix of \( H \) corresponding to \( A_2 \). Let \( \varphi_1 \) denotes \( \varphi : Z_n \rightarrow Z_{n_1} \) and \( \varphi_2 \) denotes \( \varphi : Z_n \rightarrow Z_{n_2} \). Let \( t_1 (i) \) denotes the numbers of entries in \( \{ \varphi_1^{-1} (i) \} \subset Z_n \) when \( i \in Z_{n_1} \) and \( t_2 (j) \) denotes the numbers of entries in \( \{ \varphi_2^{-1} (j) \} \subset Z_n \) when \( j \in Z_{n_2} \). Note that \( t_2 (i) = 1 \) or \( 2 \). For the arbitrary \( l \)th column of \( H \), this construction method consists of there steps:

**Step 1:** Replace the level \( i \in H_2 (; l) \) satisfies \( t_2 (i) = 1 \) with the single integer \( c \in \{ \varphi_2^{-1} (i) \} \), then let \( t_2 (i) = 0, t_1 [\varphi_1 (c)] = t_1 [\varphi_1 (c)] - 1 \) and delete \( c \) from \( \{ \varphi_1^{-1} [\varphi_1 (c)] \} \). Repeat step 1 until there is no \( i \in H_2 (; l) \) satisfies the condition. If there exists \( j \in H_1 (; l) \) satisfies \( t_1 (j) = 1 \), go to step 2, else go to step 3;

**Step 2:** Replace the level \( j \in H_1 (; l) \) satisfies \( t_1 (j) = 1 \) with the single integer \( c \in \{ \varphi_2^{-1} (j) \} \), then let \( t_1 (j) = 0, t_2 [\varphi_2 (c)] = t_2 [\varphi_2 (c)] - 1 \) and delete \( c \) from \( \{ \varphi_2^{-1} [\varphi_2 (c)] \} \). Repeat step 2 until there is no \( j \in H_1 (; l) \) satisfies the condition. If there exists \( i \in H_2 (; l) \) satisfies \( t_2 (i) = 1 \), return to step 1, else go to step 3;

**Step 3:** Choose any integer \( i \in H_2 (; l) \) satisfies \( t_2 (i) = 2 \), and replace it with the any integer \( c \in \{ \varphi_2^{-1} (i) \} \). For another element \( d \in \{ \varphi_2^{-1} (i) \} \), let \( t_1 (\varphi (d)) = t_1 (\varphi (d)) \) and delete \( d \) from \( \{ \varphi_1^{-1} (\varphi (d)) \} \). If \( t_1 (\varphi (d)) = 1 \), go to step 2. If for any \( i \in H_2 (; l) \), the \( t_2 (i) = 0 \) is established, the achieved \( H \) is the wanted NDF.

We must have the conclusions shown as follows easily:

**Theorem 1** Consider \( H, H_1 \) and \( H_2 \) as constructed above. Then we have that

1. the matrix \( H \) is a LHD(\( n, m \));

2. the submatrix \( H_1 \) and \( H_2 \) of \( H \) become LHD(\( n_1, s \)) and LHD(\( n_1, s \)) when the \( n \) levels are collapsed into \( n_1 \) and \( n_2 \) levels according to the scheme: \( i \rightarrow \varphi_1 (i) \) and \( i \rightarrow \varphi_1 (i) \) respectively.
Some examples are provided below to illustrate the proposed procedure.

**Example 1** We generate two LHD with \( n_1 = 3 \), \( n_2 = 6 \), and \( m = 3 \), then the LHD is as follows.

\[
H = \begin{bmatrix}
2 & 1 & 1 \\
1 & 3 & 2 \\
3 & 2 & 3 \\
6 & 5 & 1 \\
3 & 3 & 6 \\
2 & 2 & 2 \\
4 & 4 & 3 \\
5 & 6 & 5 \\
1 & 1 & 4 \\
\end{bmatrix}
\rightarrow
\begin{bmatrix}
6 & 2 & 3 \\
2 & 8 & 5 \\
9 & 5 & 8 \\
8 & 7 & 1 \\
4 & 4 & 9 \\
3 & 3 & 2 \\
5 & 6 & 4 \\
7 & 9 & 7 \\
1 & 1 & 6 \\
\end{bmatrix}
\]

Figure 1 is the NDF proposed above and the nested Latin hypercube design proposed in [3], the second part of Latin hypercube design \( H \) could achieve more uniformity in one dimension. Meanwhile, the new nested Latin hypercube could deal with arbitrarily parameters which have not been considered in other paper and the two parts of LHD could achieve best uniformity in one dimension, which is shown in Figure 2.

**Example 2** For \( n_1 = 3 \), \( n_2 = 5 \), and \( m = 3 \), then the NDF is as follows.

\[
H = \begin{bmatrix}
3 & 1 & 2 \\
2 & 3 & 1 \\
1 & 2 & 3 \\
5 & 1 & 1 \\
2 & 2 & 4 \\
4 & 5 & 3 \\
3 & 3 & 2 \\
1 & 4 & 5 \\
\end{bmatrix}
\rightarrow
\begin{bmatrix}
8 & 2 & 5 \\
5 & 7 & 2 \\
2 & 5 & 7 \\
7 & 1 & 1 \\
3 & 3 & 6 \\
6 & 8 & 4 \\
4 & 4 & 3 \\
1 & 6 & 8 \\
\end{bmatrix}
\]
3 Numerical illustration

For the motivating problems described in Section 2, i.e., run a computer model in batches and multiple computer models based on similar mathematics where using different run size. Actually, the first problem can be viewed the experiments using multiple computer models based on the same mathematics.

More generally, the NDF. For experiments using $t$ similar computer models, $f_1, \cdots, f_t$, where the inputs of each $f_i$ are $X = (x_1, \cdots, x_m)$ whose distribution is the uniform measure on $(0, 1]^m$. For $i = 1, \cdots, t$, define $\mu_i = E[f_1(X)]$. The goal here is to run $f_1, \cdots, f_t$, at $n_1, \cdots, n_t$ selected input values, to estimate $\mu_1, \cdots, \mu_t$. For $0 \leq \lambda_i \leq 1$, $i = 1, \cdots, t$, a linear combination of $\mu_1, \cdots, \mu_t$, given by

$$\eta = \sum_{i=1}^{t} \lambda_i \mu_i$$

can also be of interest in practice. Now, we introduce four different sampling schemes to achieve this goal as follows:

**Definition** Suppose that $n_1, \cdots, n_t, t, n$ are positive integers, with $n = \sum_{i=1}^{t} n_i$.

1. Let IID denote a scheme that takes an independent and identically distributed sample of $n_i$ runs for $f_i$, with the $t$ samples generated independently;

2. Let SPLIT denote a scheme that randomly takes an Latin hypercube sample of $n$ runs and splits it into $t$ subdesigns which run sizes are $n_1, \cdots, n_t$, corresponding to $f_1, \cdots, f_t$.

3. Let COMBINE denote a scheme that obtains $t$ independent Latin hypercube designs with $n_1, \cdots, n_t, t$ runs, each of which is associated with one $f_i$;

4. Let NFD denote a scheme that produces designs by using the method in Section 2, where each part, assigned to one $f_i$, is a smaller Latin hypercube design of $n_i$ levels.

**Example 3** According to the physical background, we use the function

$$f_1(X) = \frac{x_1}{2500} + \frac{x_2}{500} + \frac{1}{20\sqrt{x_3}}$$
to act as the stage 1,2 where the distribution of \( x_1, x_2, x_3 \) is the uniform measure on \((300, 400], (0, 30], (1.5, 2.5] \cup (3, 5]\). Then, we regard the stage 3,4 as

\[
f_2(x) = \frac{x_1}{5000} + \frac{x_2}{486} + \frac{1}{18\sqrt{x_3}}
\]

where the distribution of \( x_1, x_2, x_3 \) is the uniform measure on \((500, 900], (0, 300], (1.5, 2.5] \cup (3, 5]\). For each of the four methods described above, we compute \( \hat{\eta} \), with \( \lambda_1 = \lambda_2 = 1/2, 2000 \) times for \( n_1 = 5 \) and \( n_2 = 5, 6, \cdots , 10 \). Table 2 presents the RMSE (root mean square error) of \( \hat{\eta} \) over the 2000 replicates for every method. The table clearly indicates that for every value of \( n_2 \), with \( n_1 = 5 \), the \( N \) method achieves greater degree of variance reduction than the other three methods for \( \hat{\eta} \).

### 4 CONCLUSIONS

In this paper, we propose a new construction of Latin hypercube designs which first generates two small Latin hypercube and then arrange them together to form a new Latin hypercube design. The proposed construction is easy to implement, capable of accommodating any number of run size. For better uniformity in low dimension, the orthogonal array may be considered to constict the new Latin hypercube designs for the experiments with different ranges of factors.

### REFERENCES


---

**Table 2: RMSEs of \( \hat{\eta} \) of the four methods for Example 3.**

<table>
<thead>
<tr>
<th>( n_1 = 5 )</th>
<th>IID</th>
<th>COMBINE</th>
<th>SPLIT</th>
<th>NDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_2 = 5 )</td>
<td>0.02200</td>
<td>0.01070</td>
<td>0.00486</td>
<td>0.00387</td>
</tr>
<tr>
<td>( n_2 = 6 )</td>
<td>0.02133</td>
<td>0.01223</td>
<td>0.00472</td>
<td>0.00258</td>
</tr>
<tr>
<td>( n_2 = 7 )</td>
<td>0.02131</td>
<td>0.01272</td>
<td>0.00471</td>
<td>0.00210</td>
</tr>
<tr>
<td>( n_2 = 8 )</td>
<td>0.02137</td>
<td>0.01330</td>
<td>0.00465</td>
<td>0.00203</td>
</tr>
<tr>
<td>( n_2 = 9 )</td>
<td>0.02095</td>
<td>0.01428</td>
<td>0.00465</td>
<td>0.00244</td>
</tr>
<tr>
<td>( n_2 = 10 )</td>
<td>0.02088</td>
<td>0.01459</td>
<td>0.00453</td>
<td>0.00341</td>
</tr>
</tbody>
</table>


SEISMIC SAFETY ASSESSMENT BY AN ENHANCED MONTE CARLO METHOD

Christian Gasser\textsuperscript{1} and Christian Bucher\textsuperscript{1}

\textsuperscript{1}Forschungsbereich fuer Baumechanik und Baudynamik, Technische Universitaet Wien
Karlsplatz 13/206-3, A-1040 Vienna, Austria
\texttt{e-mail: Christian.Gasser@tuwien.ac.at}

\textbf{Keywords:} Probabilistic Safety Analysis, Seismic Hazard, Fragility Curve, Monte Carlo Simulation, Failure Probability, Arch Dam.

\textbf{Abstract.} A probabilistic seismic safety assessment is performed by means of Monte Carlo simulation. Emphasis is on the identification of response uncertainty as a result of the stochastic nature of ground motions in time domain. Especially, the relevance of low failure probabilities at moderate earthquakes is investigated. As application example an arch dam with nonlinear material properties and fluid-structure interaction is analysed. Exceedance probabilities for selected peak ground acceleration levels are obtained from a site-specific hazard curve. Stochastic ground motions are generated accordingly by a Kanai-Tajimi model. The fragility curve is calculated by performing Monte Carlo simulations. Its tail probabilities are estimated by a new extrapolation technique.
1 INTRODUCTION

Since the characteristics of earthquakes are highly random, seismic safety assessments reasonably have to be based on stochastic methods somehow. In engineering practice, usually, major structures are designed to withstand the so-called maximum credible earthquake, which depends on the importance of the structure and is defined by probabilistic methods. However, the ultimate limit state is then calculated in a deterministic way, assuming that failure can only happen at or above the maximum credible earthquake.

In fact, the probability that a certain failure state happens in a given period of time depends on both, the occurrence probabilities of all the different earthquake intensities and the respective conditional failure probabilities. For \( n \) intensity levels the failure probability is calculated by

\[
P_F = \sum_{k=1}^{n} P(F|I_k)Q(I_k)
\]

in which \( P(F|I_k) \) is the conditional failure probability given the earthquake intensity \( I_k \) and \( Q(I_k) \) is the occurrence probability of the intensity.

In the present work, \( P(F|I_k) \) is estimated by Monte Carlo simulation. For that purpose, the intensity levels are replaced by peak ground accelerations \( a_{g,\text{max}} \). For many regions of the earth seismological services provide diagrams in which the seismic hazard is indicated as exceedance probability in (e.g.) 50 years as a function of the peak ground acceleration, \( H(a_{g,\text{max}}) \). Hence, it is straightforward to replace the sum in equation 1 by an integral. \( Q(I_k) \) is then replaced by the probability density function of \( a_{g,\text{max}} \). This corresponds to the negative derivative of the exceedance probabilities provided by seismological services:

\[
h(a_{g,\text{max}}) = -\frac{dH(a_{g,\text{max}})}{da_{g,\text{max}}}
\]

The failure probability is then calculated by

\[
P_F = \int_0^{\infty} P(F|a_{g,\text{max}})h(a_{g,\text{max}})da_{g,\text{max}}.
\]

Obviously, it is referred to the same period of time as is the seismic hazard function \( H(a_{g,\text{max}}) \), typically to 50 years.

It is shown how much different earthquake intensities (corresponding to certain return periods) contribute to the overall failure probability. As application example nonlinear analysis of an arch dam is performed. Ground accelerations are created by the Kanai-Tajimi model and serve as stochastic excitation for the Monte Carlo simulations. The efficiency is increased by applying a new extrapolation technique, which is particularly suitable to estimate small failure probabilities more accurately without need for more simulation runs.

2 THE ARCH DAM

As application object for the reliability analysis a 220 m high arch dam is chosen. The earthquake safety of this (fictitious) structure has already been studied und documented in detail by Goldgruber in his PhD thesis [1]. The FE-model of the dam, see fig. [1] has kindly been made available to the authors and is used for this research. The focus of this work, however, is a Monte Carlo-based probabilistic analysis of the earthquake safety of structures in general, rather than a structural analysis of a certain dam. Hence, some changes are made on the dam described in [1]:

---

1 INTRODUCTION

Since the characteristics of earthquakes are highly random, seismic safety assessments reasonably have to be based on stochastic methods somehow. In engineering practice, usually, major structures are designed to withstand the so-called maximum credible earthquake, which depends on the importance of the structure and is defined by probabilistic methods. However, the ultimate limit state is then calculated in a deterministic way, assuming that failure can only happen at or above the maximum credible earthquake.

In fact, the probability that a certain failure state happens in a given period of time depends on both, the occurrence probabilities of all the different earthquake intensities and the respective conditional failure probabilities. For \( n \) intensity levels the failure probability is calculated by

\[
P_F = \sum_{k=1}^{n} P(F|I_k)Q(I_k)
\]

in which \( P(F|I_k) \) is the conditional failure probability given the earthquake intensity \( I_k \) and \( Q(I_k) \) is the occurrence probability of the intensity.

In the present work, \( P(F|I_k) \) is estimated by Monte Carlo simulation. For that purpose, the intensity levels are replaced by peak ground accelerations \( a_{g,\text{max}} \). For many regions of the earth seismological services provide diagrams in which the seismic hazard is indicated as exceedance probability in (e.g.) 50 years as a function of the peak ground acceleration, \( H(a_{g,\text{max}}) \). Hence, it is straightforward to replace the sum in equation 1 by an integral. \( Q(I_k) \) is then replaced by the probability density function of \( a_{g,\text{max}} \). This corresponds to the negative derivative of the exceedance probabilities provided by seismological services:

\[
h(a_{g,\text{max}}) = -\frac{dH(a_{g,\text{max}})}{da_{g,\text{max}}}
\]

The failure probability is then calculated by

\[
P_F = \int_0^{\infty} P(F|a_{g,\text{max}})h(a_{g,\text{max}})da_{g,\text{max}}.
\]

Obviously, it is referred to the same period of time as is the seismic hazard function \( H(a_{g,\text{max}}) \), typically to 50 years.

It is shown how much different earthquake intensities (corresponding to certain return periods) contribute to the overall failure probability. As application example nonlinear analysis of an arch dam is performed. Ground accelerations are created by the Kanai-Tajimi model and serve as stochastic excitation for the Monte Carlo simulations. The efficiency is increased by applying a new extrapolation technique, which is particularly suitable to estimate small failure probabilities more accurately without need for more simulation runs.

2 THE ARCH DAM

As application object for the reliability analysis a 220 m high arch dam is chosen. The earthquake safety of this (fictitious) structure has already been studied und documented in detail by Goldgruber in his PhD thesis [1]. The FE-model of the dam, see fig. [1] has kindly been made available to the authors and is used for this research. The focus of this work, however, is a Monte Carlo-based probabilistic analysis of the earthquake safety of structures in general, rather than a structural analysis of a certain dam. Hence, some changes are made on the dam described in [1]:
• Crack simulation by XFEM is removed
• The block joints are removed, hence the dam behaves as a monolithic block
• Tied contact is introduced between dam and the rock foundation
• Damping of the rock mass is increased by multiplying the Rayleigh damping factors $\alpha$ and $\beta$ by factor five.

The first three changes aim at reducing the computational effort for the Monte Carlo simulations and at guaranteeing convergence in every simulation run. The significant increase of the damping of the rock foundation is due to the necessity, in this research, to exactly excite the structure by earthquakes of certain predefined intensities. When ground motions are applied as accelerations on the model boundaries, they can be altered while travelling through the rock and possibly be intensified by reflections on the boundaries. As a result, the accelerations at the foot of the dam can significantly differ from the ones applied on the model boundaries. To avoid this, the mass density of the rock foundation is, just as in [1], chosen to be nearly zero, and, additionally in this research, a rather strong damping is implemented.

All other properties, including fluid-structure interaction with acoustic elements and nonlinear behaviour of the concrete, are equal as described in [1].

For the reliability analysis a failure criterion has to be defined. Critical zones and limiting values of quantities that lead to failure could be identified in a deterministic analysis, see [5]. Since critical values of stresses and displacements are strongly correlated with the crest displacements, the simplified assumption is made that failure (or a certain damage state) happens when the radial displacement of the crest midpoint exceeds 0.3 m. Naturally, also any other criterion, including sets of criteria, could be used.

3 SEISMIC HAZARD

It was sought to locate the dam in a region whose seismicity is well documented by a hazard curve. Hence, the city of Basle in Switzerland was chosen. Basle is the location of the largest historical earthquake on record in central Europe, with $M_w$ 6.6. The Swiss Seismological Service recently published updated hazard curves for this city in [2]. The hazard is expressed as
Christian Gasser and Christian Bucher

the probability of exceedance of $a_{g,max}$ in 50 years. In this study the hazard curve referred to as “stochastic model” in [2] is used.

The earthquake excitation is applied as acceleration boundary condition in all three directions. It is modelled as an amplitude-modulated random process

$$a(t) = e(t) \cdot b(t)$$

where $e(t)$ is the modulating function as given by

$$e(t) = 4 \cdot [exp(-0.25t) - exp(-0.5t)]$$

and $b(t)$ denotes a stationary zero-mean Gaussian random process with power spectral density according to Kanai-Tajimi:

$$S_{bb}(\omega) = S_0 \cdot \frac{4\zeta_g^2 \omega_g^2 \omega^2 + \omega_g^4}{(\omega_g^2 - \omega^2)^2 + 4\zeta_g^2 \omega_g^2 \omega^2}.$$  (6)

The numerical values are $\omega_g = 17$ rad/s and $\zeta_g = 0.3$. The spectral intensity $S_0$ is obtained by converting the values of $a_{g,max}$ from the hazard curve in [2]. Since there is a simple relationship between these two intensity quantities, namely $a_{g,max} = const \cdot \sqrt{S_0}$, this was done by repeatedly calculating the average $a_{g,max}$ from a thousand of trials and adjusting in that way the input $S_0$ to the desired $a_{g,max}$. In total, twelve intensity levels are analysed. They are listed in table 1 with the corresponding probabilities of exceedance and return periods.

<table>
<thead>
<tr>
<th>$a_{g,max}$ [g]</th>
<th>$S_0$ [m$^2$/s]</th>
<th>Prob. of exceedance</th>
<th>Return period [years]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.105</td>
<td>0.00133</td>
<td>0.04</td>
<td>1225</td>
</tr>
<tr>
<td>0.124</td>
<td>0.00185</td>
<td>0.03</td>
<td>1642</td>
</tr>
<tr>
<td>0.155</td>
<td>0.00289</td>
<td>0.02</td>
<td>2475</td>
</tr>
<tr>
<td>0.185</td>
<td>0.00412</td>
<td>0.015</td>
<td>3308</td>
</tr>
<tr>
<td>0.230</td>
<td>0.00636</td>
<td>0.01</td>
<td>4975</td>
</tr>
<tr>
<td>0.260</td>
<td>0.00813</td>
<td>0.008</td>
<td>6225</td>
</tr>
<tr>
<td>0.280</td>
<td>0.00943</td>
<td>0.007</td>
<td>7118</td>
</tr>
<tr>
<td>0.300</td>
<td>0.01083</td>
<td>0.006</td>
<td>8308</td>
</tr>
<tr>
<td>0.365</td>
<td>0.01603</td>
<td>0.004</td>
<td>12475</td>
</tr>
<tr>
<td>0.500</td>
<td>0.03007</td>
<td>0.002</td>
<td>24975</td>
</tr>
<tr>
<td>0.670</td>
<td>0.05400</td>
<td>0.001</td>
<td>49975</td>
</tr>
<tr>
<td>1.000</td>
<td>0.12030</td>
<td>0.00032</td>
<td>156225</td>
</tr>
</tbody>
</table>

Table 1: Intensity levels analysed

The time interval considered is 20 s, the time step $dt = 0.01$ s, giving a total of 2000 time steps. The excitation in vertical direction was reduced by factor 0.7 in time domain.

In general, besides the seismic load also the other loads as well as the resistances of the structure are random by nature. However, in case of the examined arch dam the variability of the structural behaviour will be primarily due to the inherent randomness of strong ground motions. As for the other two relevant loads acting on the dam, the self-weight and the hydrostatic load, there is virtually no uncertainty at all. There is some randomness in the material properties of the concrete and quite much uncertainty about the properties of the rock foundation, however, these parameters are regarded as deterministic here.
4 RESULTS

For each peak ground acceleration level listed in Table 1, 200 simulation runs are performed. The maximum crest displacements during the earthquakes are shown in Fig. 2 in increasing order. Besides the failure level of 0.3 m, also the displacement caused by the hydrostatic load (0.075 m) is plotted. The simulations with the three smallest earthquake intensities yield no event in which the failure level is exceeded. On the other hand, if ground motions with an average $a_{g,max}$ of 0.67 g are applied, the failure level is exceeded in each simulation run. Note that for that reason, simulations for the highest intensity level (1 g) could be avoided, since the failure level would be exceeded in each run anyway. Simulations with earthquake intensities in between lead to failure in a number of cases.

It is apparent that the variability of results based on identical ground motion intensities, but different time histories, is very high. Furthermore, it is interesting to note that the variability of the outcomes increases as the ground motion intensity increases. This is because stronger ground motions provoke more nonlinear effects.

Obviously, earthquakes with smaller $a_{g,max}$ rarely lead to failure. However, since smaller earthquakes occur more frequently, they might contribute substantially to the comprehensive failure probability according to equation (1). This means, in the present case, that 200 simulation runs might not be sufficient to estimate these small failure probabilities. Different methods exist to estimate small failure probabilities by Monte Carlo simulation at reasonable computational costs, e.g. Asymptotic Sampling [3] or Subset Simulation [4].

However, for this research, another, very simple approach is used. It is applied to the results of the crude Monte Carlo simulations for the four smallest intensities, i.e. for the simulations
which didn’t yield any failure event, resp. just one (i.e. for $a_{g,\text{max}} = 0.185\text{ g}$). For that purpose, the safety margin of each simulation run is calculated, which is simply the difference between the failure level and the maximum crest displacements. This is plotted versus its cumulative distribution function (CDF) expressed as the reliability index $\beta$, see fig. 3. The reliability index $\beta$ is related to the failure probability by the standardized Gaussian distribution function: $P_F = \phi(\beta)$. Apparently, the failure probabilities cannot be determined from the diagrams in fig. 3 directly, since no simulation run yielded a negative safety margin, i.e. resulted in failure (except one in diagram (d), which, thus, cannot be regarded as very representative). However, a new regression technique is applied, which approximates the CDF by a logarithmic function:

$$\beta = A \cdot \ln(B \cdot \text{safety margin} + C) + D$$  \hspace{1cm} (7)

The logarithmic approach guarantees monotonicity. The four coefficients $A, B, C$ and $D$ are obtained by an optimization algorithm. It is evident that the regression curve retraces the Monte Carlo results very well. The failure probabilities are then obtained by extrapolation to the value of $\beta$ for which the safety margin is zero.

It can be assumed that the failure probabilities can be estimated relatively accurately with this simple technique. Considerably inaccurate estimates could potentially result from a sudden change of the structural behaviour, such as change of boundary conditions, at large displacements which did not appear in the Monte Carlo simulation. However, there is no reason to assume such a behaviour for the analysed dam in case of excitation by moderate earthquakes.

The conditional failure probabilities $P(F|a_{g,\text{max}})$ obtained by extrapolation as well as the ones obtained from crude Monte Carlo simulation are shown in fig. 4. The dots mark the evaluation points. This object-related function is called fragility curve. The diagram also shows the location-related exceedance probability in 50 years, $H(a_{g,\text{max}})$, as given in [2], and its negative derivative, the probability density function, $h(a_{g,\text{max}})$. It can be seen that ground motions with $a_{g,\text{max}} = 0.2\text{ g}$ almost never lead to failure, while the failure limit is exceeded nearly every time if ground accelerations reach up to 0.5 g. Obviously, for the peak ground acceleration $a_{g,\text{max}}$, the probabilities are distributed in the opposite way.

The comprehensive failure probability $P_F$ is calculated by the convolution integral given in equation [3]. It is of high interest to which extend different sections of $a_{g,\text{max}}$ contribute to $P_F$. Hence, the product $P(F|a_{g,\text{max}}) \cdot h(a_{g,\text{max}})$ is shown in fig. 5. Apparently, the peak is at $a_{g,\text{max}} = 0.3\text{ g}$. This means that the ground motions associated with a return period of 8303 years contribute for the largest part of the long term failure probability. It is also interesting to note that the probability distribution in fig. 5 is very widebanded. This means that a Monte Carlo-based reliability analysis cannot reasonably be done for any single ground motion intensity, albeit, obviously, this probability distribution depends on the particular structure and location. Another outcome is that the contribution of the four smaller ground motion intensities, for which the extrapolation technique has been applied, is very small. However, this might be different in other applications.

For the evaluation of the integral, only peak ground accelerations up to 1 g are considered; higher values are assumed to be geophysically impossible. The calculated failure (resp. damage) probability for a period of 50 years is 0.00609, which is equivalent to one failure in 8186 years.

5 CONCLUSION

The fragility curve of an arch dam was determined by Monte Carlo simulation. It was shown that earthquakes with identical intensities but different time histories cause substantially differ-
Figure 3: CDFs of safety margins with regression functions

(a) $a_{g,max} = 0.105 \text{ g}$

(b) $a_{g,max} = 0.124 \text{ g}$

(c) $a_{g,max} = 0.155 \text{ g}$

(d) $a_{g,max} = 0.185 \text{ g}$
Figure 4: Fragility curve together with probability distribution and density of peak ground acceleration

Figure 5: Convolution of fragility curve and probability density of peak ground acceleration
ent structural responses. This applies especially for strong motions which lead to strong non-linear behaviour. Hence, it can be concluded that earthquakes are the major source of uncertainty for the performance of structures in seismically active regions.

It has been shown that the range of intensities which contribute significantly to the overall seismic endangerment is wide. This means that a probabilistic seismic safety assessment has to include the whole range of intensities, i.e. return periods. The conventional method to assess the seismic safety as a result of one single acceleration time history is a completely different approach. Obviously, it is not able to reflect the large variability of the response mentioned before.

A new extrapolation technique was introduced to efficiently estimate the small failure probabilities associated with earthquakes of moderate intensities. For the examined structure it has become clear that these small failure probabilities contribute little to the overall failure probability. However, it may be assumed that this is not the case for other structures and other hazard curves. Moreover, the extrapolation technique could be effectively used to estimate failure probabilities when the Monte Carlo simulations are based on smaller sample sizes.

ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Markus Goldgruber of Dynardo Vienna for making available the FE-model and for his professional assistance and Prof. Dimitar Kisliakov of UACEG Sofia for fruitful discussions.

The authors would like to acknowledge financial support from the Austrian Science Fund (FWF) as part of the Vienna Doctoral Programme on Water Resource Systems (DK-plus W1219-N22).

REFERENCES


VULNERABILITY OF HYDROPOWER INSTALLATIONS TO CLIMATE CHANGE: PRELIMINARY STUDY

Silvia Tolo¹, Edoardo Patelli¹, and Diyi Chen²

¹Institute for Risk and Uncertainty
University of Liverpool, UK
e-mail: {s.tolo, edoardo.patelli}@liverpool.ac.uk

²College of Water Resources and Architectural Engineering
Northwest A&F University, CH
e-mail: diyichen@nwsuaf.edu.cn

Keywords: Hydropower, Imprecise Probabilities, Bayesian Networks, Credal Networks, Climate Change

Abstract. The climate trends observed worldwide over the past few decades appear to corroborate the concerns of the scientific community about the many threats posed by global warming. Future changes of the current climate are expected to occur on different scales all around the globe, hence modifying the environmental background on the basis of which technological installations have been designed and operated. This can potentially threat the safety of the installations as well as their. The development of suitable tools aiming to predict the impact of climate change on technological installations is then essential in the wider context of climate change mitigation.

Hydropower installations play often a crucial role not only as a long-term renewable resource of energy but also for flood control and water supply in the case of droughts. All these aspects highlight the increasing importance of such installations as well as their growing vulnerability to natural hazards. It is hence essential to enlarge the current understanding of the interaction mechanisms between such installations and the changing surrounding environment in order to take adequate measures for climate change adaptation and ensure the future safety and productivity of hydropower production.

The current study aims to provide a novel model for the evaluation of the impact of climate change on the safety of hydropower stations. The approach adopted allows to include in the model the uncertainty inevitably associated with the input variables and to propagate such uncertainty within the analysis. The model proposed is finally applied to a realistic case-study in order to highlight its potential and limitations.

680
1 INTRODUCTION

Hydropower is currently the most common source of renewable energy, accounting for about 20% of global electricity production and for over 50% of the electricity generation of one third of all countries worldwide [8]. The large success of water resources for the electricity production, which has been often accompanied by the rapid growth of developing countries such as Brazil and China (today among the six biggest producers worldwide), has several reasons. First of all, it is a long-term renewable resource able to provide predictable and non-polluting energy. Secondly, hydropower systems such as dams and reservoirs can assist in flood control or, by the same token, provide water supply in the case of droughts [7]. Moreover, the conversion efficiency of hydroelectric generation can reach values over 90% (with peaks of 95% for large installations), greatly larger than fossil-fuelled production whose efficiency is generally around 30-40%. Finally the low operating and maintenance cost of hydropower and the long projected life span (up to 70 years) make this technology attractive and extremely cost-effective. On the other hand, these same strengths (e.g. the use of natural resources, the flood prevention purpose, the long operational life) make hydropower potentially sensitive to climate change, threatening the efficiency and safety of installations [4]. Indeed, the threat of climate change of worsening the occurrence of extreme weather events, both in terms of intensity and frequency, as well as the rise of temperatures are expected to have a significant impact on the water resources worldwide, affecting negatively the performance of hydroelectric systems as well as subjecting them to new and unquantified risks. Despite the importance of hydropower production and the further hazards introduced by climatic change, relatively few studies have addressed these issues and no methodology for the analysis of such impact has ever reached a general consensus [16].

The main aim of the current work is to implement a probabilistic framework for the evaluation of the impact of extreme weather events on hydroelectric systems, focusing on the scale of a single station. The climate variables are included in the analysis in light of future projections, allowing to carry the analysis with regards to future possible scenarios and trends. At this phase, the study is limited to the analysis of the impact of future precipitation trends, while other variables, such as temperature rise or occurrence of droughts are currently behind the purpose of the implementation.

In the following sections, the theoretical background of the methodology used is presented together with a brief overview of the computational tools adopted for the implementation. The proposed model, and its application to a simple case-study, are then described in Section 3.

2 Methodology and computational tools

The methodology selected for the implementation of such a model consists of the generalization of the so called Enhanced Bayesian Networks approach [17]. This latter is based on the integration of the well-known Bayesian Networks (BNs) with structural reliability methods (SRMs). Similarly, the methodology adopted integrates the use of Credal Networks (CNs), which can be regarded as a generalization of BNs able to include imprecise probabilities in the framework, with cutting-edge and robust SRMs. The choice of this particular methodology is justified by its large potential in the representation of the interaction between weather events and technological installations, as proved in former studies [19][20]. Indeed the approach allows to embody the aleatory character of natural events as well as the epistemic uncertainty associated (in particular in the case of climate projections), through the use of probabilistic models, intervals or imprecise random variables. Also, the uncertainty affecting the output (as well as discrete probabilities) is quantified and expressed in terms of probability bounds. These latter
are a crucial information for decision-makers or more generally analysts using the model: not only they express a measure of the meaningfulness of the results computed but also highlight the degree of ignorance which characterizes the estimation. Conversely, to exclude uncertainties from the analysis would lead to an unsound approach and unreliable results. Finally, a further advantage of the methodology is the use of external and existent numerical and physical models which can be easily integrated in the overall framework without affecting the congruence of the model. In the following sections, the basics of such methodology are briefly introduced.

2.1 Credal and Bayesian Networks

Credal Networks are a generalization of the well known Bayesian Networks. These latter present a graphical structure which consists of nodes and directed edges. Each node represents a variable of the model while the directed edges connecting the nodes capture the relationship (generally of a causal nature) existent between the variables. Commonly, the network structure is referred to employing a family metaphor: a node $X_2$ is said to be a *child* of a node $X_1$ if the edge connecting the two is originated in $X_2$ and points toward $X_1$; in this case, $X_2$ is also said to be a *parent* of $X_1$. Nodes without parents are said to be *roots* of the network, while nodes without children and not receiving any evidence are referred to as *barren nodes*.

These considerations regarding the structure of BNs are valid also for CNs, which present the same graphical framework. Indeed, the difference between CNs and BNs does not concern the graphical nature of the approach but its numerical aspect. The main aim of both BNs and CNs is to factorize the joint probability of complex events exploiting the information available regarding the single variables of the problem under study and their mutual relationships. Moreover, the most attractive aspect of this methodology is the capability of updating the belief regarding the event of interest when more information is available about other variables of the network. Hence, once selected the event of interest, its probability can be computed taking into account the knowledge available regarding the remaining variables of the model. Such knowledge is referred to as *evidence* and the overall process is known as *belief updating*. This task can be fulfilled through the use of inference algorithms, which can be of an analytical (i.e. exact inference algorithms) or approximate (i.e. approximate inference algorithms) nature. For both BNs and CNs, the input of such algorithms are, beyond the structure of the network, the conditional probability distributions (CPDs) that are associated with each node and quantify the strength of the relationships existent between the node itself and its parents (not having any parent, root nodes are associated with marginal probability distributions). In BNs such conditional probabilities are either crisp values (in this case the CPDs are referred to as CPTs, i.e. conditional probability tables) or continuous probabilistic distributions. In CNs the requirements regarding the nature of the input parameters are relaxed and embrace also imprecise probabilities and hence probability values known only with a certain degree of uncertainty (i.e. interval probabilities). This type of representation enhanced significantly the robustness and accuracy of the approach since allows to capture the uncertainty in input and to propagate it within the model avoiding the introduction of biases and assumptions not fully justified by the available data. Moreover, computing the uncertainty of the output, CNs provide a numerical context for the answer of the model which allows analysts to make decision acknowledging the significance and accuracy of the information obtained. On the other hand, the adoption of input parameters such as continuous distributions or imprecise probabilities increases the computational effort associated with the inference computation and can limit the availability of suitable algorithms. In many cases, in particular for CNs or more generally when continuous probabilistic distribution are involved, it is necessary to adopt approximate inference algorithms in order to deal with the complexity.
of the model at the cost of a lower (or even unknown) degree of accuracy.

2.2 CNs reduction with SRMs

Several numerical strategies and algorithms have been suggested in the scientific literature in order to overcome the limitations associated with the inference computation over CNs. The approach adopted in the current study aims to preserve the flexibility and accuracy of such methodology at the lowest computational cost. This is obtained integrating the CN framework with numerical methods imported from the field of structural reliability and involved in the reduction of the initial model. The main idea is to allow the initial network to embrace a large variety of mathematical frameworks for the representation of the data in input, such as crisp probability values, probability bounds, continuous probabilistic distributions, intervals and imprecise probabilistic distributions (i.e. continuous distributions whose parameters are known with some uncertainty and hence represented as intervals). Such initial model is then subject to a reduction procedure that results in an equivalent network containing only discrete nodes associated with either crisp or interval probability values and on which inference is finally computed. Hence, the reduction procedure consists of eliminating all the non-discrete nodes present in the initial model. The initial assumption at the basis of such procedure is that every node child of at least one non-discrete node must be defined as a domain in the outcome space of its continuous parents: when such condition is verified the elimination of the mentioned nodes can be obtained by integration. In other words, the elimination of a non-discrete node implies the reconstruction of its children’s CPDs in which the result of the integration is stored. This ensure the preservation of the information associated with the initial network that is not lost but simply reformulated and stored in the nodes of the final model.

The efficiency of the reduction procedure is obtained through the adoption of robust numerical methods commonly used in the field of structural reliability.

2.3 Computational Tool

The algorithms related to the methodology outlined in the former section have been implemented computationally in the general purpose software OpenCossan [15][6], obtaining a complete tool for the reduction and analysis of CNs. The toolbox provides the graphical and numerical implementation of the models as well as the reduction of EBNs to traditional BNs. The software allows four main types of nodes: discrete, probabilistic, bounded and hybrid. On the basis of the variables involved and the level of accuracy required, several options are available for the structural reliability methods to be adopted. In more detail, traditional and advanced Monte Carlo methods (such as Line Sampling, Subset Simulations) are suitable when no interval or hybrid nodes are involved in the computation. Conversely, the most generalized approach available consists of the adoption of Advanced Line Sampling which can compute all the possible combination of variables kinds in input [2]. Different options are also available for performing the inference computation on the reduced network. Such algorithms, both built-in or from third parties [12], are characterized by different degrees of accuracy and different computational costs. The toolbox integrates also methods for the sensitivity analysis of the models, allowing to identify the best possible strategy to effectively tune single network parameters in order to obtain the required level of imprecision in output.
3 Proposed Model

The model proposed refers to the vulnerability analysis of an impoundment facility and takes into account the structural safety of the dam as well as of the systems dedicated to the power generation. The network, shown in Fig. 1, has been applied to a simple case-study, whose implementation and analysis is described in the following section. The case-study refers to a facility located in the north-western region of China and considers a single turbine unit. The data adopted where partly collected from a real installation of compatible size and location (such as for the reservoir, dam and penstock structural parameters as well as for the water levels, in-flow, consumption, rainfall records etc). In the case of missing information (e.g. the unit transformer failure rate), realistic data have been adopted on the basis of the existing literature and represented through suitable mathematical frameworks able to account for the uncertainty of the information. The network provides the modelling of the reservoir in order to evaluate the probability of overtopping of the dam, as described in Section 3.2 under the different climate scenarios described in Section 3.1. The overtopping of the dam is assumed to potentially result in the load rejection, which can also be caused by the failure of the external power grid, on-site transformer malfunction, human action or mechanical accidents in turbine units [21]. The section of the network modelling these failure mechanisms (Section 3.3) considers also the eventual occurrence of water hammer in the penstock of the station due to load rejection.

3.1 Climate Scenarios

Two time scenarios, one related to the period 2021-2050 the other to 2071-2100, have been analyzed. Similarly, different assumptions have been considered regarding future trends for the emission of greenhouse gases: the first (SRES B1 according to the IPCC nomenclature [13]) considers the combination of a rapid economic growth with the introduction of clean and efficient technologies, able to ensure social and environmental stability; conversely, the second scenario (SRES A2) assumes a less ecologically-friendly and more divided world, resulting in a regionally oriented economic development and a significant increase of the greenhouse emis-
sions. The combination of the possible states of the nodes TimeScenario and EmissionScenario defines the overall scenario of reference, on the basis of which future trends for climate variables such as wind (WindTrend node) and precipitations (PrecipitationTrend node) are selected. These predictions are then combined with baseline probabilistic distributions for the relative variables obtaining the projected distribution of the wind (ProjectedWind) and rainfall (ProjectedRainfall) for all different scenarios. Due to the uncertainty unavoidably associated with future projections, the nodes ProjectedRainfall and ProjectedWindSpeed are described by imprecise random variables. Climate data projections for the north-western region of China have been deduced from the available literature [5] as well as the baseline for the wind speed distribution [1]. Conversely, the probability distribution of rainfall has been defined on the basis of the data collected close to the station of reference.

3.2 Dam Structural Safety

The overtopping of the dam occurs when the maximum level of the water in the reservoir overcomes the height of the dam. In order to compute the vulnerability of the dam to such mechanism, a structural reliability problem was defined [14]. According to the model adopted for such purpose, the overtopping mechanism can be triggered by the combination of several factors, such as the reservoir level (which can be influenced by precipitations) and the correct operation of the spillways. Moreover, the effect of waves originated in the reservoir by the wind is considered, assuming the wind to be uniform and constant over the fetch. Due to the lack of data the value of the fetch length has been assumed to be an interval (WindFetch node) whose bounds have been calculated on the basis of a simplified shape of the reservoir. Similarly, the dam slope has been assumed to be smooth (EmbankmentRoughness) and have an inclination (DamSlope) in the interval [0.78, 1.05] rad. On the contrary, the amount of water in-flow (ReservoirInFlow), the initial level of the reservoir (InitialLevel), the spillways flow (SpillwayFlow) and the water flow to the turbine (TurbineRelease) are described by probabilistic nodes designed on the basis of the real data available for the station of reference. The dam modelled has three spillways: the correct operation of these (SpillwayState) depends on the correct functioning of the gates associated to each spillways (Gate1, Gate2, Gate3). Each of these is assumed operational only if the associated control system (ControlSystem1, ControlSystem2, ControlSystem3) and lifting mechanism (LiftingMechanism1, LiftingMechanism2, LiftingMechanism3) are available. The failure rate for the control systems has been deduced from previous literature as for the lifting mechanisms, whose unavailability is considered in terms of failure of the related machinery [10]. Moreover, the operation of the control system is bound to the availability of electrical power produced on-site (OnSitePower) or available either from the external grid (ExternalPowerGrid) or backup diesels (BackupDiesel). The failure rates for the nodes ExternalPowerGrid and BackupDiesel have been deduced from the literature, adopting probability bounds in order to integrate multiple sources and possible set-up of the systems [3][11]. Similarly, the production of OnSitePower is associated with the availability of the on-site transformer and hence to the node UnitTransformer for which the probability bounds have been also deduced from former studies [18]. Clearly, the production of power on-site depends also on the availability of the turbine unit (TurbineOperationalState) but, as it is possible to see in Fig. [3] the two events have not been considered directly dependent. Indeed, the link between the nodes TurbineOperationalState and PowerSupplies would originate a cycle within the graph, contradicting the basic rules of BNs (and CNs) that rely on the use of acyclic directed graph. Graph theory offers possible solutions for this kind of issues, such as arc inversion, but in the current phase of the study these strategies have not been implemented.
3.3 Penstock Structural Reliability and Turbine Operation

Rapid load rejection, can cause the occurrence of pressure waves within the pipelines of the facility. This phenomenon, generally known as water hammer, threatens the structural safety of the station penstock and can trigger more critical accident scenarios jeopardizing the safety of the entire facility. In order to take into account the occurrence of the water hammer in the model, a stochastic approach was adopted on the basis of former studies [21]. Structural damages are assumed to occur when the combination loads acting on the penstock results higher than the yielding strength of the pipe material. In the case-study analyzed, this latter has been assumed to be 16 MN steel and to have a yielding coefficient normally distributed with mean equal to 390 MPa and standard deviation of 25.3 MPa. Moreover, the pipe is considered to be 216 m long and to have a diameter of 5 m. In order to take into account the effect of internal erosion occurring in the penstock, the thickness of the pipe has been assume uncertain, with values included between 12 mm and 16 mm. If the occurrence of water hammer results in the structural damage of the pipe, the facility is assumed to shut-down and hence the turbine would stop being operational (TurbineOperationalState). Apart from the occurrence of this kind of accident scenario, the turbine is assumed to be out of service also in the case of simple load rejection (LoadRejection) not resulting in the occurrence of water hammer or structural damages and in the case of maintenance (MaintenanceOutage).

3.4 Results

Fig. 3.4 shows the reduced CN obtained from the initial model (Fig. 3). Inference has been computed in the network, mainly aiming to quantify the probability of dam overtopping and of outage of the turbine unit for the different cases considered. In all of the scenarios analyzed
the probability of the occurrence of structural damages in the penstock remained largely neglectable. Also the probability of overtopping of the dam remains quite low for both the time scenarios considered, as shown in Table 1 while the probability of load rejection results quite significant. This value is coherent with the historical average frequency of load rejection calculated on the basis of eight hydroelectric stations in China [21]. The overall results can be interpreted as proof of the robustness of the facility to rapid transients: even if a sudden rejection of the load is required the structural safety of the penstock is guaranteed. Hence the design of the pipe results robust and the occurrence of pressure waves within the penstocks does not significantly threaten the overall safety of the facility. The probability of overtopping of the dam, even if it remains low in all the scenarios considered, appears to be strongly affected by uncertainty for the time period 2071-2100: this highlights the lower quality of the information available for this time scenario. The probability of overtopping was also quantified for the specific emission scenarios considered, as shown in Table 2: it results higher for the high emission scenario in the first part of the time domain, while the upper bound of the probability appears lower for the SRES A2 scenario in the time period 2071 – 2100. Nevertheless, due to the large uncertainty affecting the output, the difference between the two trends appears not significant enough to deduce a clear path for the risk of overtopping in the two emission scenarios: more investigation, including further climate factors, is needed. In order to better comprehend the behaviour of the system, the probability of overtopping has been computed also in the case of several what-if scenarios, such as the occurrence of load rejection, the malfunctioning of the spillways and the loss of electric power from the external grid. In all these cases, shown in Table 3, the probability of overtopping grows slightly with respect to the values shown in Table 1 demonstrating the robustness of the system with regards to these events.

Table 1: Inference computation results for the different time scenarios considered

<table>
<thead>
<tr>
<th>Event</th>
<th>2021-2050</th>
<th>2071-2100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dam Overtopping</td>
<td>$[1.84 \cdot 10^{-6}, 2.35 \cdot 10^{-4}]$</td>
<td>$&lt; 10^{-22}, 2.67 \cdot 10^{-4}$</td>
</tr>
<tr>
<td>Load Rejection</td>
<td>$[6.92 \cdot 10^{-1}, 6.98 \cdot 10^{-1}]$</td>
<td>$[6.92 \cdot 10^{-1}, 6.98 \cdot 10^{-1}]$</td>
</tr>
<tr>
<td>Turbine not Operational</td>
<td>$[6.98 \cdot 10^{-1}, 7.0 \cdot 10^{-1}]$</td>
<td>$[6.98 \cdot 10^{-1}, 7.0 \cdot 10^{-1}]$</td>
</tr>
</tbody>
</table>

Table 2: Inference computation results for the different time and emission scenarios considered

<table>
<thead>
<tr>
<th>Event</th>
<th>Medium Emission Scenario</th>
<th>High Emission Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dam Overtopping</td>
<td>$[1.29 \cdot 10^{-3}, 1.06 \cdot 10^{-4}]$</td>
<td>$&lt; 10^{-22}, 2.96 \cdot 10^{-4}$</td>
</tr>
</tbody>
</table>

Table 3: Probability of dam overtopping in the case of malfunction of specific subsystems

<table>
<thead>
<tr>
<th>What if...?</th>
<th>2021-2050</th>
<th>2071-2100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load Rejection</td>
<td>$[3.24 \cdot 10^{-3}, 2.34 \cdot 10^{-4}]$</td>
<td>$&lt; 10^{-22}, 3.33 \cdot 10^{-4}$</td>
</tr>
<tr>
<td>All Spillways out of Order</td>
<td>$[6.00 \cdot 10^{-5}, 2.20 \cdot 10^{-4}]$</td>
<td>$&lt; 10^{-22}, 3.40 \cdot 10^{-4}$</td>
</tr>
<tr>
<td>Power Grid Out of Order</td>
<td>$[2.15 \cdot 10^{-5}, 1.64 \cdot 10^{-4}]$</td>
<td>$&lt; 10^{-22}, 2.67 \cdot 10^{-4}$</td>
</tr>
</tbody>
</table>
4 Conclusions

A model for the vulnerability analysis of hydropower impoundment facilities was proposed. The suggested approach is based on the combination of Credal Networks with system reliability methods and takes into account the risk of overtopping of the dam, the structural safety of the penstock as well as the reliability of the power generation and connection systems. Furthermore, thanks to the flexibility of the methodology adopted, the model allows to estimate the future risks related to the facility in view of climate change projections.

To test the validity of the model this has been applied to a simple case-study based on an existent hydropower station located in the north-west region of China. The analysis was carried out for the time periods 2021 – 2050 and 2071 – 2100 and for two different emission scenarios (SRES B1 and A2). The results highlight the robustness of the facility to rapid transients (and hence to the risk of water hammer occurring in the penstock) on the one hand but result affected by strong uncertainty on the other. This can be attributed mainly to the high imprecision associated with the climate change projections.

Future research will focus on overcoming the current limitations of the model and on its further development in order to include other factors having the potential to affect the safety and productivity of the stations (e.g. temperature rise, droughts etc.). Indeed, the inclusion of further climate variables could result helpful in order to better track the evolution of the risk in time and hence in defining suitable precautionary measures able to effectively tackle future risks threatening both existing and new facilities.

5 Acknowledgement

This work has been supported by the the Newton Fund PhD Placement Programme and carried out with the collaboration of Northwest A&F University, YangLing, CH.

REFERENCES


STOCHASTIC MULTI-OBJECTIVE OPTIMISATION OF COMPOSITES MANUFACTURING PROCESS

K. I. Tifkitsis¹, A. A. Skordos¹

¹School of Aerospace, Transport and Manufacturing, Cranfield University
Bedford, MK430AL, UK
e-mail: {k.tifkitsis, a.a.skordos}@cranfield.ac.uk

Keywords: composite materials, uncertainty quantification, stochastic simulation, surrogate model, stochastic multi-objective optimisation

Abstract. This paper addresses the development of a stochastic multi-objective optimisation methodology and its implementation in the manufacturing of thick composite parts. Boundary conditions variability was quantified conducting a series of experiments and stochastic objects have been developed representing these uncertainties. The stochastic optimisation scheme takes into account the uncertainty of process parameters and boundary conditions and identifies optimal solutions that minimise process outcomes such as process duration and extent of defect formation and their uncertainty. The Kriging method was implemented to construct a computationally efficient surrogate model of manufacturing based on sample points selected by the Latin Hypercube Sampling (LHS) method and generated by a Finite Element (FE) model of the process. Response surfaces were constructed to test the accuracy of the surrogate model against the FE solution. A Genetic Algorithm (GA) was utilised to solve the multi-objective optimisation problem. The surrogate model was coupled with Monte Carlo (MC) and integrated into the stochastic multi-objective optimisation framework. The results show that a significant reduction in process duration and process induced defects variability in comparison with conventional processing conditions of up to 80% and 40% respectively can be achieved by the optimisation.
1 INTRODUCTION

Thermosetting matrix fibrous composites have been widely utilised in a variety of applications in industry due to their excellent mechanical properties combined with low weight. The cure process, which is the last stage of thermosetting composite fabrication, is a non-linear heat transfer phenomenon, during which the exothermic reaction transforms the matrix from an oligomeric liquid to a glassy solid. This spontaneous cross linking reaction results in the generation of a significant amount of heat which, in addition to the low thermal conductivity of the material in the through thickness direction, can lead to significant thermal gradients. Temperature overshoots can result in material degradation and excessive thermal stresses which can be critical in the case of high temperature operation [1]. Also, thermal gradients through the thickness lead to non-uniformity of degree of cure.

Conventional manufacturing processes reduce the probability of significant exothermic phenomena by using conservative thermal profiles. These usually involve long dwell at relatively lower temperatures, which naturally result in long process durations and high manufacturing costs. In this context, multi-objective optimisation of composites manufacturing offers an opportunity for the minimisation of potential cure process-induced defects and cost. The selection of optimal cure profiles in order to minimise cure time and cure process-induced defects has been investigated as a single-objective optimisation problem [2-4] whilst the multi-objective problem has been addressed recently [3] showing a trade-off behaviour with an L shape Pareto front formed by process designs that minimise temperature overshoot and process duration. A number of these solutions represent considerable improvements in both process time and temperature overshoot.

The manufacture of composite materials involves many parameters presenting considerable variability [4]. This variability induces uncertainty into the manufacturing process and affects its outcome. These uncertainties can also initiate process defects resulting in significantly amount of rejected parts associated with considerable cost. Stochastic simulation methodologies have recently started to be developed in order to address the uncertainties in composites manufacture and to investigate their influence in process outcomes such as cure time, geometrical distortion and temperature overshoot. The cure process includes different sources of variability in process parameters (cure kinetics) and boundary conditions (heat transfer coefficient, tool temperature). Cure kinetics parameters variability has been quantified experimentally using Differential Scanning Calorimetry (DSC) [5]. The initial degree of cure, activation energy and reaction order present significant variability introducing a coefficient of variation of approximately 30% in temperature overshoot and resulting in potential defects into the cured part. In terms of boundary conditions, experiments and stochastic models have shown carried out for the quantification of boundary conditions variability in an infusion process [6]. Stochastic models developed representing this variability integrated with an existing model of cure kinetics uncertainty to examine variability impact in the process outcomes. The surface heat transfer and tool temperature variability cause significant variability in cure time reaching a coefficient of variation of approximately 20% [6]. Tool temperature variability has the greatest influence on process outcome [7] whilst, higher levels of uncertainty increase the optimal cure time [8].

A stochastic multi-objective optimisation framework has been developed in the present study integrating a stochastic simulation method (Monte Carlo scheme) and a Genetic Algorithm (GA) in order to minimise the cure process time and temperature overshoot as well as their uncertainty. The manufacturing process boundary conditions (convection heat transfer coefficient, surface temperature) variability has been quantified experimentally. Stochastic objects representing the boundary conditions variability have been developed and incorporated into the stochastic simulation scheme. A surrogate model was developed using the Kriging method.
replacing the FE model and reducing significantly the computational effort. The methodology is applied to the case of a thick flat carbon fibre-epoxy laminate.

2 METHODOLOGY

2.1 Cure Simulation

A thermal cure simulation model was developed in the finite element solver MSC.Marc to simulate the curing stage of an infusion process in an oven. Figure 1 depicts a schematic representation of the model geometry and boundary conditions. The model represents a flat composite panel and comprises 26 3D iso-parametric eight-node composite brick elements of type 175 in MSC.Marc for thermal analysis [9]. Each element comprises two plies of Hexcel G1157 pseudo unidirectional carbon fibre reinforcement with 0.3 mm nominal thickness. Consequently, the overall thickness of the flat laminate is 15.6 mm. The matrix is Hexcel RTM6 epoxy resin. The boundary conditions illustrated in Figure 1 are implemented using user subroutines FORCDT and UFILM for time dependent prescribed temperature and forced air convection respectively [10]. A two dwell cure profile as shown in Figure 2 is applied on the lower surface of the composite part by selecting the following parameters; 1\textsuperscript{st} dwell temperature \( T_1 \), 2\textsuperscript{nd} dwell temperature \( T_2 \), 1\textsuperscript{st} and 2\textsuperscript{nd} dwell time \( d_1 \) and \( d_2 \) respectively, and ramp rate \( r \). Due to the symmetry across the width of the part, the heat transfer model was solved as a transient 1D heat transfer problem. The initial condition was considered to be 2\% degree of cure and uniform temperature after the end of filling. User subroutines UCURE, USPCHT, and ANKOND were utilised for the integration of material sub-models, cure reaction kinetics, specific heat capacity and thermal conductivity respectively [10].

The cure kinetics model is a combination of an \( n \)\textsuperscript{th} order model an autocatalytic model [11]. The cure reaction rate in is calculated as follows:

\[
\frac{da}{dt} = k_1 (1 - a)^{n_1} + k_2 (1 - a)^{n_2}a^m
\]  

(1)

where \( a \) is the current degree of cure, \( m, n_1, n_2 \) the reaction orders, \( k_1 \) and \( k_2 \) the reaction rate constants following an Arrhenius law:

\[
k_1 = A_1 e^{(-E_1/RT)}
\]

(2)

\[
k_2 = A_2 e^{(-E_2/RT)}
\]

(3)

where \( A_1, A_2 \) denote pre-exponential factors, \( E \) the activation energy and \( R \) the universal gas constant. Model constants values are reported in [12].

![Model schematic representation.](image-url)
The specific heat capacity of the resin and the fibre is calculated using experimental data obtained by modulated differential scanning calorimetry [2]. The specific heat capacity of the composite is computed making use of the rule of mixtures as follows:

\[ c_p = w_f c_{pf} + (1 - w_f) c_{pr} \]  \hspace{2cm} (4)

where \( w_f \) is the fibre weight fraction, \( c_{pf} \) the fibre specific heat capacity and \( c_{pr} \) the specific heat capacity of the resin. The thermal conductivity of the anisotropic composite material in the longitudinal direction is computed using an appropriate geometry-based model [13] and can be expressed as follows:

\[ K_{11} = v_f K_{lf} + (1 - v_f) K_r \]  \hspace{2cm} (5)

where \( v_f \) is the fibre volume fraction, \( K_{lf} \) and \( K_r \) are the thermal conductivity of the fibre in the longitudinal direction and of the resin, respectively. In the transverse direction the thermal conductivity is calculated as follows:

\[ K_{22} = K_{33} = v_f K_r \left( \frac{K_{lf}}{K_r} - 1 \right) + K_r \left( \frac{1}{2} - \frac{K_{lf}}{2 K_r} \right) + K_r \left( \frac{K_{lf}}{K_r} 1 \right) \sqrt{v_f^2 - v_f + \left( \frac{K_{lf}}{K_r} + 1 \right)^2} \left( 2 \frac{K_{lf}}{K_r} - 2 \right)^2 \]  \hspace{2cm} (6)

where \( K_{lf} \) is the thermal conductivity of the fibre in the transverse direction.

2.2 Surrogate model

The utilisation of FE analysis for the simulation of cure requires high computational time. In terms of stochastic simulation and multi-objective optimisation problems, the use of FE model becomes computationally highly demanding. A surrogate model was constructed using the Kriging method to address this high computational effort replacing the FE-model. Figure 3 shows the procedure of surrogate model development. The Kriging method requires a set of input points and their responses generated using FE analysis. Latin Hypercube Sampling (LHS) [14], a random sample generation method, was chosen for generating a sample of 15000 input values and their responses. In this study, the 1st and 2nd dwell temperature, 1st dwell time, heat
transfer coefficient, reaction order and activation energy have been considered as input parameters. Table 1 presents the input ranges of the initial design space. The outputs of the surrogate model are the cure time and temperature overshoot. Cure time is considered as the time at which the minimum degree of cure of the part is greater than 0.88, which is the degree of cure at which the epoxy resin (RTM6) reaches during an isothermal cure at 180°C, whilst temperature overshoot is the maximum difference between tool temperature and local temperature of the laminate during curing.

Kriging allows the estimation of untried parameters values to be made without bias and with minimum variance and more accurately in comparison with low order polynomial regression models [15].

Given a set of \( m \) design sites

\[
S = [s_1 \ s_2 \ \cdots \ s_m]^T \quad \text{with} \quad s_i \in \mathbb{R}^n
\]

and responses

\[
Y = [y_1 \ y_2 \ \cdots \ y_m]^T \quad \text{with} \quad y_l \in \mathbb{R}^q
\]

The data is assumed to satisfy the normalisation conditions

\[
\mu[S_{:,j}] = 0, \quad V[S_{:,j}, S_{:,j}] = 1, j = 1, \ldots, n
\]

\[
\mu[Y_{:,j}] = 0, \quad V[Y_{:,j}, Y_{:,j}] = 1, j = 1, \ldots, q
\]

where \( \mu[\cdot] \) and \( V[\cdot, \cdot] \) denote the mean and the covariance respectively.

The Kriging model treats the deterministic response vector \( y(x) \in \mathbb{R}^q \), for a \( n \) dimensional input \( x \in D \subseteq \mathbb{R}^n \) as a realisation of a regression model \( \mathcal{F} \) and a random field,

\[
\hat{y}_l(x) = \mathcal{F}(\beta_{:,l}, x) + z_l(x), \quad l = 1, \ldots, q
\]

The regression model \( \mathcal{F} \) is a linear combination of \( p \) chosen functions \( f_j(x) : \mathbb{R}^n \mapsto \mathbb{R} \),

\[
\mathcal{F}(\beta_{:,l}, x) = \beta_{1,l}f_1(x) + \cdots + \beta_{p,l}f_p(x) = [f_1(x) \cdots f_p(x)]\beta_{:,l}
\]

\[
\equiv f(x)^T \beta_{:,l}
\]

where the coefficients \( \{\beta_{p,l}\} \) are regression parameters.

The random field \( z \) is assumed to have mean zero and covariance

\[
E[z_l(w)z_l(x)] = \sigma_l^2R(\theta, w, x), \quad l = 1, \ldots, q
\]

where \( \sigma_l^2 \) is the field variance for the \( l \)th component of the response and \( R(\theta, w, x) \) is the correlation surface with parameter vector \( \theta \).

\[
\Delta \tau_{\max} = f(T_1, T_2, dt_1, h, E_2, m) \quad \tau_{\text{cure}} = f(T_1, T_2, dt_1, h, E_2, m)
\]
Table 1 Range of surrogate model input parameters.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1[°C]$</td>
<td>135-175</td>
</tr>
<tr>
<td>$T_2[°C]$</td>
<td>175-215</td>
</tr>
<tr>
<td>$dt_1[\text{min}]$</td>
<td>42-300</td>
</tr>
<tr>
<td>$h[\text{W/m}^2/°C]$</td>
<td>13.8-21.8</td>
</tr>
<tr>
<td>$E_2[J/\text{mol}]$</td>
<td>56020-59620</td>
</tr>
<tr>
<td>$m$</td>
<td>1.008-1.572</td>
</tr>
</tbody>
</table>

For the set $S$ of design sites, an $m \times p$ design matrix $F$ can be constructed with $F_{ij} = f_j(s_i)$,

$$F = [f(s_1) \cdots f(s_m)]^T$$

(14)

The $m \times p$ correlation matrix $R$ can be constructed as

$$R_{ij} = \mathcal{R}(\theta, s_i, s_j), \quad i, j = 1, \ldots, m$$

(15)

The fitted regression parameter $\beta^*$, a $p \times q$ matrix, can be calculated considering matrices $F$ and $R$, using least squares as follows:

$$\beta^* = (F^T R^{-1} F)^{-1} F^T R^{-1} Y$$

(16)

For any untried design point $x$, the vector $r(x)$ of correlations between different $z$ at design sites and $x$, can be defined as

$$r(x) = [\mathcal{R}(\theta, s_1, x) \cdots \mathcal{R}(\theta, s_m, x)]^T$$

(17)

Therefore, the Kriging predictor is

$$\hat{y}(x) = f(x)^T \beta^* + r(x)^T \gamma^*$$

(18)

where the $m \times q$ matrix $\gamma^*$ can be calculated through the residuals,

$$\gamma^* = Y - F \beta^*$$

(19)

Matrices $\beta^*$ and $\gamma^*$ are fixed for a fixed set of design data. Only vectors $f(x) \in \mathbb{R}^p$ and $r(x) \in \mathbb{R}^m$ have to be computed for every new $x$.

Correlation models can be divided into two groups, those containing functions that have a parabolic behaviour near the origin (Gaussian, Cubic and Spline), and those including functions with a linear behaviour near the origin (exponential, linear and spherical). The curing stage is a continuously differentiable phenomenon resulting in a parabolic behaviour of correlation function close to origin. Therefore, a Gaussian function and a 2nd order polynomial were selected for the correlation and the regression model respectively. Coefficients $\beta^*$ and $\gamma^*$ of Eq. (19), the 2nd order regression and the Gaussian correlation function respectively, were calculated using the Matlab toolbox for Kriging modelling [16]. The surrogate model was implemented in C++.

2.3 Experimental set-up for the quantification of boundary conditions uncertainty

The uncertainty of boundary conditions was investigated by carrying out a series of experiments using an infusion set-up placed inside in an oven. These experiments aim to quantify the variability of the air temperature in the oven and of the surface heat transfer coefficient between the vacuum bag and the air.

Ten experiments were conducted, utilising the set-up depicted in Figures 4 and 5. It comprises a CALTHERM E9321V2 oven with an EUROTERM 2408P4 PID controller, a 10 mm aluminium plate, a nylon N64PS-x VAC INNOVATION peel ply fabric, a nylon xR1.2 VAC INNOVATION vacuum bag, two K-type thermocouples and two RdF micro-foil heat flux sensors [17].
A carbon fibre-epoxy flat panel fabricated by infusion process was utilised to produce thermal conditions similar to those during the cure of a part. The resin system of the panel was Hexcel RTM6, whilst the preform was Hexcel AS7 6k carbon fibre [18] with an areal density of 280 g/m². The composite part was placed on the tooling plate, covered with the peel ply and the vacuum bag and sealed before experimental runs.

The flux sensors were placed on the vacuum bag to measure the forced convection variability as well as its spatial dependence. A K-type thermocouple was mounted on the bag to measure the surface temperature, whilst a second sensor was placed close to surface but outside the thermal boundary layer to monitor air temperature. The temperature was set-up at 160 °C in all tests. A National Instruments LabVIEW in house code was utilised for data acquisition and the data were acquired with a frequency of 0.8 Hz.

The micro-foil flux sensor consists of a thin layer, and is a differential thermocouple sensor using T-type thermocouples [17]. The sensor measures the temperature on both sides of the thin layer, which are used to calculate the heat gain or loss through the thin layer. The same heat flux should flow through the sensor and the surface where the sensor is placed. The sensor output is a voltage signal which is proportional to heat flux. Specifically, the heat flux $\dot{Q}$ is calculated by the following relation:

$$\dot{Q} = (H/(C \cdot TF))$$

where $H$ is the sensor output, whilst $C$ and $TF$ are a calibration multiplier and a temperature multiplication factor respectively, and are provided by the manufacturer. The sensors used in these experiments have a calibration multiplier of 0.15 μV/W/m². The temperature multiplication factor is a function of temperature and can be found in [17]. The heat transfer coefficient $h$ was calculated using the temperatures of the surface $T_s$ and air in the oven $T_{air}$ as follows:

$$h = \frac{\dot{Q}}{T_s - T_{air}}$$

Figure 4 Schematic representation of experimental set-up.

Figure 5 a) CALTHERM E9321V2 oven with a EUROTHERM 2408P4 PID controller and data acquisition system b) infusion set up with the sensors.
2.4 Stochastic simulation

The stochastic simulation was carried out using a Monte Carlo (MC) scheme. This method involves generation of random input stochastic variables using a random number generator. The deterministic model is executed a number of times for each set of inputs and the process outcomes are analysed in term of their probability distributions. The simulation requires a certain number of iterations to ensure the convergence in mean and standard deviation.

2.5 Stochastic multi-objective optimisation

The aim of stochastic optimisation is to minimise cure time and temperature overshoot and their variability by optimising the cure profile parameters. An interface has been developed incorporating MC simulation into a Genetic algorithm (GA) for multi-objective optimisation. A two dwell cure profile depicted in Figure 2 has been considered and parametrised using three parameters; temperature of first and second dwell and the duration of the first dwell. The ranges of potential values for each parameter used in optimisation are summarised in Table 2. It should be noted that the ramp rate $r$ and second dwell time $dt_2$ are not considered as design parameters in order to reduce the dimensionality of the problem and are equal to 2 °C/min and $dt_1$ respectively. The minimisation objectives of the optimisation are the mean values and standard deviations of $t_{\text{cure}}$ and $\Delta T_{\text{max}}$. Figure 6 summarises the steps the stochastic optimisation methodology. The GA begins by generating the first population of individuals. A Monte Carlo simulation is carried out using the surrogate model for each individual calculating the mean value and the standard deviation of cure time and maximum temperature overshoot. A new population is generated by a number of operations (crossover, mutation) based on the previous population individuals. The GA finishes when a convergence criterion is satisfied and the output is a matrix of the optimal points called Pareto front.

### Table 2 Design parameters ranges

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1[\degree\text{C}]$</td>
<td>135-175</td>
</tr>
<tr>
<td>$T_2[\degree\text{C}]$</td>
<td>175-215</td>
</tr>
<tr>
<td>$dt_1[\text{min}]$</td>
<td>42-300</td>
</tr>
</tbody>
</table>

**Figure 6 Stochastic multi-objective optimisation framework.**
3 RESULTS

3.1 Validation of surrogate model

Figure 7a shows the dependence of $t_{\text{cure}}$ on $T_1$ and $dt_1$, for constant $T_2, h, m$ and $E_2$ equal to 195 °C, 17.82 W/m$^2$/°C, 1.29 and 57820 J/mol respectively. The response of the surrogate model is in good agreement with the response of the FE model, with an average absolute difference of 0.56 min. The response surface shows a trade-off between $dt_1$ and $t_{\text{cure}}$ especially in the region of low $T_1$. Figure 7b illustrates the response surface of temperature overshoot as a function of the first and second dwell temperature using the surrogate and FE model. In this case the dwell time is equal to 83 min and the remaining input parameters $h$, $m$ and $E_2$ as reported in the previous case. The average absolute difference is 0.69 °C, showing the high accuracy of the surrogate model. The overshoot increases as a function of $T_1$. A global minimum of temperature overshoot is reached for a $T_1$ in the 140-145 °C range. In this case the first dwell advances the chemical reaction as much as possible in order to minimise temperature overshoots in the second dwell. The overshoot increases slightly for first dwell temperatures below 140 °C as the temperature is too low to advance the resin reaction during the first dwell sufficiently to suppress generation of higher exothermic effects during the second dwell.

The use of the surrogate model addresses problems related to computational time, specifically in the case of stochastic simulations where a large number of iterations is required. A typical two-dwell cure profile was utilised as input for Monte Carlo where the number of model runs was equal to 1000. Figures 8a and 8b show the cumulative density functions (CDFs) of cure time and temperature overshoot respectively, calculated both by the FE and surrogate models. It can be observed that the two CDFs are in a good agreement.

![Figure 7 Response surfaces](image-url)
3.2 Experimental results

Figures 9a and 9b summarise the experimental data of heat transfer coefficient and surface temperature evolution over time respectively for ten different runs. Both $T_s$ and $h$ present significant variability across different runs. The short term variability can be attributed to the motion of the air streams inside the oven due to the fan rather than to signal noise effects. The surface temperature presents a periodic term over time caused by the temperature controller of the oven and short term variability due to random variations. The results of surface heat transfer coefficient show only short term variability and a variable level across the runs. These variations can be attributed to the fact that the air movement is forced by the fan in the oven.
3.3 Stochastic objects development

Fast Fourier Transformation (FFT) implemented in MATLAB was utilised to calculate the frequency of $T_s$. Then, a cosinusoidal fit as shown in Figure 10 was performed making use of generalised reduced gradient non-linear optimisation method implemented in MS Excel [19] calculating the amplitude of the periodic trend. Consequently, detrending was applied in order to generate a stationary process of surface temperature. The Ornstein-Uhlenbeck process (OU), which is a mean reverting second order stationary Gaussian process, was utilised to represent the surface temperature after detrending. The stochastic differential equation of OU process can be written as follows [20]:

$$dS = \lambda (\mu - S) dt + \sigma dW_t$$  \hspace{1cm} (22)

where $S$ is the OU process, $W_t$ a Brownian motion following a normal distribution with mean 0 and standard deviation 1 so that $W_t \sim N(0,1)$, whilst $dW_t$ follows a normal distribution with average 0 and standard deviation $\sqrt{dt}$ so that $W_t \sim N(0, \sqrt{dt})$. $\lambda$ controls the speed of reversion to the long term average of the process, $\sigma$ is the process volatility and $\mu$ is the long term mean of the stochastic process. The analytical solution of Eq. (22) has been used in this study to develop the stochastic object of surface temperature and can be written as follows [20]:

$$S_t = e^{-\lambda \Delta t} S_{t-1} + (1 - e^{-\lambda \Delta t}) \mu + \sigma \sqrt{\frac{1 - e^{-2\lambda \Delta t}}{2\lambda}} W_t$$  \hspace{1cm} (23)

where $\Delta t$ is the time increment.

This procedure was repeated for each experimental run. The stochastic process of surface temperature can be expressed as follows:

$$T_s = A_s + B_s \cos \omega_s t + S_s$$  \hspace{1cm} (24)

where $A_s$ is the level of each experimental curve, $B_s$ and $\omega_s$ the amplitude and frequency of the cosinusoidal fit respectively, and $S_s$ the mean reverting stationary stochastic process (OU) calculated by Eq. (23). The surface heat transfer coefficient does not involve a periodic trend and can be modelled as realisation of a random scalar variable as follows:

$$h = A_h + B_h y$$  \hspace{1cm} (25)

where $A_h$ is the level, $B_h$ the volatility of the process for each run, and $y$ is a standard normal variable. The statistical properties as calculated from the different experimental run are presented in Table 3.

![Cosinusoidal fit of surface temperature](image)
Table 3 Statistical properties of boundary conditions.

<table>
<thead>
<tr>
<th>Boundary conditions</th>
<th>Stochastic model parameter</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h$ ([\text{W/m}^2/\text{°C}])</td>
<td>$A_h$</td>
<td>17.82</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>$B_h$</td>
<td>2.1</td>
<td>0.1</td>
</tr>
<tr>
<td>$T_s$ [°C]</td>
<td>$A_s$</td>
<td>152</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td>$B_s$</td>
<td>0.82</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>$\omega_s$</td>
<td>0.02</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>$\lambda_s$</td>
<td>0.97</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>$\sigma_s$</td>
<td>0.4</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>$\mu_s$</td>
<td>0.007</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Figure 11 Surface temperature variability over time.

The FE model, described in section 2.1 was used with a time increment of 3 sec to represent short term phenomena. The stochastic simulation was compared with the deterministic model to evaluate the significance of short term variability. Figure 11 shows the temperature evolution with time at the bottom and the middle of the composite part for both stochastic and deterministic model. The temperature at the bottom of the part presents a short term variability with negligible influence of variability.

A number of stochastic simulation iterations have been carried out to calculate the average and standard deviation of cure time and temperature overshoot over time. It was found that the mean value of cure time and temperature overshoot are 94.9 min and 34.6 °C with standard deviations 0.07 min and 0.3 °C respectively. The cure time and temperature overshoot of the deterministic model were 95 min and 34 °C, indicating the negligible influence of short term variability of surface temperature. Similarly, heat transfer coefficient short term variability does not affect the cure time and the temperature overshoot.

According to these results, the stochastic simulation can be carried out considering the variability of the level of surface temperature and surface heat transfer coefficient. Therefore Eq. (24) and (25) can be expressed as follows:

$$T_s = A_s$$  \(26\)

$$h = A_h$$  \(27\)

Furthermore, the variability of cure kinetics is controlled by the reaction order $m$, and activation energy $E_2$ (Eqs. (1),(3)) [12]. Tables 4 and 5 summarise the statistical properties and the correlation matrix of $m$ and $E_2$ respectively.
3.4 Stochastic simulation

A Monte Carlo simulation has been carried out for the standard two dwell cure profile [21] in order to investigate the variability of cure time and temperature overshoot. The nominal two dwell cure profile comprises one dwell temperature at 160 °C with duration 75 min and a second dwell temperature at 180 °C for 75 min, whilst the ramp rate is equal to 1 °C/min. The surface temperature $T_1$ and $T_2$, surface heat transfer coefficient $h$, activation energy $E_2$ and reaction order $m$ have been considered as stochastic objects. Figure 12a illustrates the results of Monte Carlo for cure time. The mean time converges after 100 iterations and is 111 min, considerably faster than standard deviation which converges after about 500 iterations and is equal to 7.5 min. The results of temperature overshoot are illustrated in Figure 12b. The mean value converges after 100 iterations to 20.5 °C, whilst the standard deviation converges after 400 runs to 3.4 °C. Cure time and temperature overshoot of the standard cure profile present coefficients of variation of 6% and 16.5% respectively, necessitating a thorough investigation of a solution not only with the minimum value but with minimum uncertainty.

![Figure 12 Monte Carlo simulation results](image-url)
3.5 Stochastic multi-objective optimisation

A deterministic two-objective optimisation was carried out using the nominal values of input parameters. Figures 13a and 13b illustrate box plots of each of stochastic Pareto points of $t_{\text{cure}}$ and $\Delta T_{\text{max}}$ respectively. In addition, the deterministic Pareto front is illustrated with a solid grey line. Pareto fronts of both stochastic and deterministic optimisation are in the form of an L-shape curve dividing the objective space into two different regions. A horizontal region in which cure time can be reduced significantly without considerable changes in exothermic effects and a vertical region in which low temperature overshoot can be reduced significantly with small changes in cure time. The main difference between the stochastic and deterministic Pareto fronts is that the former includes points in which the mean values are dominated by other optimal points, which they dominate then in terms of variability, whilst in deterministic the domination ranking occurs only in terms of nominal values. Also, the mean values of the stochastic Pareto front present higher cure times than the points of the deterministic Pareto front in the case of conservative cure profiles.

The comparison between an optimal point and a standard two dwell profile, as reported in Table 6, illustrates the improvements of stochastic optimisation both in minimisation of mean values and variability. Specifically, the optimal cure profile results in reduction of 13% and 51% in cure time and temperature overshoot in comparison with the standard cure profile, whilst standard deviations are reduced by 66% and 40% respectively.

Cure time shows an overall lower variability than temperature overshoot with the means of coefficient of variation are equal to 2% and 21% for cure time and temperature overshoot respectively. The stochastic Pareto front includes points with cure time values twice as high as for deterministic points. These points are included in the stochastic Pareto front because they present the highest stability in potential fluctuations of process parameters and boundary conditions. These individuals are generated using conservative cure profiles with low first dwell temperature and long first dwell time, therefore the thermal gradients are negligible and the duration of cure long. Moreover, the tip of vertical region of stochastic Pareto front includes one point in which cure time and temperature present high uncertainty with standard deviations of 7 min and 3.6 °C respectively.

The variability of temperature overshoot of points in the vertical region may result in shifting of the stochastic Pareto front in higher or lower values of temperature overshoot than in the deterministic case. In the horizontal region stochastic optimal points present higher temperature overshoots than deterministic points in the range of 3-4 °C. The results obtained with the stochastic optimisation methodology highlight the significant improvements in terms of minimisation both of mean value and variability of cure time and temperature overshoot in comparison with the results of standard cure profiles.

| Table 6 Parameters values for optimal points and a standard cure profiles. |
|-----------------|--------|--------|
| Parameter | Optimal point | Standard two-dwell |
| $T_1$[°C] | 145 | 160 |
| $T_2$[°C] | 211 | 180 |
| $dt_1$[min] | 65 | 75 |
4 CONCLUSIONS

The findings of this work highlight the efficient opportunities offered by the stochastic optimisation in terms of eliminating cure time and temperature overshoot uncertainty. The utilisation of stochastic multi-objective optimisation may lead to significant improvements of the composites manufacturing process accomplishing more stable solutions than convention profiles. In addition, the implementation of the stochastic optimisation in the other stages of manufacturing will result in reduction of process duration and thus in minimisation of the cost.

5 ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council, through the grant ‘Robustness performance optimisation for automated composites manufacture’ (EP/K031430/1).

REFERENCES


