



BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF CIVIL ENGINEERING

FAKULTA STAVEBNÍ

INSTITUTE OF STRUCTURAL MECHANICS

ÚSTAV STAVEBNÍ MECHANIKY

**UNCERTAINTY QUANTIFICATION IN CIVIL
ENGINEERING**

KVANTIFIKACE NEJISTOT VE STAVEBNÍM INŽENÝRSTVÍ

HABILITATION THESIS

HABILITAČNÍ PRÁCE

AUTHOR

AUTOR PRÁCE

Ing. LUKÁŠ NOVÁK, Ph.D.

BRNO 2024

ABSTRACT

The presented habilitation thesis summarizes significant recent results from uncertainty quantification in civil engineering. The presented results were obtained within the framework of various research projects investigated at the Faculty of Civil Engineering of Brno University of Technology. The habilitation consists of seven scientific articles divided into two main research topics – semi-probabilistic methods, and approximations in the form of polynomial chaos expansions. The presented articles significantly extended the possibilities of the state-of-the-art methods in both research topics and represent a coherent collection of methods for computationally efficient uncertainty quantification of costly mathematical models.

KEYWORDS

Uncertainty quantification, semi-probabilistic approach, surrogate model, polynomial chaos expansion

ABSTRAKT

Předložená habilitační práce shrnuje nedávné významné výsledky z oblasti kvantifikace nejistot ve stavebním inženýrství. Prezentované výsledky byly získány v průběhu řešení řady výzkumných projektů na Fakultě stavební Vysokého učení technického v Brně. Habilitační práce je tvořena sedmi vědeckými články rozdělenými do dvou stěžejních výzkumných témat – polo-pravděpodobnostní metody návrhu a aproximační metody ve formě polynomiálního chaosu. Prezentované publikace významně rozšiřují možnosti metod aktuálního stavu poznání v obou tématických blocích a představují ucelenou kolekci metod umožňující výpočetně efektivní kvantifikaci nejistot výpočetně náročných matematických modelů.

KLÍČOVÁ SLOVA

Kvantifikace nejistot, polo-pravděpodobnostní přístup, náhradní plocha odezvy, rozvoj polynomiálního chaosu

NOVÁK, Lukáš. *Uncertainty Quantification in Civil Engineering*. Brno, 2024, 157 p. Habilitation thesis. Brno University of Technology, Faculty of Civil Engineering, Institute of Structural Mechanics.

Copyright © 2024 Lukáš Novák. All rights reserved.

Institute of Structural Mechanics
Faculty of Civil Engineering
Brno University of Technology

Typeset by L^AT_EX

DECLARATION

I declare that I have written the Habilitation thesis titled “Uncertainty Quantification in Civil Engineering” independently, and using exclusively the technical references and other sources of information cited in the thesis and listed in the comprehensive bibliography at the end of the thesis.

As the author I furthermore declare that, with respect to the creation of this Habilitation thesis, I have not infringed any copyright or violated anyone’s personal and/or ownership rights. In this context, I am fully aware of the consequences of breaking Regulation § 11 of the Copyright Act No. 121/2000 Coll. of the Czech Republic, as amended, and of any breach of rights related to intellectual property or introduced within amendments to relevant Acts such as the Intellectual Property Act or the Criminal Code, Act No. 40/2009 Coll., Section 2, Head VI, Part 4.

Brno

.....

author’s signature

ACKNOWLEDGEMENT

First of all, I am very grateful to all my colleagues from the Institute of Structural Mechanics. I especially appreciate our enthusiastic discussions with prof. Drahomír Novák, prof. Miroslav Vořechovský, prof. David Lehký, prof. Jan Eliáš and Dr. Petr Miarka about various aspects of structural mechanics, uncertainty quantification and structural reliability. My heartfelt gratitude goes to my family. I sincerely thank my wife Michaela, my son Adam, and my parents.

The presented work is a collection of outcomes of several research projects. The financial support was provided by the *Czech Science Foundation (GA)* and the *Ministry of Education, Youth and Sports of the Czech Republic* under project numbers:

- GA20-01734S Probability oriented global sensitivity measures of structural reliability
- GA20-01781S Uncertainty modelling in safety formats of concrete structures
- GA22-00774S Metamodel-assisted probabilistic assessment in bridge structural engineering
- GA23-04712S Importance of Stochastic Interactions in Computational Models of Structural Mechanics
- LTAUSA19058 Development of theory and advanced algorithms for uncertainty analyses in Engineering problems

Additional financial support was provided by internal BUT mobility projects in the framework of International Mobility of Researchers of Brno University of Technology, Czechia, under project No. EF18_053/0016962, and stipend and partnership program of Brno University of Technology for Excellence 2023.

Brno

.....

author's signature

Contents

1	Introduction	7
2	Uncertainty Quantification and Semi-probabilistic Approach	10
2.1	Estimation of Coefficient of Variation for Structural Analysis: The Correlation Interval Approach	15
2.2	Comparison of Advanced Semi-probabilistic Methods for Design and Assessment of Concrete Structures	27
3	Polynomial Chaos Expansion in Structural Mechanics	45
3.1	Variance-based Adaptive Sequential Sampling for Polynomial Chaos Expansion	53
3.2	Active Learning-based Domain Adaptive Localized Polynomial Chaos Expansion	79
3.3	On Distribution-based Global Sensitivity Analysis by Polynomial Chaos Expansion	102
3.4	Physics-Informed Polynomial Chaos Expansions	117
3.5	UQpy v4.1: Uncertainty Quantification with Python	140
4	Discussion & Concluding Remarks	148
	Bibliography	152

1 Introduction

All models are wrong, but some are useful.

— George Box

Increasing economical and safety requirements placed on engineering in today's society inevitably increase the demands for accuracy and efficiency of mathematical models representing physical systems. As a result of these needs and availability of a higher computational power than ever, there is a rapid development of advanced numerical methods designed for the analysis of complex engineering structures. Such methods are typically employed for the analysis of structures constructed from non-traditional materials or complicated systems affected by multi-physics phenomena. Moreover, there are a lot of structures, especially bridges, built in the last century, which must often be enhanced for higher loads taking actual conditions of the structures into account.

Although the advanced numerical methods, such as the non-linear finite element method (NLFEM), are more often employed in industry, their routine applications are still limited by their extensive computational costs. The computational burden is further accented in the presence of various uncertainties associated with an investigated physical system as well as with its mathematical model. Uncertainties arise from the lack of knowledge of material characteristics (e.g. fracture energy), actual geometrical properties (e.g. position of reinforcement), and even mathematical models of some physical phenomena (e.g. fracture mechanics of quasi-brittle materials). This lack of knowledge may generally lead to inaccurate results and even fatal failures despite the advanced numerical analysis. The impact of uncertainties is especially significant in the safety of systems containing a large number of unknown parameters, which is typical for concrete structures. Although a structural safety is implicitly implemented in codes by partial safety factors [1], their compatibility with advanced numerical methods is not generally guaranteed, and thus adequate stochastic analysis should be performed. Therefore, the crucial task here is the propagation of uncertainties through the numerical mathematical model and quantifying the uncertainty of the response as illustrated in Fig. 1.1.

Uncertainty quantification (UQ) covers quantifying and characterizing uncertainty in computational models [2]. In modern engineering, uncertainties are represented by random variables or random vectors described by certain probability measures. UQ then contains various tasks such as the estimation of statistical moments or probability distribution of a quantity of interest (QoI), and sensitivity analysis characterizing the role of uncertain inputs with respect to QoI. UQ as a separate scientific area combining various fields of applied mathematics is nowadays

present in almost every branch of engineering and science [3]. The importance of UQ is clearly visible from the exponentially growing number of published scientific works presenting novel techniques as well as their applications in real-life systems.

A successful uncertainty quantification allows for direct probabilistic assessment of structures, i.e. a calculation of the probability of structural failure. The concept of failure probability was already implemented into the general design standards for structures by the simplified semi-probabilistic approach. Instead of failure probability, the semi-probabilistic approach is focused on the estimation of design values of load effect and structural resistance satisfying the prescribed safety requirements. Design values can be seen as a specific quantile of the QoI's probability distribution, and thus UQ plays a crucial role in this process.

Although semi-probabilistic methods offer computationally efficient tools for the design and assessment of structures, they are only valid for relatively simple physical systems and stochastic models. Otherwise, it is necessary to employ advanced techniques offering higher versatility, though associated with a higher computational cost. The only generally applicable method for UQ is a Monte Carlo (MC) type sampling, which is based on a large number of repetitive deterministic simulations with realizations of the input random vector randomly generated according to its probability distribution. Unfortunately, the combination of MC and computationally costly mathematical models prevents its use in industrial applications. The natural solution of this problem is an approximation of the costly mathematical models by computationally efficient surrogate models.

Besides the traditional surrogate models such as Taylor series expansion, there are more efficient methods often employed specifically for UQ: support vector machines, neural networks, kriging, or polynomial chaos expansion (PCE). PCE is especially suitable for UQ thanks to its convenient properties that derive from its orthogonality properties with respect to the probability measures of the input variables. Interestingly, PCE was originally proposed by the famous mathematician Norbert Wiener in 1938 [4], but it took more than five decades to be employed in UQ of engineering systems [5]. Ever since then, there has been a rapid development of this method and various adaptations have been proposed. Nonetheless, although PCE represents a general UQ tool, its theoretical background is still under development, and it should be adapted for applications in structural mechanics and engineering further.

This habilitation thesis is presented through a series of journal articles thematically divided into two separate chapters. The first chapter is focused on the semi-probabilistic approach, specifically the estimation of the coefficient of variation of QoI, and the second chapter is focused on PCE and its modifications. The thesis contains significant recent developments in both topics – simplified methods for

the semi-probabilistic approach based on the estimation of the first two statistical moments presented 2, and novel theoretical methods developed for an efficient construction and analysis of surrogate models in the form of PCE in chapter 3. Both approaches tackle the identical problem – UQ of engineering systems –, but they are both employed in different contexts. While simplified methods should be applied for a routine analysis of relatively simple mathematical models in engineering practice, PCE provides an access to a broad palette of measures for UQ including statistical and sensitivity analysis. The discussion of the obtained results together with concluding remarks can be found in the last chapter of this thesis.

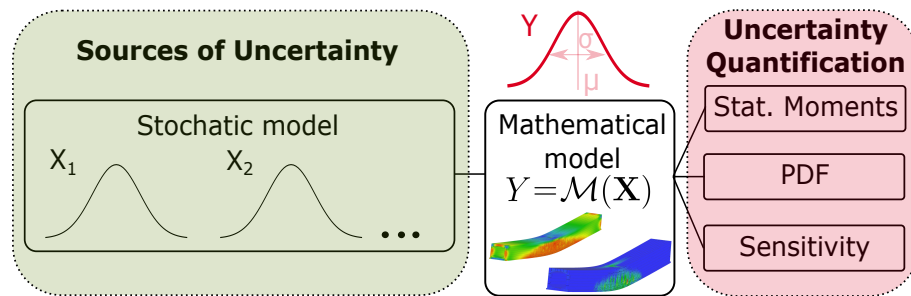


Fig. 1.1: Propagation of uncertainties through a mathematical model and UQ.

2 Uncertainty Quantification and Semi-probabilistic Approach

With thermodynamics, one can calculate almost everything crudely; with kinetic theory, one can calculate fewer things, but more accurately; and with statistical mechanics, one can calculate almost nothing exactly.

— Eugene Wigner

Assume a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ where Ω is an event space, \mathcal{F} is a σ -algebra on Ω and \mathcal{P} is a probability measure on \mathcal{F} . If the Doob-Dynkin lemma is satisfied and the input variable of a costly mathematical model $\mathcal{M}(X)$ is a random variable $X(\omega)$, $\omega \in \Omega$ with finite variance, the output QoI $Y = \mathcal{M}(X)(\omega)$ is also a random variable with finite variance. The deterministic mathematical model $\mathcal{M}(X)$ here represents an analyzed physical system, and it is typically created in commercial software. Thus, it can be seen as a black box in the context of UQ. Moreover, these systems can be seen as complex models of a large number of input parameters represented by a random vector \mathbf{X} consisting of M marginal random variables and described by a joint probability distribution function $f_{\mathbf{X}}$. The ultimate goal of a probabilistic analysis is thus UQ of $\mathcal{M}(\mathbf{X})$ in order to quantify statistical moments and estimate a probability distribution of Y . Moreover, it is often desired to perform a sensitivity and reliability analysis for a further analysis of the model.

Reliability analysis is a crucial topic in engineering, since it is focused on the estimation of the safety margin Z given as a difference between structural resistance R and action effect E . The probability of a negative safety margin – probability of failure $p_f = P(E > R) = P(Z < 0)$, quantifies the safety of structures. The failure probability can be generally obtained by the integral as follows:

$$p_f = \int_{\mathbb{R}} f_E(t) F_R(t) dt \quad (2.1)$$

where f_E is a probability density function of E , and F_R is a cumulative distribution function of R . A direct analytical evaluation of the p_f is extremely complicated or impossible in engineering applications, and thus it is necessary to employ numerical methods. The traditional approach is a transformation of the stochastic problem into a large set of deterministic calculations using well-known Monte Carlo (MC) method. MC is based on a large number of repetitive simulations with randomly generated realizations of input random vector \mathbf{x} according to its probability distribution law:

$$p_f = \int_{\Omega_f} f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x} = \int_{\mathbb{R}^M} I(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}, \quad (2.2)$$

where Ω_f is a failure domain and $I(\mathbf{x})$ is an indicator function defined as:

$$I(\mathbf{x}) = \begin{cases} 1 & \text{if } g(\mathbf{x}) \leq 0 \\ 0 & \text{otherwise} \end{cases}, \quad (2.3)$$

with $g(\mathbf{x})$ as the limit state function.

Naturally, in order to get a reliable estimation of p_f , it is necessary to evaluate a very large set of deterministic simulations, and this number of simulations is dependent on p_f . Note that civil engineering is characterized by a very low target p_f around 10^{-6} (depending on the specific category of a structure) and the fact that in practical applications, models solved by NLFEM could be computationally expensive. Due to the combination of both aspects, MC is rarely employed in industrial applications due to its enormous computational cost. It was necessary to develop simplified safety formats based on the semi-probabilistic approach focused on the estimation of coefficient of variation (ECoV) as described in the following paragraphs.

Semi-probabilistic methods were developed as an alternative to a complex reliability analysis for relatively simple mathematical models, and they are implemented in today's design codes [1]. The concept of a semi-probabilistic approach is based on the separation of R and E , and the estimation of their *design values* R_d and E_d satisfying the given safety requirements instead of the direct calculation of the failure probability p_f . The safety of a physical system is then assessed simply as $R_d \geq E_d$. The following paragraphs are only focused on the resistance side R , since the load effects are often not known a priori. The design value of resistance R_d in Eurocode is completely described by the sensitivity factor derived from First Order Reliability Method (usually simplified by the absolute value of $\alpha_R = 0.8$), the target reliability index β , and finally the first two central statistical moments together with the assumption of Gaussian or Lognormal probability distribution of R [1]. In other words, instead of a computationally costly estimation of p_F by various numerical methods, structural safety is assessed by the semi-probabilistic approach based on a statistical analysis of QoI – structural resistance $R = \mathcal{M}(\mathbf{X})$. Moreover, statistical analysis is further simplified by the assumption of a two-parametric probability distribution, and thus it is possible to employ simplified, but computationally efficient ECoV methods.

In the context of the semi-probabilistic approach and the corresponding safety formats, several methods were developed or adapted for the estimation of CoV of R : numerical quadrature [6, 7], ECoV methods [8, 9], Taylor series expansion (TSE)

[10], and Latin Hypercube Sampling (LHS) [11, 12]. These methods generally differ in the statistical sampling of input variables. For the sake of completeness, there are also two methods implemented in the Eurocode: Partial Safety Factor (PSF) method commonly used for structural design, and global safety factor method for a non-linear analysis of concrete structures according to EN 1992-2 (EN 1992-2). Note that these methods implemented in the Eurocode are not focused on the estimation of statistical moments. These methods try to directly estimate the design quantile of resistance from a single numerical simulation and several strong simplifications. Despite the success of such an approach in linear calculations, its utilization in NLFEM is questionable, since simulations with extremely low design material characteristics in PSF may lead to unrealistic results, or an implicit assumption of the CoV of R in the global safety factor could lead to a significant deviation from the reality as has already been shown by the author of this thesis in several studies [13, 14, 15, 16].

The standard ECoV method [9] proposed by Červenka was implemented in *fib* Model Code [17], and it is worthy to note that it will be implemented also in the new version of Eurocode prEN 1990:2022 and replace the global safety factor method in prEN1992-1-1:2022. Despite the success of the method in various industrial applications, it is still significantly limited by two aspects: (a) an assumption of Lognormal distribution of QoI, and (b) estimation of CoV from two numerical simulations with mean values and characteristic values (5% quantile) of input variables. The author of this thesis thus proposed a generalization of this standard ECoV method.

In order to derive a simplified method for statistical analysis based on a solid theoretical background, it was necessary to review the classic TSE method and propose several differencing schemes suitable for semi-probabilistic methods [18, 8]. Unfortunately, TSE suffers from the curse-of-dimensionality, since the number of terms grows rapidly with the number of input variables and the order of TSE. Therefore, it was necessary to assume another simplification – a correlation structure among input random variables, which is not typically known.

The role of correlation is thoroughly investigated in the first journal publication ***Estimation of Coefficient of Variation for Structural Analysis: The Correlation Interval Approach*** [19] attached to this chapter. First of all, it is shown that standard ECoV is implicitly based on the very strong assumption of fully correlated input random variables. Further, it is proved that standard ECoV coincides with TSE based on simple backward differencing. Identical methodology can be easily utilized for an adaptation of various differencing schemes developed for TSE and circumvent the curse-of-dimensionality. The proposed methodology is referenced as Eigen ECoV.

The Eigen ECoV fills the gap between the existing over-simplified methods implemented in codes commonly employed by civil engineers and the advanced costly

techniques based on TSE. Naturally, it is necessary to carefully consider the applicability of these assumptions in industrial applications and their possible impact. However, it is important to emphasize that identical simplifications are also implicitly assumed in standard ECoV, though it is not often discussed. The significant benefits of Eigen ECoV are its versatility and adaptivity allowing for easy modifications reflecting various aspects of specific mathematical models. The comparison of the existing semi-probabilistic methods and the Eigen ECoV is schematically depicted in Fig.2.1 together with LHS.

The estimated variance of QoI is significantly affected by the assumed correlation among input random variables, though the definition of correlation matrix among material characteristics is still challenging and there aren't any recommendations in codes. Although fully correlated input variables assumed by ECoV methods are not physically realistic, they significantly reduce the computational cost. On the other side of the spectrum, it is common to assume the also unrealistic uncorrelated input variables. In the second case, it is necessary to employ methods using a significantly higher number of simulations such as standard TSE or LHS. These two limit cases define the interval of variance reflecting vague or incomplete information about the correlation structure among input random variables. Practically speaking, one can suggest starting with the case of fully correlated random variables solved by computationally efficient but conservative Eigen ECoV. If the estimated CoV leads to a too conservative R_d , one should employ TSE or LHS assuming uncorrelated input random variables.

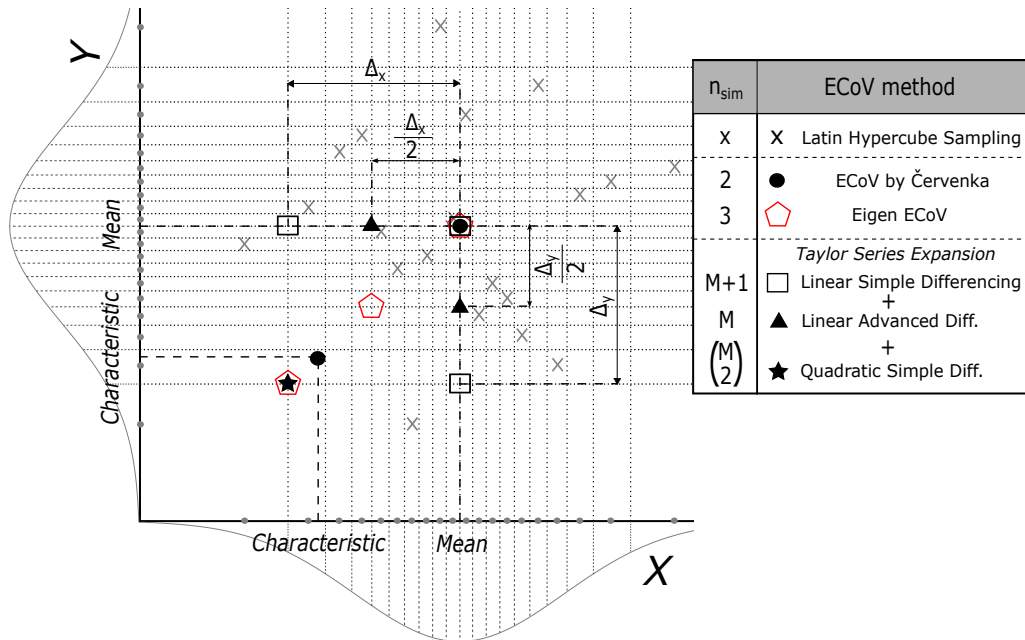


Fig. 2.1: Graphical interpretation of ECoV methods. (Adapted from [19])

The second paper attached to this chapter entitled *Comparison of Advanced Semi-probabilistic Methods for Design and Assessment of Concrete Structures* [20] presents a thorough comparison of existing semi-probabilistic methods together with a numerical validation of the correlation interval approach. The paper serves as a guideline for civil engineers and complements the theoretical results obtained from the first paper presenting Eigen ECoV. Numerical examples solved by NLFEM represent three typical concrete structural elements failing in different mechanisms. The mathematical models replicated the destructive tests found in the literature with credibility. All numerical examples are analyzed by PSF, TSE, Eigen ECoV, and standard ECoV in order to estimate CoV of resistance, and consequently R_d . The obtained results are compared to the correlation intervals obtained by LHS assuming uncorrelated and fully correlated variables. Moreover, there is also a result assuming a realistic correlation matrix obtained from an experimental campaign [13].

The numerical results presented in this paper are very important for a further development of the semi-probabilistic methods, since they generally show a very good agreement with theoretical expectations, and thus support the whole methodology. Specifically, it is evident that Eigen ECoV consistently leads to accurate and robust estimations, even for a significantly non-linear example, as opposed to standard ECoV, which fails in the case of shear failure due to limited information obtained from the simple backward differencing scheme. Furthermore, one can see that the concept of the correlation interval works for practical applications very well. Finally, it can be concluded that the whole methodology of Eigen ECoV serves as a basis for the development of the class of ECoV methods. An analyst can choose a suitable differencing scheme, an order of Taylor series expansion, and follow the process presented in the first paper in order to create an ECoV method suitable for a specific application. The obtained ECoV scheme can be further utilized for an estimation of CoV of R , and ultimately in safety formats based on the semi-probabilistic approach.

2.1 Estimation of Coefficient of Variation for Structural Analysis: The Correlation Interval Approach

NOVÁK, L.; NOVÁK, D. Estimation of coefficient of variation for structural analysis: The correlation interval approach. *Structural Safety*, 2021, vol. 92, no. 1. ISSN: 0167-4730. (WoS-AIS: D1)
DOI: 10.1016/j.strusafe.2021.102101

Description

The paper presents a theoretical background of the novel semi-probabilistic method referenced as Eigen ECoV, and a concept of correlation interval approach for design and assessment of structures represented by costly mathematical models solved by NLFEM. The existing safety formats are critically reviewed in the first part of the paper, and their limitations are theoretically investigated. Further, the role of statistical correlation among input random variables is thoroughly investigated. The core of the paper presents a general methodology for the derivation of various ECoV methods combining TSE of the first and the second order with simple and advanced differencing schemes. Finally, a concept of a correlation interval for structural assessment is presented. The correlation interval approach is based on an estimation of two physically unrealistic limit states – uncorrelated input random variables and fully correlated random variables. Both limit states lead to different variances of QoI and clearly show the consequences of the imprecise determination of correlation matrix.

Role of the author

Percentage of contribution: 80%

Lukáš Novák is the main author of this paper responsible for the concept, the methodology, and the numerical results of the presented research. Furthermore, he prepared the original draft of the paper, which was later reviewed by the second author, Drahomír Novák.



Estimation of coefficient of variation for structural analysis: The correlation interval approach

Lukáš Novák*, Drahomír Novák

Faculty of Civil Engineering, Brno University of Technology, Veverí 331/95, Brno 60200, Czech Republic

ARTICLE INFO

Keywords:

Semi-probabilistic approach
Estimation of coefficient of variation
Taylor series expansion
Correlation among random variables
Nataf transformation

ABSTRACT

The paper is focused on the efficient estimation of the coefficient of variation for functions of correlated and uncorrelated random variables. Specifically, the paper deals with time-consuming functions solved by the non-linear finite element method. In this case, the semi-probabilistic methods must reduce the number of simulations as much as possible under several simplifying assumptions while preserving the accuracy of the obtained results. The selected commonly used methods are reviewed with the intent of investigating their theoretical background, assumptions and limitations. It is shown, that Taylor series expansion can be modified for fully correlated random variables, which leads to a significant reduction in the number of simulations independent of the dimension of the stochastic model (the number of input random variables). The concept of the interval estimation of the coefficient of variation using Taylor series expansion is proposed and applied to numerical examples of increasing complexity. It is shown that the obtained results correspond to the theoretical conclusions of the proposed method.

1. Introduction

Today, non-linear finite element analysis (NLFEA) is employed ever more frequently for the design and assessment of structures, especially concrete structures with significant non-linear behaviour. Moreover in the last decade, it has become more common to use reliability analysis of real structures. This trend reflects the higher economical and safety requirements placed on engineering in today's society. Therefore, it is natural to connect NLFEA and reliability analysis in order to obtain accurate results [1–4]. Although the combination of NLFEA and reliability analysis is a strong tool for the realistic modelling of structures, it is also still highly time consuming to perform the reliability analysis of large non-linear mathematical models with many input random variables.

This paper is focused on the semi-probabilistic approach, which is well known from EN 1990 and partial safety factors [5]. This approach is able to greatly reduce the number of non-linear calculations necessary in order to estimate the design value of resistance satisfying the target reliability when the approach is used instead of the direct calculation of failure probability. However, it is still challenging to apply the semi-probabilistic approach to non-linear mathematical models solved by finite element software, when one deals with the non-linearity of functions combined with highly computationally demanding calculations.

Assuming a mathematical model of input random vector \mathbf{X} described by a specific joint probability distribution, the basic reliability

concept is given as $Z(\mathbf{X}) = R - E$, where $Z(\mathbf{X})$ represents safety margin, which is defined as the difference between structural resistance R and action effect E . Failure of the structure is represented by condition $Z(\mathbf{X}) < 0$. In the semi-probabilistic approach, the resistance of structure R is separated and the design value of resistance R_d that satisfies safety requirements is evaluated, instead of the direct calculation of failure probability $p_f = P(Z(\mathbf{X}) < 0)$. The typical formula for the estimation of R_d , assuming a lognormal distribution of R , is

$$R_d = \mu_R \cdot \exp(-\alpha_R \beta v_R), \quad (1)$$

where μ_R is the mean value, v_R is the coefficient of variation (CoV) and α_R represents sensitivity factor derived from First Order Reliability Method (FORM) [6,7]; the recommended value is $\alpha_R = 0.8$ according to [5]. The target reliability index β for the ultimate limit state, moderate consequences of failure and a reference period of 50 years is set at $\beta = 3.8$ according to the Eurocode. Note that, from a probabilistic point of view, the whole process represents the estimation of a quantile satisfying the given safety requirements under the prescribed assumption of lognormal distribution.

Obviously, for the determination of a design value by a semi-probabilistic approach, it is crucial to estimate the mean value and variance of structural resistance $R = r(\mathbf{X})$ accurately. This can be done via various techniques, such as numerical quadrature [8], simplified

* Corresponding author.

E-mail addresses: novak.l@fce.vutbr.cz (L. Novák), novak.d@fce.vutbr.cz (D. Novák).

methods for the estimation of the coefficient of variation (ECoV methods) [9], or stratified sampling [10]. Although simplified ECoV methods are often discussed at conferences, e.g. [11–14], and recommendations already exists in *fib* Model Code 2010 [15] and such methods are expected to be included in the Eurocode and *fib* Model Code 2020, there are still no significant scientific publications presenting the theoretical background and more importantly the limitations of the existing methods. Therefore, this paper contains a brief review of commonly used methods and also an investigation of the theoretical background of selected methods and their connection to well-known mathematical concepts.

ECoV methods offer a balance between computational cost and accuracy. However ECoV methods also have several limitations due to the assumed simplifications. Therefore besides an overview of ECoV methodology, this paper presents a novel generalization of ECoV methods for correlated random variables, since material characteristics (especially in the case of concrete structures) are correlated and this may play a crucial role in probabilistic analysis. In most cases, there is a lack of information on statistical correlation. Therefore, it is useful to examine two extreme cases (uncorrelated and fully correlated random variables). The interval ECoV approach and novel Eigen ECoV for fully correlated case are thus proposed in this paper.

2. Safety formats for NLFEA

The safety formats that include ECoV methods can be sorted by type of simplification into three levels as will be described in this section. Since non-linear mathematical models are generally not proportional, the standard quantile-based approach (level I) may lead to incorrect results. Accurate results are only guaranteed if the probability distribution of resistance f_R is identified together with statistical moments (Level III methods), which might not be a simple matter in general cases and is definitely time-consuming. Therefore, it is beneficial to assume several simplifications and employ Level II methods representing a compromise between accuracy and efficiency.

2.1. Level I: Quantile-based methods

Quantile-based methods are based on the very strict assumption that $r(\mathbf{X}_d) = R_d$, i.e. a numerical simulation with input variables set to a generally desired quantile (e.g. design \mathbf{X}_d) leads to a result corresponding to the identical desired quantile of response distribution R . Of course, this might be a severe problem in case of NLFEA, where a simulation with extreme input variable values may lead to the unrealistic behaviour of the computational model, which is usually verified within a specific range of input variables. However, such an approach can still be acceptable for simple structural members with a single almost linear failure mode and low v_R , e.g. the bending of a simple beam.

2.1.1. Partial safety factors

According to Partial Safety Factors (PSF) method proposed in EN 1990 [5], NLFEA is computed with design values of input random variables and it is assumed that the obtained result corresponds to the design value of resistance R_d [16]. The design values of input variables are typically derived from characteristic values using normative coefficients γ_M , which consider material and model uncertainty:

$$R_d = r(X_1/\gamma_M, X_2/\gamma_M, \dots). \quad (2)$$

Note that, the design values in the partial safety factors method are extremely low, thus possibly leading to the unrealistic redistribution of internal forces and even to different structural failure modes. One solution might be the calibration of partial safety factors based on laboratory experiments involving material and structural measurements [17].

2.1.2. Global safety factor according to EN 1992-2

In the global safety factor concept according to EN 1992-2 [18], the design value is estimated as follows:

$$R_d = \frac{r(f_{ym}, \tilde{f}_{cm}, \dots)}{\gamma_R}, \quad (3)$$

where $f_{ym} = 1.1f_{yk}$ is the mean value of the yield strength of steel reinforcement and f_{yk} represents its characteristic value (5% quantile), \tilde{f}_{cm} is the reduced mean value of concrete because of its higher variability and the idea shown in Eq. (4) that design values should correspond to the same probability and reflect the safety of normative material partial safety factors $\gamma_s = 1.15$ and $\gamma_c = 1.5$. The global safety factor for resistance is set as $\gamma_R = 1.27$ including model uncertainty.

$$\tilde{f}_{cm} = \gamma_s \cdot 1.1 \frac{f_{ck}}{\gamma_c} \approx 0.85 f_{ck}. \quad (4)$$

It is assumed that design values of concrete and steel should correspond to an identical quantile of probability. Furthermore, it is assumed that the mean value of steel can be obtained as $f_{ym} = 1.1f_{yk}$, and thus $f_{yd} = f_{ym}/1.27$.

Therefore, \tilde{f}_{cm} , according to Eq. (4), reflects the partial safety factors by virtue of the presented rationale. Note that \tilde{f}_{cm} does not represent the mean value of concrete material characteristics and it is lower than the characteristic values. As a result, it includes additional safety due to the higher variability of concrete. Also note that for concrete characteristics, Eurocode 2 allows only the compressive type of failure.

2.2. Level II: Simplified probabilistic methods

The task of Level II methods is reduced to the estimation of the mean value μ_R and variance of R , represented by the coefficient of variation v_R , which can be further decomposed as:

$$v_R = \sqrt{v_g^2 + v_m^2 + v_f^2}, \quad (5)$$

where v_g or v_m represents the coefficient of variation caused by geometrical or model uncertainties and the v_f coefficient of variation caused by the material. There are several studies dealing with model uncertainties and it is necessary to adopt v_g or v_m for specific structures [16, 19, 20]. Therefore for the sake of generality, the paper is focused only on the estimation of the coefficient of variation of mathematical model caused by the uncertainty of material parameters v_f .

2.2.1. Numerical quadrature

A classic method to estimate moments of function R , was proposed in 1975 by Rosenblueth [8]. This point estimate method is simple and direct, thus the method can be easily employed in practical applications. Moreover, Christian and Baecher [21] have shown its robustness and mathematical background in numerical quadrature. The expected value of the m -th moment of function $r(\mathbf{X})$ can be estimated as:

$$\mathbb{E}[R^m] \approx \sum_{i=1}^{2^N} P_i \cdot r_i^m, \quad (6)$$

where $P_i = \frac{1}{2^N}$ are weighting factors and r_i is a result of mathematical model. Since the function $r(\mathbf{X})$ is computed in 2^N points with coordinates plus/minus one standard deviation σ_{X_i} for the i -th random variable, the computational requirements increase rapidly with the number of input random variables – 2^N simulations are needed to estimate the statistical moments of R and thus it cannot be recommended for typical engineering applications.

2.2.2. ECoV by Červenka

The computationally efficient ECoV method was proposed in 2008 by Červenka [22]. The method is based on a simplified formula for the estimation of a characteristic value corresponding to a lognormal variable with the mean value μ_R and v_f :

$$R_k = \mu_R \exp(-1.645 v_f), \quad (7)$$

where -1.645 corresponds to the 5% quantile of standardized Gaussian distribution $\Phi(0.05)$. After simple mathematical operations and under the assumption that $R_m \approx \mu_R$, the coefficient of variation of R associated with material uncertainties v_f can be estimated as:

$$v_f = \frac{1}{1.645} \ln \left(\frac{R_m}{R_k} \right), \quad (8)$$

and the global resistance safety factor is calculated as:

$$\gamma_R = \exp(\alpha_R \beta v_f). \quad (9)$$

Note that, just 2 NLFEA simulations are needed in this approach independent of the size of the stochastic model – $R_m \approx r(\mu_X)$ with mean values of input random variables and R_k using characteristic values (5% percentile) of input variables. Obviously, there is the strong assumption that $R_k \approx r(X_k)$. However there is significant advantage in comparison to previous methods, since it estimates v_f and thus the design value R_d corresponds to target safety requirements and the specific distribution of R . Moreover the characteristic values of material parameters are not as extremely low as in the case of PSF and may not lead to structural system or material model exhibiting unrealistic behaviour, which can be considered as a significant advantage. Note that, the described concept was adopted in the *fib* Model Code 2010 [15] and is widely accepted in the engineering community today [23]. However, in spite of the success of this method, its theoretical background has not yet been sufficiently investigated. In this paper an attempt to fill this gap has been made, primarily in Section 4.1.

2.2.3. Taylor series expansion

Let us assume the mathematical model $r(\mathbf{X})$ is infinitely differentiable in an open interval around the mean values. Under this assumption, it is possible to expand the original model into an infinite Taylor series:

$$R = r(\mathbf{X}) = r(\mu_X) + \nabla r(\mu_X) \cdot (\mathbf{X} - \mu_X) + \frac{1}{2} (\mathbf{X} - \mu_X) \cdot \nabla \nabla r(\mu_X) \cdot (\mathbf{X} - \mu_X) + \dots \quad (10)$$

where the derivatives are evaluated at μ_X . In engineering applications, it is common to assume that the terms of TSE are only linear and that input random variables are independent. For the sake of clarity, the commonly known analytical expressions for the estimation of the expected value $\mathbb{E}[R]$ and variance $\text{VAR}[R]$ of a function $r(\mathbf{X})$ of N independent random variables, approximated by linear terms of the TSE, are as follows:

$$\mathbb{E}[R] \approx r(\mu_{X_1}, \mu_{X_2}, \dots, \mu_{X_n}), \quad (11)$$

and

$$\text{VAR}[R] \approx \sum_{i=1}^N \left(\frac{\partial r(X)}{\partial X_i} \right)^2 \sigma_{X_i}^2, \quad (12)$$

As can be seen from the equations, the efficiency and accuracy of TSE depends on the number of used terms and the differencing scheme for the practical computation of derivatives. A practical example of TSE utilization is the ECoV method proposed by Schlune et al. [24], which can be seen as a TSE in which, derivatives are approximated by one-sided differencing as:

$$\frac{\partial r(X)}{\partial X_i} = \frac{R_m - R_{X_i \Delta}}{\Delta X_i}. \quad (13)$$

where the response of mathematical model R_m is determined by a calculation with mean values, and $R_{X_i \Delta}$ is the result of a model using mean values of input random variables and a value of the i -th random variable which has been reduced by ΔX_i . This differencing scheme has been adapted for structural design according to Schlune et al. using step size parameter $c = (\alpha_R \beta) / \sqrt{2}$ and $X_{i \Delta} = F_i^{-1}(\Phi(-c))$, where F_i^{-1} is an inverse cumulative distribution function of the i -th variable and Φ is the cumulative distribution function of the standardized Gaussian distribution. For the sake of clarity, the difference is calculated as $\Delta X_i = \mu_{X_i} - X_{i \Delta}$.

Schlune et al. thus proposed a simple formula [24] for the coefficient of variation caused by material uncertainty v_f if material parameters are not correlated as:

$$v_f \approx \frac{1}{R_m} \sqrt{\sum_{i=1}^N \left(\frac{R_m - R_{X_i \Delta}}{\Delta X_i} \sigma_{X_i} \right)^2}. \quad (14)$$

Note that this approach requires $N + 1$ simulations of NLFEA, where N is the number of random variables. However, all simulations act as a parametric study of a numerical model, which is usually performed during the development of a model in industrial applications. As a result, TSE can be recommended due to its medium computational cost and strong theoretical background.

Of course, one can use various differencing schemes instead of Eq. (13) depending on ones computational possibilities, as was proposed by the authors of this paper in [25]. Assuming linear TSE, one of the most promising advanced differencing schemes using $n_{sim} = 2N + 1$ simulations is defined as:

$$\frac{\partial r(X)}{\partial X_i} = \frac{3R_m - 4R_{X_i \frac{\Delta}{2}} + R_{X_i \Delta}}{\Delta X_i}, \quad (15)$$

where the middle additional term $R_{X_i \frac{\Delta}{2}}$ is obtained via the evaluation of the original mathematical model with mean values and a reduced i -th variable $X_{i \frac{\Delta}{2}} = \mu_{X_i} - \Delta X_i / 2$.

2.3. Level III: Monte Carlo methods

Monte Carlo (MC) type sampling methods are the only general tools available for reliability or statistical analysis. However, it is necessary to perform large number of calculations if they are used. Nonetheless, the number of simulations is still lower than in the case of Level II methods for large stochastic models and thus Level III for ECoV [10] is recommended for large stochastic models or computationally cheap computational models.

The main feature of MC techniques is their use of pseudo-random sampling and the statistical analysis of performed deterministic simulations. Crude Monte Carlo is not efficient because thousands of simulations are needed and the use of this approach in combination with NLFEA, is not feasible in industrial applications. A stratified sampling technique called Latin Hypercube Sampling (LHS) was developed for the efficient estimation of statistical moments [26,27]. It drastically reduces the number of needed simulations. LHS is not dependent on the size of the stochastic model, and thus it is recommended for extensive stochastic models. The cumulative distribution function of the input variable is divided into n_{sim} equal intervals, where n_{sim} is the number of simulations. Every value is picked within each segment. There are several ways to choose the probability of picked value — mean value of interval, median or random value. Once the values are chosen, the random permutation of realizations is performed and random vectors of input variables are generated. The described approach leads to uniform distribution within a design domain. Additionally, MC type simulation techniques are able to take the correlation among input random variables into account. Several methods have been developed for this purpose, e.g. generalized Nataf transformation [28] and optimization techniques [29,30].

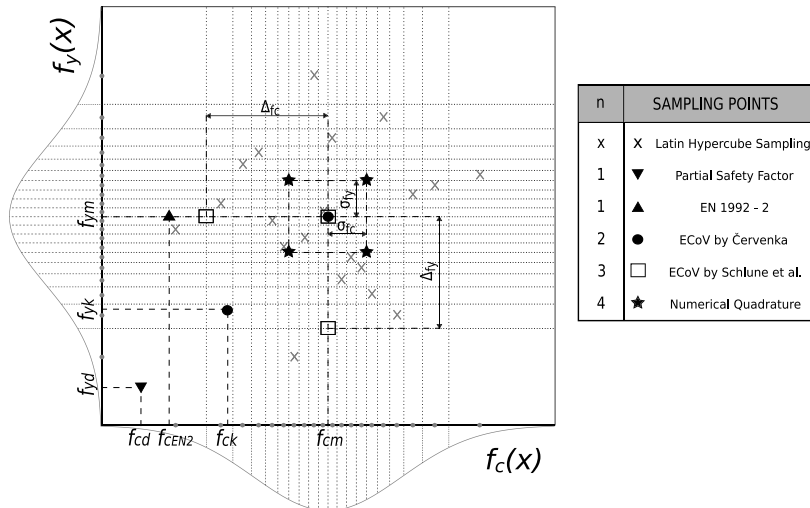


Fig. 1. The presented semi-probabilistic methods together with their computational cost n are depicted in 2-dimensional space (figure adapted from [31]).

2.4. Graphical comparison of safety formats

As was already shown by Pimentel et al. [31], each NLFEA simulation can be represented by a point in N -dimensional space called the design domain. The design domain is considered to be the domain of sampling probabilities, where coordinates of sampling points are described by an input random vector. The 2-dimensional case can be seen in Fig. 1 together with the sampling points used for the presented safety formats. The stochastic model contains only 2 typical input material characteristics – the yield strength of reinforcement f_y and the compressive strength of concrete f_c , which are considered to be random variables described by specific probability distributions.

Level I: The PSF and the EN-1992 methods represented by triangles are very efficient from the computational point of view — only one simulation is needed but CoV is not estimated. Although quantile-based methods are sufficient for a linear mathematical model, they may lead to severe problems depending on the degree of non-linearity of mathematical models solved by the NLFEA. Moreover, design values of input random variables are defined only for selected material parameters (excluding e.g. fracture energy).

Level II: ECoV by Červenka (represented in Fig. 1 by circles) always works with 2 simulations — the mean and the characteristic values of input random variables. On the other hand, TSE with differencing according to Schlune et al. (represented in Fig. 1 by squares) is linearly dependent ($n_{sim} = N + 1$), and the Numerical Quadrature (stars) is exponentially dependent ($n_{sim} = 2^N$) on the number of input random variables. Although such dependency is not a problem for low dimensional space, it can play a crucial role in industrial applications with many input variables and thus it might be more efficient to employ Monte Carlo type methods. The level III Monte Carlo type method is represented by LHS (crosses) in the figure. LHS can be used for general stochastic models and will be employed in this paper as a reference solution.

3. Correlation among random variables

The presented ECoV methods are recommended for the practical assessment of structures assuming that material characteristics are independent, which is usually incorrect. This is especially true in the case of concrete structures, where a correlation among compressive strength, tensile strength and fracture energy is usually assumed [3, 32]. Therefore, the paper is focused on the generalization of ECoV methods for structures with dependent input material characteristics, which are typically obtained from laboratory experiments or assumed according to the literature. General transformation between correlated and uncorrelated space is briefly described in this section.

3.1. Nataf transformation

In a general case involving non-normal correlated random variables, it is necessary to utilize what is known as the Rosenblatt transformation [33]. However, in practical applications only the marginal distributions and the correlation matrix are usually known, which does not provide complete information about the joint probability distribution [34]. Therefore, it is necessary to assume a specific copula [35] or construct an arbitrary joint distribution using vine copulas [36], which is beyond the scope of this paper. A special case of Rosenblatt transformation that assumes Gaussian copula [37] is also known as the Nataf transformation [38], which is usually utilized in reliability applications. Nataf transformation to ξ space is composed of three steps:

$$\xi = T_{Nataf}(\mathbf{X}) = T_3 \circ T_2 \circ T_1(\mathbf{X}). \quad (16)$$

The first two steps are commonly known as iso-probabilistic transformation, which uses the cumulative distribution function of variables F_x and the Gaussian inverse cumulative distribution function Φ^{-1} as follows:

$$T_2 \circ T_1(\mathbf{X}) : \mathbf{X} \mapsto \mathbf{Z} = \Phi^{-1}(F_x(\mathbf{X})). \quad (17)$$

The last step represents a transformation to uncorrelated space using linear transformation. For this procedure, we can use the Cholesky decomposition or the Eigen decomposition of the fictive correlation matrix \mathbf{R}_Z . Using Cholesky decomposition, the decomposition is

$$\mathbf{R}_Z = \mathbf{L}\mathbf{L}^T, \quad (18)$$

and final transformation using $\Gamma = \mathbf{L}^{-1}$ thus reads

$$T_3 : \mathbf{Z} \mapsto \xi = \Gamma\mathbf{Z}. \quad (19)$$

The Nataf transformation, can be easily inverted in order to transform $\xi \mapsto \mathbf{X}$. Note that the transformation matrix \mathbf{L} is a lower triangular matrix with a unit on the first entry of the main diagonal and therefore the first coordinate x_1 remains unchanged. This complication can be circumvented by using Eigen decomposition instead of Cholesky decomposition. The target covariance matrix Σ can be decomposed using Eigen decomposition as:

$$\Sigma = \Theta \lambda^{\frac{1}{2}} \lambda^{\frac{1}{2}} \Theta^T, \quad (20)$$

where λ is the diagonal matrix of eigenvalues of Σ and Θ is the eigenvector matrix associated with the eigenvalues. Instead of the transformation matrix \mathbf{L} , one can then use $\Theta \lambda^{\frac{1}{2}}$.

3.2. Fictive correlation matrix

A critical task for Nataf transformation is the determination of \mathbf{R}_Z . The fictive correlation matrix \mathbf{R}_Z is a square symmetric positive-definite matrix, and thus it is possible to perform Cholesky decomposition. The assumed Gaussian copula is parametrized by elements ρ_{zij} of \mathbf{R}_Z . Note that $\rho_z = 0 \leftrightarrow \rho_x = 0$ and $|\rho_x| \leq |\rho_z|$ as was shown in [34]. The relationship between fictive correlation coefficients ρ_{zij} and ρ_{xij} determined for \mathbf{X} is defined by the following integral equation:

$$\rho_{xij} = \frac{1}{\sigma_i \sigma_j} \iint_{\mathbb{R}^2} \{F_i^{-1}[\Phi(z_i) - \mu_i] F_j^{-1}[\Phi(z_j) - \mu_j] \times \phi_2(z_i, z_j, \rho_{zij})\}, \quad (21)$$

where μ is the mean value, σ is the standard deviation and ϕ_2 is the bivariate standard normal probability density function parametrized by fictive correlation coefficients ρ_{zij} . The computation of Eq. (21) might be complicated for practical usage. Moreover, for specific combinations of input parameters there is not a guaranteed solution (more details about the limitations of Nataf transformation can be found in [35]). Generally, a simplification of Eq. (21) according to Liu & Kiureghian [39] can be in the form $\rho_z = t \cdot \rho_x$, where t is known for several combinations of probability distributions of random variables. The material characteristics are often assumed to be lognormally distributed with coefficient of variation $v \leq 0.5$ in practical applications. Assuming both random variables to be lognormally distributed, Liu & Kiureghian derived t in the following form:

$$t = \frac{\ln(1 + \rho_x v_1 v_2)}{\rho_x \sqrt{\ln(1 + v_1^2) \ln(1 + v_2^2)}}, \quad (22)$$

where v_1 and v_2 are coefficients of variation of the first and second random variable respectively. As can be found in [39], the derived formula is exact in this specific case. Additionally, in common practical applications, one can assume there is a positive correlation among material characteristics which leads to negligible differences between ρ_z and ρ_x [34].

4. ECoV for functions of correlated random variables

Besides the general probabilistic methods in level III, it is also possible to extend Level II of safety formats for correlated random variables. The numerical quadrature can be easily extended for correlated variables via the modification of weighting factors P_i as follows [21]:

$$P_{(s_1, s_2, \dots, s_n)} = \frac{1}{2^n} \left[1 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n (s_i)(s_j) \rho_{ij} \right], \quad (23)$$

where s_i is a positive sign when the value of the i -th variable is the mean plus the standard deviation σ and negative for points with a coordinate mean value minus the standard deviation. Although it is generally possible to use numerical quadrature, it is highly computationally demanding ($n_{sim} = 2^N$) and thus its potential for industrial applications is limited, and it will not be employed in numerical examples.

The next presented Level II method – TSE – is more interesting for industrial applications, since it is not highly computationally demanding. It is possible to generalize the TSE for correlated variables using additional terms of the expansion. Specifically, an extension of the method for dependent random variables can generally be obtained from a first order TSE assuming correlation among random variables represented by the correlation coefficient ρ in analytical form as

$$\text{VAR}[R] \approx \sum_{i=1}^N \left(\frac{\partial r(\mathbf{X})}{\partial X_i} \right)^2 \sigma_{X_i}^2 + \sum_{i,j=1, \dots, N, i \neq j} \rho_{i,j} \sigma_{X_i} \sigma_{X_j} \frac{\partial r(\mathbf{X})}{\partial X_i} \frac{\partial r(\mathbf{X})}{\partial X_j}. \quad (24)$$

However, higher terms of the TSE or more accurate approximations of derivatives should be considered for the correct estimation of variance in the case of dependent input random variables and non-linear

functions. The authors of this paper recently proposed a methodology consisting of three levels of increasing accuracy and complexity described from the mathematical point of view in [25]. This methodology can be used for an arbitrary correlation matrix.

There is no theoretical background available in literature for the second presented Level II method (ECoV by Červenka), and thus it is not possible to directly generalize it for any correlation matrix. Therefore, the TSE for functions of fully correlated random variables and its connection to ECoV by Červenka are investigated in the next subsection.

4.1. Special case: ECoV for fully correlated random variables

There is often a strong assumption of fully correlated input random variables in industrial applications, which will be adopted for the further investigation of TSE. Without loss of generality, let us investigate the situation in Gaussian space. Similarly as in the case of Eq. (7) for lognormal distribution, ECoV by Červenka, which assumes Gaussian distribution, is based on the following formula:

$$R_k = \mu_R (1 - 1.645 v_f), \quad (25)$$

and v_f is therefore obtained as:

$$v_f = \frac{R_m - R_k}{1.645 R_m}, \quad (26)$$

where $R_k = r(\mathbf{X}_k)$ and $R_m = r(\mathbf{X}_m) \approx \mu_R$.

For further comparisons, let us set up the step size parameter of TSE as $c = -\Phi(0.05) \approx 1.645$, which corresponds to the same quantile as in ECoV by Červenka. The differencing scheme defined for the uncorrelated case in Eq. (13) can be transformed by a Nataf transformation which has been parametrized by arbitrary correlation coefficients, as can be seen in Fig. 2 (left).

The N -dimensional ellipsoid corresponding to σ -distance is described by eigenvectors $(\theta_1, \dots, \theta_N)$ and eigenvalues $(\lambda_1, \dots, \lambda_N)$ obtained from the Eigen decomposition of a covariance matrix. Note that in the limit case $\lim_{\rho \rightarrow 1} \lambda_1 = \text{tr}(\Sigma) = \sigma_\theta^2$ and $\lim_{\rho \rightarrow 1} \lambda_i = 0 \forall i > 1$. In other words, the N -dimensional joint probability distribution is reduced to a 1-dimensional projection with the distribution $X_\theta \sim \mathcal{N}(\mu_\theta, \lambda_1)$. The Nataf transformation of $\mathbf{X}_{i\Delta}$ is depicted in standardized Gaussian space ξ for a 2D case with increasing positive $\rho \in (0, 1)$ together with isolines of bivariate Gaussian distribution in $c \cdot \sigma$ -distance. As can be seen in Fig. 2 (left), with increasing $\rho \rightarrow 1$ the coordinates of $\mathbf{X}_{i\Delta}$ transform to \mathbf{X}_k and $\mathbf{X}_{i\Delta} \forall i > 1$ to \mathbf{X}_m , and thus:

$$\frac{\partial r(\mathbf{X})}{\partial X_i} = \frac{R_m - R_{\mathbf{X}_{i\Delta}}}{\Delta X_i} = 0 \forall i > 1. \quad (27)$$

The limit cases $\rho = 0$ and $\rho = 1$ are compared in Fig. 2 (right). As can be seen, the iso-lines of bivariate Gaussian distribution in σ -distance (grey) and $c \cdot \sigma$ -distance (red) are reduced to a single line. From the simple geometry, one can derive the following expressions:

$$A_\theta = c \cdot \sqrt{\lambda_1} = \sqrt{\sum_{i=1}^N (\mu_{X_i} - X_{i\Delta})^2}. \quad (28)$$

Finally, the variance estimated by linear TSE for fully correlated input variables can be estimated as:

$$\text{VAR}[R] = \left(\frac{\partial r(\mathbf{X})}{\partial X_\theta} \right)^2 \lambda_1, \quad (29)$$

where the derivative is obtained from two simulations $R_m = r(\mathbf{X}_m)$ and $R_{\theta\Delta} = R_k = r(\mathbf{X}_{i\Delta})$ as

$$\frac{\partial r(\mathbf{X})}{\partial X_\theta} = \frac{R_m - R_{\theta\Delta}}{\Delta_\theta}. \quad (30)$$

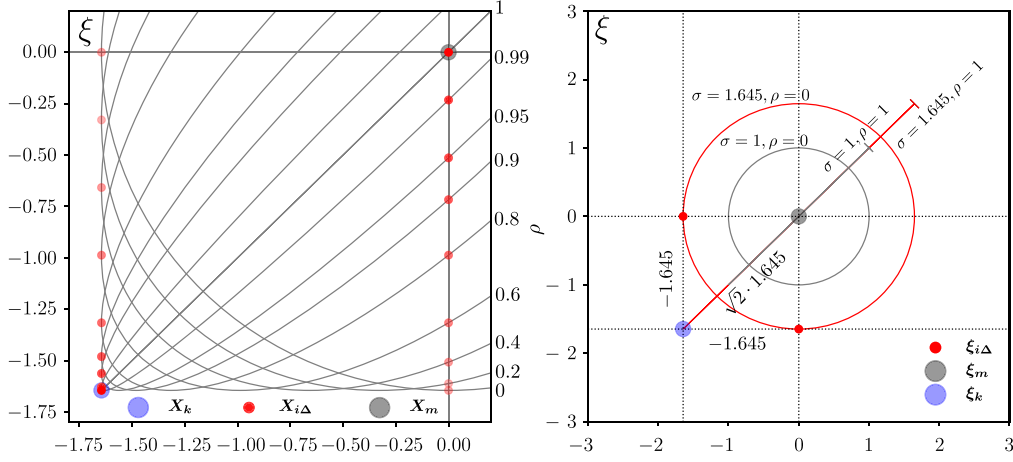


Fig. 2. Nataf transformation of a TSE in standardized Gaussian space. Transformation of $X_{i\Delta}$ with increasing ρ together with isolines of bivariate Gaussian distribution (left). Comparison of limit cases for $\rho = 0$ and $\rho = 1$ (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

For a direct comparison with ECoV by Červenka, v_f from the equation above is obtained as:

$$v_f = \frac{\sigma_R}{\mu_R} = \sqrt{\left(\frac{R_m - R_{\theta\Delta}}{c \cdot \sqrt{\lambda_1}}\right)^2} \lambda_1 \frac{1}{\mu_R} = \frac{R_m - R_k}{1.645 R_m}. \quad (31)$$

Therefore, ECoV by Červenka (see Eq. (26)) can be seen as a special case of the TSE for fully correlated random variables that assumes linearity of the mathematical model together with a specific distribution of R (typically lognormal). If these assumptions are fulfilled ECoV by Červenka is a highly efficient method. In the opposite case, it may lead to inaccurate results. Note that this method is widely used without knowledge of assumed fully correlated input random variables, which usually increases the variance of the function in practical applications and thus obtains conservative results.

As significant disadvantage of ECoV by Červenka is its theoretical background based on the simplified formula Eq. (7), which is only accurate for low v_R . Moreover, it cannot be easily generalized and thus more complex ECoV formulas using different derivative schemes are derived from the TSE for fully correlated variables in the following section.

4.2. Eigen ECoV

Using differencing schemes proposed by the authors of this paper in [25], one can create several formulas similar to ECoV by Červenka directly from a TSE transformed by Nataf transformation for fully correlated random variables, which is depicted in Fig. 3 for Gaussian input random variables. In the special case that $\rho \rightarrow 1$, the joint probability distribution is reduced to the 1D distribution $X_\theta \sim \mathcal{N}(\mu_\theta, \lambda_1)$, which can be expanded by the TSE.

If there is no assumption of Gaussian input random variables, one has to use a corresponding probability distribution of input random variables, i.e. $X_{i\Delta} = F_i^{-1}(\Phi(-c))$. Moreover, the geometrical properties become more complex and thus one has to assume a specific distribution of X_θ in order to calculate Δ_θ in physical space. Typically, one can assume lognormal distribution and thus Δ_θ can be estimated as follows:

$$\Delta_\theta = \mu_\theta - \mu_\theta \cdot \exp(-c \cdot \frac{\sqrt{\lambda_1}}{\mu_\theta}), \quad (32)$$

where μ_θ is calculated as:

$$\mu_\theta = \sqrt{\sum_{i=1}^N (\mu_{X_i})^2}. \quad (33)$$

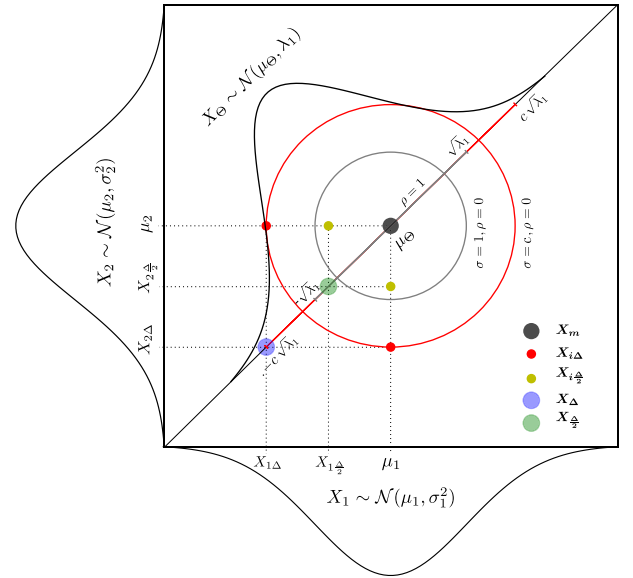


Fig. 3. Graphical interpretation of Eigen ECoV.

The first order TSE leads to well known expressions for the variance (Eq. (29)) and mean value $R_m \approx r(\mathbf{X}_m)$ of structural resistance R . Furthermore, one can use an arbitrary differencing scheme and step-size parameter c . As was shown, simple backward differencing is used with different c by Schlune et al. [24] ($c = (\alpha_R \beta) / \sqrt{2}$) and Červenka [22] ($c = 1.645$). Note that $c = 1.645$ is assumed in Fig. 3. For the sake of clarity, let us recall the following notation: $R_{\theta\Delta} = r(\mathbf{X}_{\theta\Delta})$ and $\mathbf{X}_{\theta\Delta} = (X_{1\Delta}, \dots, X_{N\Delta})$. The input vector consists of reduced values of input random variables $X_{i\Delta} = F_i^{-1}(\Phi(-c))$. The Δ_θ is calculated according to Eq. (32) under the assumption of lognormally distributed X_θ or Eq. (28) under the assumption of Gaussian X_θ (Gaussian input random variables). The variance of 1D Eigen distribution is $\lambda_1 = \text{tr}(\Sigma) = \sum_{i=1}^N \sigma_{X_i}^2$. Based on the presented theory and notation, the following variants of Eigen ECoV are proposed:

- (a) The Eigen ECoV derived from the first order TSE using simple backward differencing leads to the following expression for the expected value, variance and CoV using two simulations:

$$\mathbb{E}[R] = R_m \approx r(\mathbf{X}_m), \quad (34)$$

$$\text{VAR}[R] \approx \left(\frac{R_m - R_{\theta\Delta}}{\Delta_\theta}\right)^2 \cdot \lambda_1, \quad (35)$$

$$v_f \approx \frac{R_m - R_{\theta\Delta}}{\Delta_\theta} \cdot \frac{\sqrt{\lambda_1}}{R_m}. \quad (36)$$

(b) Furthermore the Eigen ECoV derived from the first order TSE using advanced backward differencing leads to the following expression for mean, variance and CoV using three simulations:

$$\mathbb{E}[R] = R_m \approx r(\mathbf{X}_m), \quad (37)$$

$$\text{VAR}[R] \approx \left(\frac{3R_m - 4R_{\theta\frac{\Delta}{2}} + R_{\theta\Delta}}{\Delta_\theta} \right)^2 \cdot \lambda_1, \quad (38)$$

$$v_f \approx \frac{3R_m - 4R_{\theta\frac{\Delta}{2}} + R_{\theta\Delta}}{\Delta_\theta} \cdot \frac{\sqrt{\lambda_1}}{R_m}, \quad (39)$$

where $R_{\theta\frac{\Delta}{2}} = r(\mathbf{X}_{\theta\frac{\Delta}{2}})$ and position of $\mathbf{X}_{\theta\frac{\Delta}{2}} = (X_{1\frac{\Delta}{2}}, \dots, X_{N\frac{\Delta}{2}})$ is depicted in Fig. 3. The input vector consists of reduced values of input random variables:

$$X_{i\frac{\Delta}{2}} = \mu_{X_i} - \frac{\mu_{X_i} - X_{i\Delta}}{2} = \mu_{X_i} - \frac{\Delta X_i}{2}. \quad (40)$$

(c) Additionally, using three identical simulations to those used in case (b), one can derive the Eigen ECoV from the second order TSE and thus obtain more accurate expressions. However it is necessary to include the information of higher statistical moments into the expression for variance in the case that X_θ has an assumed lognormal distribution (lognormally distributed input random variables). The following expressions are derived:

$$\mathbb{E}[R] \approx R_m + \frac{R_m - 2R_{\theta\frac{\Delta}{2}} + R_{\theta\Delta}}{\Delta_\theta^2} \cdot \frac{\lambda_1}{2}, \quad (41)$$

$$v_f \approx \frac{\sqrt{\text{VAR}[R]}}{\mathbb{E}[R]}. \quad (42)$$

- Assuming Gaussian X_θ , the third central moment $\mu_3 = 0$ and the fourth central moment $\mu_4 \approx 3\lambda_1^2$, and thus the variance is obtained via the following expression:

$$\text{VAR}[R] \approx \left(\frac{3R_m - 4R_{\theta\frac{\Delta}{2}} + R_{\theta\Delta}}{\Delta_\theta} \right)^2 \lambda_1 + \left(\frac{R_m - 2R_{\theta\frac{\Delta}{2}} + R_{\theta\Delta}}{\Delta_\theta^2} \right)^2 \frac{\lambda_1^2}{2}. \quad (43)$$

- Assuming lognormally distributed X_θ , higher central moments must be included in the expression as follows:

$$\text{VAR}[R] \approx \left(\frac{3R_m - 4R_{\theta\frac{\Delta}{2}} + R_{\theta\Delta}}{\Delta_\theta} \right)^2 \lambda_1 + \left(\frac{R_m - 2R_{\theta\frac{\Delta}{2}} + R_{\theta\Delta}}{\Delta_\theta^2} \right)^2 \frac{\mu_{4\theta} - \lambda_1^2}{4} + \mu_{3\theta} \left(\frac{3R_m - 4R_{\theta\frac{\Delta}{2}} + R_{\theta\Delta}}{\Delta_\theta} \right) \left(\frac{R_m - 2R_{\theta\frac{\Delta}{2}} + R_{\theta\Delta}}{\Delta_\theta^2} \right). \quad (44)$$

The third and fourth central moments can be derived for lognormal distribution $X_\theta \sim \mathcal{LN}(\mu_{LN}, \sigma_{LN})$ directly from the shape parameter $\sigma_{LN}^2 = \ln(1 + \frac{\lambda_1}{\mu_\theta^2})$ as:

$$\mu_{3\theta} = \left[\left(e^{\sigma_{LN}^2} + 2 \right) \sqrt{e^{\sigma_{LN}^2} - 1} \right] \lambda_1^{\frac{3}{2}}, \quad (45)$$

$$\mu_{4\theta} = \left[\left(e^{4\sigma_{LN}^2} \right) + 2 \left(e^{3\sigma_{LN}^2} \right) + 3 \left(e^{2\sigma_{LN}^2} \right) - 3 \right] \lambda_1^2. \quad (46)$$

4.3. Correlation interval ECoV

Since the information about correlation among input random variables is often vague and usually based on expert judgement, it is beneficial to study two limit cases: uncorrelated variables and fully correlated variables. The obtained results can be used for the reliable estimation of variance or CoV. Moreover, an analyst can clearly see

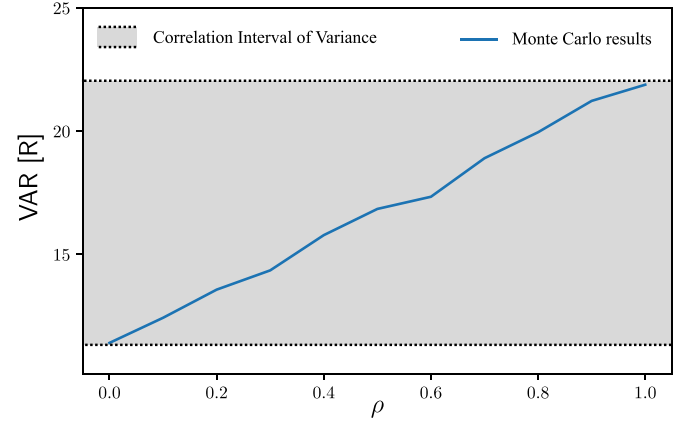


Fig. 4. Interval ECoV approach assuming identical increasing ρ among all variables. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the consequences of the imprecise determination of correlation matrix, which is often neglected. In the case of uncorrelated input random variables, TSE with simple ($n_{sim} = N + 1$) and advanced ($n_{sim} = 2N + 1$) differencing can be used according to the TSE methodology presented in [25]. The fully correlated case is examined via the proposed Eigen ECoV methods in three variants together with ECoV by Červenka. This methods are utilized in numerical examples in order to compare and show the limitations of the existing and proposed ECoV methods. From the practical point of view, it is beneficial to estimate variance while assuming fully correlated random variables in the case of a limited computational budget (large mathematical models) and only make further use of the TSE in order to obtain an accurate estimate of the role of correlation.

5. Numerical examples

The results of the numerical examples in this section are presented for two variants of stochastic models: under the assumption of Gaussian input variables and under the assumption of lognormal input variables. The reference solution is obtained by LHS with $n_{sim} = 10^4$ for uncorrelated random variables and also for increasing $\rho = (0, 1)$ with step 0.1 (identical ρ is assumed among all input variables). The two extremes (fully correlated and uncorrelated) define the boundaries for the interval of variance as can be seen in Fig. 4, where the blue line represents the reference solution obtained by a Monte Carlo type simulation technique for increasing ρ , and the interval is highlighted in grey. The depicted results correspond to Example 3, though the approach was used for all examples.

The variance of the uncorrelated case is estimated via a linear TSE with a simple derivative scheme (Eq. (13)) represented in the figures by a dashed line, and with an advanced differencing scheme (Eq. (15)) in figures represented by a dot-and-dash line. Note that these two methods represent the first and second order of the methodology proposed in [25]. The estimation of variance for a fully correlated limit case is obtained by the proposed Eigen ECoV and ECoV by Červenka, which is equal to Eigen ECoV (a) in Gaussian space, though there is a difference due to the approximation of lognormal distribution by Eq. (8).

5.1. Example 1: Ultimate bending moment

The very first example is a classical mathematical model of the ultimate bending moment of a reinforced section taken from Ditlevsen [40]:

$$R = r(\mathbf{X}) = X_1 X_2 X_3 - X_4 \frac{X_1^2 X_2^2}{X_5 X_6}, \quad (47)$$

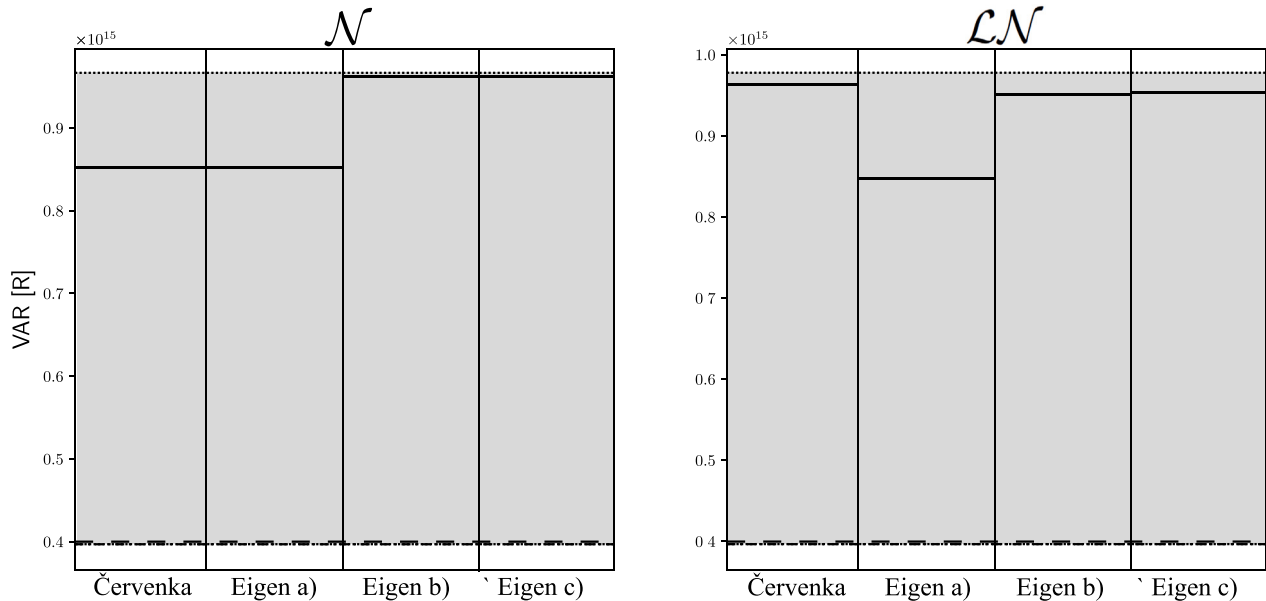


Fig. 5. Results of example 1 assuming Normal (left) and Lognormal (right) distribution of input variables. The variance of the uncorrelated case is estimated by the TSE with simple (dashed) and advanced (dot-and-dash) differencing. The variance of the correlated case is estimated by ECoV methods, which are depicted in the corresponding columns by solid lines. The reference solution (grey interval) is estimated by LHS.

Table 1

Stochastic model of the first example described by the first and the second statistical moments of input random variables.

Variable	X_1	X_2	X_3	X_4	X_5	X_6
Mean value	1260 mm ²	250 N/mm	770 mm	0.55	30 N/mm ²	250 mm
Standard deviation	63 mm ²	17.5 N/mm	10 mm	0.055	4.5 N/mm ²	5 mm

and the stochastic model contains six input random variables summarized in Table 1.

The results are compared in Fig. 5. The X-axis is divided into four columns representing the results of the ECoV methods used for the estimation of variance for fully correlated variables, and the Y-axis represents VAR [R]. Note that the grey colour corresponds to the variance interval determined by LHS (for the sake of clarity, results for intermediate ρ are not depicted). The TSE with both differencing schemes for uncorrelated random variables estimated almost identical variance, which reflects the linearity of the mathematical model. From the obtained results for fully correlated variables it is clear that ECoV by Červenka provided a very good estimate of the variance of the lognormal case, since this case fulfils both assumptions of the method: that the mathematical model is almost linear and the distribution of R is close to lognormal. In such cases, ECoV by Červenka represents the most efficient method. However, if the stochastic model contains Gaussian distribution, ECoV by Červenka fails. Note that ECoV (a) leads to inaccurate results in both cases due to the simple derivative scheme employed. However, using one more simulation ($n_{sim} = 3$) and expressions (b) and (c) of Eigen ECoV leads to accurate results that are independent of the distribution of input variables.

5.2. Example 2: Approximation of an industrial example

The second example is motivated by the industrial applications in civil engineering that are often represented by non-linear finite element models — typically the ultimate resistance given by the peak of the load-deflection curve of a concrete structural element. The behaviour of such physical system is often monotone with a slightly non-linear progression. A typical function solved by the FEM can be found for example in [24], and due to the computational demands of FEM, its

shape was replicated by the following artificial function suitable for the purposes of our tests:

$$R = r(\mathbf{X}) = X_1 X_2 - X_1^2 - \left(\frac{X_2^2}{30} \right) - (X_1 - 30)(X_2 - 200). \quad (48)$$

This function is significantly non-linear, and the stochastic model contains two input variables with the vector of mean values $\mu = [40, 300]$ and the corresponding vector $\mathbf{CoV} = [0.10, 0.15]$.

The non-linearity of the second mathematical model can be clearly seen from the difference between both TSE approximations used for the uncorrelated case. Generally, the difference between the two results is more significant with increasing non-linearity of the mathematical model, which is additionally highlighted by non-Gaussian distribution. Of course, Eigen ECoV (a) leads to a value identical to that obtained by ECoV by Červenka in Gaussian space, and a similar result is gained in lognormal space. With only one additional calculation, the results obtained by Eigen ECoV (b) and (c) are far more accurate. The superiority of these two methods is obvious from Fig. 6. In Gaussian space the results are almost exact and identical to each other, since higher central moments have negligible influence. However, in the lognormal case there is an obvious difference between both methods using identical calculations of the original mathematical model.

5.3. Example 3: Truss structure 2D NLFEA

The third example is represented by the 2D truss structure shown in Fig. 7. The ultimate load F for the allowed midspan deflection (the blue point in the figure) of $v = 10$ cm is obtained by NLFEA implemented in OpenSeesPy [41]. Uniaxial Giuffrè–Menegotto–Pinto steel material with isotropic strain hardening is used to represent all structural members. The stochastic model contains six random variables: yield strength f_y and initial elastic tangent E for the top chords (1), web members (2) and bottom chords (3), with the mean values $\mu_{f_y} = 255$ MPa, $\mu_E = 210$ GPa and CoVs: $CoV_{f_y} = 0.10$, $CoV_E = 0.05$.

The reference solution was obtained by LHS with $n_{sim} = 10^4$ simulations for each ρ and both variants (Gaussian and lognormal). The results gained by TSE and Eigen ECoV for this NLFEA are depicted in Fig. 8. Note that the function exhibits significant non-linearity, since the variance estimated by the linear TSE with simple differencing is

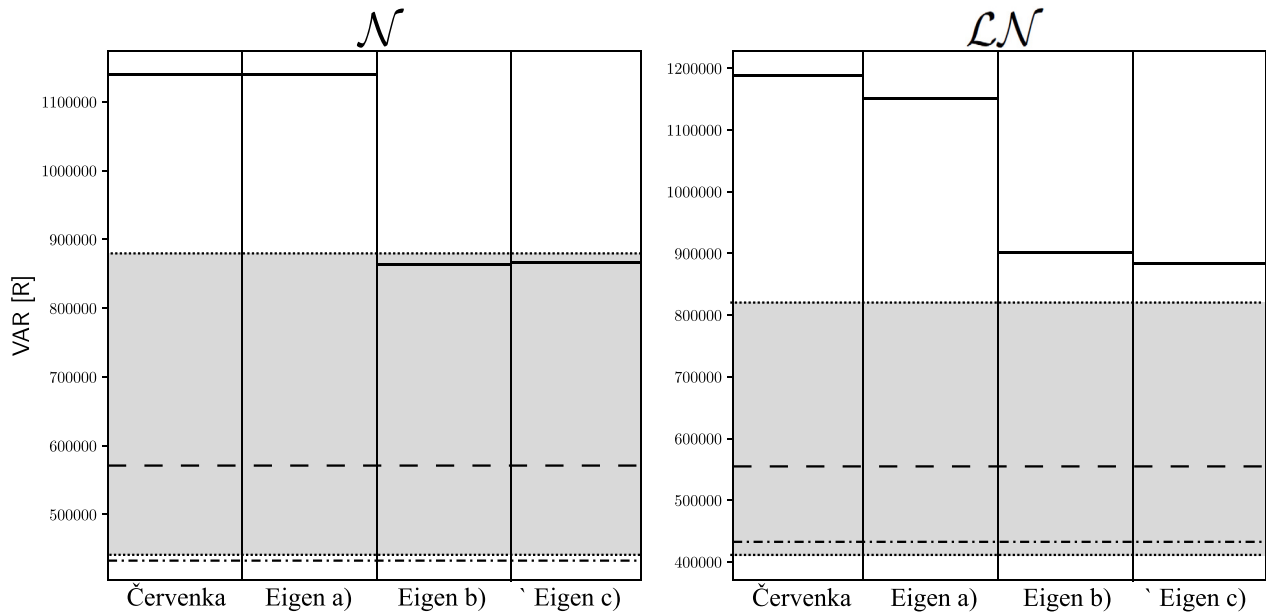


Fig. 6. Results of example 2 assuming Normal (left) and Lognormal (right) distribution of input variables. The variance of the uncorrelated case is estimated by the TSE with simple (dashed) and advanced (dot-and-dash) differencing. The variance of the correlated case is estimated by ECoV methods, which are depicted in the corresponding columns by solid lines. The reference solution (grey interval) is estimated by LHS.

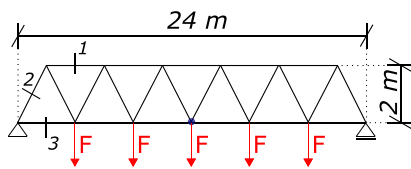


Fig. 7. Scheme of 2D truss structure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

significantly different from the reference solution and thus simple linear differencing is not able to approximate the function. On the other hand, the TSE with advanced differencing is very accurate, especially for the Gaussian variant. For the fully correlated random variables, Eigen ECoV (b) and (c) with $n_{sim} = 3$ calculations lead to very accurate results in both the Gaussian and the lognormal case. Moreover, Eigen ECoV is not limited to lognormal distribution but works generally for arbitrary distributions of response in similar manner to TSE.

6. Discussion

The proposed Eigen ECoV is derived directly from TSE, which is in compliance with PSF defined in Eurocode [7,42] and therefore can be recommended for the design and assessment of structures. Moreover, it is a general method that works without assumptions about the probability distribution of R , which stands in contrast to the widely used ECoV by Červenka implemented in Model Code 2010, which assumes Gaussian or lognormal distribution (with $v_R < 0.2$) of R . On the other hand, if these assumptions are fulfilled and the mathematical model is nearly linear, ECoV by Červenka is highly efficient.

The significant advantage of Eigen ECoV is its adaptivity using various differencing schemes. Eigen ECoV (a) represents an equivalent method to ECoV by Červenka. It is derived from the first order TSE with simple backward differencing and thus leads to identical results in Gaussian space. However, it has been shown that it is not suitable for non-linear mathematical models. Eigen ECoV (b) is derived from the first order TSE with advanced backward differencing, which leads to more accurate estimates and preserves the simplicity of the formulas for variance. Therefore, it can be easily used by civil engineers in industrial

Table 2

Comparison of estimated variance for example 2 assuming increasing uncertainty of input random variables.

CoV	LHS	Eigen ECoV b)	Eigen ECoV c)	ϵ
[0.10, 0.15]	$0.82 \cdot 10^6$	$0.90 \cdot 10^6$	$0.88 \cdot 10^6$	2%
[0.30, 0.35]	$4.63 \cdot 10^6$	$6.47 \cdot 10^6$	$5.82 \cdot 10^6$	14%
[0.40, 0.45]	$7.14 \cdot 10^6$	$11.15 \cdot 10^6$	$9.32 \cdot 10^6$	26%

applications using NLFEA. Although Eigen ECoV (c) is derived from the second order TSE, it uses identical numerical calculations of the original mathematical model to those employed by Eigen ECoV (b). It typically leads to slightly improved estimates of variance and expected values of R , taking higher moments of probability distributions of input random variables into account. However, the ECoV formula is much more complicated and should be implemented into a software application.

The difference between proposed Eigen ECoV (b) and (c) is higher with growing skewness and kurtosis of X_θ . Naturally, this plays significant role in case of lognormal distribution of X_θ with high CoV as can be clearly seen in Eq. (44). In order to amplify this difference, let us artificially increase the uncertainty of both input random variables in Example 2 as follows: $CoV = [0.30, 0.35]$ and $CoV = [0.40, 0.45]$. The estimated $VAR[R]$ assuming both input random variables lognormally distributed and fully correlated are summarized in the Table 2. Note that the percentual difference ϵ is defined as an absolute value of a difference between variance estimated by Eigen ECoV (b) and (c) divided by the reference solution estimated by LHS. We would like to note that, such high uncertainty of input variable is not common in industrial applications and thus the difference between both solutions is typically much lower.

In practical application of the correlation interval approach, one should start with Eigen ECoV for the estimation of the variance of fully correlated cases and then use standard TSE in order to obtain the variance of uncorrelated cases if necessary. The difference between variances is a direct measure of the impact of the vagueness of available information about the dependency structure among input random variables. Higher correlation among random variables typically leads to higher variance of R , and thus one can assume Eigen ECoV as a conservative estimate. However, analysts might need more accurate

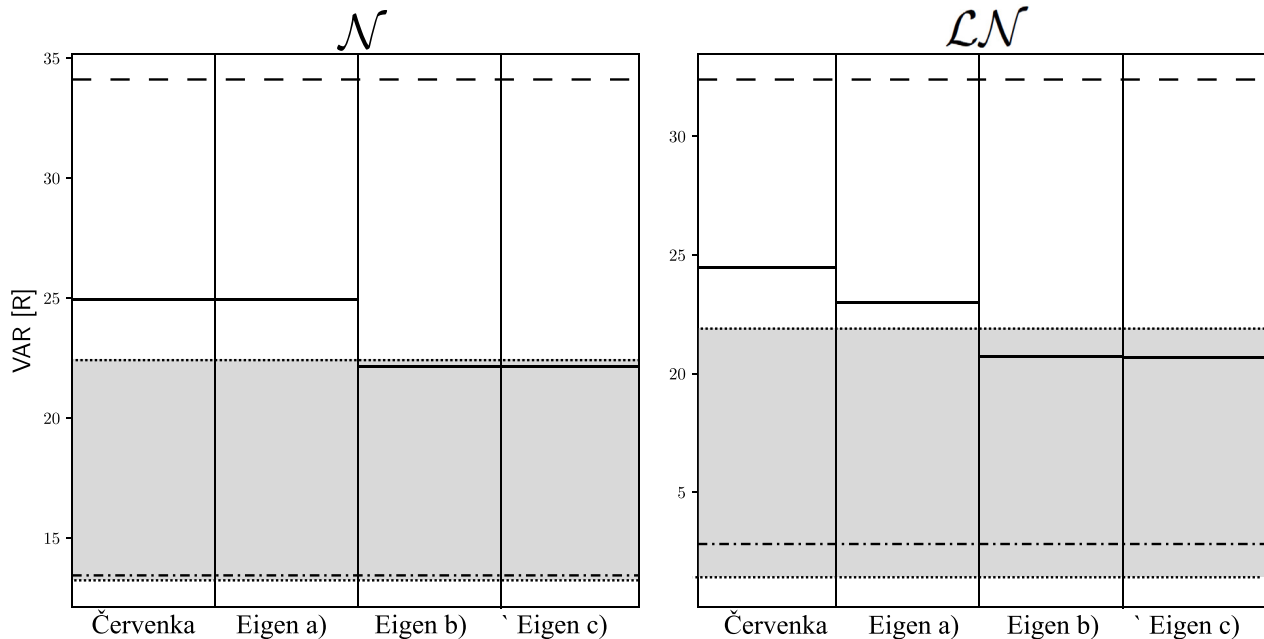


Fig. 8. Results of example 3 assuming Normal (left) and Lognormal (right) distribution of input variables. The variance of the uncorrelated case is estimated by the TSE with simple (dashed) and advanced (dot-and-dash) differencing. The variance of the correlated case is estimated by ECoV methods, which are depicted in the corresponding columns by solid lines. The reference solution (grey interval) is estimated by LHS.

results, and thus the standard TSE with simple or advanced differencing should be employed. More accurate results are especially important in case of existing structures, since the economic impact of unnecessary interventions could be significant. It naturally leads to more advanced structural analysis by NLFEA but it should be also reflected in semi-probabilistic analysis [31]. Note that for the practical design and assessment of structures, it is necessary to additionally include model uncertainty and geometrical uncertainty according to Eq. (5). The correlation interval ECoV approach then leads to minimal or maximal v_R and thus to the maximal (unsafe) design value R_d or the minimal (safe) design value R_d obtained as a corresponding quantile of the structural resistance.

7. Conclusions

The paper is focused on estimation of coefficient of variation methods for NLFEA. ECoV methods are the basis for the semi-probabilistic approach for the design and assessment of structures and thus it is crucial to use accurate and efficient methods for industrial applications. The review of existing methods and three levels of assumed simplifications is presented with attention to the theoretical mathematical background of each method. Furthermore, the influence of correlation among random variables and Nataf transformation is briefly described. Finally, the general Eigen ECoV method for functions of fully correlated random variables and the interval ECoV approach are proposed. The Eigen ECoV is analytically compared to the existing well known ECoV method by Červenka. It is shown that Eigen ECoV represents a special case of the general Taylor Series Expansion and can be directly compared to the ECoV method by Červenka in Gaussian space. However, it is a general method without any assumption regarding the probability distribution of structural resistance, which is in contrast to existing methods. The presented methods are applied for three numerical examples, and the expected behaviour of Eigen ECoV is proved. The most efficient method for industrial applications is Eigen ECoV (b) using three numerical calculations of the original mathematical model since it leads to the accurate estimation of variance, preserving the simplicity of analytical formulas that are easily applicable in industrial practice.

Acknowledgements

This research was funded by Czech Science Foundation, Czech Republic under project No. 20-01781S. The first author is Brno Ph.D. Talent Scholarship Holder - Funded by the Brno City Municipality, Czech Republic.

References

- [1] Bagge N. Demonstration and examination of a procedure for successively improved structural assessment of concrete bridges. *Structural Concrete* 2020;21(4):1321–44. <http://dx.doi.org/10.1002/suco.201900265>.
- [2] Casas JR, Wisniewski D. Safety requirements and probabilistic models of resistance in the assessment of existing railway bridges. *Structure and Infrastructure Engineering* 2013;9(6):529–45. <http://dx.doi.org/10.1080/15732479.2011.581673>.
- [3] Strauss A, Hoffmann S, Wendner R, Bergmeister K. Structural assessment and reliability analysis for existing engineering structures, applications for real structures. *Structure and Infrastructure Engineering* 2009;5(4):277–86. <http://dx.doi.org/10.1080/15732470601185638>.
- [4] Val D, Bljucer F, Yankelevsky D. Reliability evaluation in nonlinear analysis of reinforced concrete structures. *Structural Safety* 1997;19(2):203–17. [http://dx.doi.org/10.1016/S0167-4730\(96\)00025-2](http://dx.doi.org/10.1016/S0167-4730(96)00025-2).
- [5] Comité Européen de Normalisation (CEN), Brussels, Belgium, EN 1990: Eurocode: Basis Of structural design. 2002.
- [6] Hasofer AM, Lind NC. Exact and invariant second-moment code format. *Journal of Engineering Mechanics-asce* 1974;100(EM1):111–21.
- [7] Melchers RE, Beck AT. *Second-moment and transformation methods*. John Wiley & Sons, Ltd; 2017, p. 95–130 (Ch. 4).
- [8] Rosenblueth E. Point estimates for probability moments. *Proc Natl Acad Sci* 1975;72(10):3812–4. <http://dx.doi.org/10.1073/pnas.72.10.3812>.
- [9] Castaldo P, Gino D, Mancini G. Safety formats for non-linear finite element analysis of reinforced concrete structures: discussion, comparison and proposals. *Eng Struct* 2019;193:136–53. <http://dx.doi.org/10.1016/j.engstruct.2019.05.029>.
- [10] Allaix DL, Carbone VI, Mancini G. Global safety format for non-linear analysis of reinforced concrete structures. *Structural Concrete* 2013;14(1):29–42. <http://dx.doi.org/10.1002/suco.201200017>.
- [11] Allaix DL, Carbone VI, Mancini G. Global resistance factor for reinforced concrete beams. In: *Proceedings of the 5th international probabilistic workshop*. Ghent, Belgium: ACCO; 2007, p. 195–208.
- [12] Šýkora M, Červenka J, Červenka V, Mlčoch J, Novák D, Novák L. Pilot comparison of safety formats for reliability assessment of rc structures. In: *Proceedings of the fib symposium 2019: concrete - Innovations in materials, design and structures*. 2019. p. 2076–83.

- [13] Novák L, Novák D, Pukl R. Probabilistic and semi-probabilistic design of large concrete beams failing in shear. In: *Advances in engineering materials, structures and systems: innovations, mechanics and applications*. Taylor and Francis Group CRC Press; 2019.
- [14] Novák D, Novák L, Slowik O, Strauss A. Prestressed concrete roof girders: Part iii – semi-probabilistic design. In: *Proceedings of the sixth international symposium on life-cycle civil engineering*. CRC Press, Taylor and Francis Group; 2018, p. 510–7.
- [15] Fib federation internationale du beton. In: *Fib model code for concrete structures 2010*. Berlin, Heidelberg: John Wiley & Sons; 2013.
- [16] Castaldo P, Gino D, Bertagnoli G, Mancini G. Partial safety factor for resistance model uncertainties in 2d non-linear finite element analysis of reinforced concrete structures. *Eng Struct* 2018;176:746–62. <http://dx.doi.org/10.1016/j.engstruct.2018.09.041>.
- [17] Gino D, Castaldo P, Bertagnoli G, Giordano L, Mancini G. Partial factor methods for existing structures according to fib bulletin 80: Assessment of an existing prestressed concrete bridge. *Structural Concrete* 2020;21(1):15–31. <http://dx.doi.org/10.1002/suco.201900231>.
- [18] Comité Européen de Normalisation (CEN), Brussels, Belgium, EN 1992: Eurocode 2: Design of concrete structures. 2004.
- [19] Engen M, Hendriks MA, Kohler J, Øverli JA, Åldstedt E. A quantification of the modelling uncertainty of non-linear finite element analyses of large concrete structures. *Structural Safety* 2017;64:1–8. <http://dx.doi.org/10.1016/j.strusafe.2016.08.003>.
- [20] Holický M, Retief JV, Sýkora M. Assessment of model uncertainties for structural resistance. *Probabilistic Engineering Mechanics* 2016;45:188–97. <http://dx.doi.org/10.1016/j.probengmech.2015.09.008>.
- [21] Christian JT, Baecher GB. Point-estimate method as numerical quadrature. *Journal of Geotechnical and Geoenvironmental Engineering* 1999;125(9):779–86. [http://dx.doi.org/10.1061/\(ASCE\)1090-0241\(1999\)125:9\(779\)](http://dx.doi.org/10.1061/(ASCE)1090-0241(1999)125:9(779)).
- [22] Červenka V. Global safety format for nonlinear calculation of reinforced concrete. *Beton- und Stahlbetonbau* 2008;103(S1):37–42. <http://dx.doi.org/10.1002/best.200810117>.
- [23] Červenka V. Reliability-based non-linear analysis according to fib model code 2010. *Structural Concrete* 2013;14(1):19–28. <http://dx.doi.org/10.1002/suco.201200022>.
- [24] Schlune H, Plos M, Gylltoft K. Safety formats for nonlinear analysis tested on concrete beams subjected to shear forces and bending moments. *Eng Struct* 2011;33(8):2350–6.
- [25] Novák L, Novák D. On Taylor series expansion for statistical moments of functions of correlated random variables. *Symmetry* 2020;12:1379. <http://dx.doi.org/10.3390/sym12081379>.
- [26] McKay MD. Latin hypercube sampling as a tool in uncertainty analysis of computer models. In: *Proceedings of the 24th conference on winter simulation*. New York, NY, USA: 1992. p. 557–64.
- [27] Iman RL, Conover W. Small sample sensitivity analysis techniques for computer models. with an application to risk assessment. *Comm Statist Theory Methods* 1980;9(17):1749–842.
- [28] Lebrun R, Dutfoy A. A generalization of the Nataf transformation to distribution with elliptical copula. *Probabilistic Engineering Mechanics* 2009;24(2):172–8. <http://dx.doi.org/10.1016/j.probengmech.2008.05.001>.
- [29] Vořechovský M, Eliáš J. Modification of the maximin and ϕ_p (phi) criteria to achieve statistically uniform distribution of sampling points. *Technometrics* 2020;62(3):371–86. <http://dx.doi.org/10.1080/00401706.2019.1639550>.
- [30] Vořechovský M, Novák D. Correlation control in small-sample Monte Carlo type simulations i: A simulated annealing approach. *Probabilistic Engineering Mechanics* 2009;24(3):452–62. <http://dx.doi.org/10.1016/j.probengmech.2009.01.004>.
- [31] Pimentel M, Brühwiler E, Figueiras J. Safety examination of existing concrete structures using the global resistance safety factor concept. *Eng Struct* 2014;70:130–43. <http://dx.doi.org/10.1016/j.engstruct.2014.04.005>.
- [32] Zimmermann T, Lehký D, Strauss A. Correlation among selected fracture-mechanical parameters of concrete obtained from experiments and inverse analyses. *Structural Concrete* 2016;17(6):1094–103. <http://dx.doi.org/10.1002/suco.201500147>.
- [33] Rosenblatt M. Remarks on a multivariate transformation. *Ann Math Stat* 1952;23(3):470–2.
- [34] Kiureghian AD, Liu P. Structural reliability under incomplete probability information. *J Eng Mech* 1986;112(1):85–104. [http://dx.doi.org/10.1061/\(ASCE\)0733-9399\(1986\)112:1\(85\)](http://dx.doi.org/10.1061/(ASCE)0733-9399(1986)112:1(85)).
- [35] Lebrun R, Dutfoy A. An innovating analysis of the Nataf transformation from the copula viewpoint. *Probabilistic Engineering Mechanics* 2009;24(3):312–20. <http://dx.doi.org/10.1016/j.probengmech.2008.08.001>.
- [36] Torre E, Marelli S, Embrechts P, Sudret B. A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas. *Probabilistic Engineering Mechanics* 2019;55:1–16. <http://dx.doi.org/10.1016/j.probengmech.2018.08.001>.
- [37] Lebrun R, Dutfoy A. Do Rosenblatt and Nataf isoprobabilistic transformations really differ? *Probabilistic Engineering Mechanics* 2009;24(4):577–84. <http://dx.doi.org/10.1016/j.probengmech.2009.04.006>.
- [38] Nataf A. Determination des distributions de probabilités dont les marges sont données. *C R Acad Sci* 1962;225:42–3.
- [39] Liu P-L, Der Kiureghian A. Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics* 1986;1(2):105–12. [http://dx.doi.org/10.1016/0266-8920\(86\)90033-0](http://dx.doi.org/10.1016/0266-8920(86)90033-0).
- [40] Ditlevsen O. *Uncertainty modeling with applications to multidimensional civil engineering systems*. Advanced Book Program, McGraw-Hill International Book Company; 1981.
- [41] Zhu M, McKenna F, Scott MH. *Openseespy: Python library for the opensees finite element framework*. SoftwareX 2018;7:6–11.
- [42] Cornell CA. A probability based structural code. *J Amer Concr Inst* 1969;66(12):974–85.

2.2 Comparison of Advanced Semi-probabilistic Methods for Design and Assessment of Concrete Structures

NOVÁK, L.; ČERVENKA, J.; ČERVENKA, V.; NOVÁK, D.; SÝKORA, M. Comparison of advanced semi-probabilistic methods for design and assessment of concrete structures. *Structural Concrete*, 2022, vol. 1, no. 4, ISSN: 1751-7648. (WoS-AIS: Q2)

DOI: 10.1002/suco.202200179

Description

The paper presents a thorough comparison of existing semi-probabilistic methods and safety formats for the design and assessment of concrete structures represented by costly mathematical models solved by NLFEM. A special attention is given to the recently proposed Eigen ECoV method and the correlation interval approach. Numerical examples represent carefully selected structural elements failing in different modes. From the obtained numerical results, it can be concluded that Eigen ECoV leads to superior results, though its benefit for an almost linear function (bending failure) is negligible. Furthermore, it is shown that the correlation interval can be efficiently estimated by ECoV methods and TSE, and it clearly shows the impact of the assumed correlation among input variables. The final part of the paper shows some preliminary results for physical systems with multiple failure modes. Multiple failure modes could lead to multimodal probability distribution of QoI, and thus the existing safety formats assuming Lognormal distributions fail. This topic should be an object of a future investigation, since it might play a crucial role in the analysis of complex structural systems by NLFEM.






Role of the author

Percentage of contribution: 40%

Lukáš Novák is the main author of this paper responsible for the concept, the methodology, and the numerical results of the presented research. Non-linear finite element models were created by Jan Červenka and Vladimír Červenka. Drahomír Novák and Miroslav Sýkora prepared the methodology and theoretical descriptions of the compared methods in cooperation with Lukáš Novák.

ARTICLE

Comparison of advanced semi-probabilistic methods for design and assessment of concrete structures

Lukáš Novák¹  | Jan Červenka²  | Vladimír Červenka²  |
 Drahomír Novák¹  | Miroslav Sýkora³ 

¹Faculty of Civil Engineering, Brno University of Technology, Brno, Czech Republic

²Červenka Consulting, Prague, Czech Republic

³Klokner Institute, Czech Technical University in Prague, Prague, Czech Republic

Correspondence

Lukáš Novák, Faculty of Civil Engineering, Brno University of Technology, Veveří 331/95, 602 00 Brno, Czech Republic.
 Email: novak.l@fce.vutbr.cz

Funding information

Grantová Agentura České Republiky; Czech Science Foundation, Grant/Award Number: 2001781S

Abstract

This paper presents the comparison of advanced semi-probabilistic methods for the design and assessment of concrete structures represented by mathematical models solved by non-linear finite element methods. The special attention is given to the advanced methods focused on the estimation of the coefficient of variation of structural resistance. Numerical examples represent a replication of laboratory experiments of beams with different failure modes. The obtained results are discussed with respect to the accuracy of the employed methods and the influence of the assumed statistical correlation among basic variables. Simplified methods give a good estimation of the design values, though their accuracy is dependent on the type of the failure mechanism. Moreover, it is shown that mutual correlations among random variables may significantly affect the design value of resistance, and they should be carefully defined and modeled.

KEYWORDS

non-linear finite element method, reliability of concrete structures, safety formats, semi-probabilistic methods, statistical analysis, Taylor series expansion

Abbreviations: R , structural resistance; E , action effect; F^{-1} , inverse cumulative distribution function; α , sensitivity factor; β , target reliability index; R_d , design value of resistance; γ_M , partial safety factor for a material property including model uncertainty; γ_m , partial safety factor for a material property without model uncertainty; γ_{R_d} , model uncertainty in γ_M ; v_X , CoV of resistance variable X in PSF; μ_X , mean of resistance variable X in PSF; d , effective depth; f_y , yield strength of reinforcement; f_c , compressive strength of concrete; A_c , concrete area; η , conversion factor; θ , model uncertainty; f_{ct} , tensile strength; G_f , fracture energy; R_m , simulation with mean values of basic variables; R_k , simulation with characteristic values; X_k , characteristic values (5% quantile); X_m , mean values of basic variables; X_k^* , characteristic values including uncertainty according to PSF; $r(\mathbf{X})$, the function of structural resistance; \mathbf{X} , input random vector of N basic variables; μ , mean value; v , coefficient of variation; c , step size parameter; Θ , 1D eigen distribution; σ^2 , variance; λ_1 , the first eigenvalue of input covariance matrix; $X_{i,d}$, reduced basic variable for Eigen ECoV; ρ , reinforcement ratio; Δ_θ , distance between μ_θ and desired quantile $F_\theta^{-1}(\Phi(-c))$; Φ , cumulative distribution function of the standardized normal distribution.

Discussion on this paper must be submitted within two months of the print publication. The discussion will then be published in print, along with the authors' closure, if any, approximately nine months after the print publication.

1 | INTRODUCTION

The determination of a design value of resistance ensuring a target level of structural reliability represents a key task for engineers. Design codes like Eurocodes offer a clear background for that, making it possible to separate structural resistance and action effect in most design situations. A relative importance of resistance and of the load effect is expressed through the respective sensitivity factors whose recommended values commonly ensure reasonably conservative design solutions. As structural analysis can then be fully focused on resistance, the approach is called semi-probabilistic.

Various procedures to determine the design value of resistance are generally termed as safety formats. They are elaborated at different levels of simplification and accuracy. Safety formats according to Eurocodes fully rely on the partial factor method; also the EN 1992-2:2005¹ approach belongs to this category. The design value of resistance is obtained by one calculation of the computational structural model using the design values of basic variables. While this approach performs well and is fully justified for linear computational models, its use for non-linear models is questionable and can lead to an over-conservative design resistance^{2,3} or may even fail.⁴ In contrast, it is well known that the probabilistic analysis in combination with a non-linear finite element analysis offers a significant added value over the standard linear analysis and semi-probabilistic approach implemented in Eurocodes,⁵ and researchers are highly motivated to develop novel techniques coupling the accuracy of the non-linear finite element analysis (NLFEA) with a realistic description of the basic variables by probabilistic modeling.

The only general tool for probabilistic analysis is represented by the Monte Carlo simulation (MC), simulating uncertainties with their complete probability distribution and statistical correlation. For a large number of simulations, the approach leads to the complete information about the distribution of resistance, but the number of simulations is often limited, and the design value of resistance is estimated based on the estimates of the mean value and coefficient of variation (CoV). The accuracy depends on the quality of these estimates. Even if the advanced stratified sampling such as Latin Hypercube Sampling (LHS)⁶ is used, a number of simulations may range from tens to hundreds. There are more advanced and computationally demanding MC techniques,⁷ but their implementation is usually far more complicated, and thus they are typically utilized for scientific applications only. The computational burden of MC represents the main obstacle of the approach for time-consuming mathematical models like NLFEA, since

it is not computationally feasible for industrial applications. Significant efforts have been made to reduce the computational cost of the estimation of statistical moments and its dependency on number of input random variables. Recently, promising results were obtained by high-dimensional model representation method⁸ and its later modifications, such as the Maximum Entropy Multiplicative Dimensional Reduction Method.⁹ Although these methods represent a significant improvement for general estimating of statistical moments, it is still necessary to perform tens of numerical simulations.

That is why alternative techniques focused on the Estimation of Coefficient of Variation (ECoV) of structural resistance have been developed. They represent a compromise between the simple and, in most cases, the conservative approach of partial factors and MC. They consider uncertainties in the form of N basic (input) random variables, but under several simplifying assumptions, they reduce the computational model calculations to a very low number acceptable in practice:

- ECoV according to *fib* Model Code 2010¹⁰—2 numerical calculations for any N ,
- Eigen ECoV¹¹—3 numerical calculations for any N ,
- ECoV based on Taylor Series Expansion^{12,13}— $N + 1$ or $2N + 1$ numerical calculations, i.e. only three calculations when the concrete compressive strength and the yield strength of the reinforcement are modeled as stochastic variables for the whole structure, but an excessively increasing number when random material properties are modeled at various locations at the structure, or when the stochastic model contains additional random parameters.

The ECoV methods commonly simplify an estimation of the mean value of resistance as a result of the calculation of the computational model using the mean values of input variables. This assumption is strong (particularly for highly nonlinear models), though it can be accepted in many applications where ECoV methods achieve sufficient accuracy.^{14–17} The second strong simplification common for all ECoV methods is assuming a lognormal distribution of resistance.

Note that LHS can be used as an ECoV technique for the estimation of the mean value and CoV.¹⁷ In particular, for a low number of simulations (tens), the assumption of lognormal distribution is necessary, as a reliable estimation of distribution of resistance generally requires a higher number of simulations. A similar situation applies to the methods based on numerical quadrature,¹⁸ however, they are extremely computationally expensive for increasing N , and thus are rarely employed in industry.

This paper presents a summary of the available approaches and the comparison of the semi-probabilistic methods for practical examples of NLFEA. In contrast to

the review of ECoV methods,¹⁹ a special attention is given to the verification of the recently proposed adapted Taylor series expansion (TSE), and a special case of TSE referenced as Eigen ECoV, which does not bring a significant additional computational burden, but extends the range of applicability of ECoV according to *fib* Model Code 2010.

The aim of the paper is to contribute to the discussion and clarify the recommendations provided by the nearly complete drafts of *fib* MC 2020 and prEN 1992-1-1.²⁰ The general theoretical background of the semi-probabilistic approach, PSF, and selected ECoV methods are described in Section 2. Note that there is a strong reasoning for the selection of these particular simplified ECoV methods, since they are based on the general formulation of TSE. The purpose of this section is thus not only to summarize the selected methods, but also to clarify their similarities and assumed simplifications. Besides the simplified ECoV methods and PSF, a brief summary of stratified sampling is also presented, since we use this method for a reference solution of the numerical examples. Section 3 describes selected case studies and the methodology of the numerical examples, i.e. our strategy for a proper comparison of the selected methods. Besides the comparison of the ECoV methods, we also present their theoretical characteristics regarding the correlation among input random variables. The case studies are represented by three mathematical models replicating laboratory experiments from the literature. Each model exhibits a different failure mode, and thus a different influence of correlation among material parameters. Section 4 contains an extended discussion of the obtained results reflecting theoretical assumptions of the selected methods and their limitations. Moreover, in the second part of the discussion, we present an artificial analytical example presenting some of these limitations in the case of models with multiple failure modes. In the last section, Section 5, the main findings of this paper are summarized and the importance of the correlation among the material characteristics of concrete is emphasized.

2 | SEMI-PROBABILISTIC APPROACH

Structural reliability represents a crucial topic of civil engineering globally implemented into the design codes using semi-probabilistic approaches. The semi-probabilistic approaches assume the separation of two random variables, structural resistance R and action effect E , through their design values:

$$R_d = F_R^{-1}(-\alpha_R \beta), \quad (1)$$

and

$$E_d = F_E^{-1}(-\alpha_E \beta), \quad (2)$$

where F^{-1} represents the inverse cumulative distribution function, α is a sensitivity factor originally derived from First Order Reliability Method (FORM), and β is the target reliability index. The paper is focused on the estimation of R_d when the function of structural resistance $r(\mathbf{X})$ of input random vector (\mathbf{X} being a vector of N basic variables) is solved by an NLFEA. The recommended value of $\alpha_R = 0.8$ is then utilized, typically with a lognormal distribution of R . Based on these assumptions, the probability distribution is fully described by the mean value and CoV, and the reliability analysis reduces to the estimation of the first two statistical moments—the task of the ECoV methods.

2.1 | Partial safety factors

Although NLFEA has been employed for the design and assessment of structures more frequently in recent decades, it is still insufficiently included in Eurocodes, and its potential for a wide application in the industry is thus limited. Specifically, there is the Partial Safety Factors (PSF) method, and the global factor method for NLFEA of concrete structures according to EN 1992-2:2005¹ implemented in Eurocode. Unfortunately, both methods may provide only crude estimates in the cases with a strongly non-linear behavior, multiple failure modes, or when the assumptions adopted by these approaches do not apply.

Although the PSF implemented into Eurocodes was originally not intended for NLFEA applications, it is often employed due to its simplicity. To estimate the design value of resistance R_d in Equation (3), only one calculation must be computed with the design values of basic variables. The design values of input variables are typically derived from the characteristic values using partial factors $\gamma_M = \gamma_{R_d} \gamma_m$, which reflect the uncertainties in the material and the geometrical properties, and model uncertainties:

$$R_d = r(X_{1,k}/\gamma_M, X_{2,k}/\gamma_M, \dots) = r(f_{ck}/\gamma_C, f_{yk}/\gamma_S, \dots). \quad (3)$$

where X_k denotes a characteristic value of the basic variable X , γ_M is the partial factor for a material property, f_c is the concrete compressive strength, and f_y is the yield strength of steel reinforcement.

Note that the design values used in the PSF method might be very low, which might lead to unrealistic

Parameter	CoV	Bias factor
Yield str. f_y	$v_{f_y} = 0.045$	$f_{y,m}/f_{y,k} = \exp(1.645v_{f_y})$
Model unc. ^a	$v_\theta = 0.045$	$\mu_\theta = 1.09$
Effect. depth d^b	$v_d = 0.05$	$\mu_d = 0.95$
Resistance characteristics	$v_{f_{y,R}} = \sqrt{0.045^2 + 0.05^2} = 0.067$	$\mu_{f_{y,R}} = f_{y,m}/f_{y,k} \times \mu_d = 1.02$

Abbreviation: CoV, coefficient of variation.

^aModel uncertainties are reflected in this study by γ_{R_d} and are determined separately for each numerical example.

^bValid for $d = 200$ mm. For other effective depths: $v_d = 0.05 (200/d)^{2/3}$ and $\mu_d = 1 - 0.05 (\frac{200}{d})^{2/3}$.

results, since non-linear material models are often calibrated to specific ranges of input values only (closer to their mean and/or characteristic values). Therefore, it is beneficial to calibrate the partial factors with respect to real laboratory experiments involving material and structural measurements.²¹

Another approach for the derivation of PSF with the explicit definition of the model, material, and geometrical uncertainties was recently introduced in Annex A to prEN 1992-1-1:2021.²⁰ The main idea is to account for the biases and CoVs of various basic variables directly in the model of resistance, not only the material itself. Such an approach leads to a considerable simplification of the modeling of the geometrical uncertainties, which may be a difficult task in NLFEA. Disregarding now the model uncertainties—treated separately later by a model uncertainty factor γ_{R_d} specific to the case under consideration—allows for an unambiguous comparison of all the presented safety formats and semi-probabilistic methods (Section 3). The general formula for the definition of PSF according to the new Eurocode proposal is then:

$$\gamma_c = \frac{\exp(\alpha_R \beta v_{R_c})}{\mu_{R_c}} = \frac{\exp(\alpha_R \beta \sqrt{v_{f_c}^2 + v_\eta^2 + v_{A_c}^2})}{\frac{f_{cm}}{f_{ck}} \mu_\eta \mu_{A_c}} \quad (4)$$

$$\gamma_s = \frac{\exp(\alpha_R \beta v_{R_s})}{\mu_{R_s}} = \frac{\exp(\alpha_R \beta \sqrt{v_{f_y}^2 + v_d^2})}{\frac{f_{ym}}{f_{yk}} \mu_d}$$

where X_m and v_X denote the mean and the coefficient of variation of the basic variable X , R_c and R_s are model resistances related to concrete crushing and yielding of reinforcement respectively, and μ in Equation (4) is a bias in the basic variable X —the systematic deviation of random values of the variable from its characteristic (nominal) value, expressed as the ratio of the mean to the characteristic (nominal) value.

Equation (4) assumes that concrete resistance R_c , typically governing the resistance of non-slender columns,

and reinforcement resistance R_s , typically governing the flexural resistance, are random variables that are lognormally distributed, obtained as a linear product of the relevant resistance parameters; see Tables 1 and 2 taken from Annex A to prEN 1992-1-1:2021.²⁰

In the case of a bending failure governed by reinforcement, geometrical uncertainties relate to the most important parameter—the effective depth d as described in Table 1, along with a relevant material property and model uncertainties. In the case of compressive failure with dominating concrete strength, the CoV of resistance is similarly affected by a geometrical uncertainty through the concrete area A_c , but the uncertainty of in-situ strength is also additionally affected by the conversion factor η (see Table 2).

Non-linear material models of concrete typically consider additional material characteristics, such as tensile strength f_{ct} , and fracture energy G_f of concrete. An identical philosophy as in the case of compressive strength was adopted in order to derive the statistics for the PSF and ECoV methods. Specifically, the values according to Table 3 are taken into account.

The CoV of f_{ct} is set to $v = 0.18$ in compliance with prEN 1992-1-1.²⁰ Note that the variability of G_f is assumed to be identical as for f_{ct} . The characteristic value of tensile strength is obtained from compressive strength according to prEN 1992-1-1²⁰ as [in MPa]:

$$f_{ct,k} = 0.7 \left(0.3 f_{c,k}^{2/3} \right) \quad (5)$$

and fracture energy according to the 2021 draft of Model Code 2020 as follows [in MPa]:

$$G_{f,k} = 85 f_{c,k}^{0.15} \quad (6)$$

Note that values from Tables 1–3 are further utilized for derivation of mean and characteristic values used in advanced semi-probabilistic methods as described in the next subsection.

TABLE 1 Parameters assumed for the derivation of a partial factor for reinforcement

TABLE 2 Parameters assumed for the derivation of a partial factor for concrete

Parameter	CoV	Bias factor
Compr. str. f_c	$v_{f_c} = 0.1$	$f_{c,m}/f_{c,k} = \exp(1.645v_{f_c})$
Conversion fact. η	$v_\eta = 0.12$	$\mu_\eta = 0.95$
Conc. area A_c	$v_{A_c} = 0.04$	$\mu_{A_c} = 1.0$
Model unc. ^a	$v_\theta = 0.06$	$\mu_\theta = 1.02$
Resistance characteristics	$v_{f_{c,R}} = \sqrt{0.1^2 + 0.12^2 + 0.04^2} = 0.16$	$\mu_{f_{c,R}} = 1.18 \times 0.95 \times 1.0 = 1.12$

Abbreviation: CoV, coefficient of variation.

^aModel uncertainties are reflected by γ_{R_d} determined separately for each numerical example.

TABLE 3 Additional parameters assumed for the derivation of a partial factor for concrete

Parameter	CoV	Bias factor
f_{ct}, G_f	$v = 0.18$	$f_m/f_k = \exp(1.645v)$
Conversion fact. η	$v_\eta = 0.12$	$\mu_\eta = 0.95$
Conc. area A_c	$v_{A_c} = 0.04$	$\mu_{A_c} = 1.0$
Model unc. ^a	$v_\theta = 0.06$	$\mu_\theta = 1.02$
Resistance characteristics	$v_{X,R} = \sqrt{0.18^2 + 0.12^2 + 0.04^2} = 0.22$	$\mu_{X,R} = 1.34 \times 0.9 \times 1.0 = 1.28$

Abbreviation: CoV, coefficient of variation.

^aModel uncertainties are reflected by γ_{R_d} determined separately for each numerical example.

2.2 | Simplified methods for estimation of coefficient of variation

Besides PSF, there are alternative methods published in scientific papers and international documents such as *fib* Model Code 2010.²² The ECoV methods under consideration were developed in order to effectively estimate the first two statistical moments of function of random variables from simple formulas. Simplified ECoV methods are often applied in practical design and the assessment of structures without the knowledge of their theoretical background. However, it is essential to respect their limitations to avoid making crude estimates of the design resistance.

2.2.1 | ECoV according to *fib* model code

Probably the most frequently used method is the one developed by Červenka⁷ and implemented into *fib* Model Code 2010.²² It is based on a simplified formula for the estimation of a characteristic value corresponding to a lognormal variable with the mean value μ_R and CoV of the model resistance v_R . Based on two numerical simulations—one with the mean values of the basic variables, R_m , and the other using the characteristic values (5% fractile for material parameters) of basic variables, R_k —the following simplified formula was derived:

$$v_R = \frac{1}{1.645} \ln \left(\frac{R_m}{R_k} \right) \quad (7)$$

Based on the conventional models for basic variables provided in Tables 1–3, one can derive the mean and characteristic values summarized in Table 4. Note that the characteristic values X_k^* in Table 4 reflect the uncertainty in the basic variables assumed for the derivation of PSF.

Since there are only two numerical calculations used in Equation (7), it can be shown that ECoV according to *fib* Model Code 2010 implicitly assumes a full correlation among basic variables (including f_c, f_y or geometrical parameters).¹¹ Moreover, the simplified Equation (7) for the fractile of a lognormal distribution should be applied for a low CoV only. According to prEN 1990:2021²³ and prEN 1992-1-1:2021,²⁰ this approximation may be used for a coefficient of variation of less than 0.2; the exact formula for the fractile provided in prEN 1990:2021²³ leads to lower values of v_R .

2.2.2 | ECoV based on Taylor series expansion

The standard method for a statistical analysis of functions of random variables is the Taylor Series Expansion (TSE). The most significant advantages of ECoV based on TSE are its versatility and adaptability. TSE is generally an infinite series which must be truncated to a finite number of terms considering the non-linearity of the

Parameter	Mean value	Characteristic value
Yield strength (Table 1)	$f_{y,m} = \mu_{f_{y,R}} \times f_{y,k}$	$f_{y,k}^* = f_{y,m} \times \exp(-1.645 v_{f_{y,R}})$
Compressive strength (Table 2)	$f_{c,m} = \mu_{f_{c,R}} \times f_{c,k}$	$f_{c,k}^* = f_{c,m} \times \exp(-1.645 v_{f_{c,R}})$
Tensile strength (Table 3)	$f_{ct,m} = \mu_{f_{ct,R}} \times 0.7 \times (0.3 f_{c,k}^{2/3})$	$f_{ct,k}^* = f_{ct,m} \times \exp(-1.645 v_{f_{ct,R}})$
Fracture energy (Table 3)	$G_{f,m} = \mu_{G_{f,R}} \times 85 \times f_{c,k}^{0.15}$	$f_{c,k}^* = G_{f,m} \times \exp(-1.645 v_{G_{f,R}})$

TABLE 4 Input random variables and the defined values for safety formats and ECoV methods

Abbreviation: ECoV, estimation of coefficient of variation.

approximated function. In engineering applications, it is common to use TSE truncated to linear terms, and thus with $\mu_R \approx R_m$ and CoV:

$$v_R \approx \frac{1}{R_m} \sqrt{\sum_{i=1}^N \left(\frac{\partial r(X)}{\partial X_i} \sigma_{X_i} \right)^2}, \quad (8)$$

where the partial derivatives can be numerically computed by various differencing schemes.¹³ The simplest scheme is one-sided backward differencing:

$$\frac{\partial r(X)}{\partial X_i} = \frac{R_m - R_{X_i\Delta}}{\Delta X_i}. \quad (9)$$

$R_{X_i\Delta}$ is the result of a numerical simulation using the mean values of all random variables except the value of the i -th random variable, which is reduced by ΔX_i . Naturally, one can derive various differencing schemes adapted for different situations. Specifically, for ECoV based on TSE, the methodology based on linear and quadratic TSE was recently proposed, providing for three levels of complexity and accuracy.¹³ The balance between efficiency and accuracy is achieved by the second level based on linear TSE and the following advanced differencing scheme:

$$\frac{\partial r(X)}{\partial X_i} = \frac{3R_m - 4R_{X_{i\frac{\Delta}{2}}} + R_{X_i\Delta}}{\Delta X_i}, \quad (10)$$

where the middle term $R_{X_{i\frac{\Delta}{2}}}$ is obtained by evaluating the mathematical model with a reduced i -th variable $X_{i\frac{\Delta}{2}} = \mu_{X_i} - \Delta X_i/2$ and with the mean values of all the other variables.

The adaptivity of TSE is enhanced by introducing a step size parameter c used for defining the difference $\Delta X_i = \mu_{X_i} - X_{i\Delta}$, where $X_{i\Delta} = F_i^{-1}(\Phi(-c))$. F_i^{-1} is an inverse cumulative distribution function of the i -th variable and Φ is the cumulative distribution function of the standardized normal distribution. Schlune et al.¹² proposed to consider $c = (\alpha_R \beta) / \sqrt{2}$. Occasionally, it brings

additional computational burden when analyzing different limit states with different β , since it is necessary to calculate $N + 1$ (Equation (9)) or $2N + 1$ (Equation (10)) simulations for each limit state. It might be recommended to use $c = 1.645$ irrespective of the type of the investigated limit state, which is in accordance with the ECoV according to *fib* Model Code 2010.

2.2.3 | Eigen ECoV

The recently proposed Eigen ECoV¹¹ is derived directly from TSE. However, in contrast to TSE suitable for arbitrary correlation structures, Eigen ECoV assumes fully correlated input random variables similarly to ECoV according to *fib* Model Code 2010. Therefore, the number of simulations is significantly reduced in comparison to TSE. The reduction of the number of simulations is achieved by the projection of the differencing scheme into the fully correlated space, i.e. Eigen ECoV is based on the idea of projecting the input random vector on 1D eigen distribution Θ with the variance equal to the first eigenvalue of input covariance matrix $\sigma_\Theta^2 = \sum \sigma_{X_i}^2 = \lambda_1$, and the mean value is simply obtained as:

$$\mu_\Theta = \sqrt{\sum_{i=1}^N (\mu_{X_i})^2}. \quad (11)$$

In the original proposal, there are three levels of Eigen ECoV.¹¹ The most promising Eigen ECoV formula for the estimation of v_R offering a balance between the efficiency and accuracy (derived directly from Equation (10)) is:

$$v_R \approx \frac{3R_m - 4R_{\Theta\frac{\Delta}{2}} + R_{\Theta\Delta}}{\Delta_\Theta} \cdot \frac{\sqrt{\lambda_1}}{R_m}, \quad (12)$$

where the simulation $R_{\Theta\Delta} = r(X_{\Theta\Delta})$ with the coordinates of the input realization $X_{\Theta\Delta} = (X_{1\Delta}, \dots, X_{N\Delta})$ and $R_{\Theta\frac{\Delta}{2}} = r(X_{\Theta\frac{\Delta}{2}})$ with the coordinates $X_{\Theta\frac{\Delta}{2}} = (X_{1\frac{\Delta}{2}}, \dots, X_{N\frac{\Delta}{2}})$ are depicted together with an illustration of the Eigen ECoV method in Figure 1.

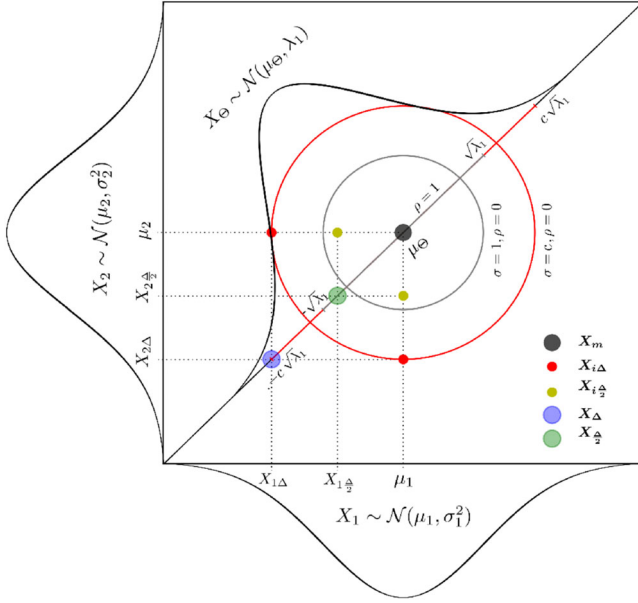


FIGURE 1 Illustration of Eigen ECoV in standardized normal space with coordinates of realizations of input random vector.¹¹ ECoV, estimation of coefficient of variation

For the sake of clarity, the input vectors consisting of reduced values of input random variables are $X_{i\Delta} = F_i^{-1}(\Phi(-c))$, and the intermediate coordinates are as follows:

$$X_{i\frac{\Delta}{2}} = \mu_{X_i} - \frac{\mu_{X_i} - X_{i\Delta}}{2} = \frac{\mu_{X_i} + X_{i\Delta}}{2}. \quad (13)$$

Δ_θ represents the distance between μ_θ and the desired quantile $F_\theta^{-1}(\Phi(-c))$ obtained as:

$$\Delta_\theta = \mu_\theta - \mu_\theta \cdot \exp\left(-c \cdot \frac{\sqrt{\lambda_1}}{\mu_\theta}\right). \quad (14)$$

The Eigen ECoV combines the versatility and adaptability of TSE through various differencing schemes and the step size parameter c , together with the efficiency of ECoV according to *fib* Model Code 2010. Note that more theoretical details can be found in the original proposal of Eigen ECoV including additional formulas based on other differencing schemes or higher TSE, which is suitable for input variables with high CoV.¹¹ Similarly, as for ECoV according to Model Code 2010, $c = 1.645$ in numerical examples, which leads to $X_{i\Delta} = X_{i,k}^*$ and $\mu_{X_i} = X_{i,m}$ as summarized later in Table 4.

2.3 | Stratified sampling for estimation of coefficient of variation

The standard approach to the statistical analysis of complex functions of random input variables is the MC

simulation consisting of a large number of repetitive deterministic calculations with randomly generated realizations of the input random vector. In order to improve the efficiency of the crude MC method, a stratified sampling (Latin Hypercube Sampling, LHS) was developed.⁶ Although the MC simulations lead to an accurate estimation of the statistical moments, it is typically necessary to perform tens to hundreds of simulations, which is often not feasible in combination with NLFEA due to an enormous computational burden. In contrast, LHS is the only general tool for a complex stochastic analysis without any simplifying assumptions (taking arbitrary correlation into account), allowing for estimating statistical characteristics from tens of simulations. This is why it will be used as a reference in the following numerical examples. Note that in order to obtain the consistency of the results and the design values of resistance, we assume lognormal distributions of input random variables with the mean values and CoVs given in Tables 1–3.

3 | CASE STUDIES

The models developed in the ATENA Science software based on non-linear fracture mechanics²⁴ are used to replicate the experimental results from the scientific literature. The nonlinear behavior of the concrete material is modeled using the fracture-plastic material model.^{25,26} Specifically, three typical structural members, each failing in a different mode, are selected. The presented advanced ECoV methods and PSF are compared to a reference LHS solution. Moreover, the specific values of input material characteristics for each of the methods are summarized in tables in order to simplify their practical application or replication of the obtained results.

3.1 | Methodology for numerical comparison of ECoV methods

The task of a probabilistic analysis is simplified to estimating the first two statistical moments, and all the described methods were employed for the comparison of the obtained results in terms of the design resistances determined by Equation (1) and considering $\alpha_R = 0.8$ and $\beta = 3.8$. Model uncertainties are included by an additional reduction factor γ_{R_d} obtained during the extensive benchmark investigation specifically for the employed ATENA Science software and for different failure modes²⁷ (assuming model uncertainty as a non-dominant resistance parameter). Note that it is recommended to perform Bayesian updating of the prior distribution of the resistance model uncertainty given in the draft Model

TABLE 5 Correlation matrix considered in the case studies²

	f_t	f_c	G_f
f_t	1	0.7	0.8
f_c	0.7	1	0.6
G_f	0.8	0.6	1

Code 2020 as recently described by Engen et al.²⁸; the application of this updating is beyond the scope of this paper.

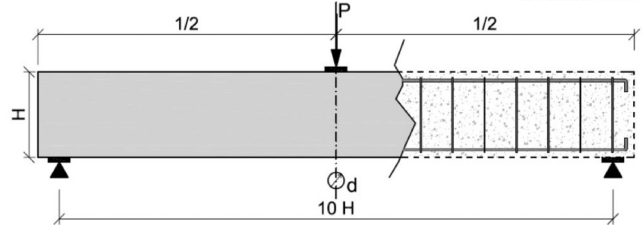
For each example, three reference solutions provided by LHS were obtained reflecting the assumed correlation structure among basic variables. The first case corresponds to the unit covariance matrix, i.e. the case with independent basic variables, which is rather unrealistic for concrete structures (though often assumed). The second limit case represents the full correlation among all input random variables. The full correlation does not represent real situations in physical systems, but it allows for considerable simplifications¹¹ and it could be considered as another reference solution for the ECoV methods. Both limit cases define the variance interval within which the design values can range depending on the actual correlation structure. The last case solved by LHS represents a realistic correlation matrix inferred from experiments.² However, this information is commonly unavailable, and thus not reflected in industrial applications. The realistic correlation matrix according to Slowik et al.² prescribes high positive Spearman correlation coefficients only among concrete parameters as can be seen in Table 5.

To sum up, the three cases solved by LHS represent the reference solutions dependent on the correlation among input random variables, i.e. addressing none, full, and realistic correlation. It was recently shown that the simplified ECoV methods implicitly assume full correlation.¹¹ In contrast to this method, the only method designed for uncorrelated (and possibly arbitrary correlated) variables is TSE, and thus its estimation should be close to the second boundary of the defined correlation interval. Note that TSE with advanced differencing determined almost identical design values as TSE with simple differencing in the following examples, and thus the results of the former are not presented.

In the following figures presenting the obtained results, the reader can find load-deflection diagrams of reference solutions consisting of 30 simulations generated by LHS. In order to clearly show the influence of correlation, three selected realizations are highlighted: the first simulation, 15th (median), and the last realization of the input random vector, where realizations are in an increasing order of the compressive strength of concrete.

Beam	H	B	L	c	c'	Tension reinforcement	Compression reinforcement	Stirrups diam/spacing
T-1	200	100	2000	24	22	1 Ø 12	1 Ø 8	Ø 6 / 150
T-2	200	100	2000	24	22	2 Ø 12	2 Ø 8	Ø 6 / 150
T-3	200	100	2000	24	22	3 Ø 12	2 Ø 8	Ø 6 / 150
T-8	600	300	6000	35	70	2 Ø 12	2 Ø 12	Ø 6 / 150
T-9	600	300	6000	35	70	4 Ø 12	2 Ø 12	Ø 6 / 150
T-10	600	300	6000	35	70	9 Ø 12	2 Ø 12	Ø 6 / 150
T-11	600	300	6000	35	70	18 Ø 12	2 Ø 12	Ø 6 / 150

*all measures in mm.

FIGURE 2 Geometry of the beam specimens by Bosco and Debenardi²⁹

Besides the load-deflection diagrams and the corresponding statistical values, one can see a comparison of the design values estimated by simplified methods and the defined correlation interval of design values (reference solution). Note that if the estimated design value is out of the interval, it is highlighted by green or red color indicating whether it is conservative or non-conservative, respectively.

For the sake of clarity, Table 4 summarizes the general formulas for the determination of the mean values and characteristic values f_k^* of all the basic random variables used for the ECoV methods in the following examples. Note that the characteristic values with the superscript * are obtained as 5% quantiles of lognormal distributions based on the conventional models adopted for PSF (Section 2.1). Note that in all computations presented here, the nominal values of geometrical variables are applied in NLFEA models. The influence of their bias and CoV, μ_{geo} and v_{geo} , is already reflected in the characteristic values f_k^* .

3.2 | Experimental program by Bosco and Debenardi

The first two examples are a replication of the tests done by Bosco and Debenardi.²⁹ The investigated structural member is a simple beam failing in bending. The geometry and reinforcement arrangement of the analyzed beams are described in Figure 2. Two tests with identical beam geometry were selected for this comparison: the T8 test with a low reinforcement ratio, which exhibits a bending failure due to reinforcement yielding, and the T11 test with the reinforcement ratio exhibiting a bending failure due to concrete crushing (see Figure 3). The reinforcement is modeled using the embedded approach assuming a perfect connection to the surrounding

FIGURE 3 Finite element model and failure modes of the T8 beam (top) with reinforcement failure, and T11 (bottom) with concrete crushing failure

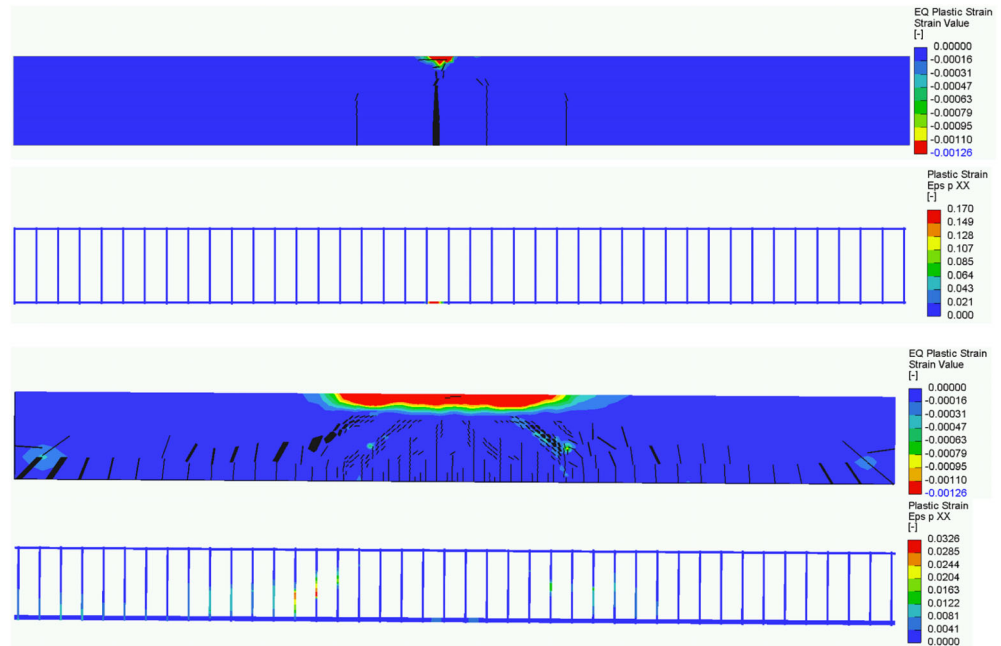
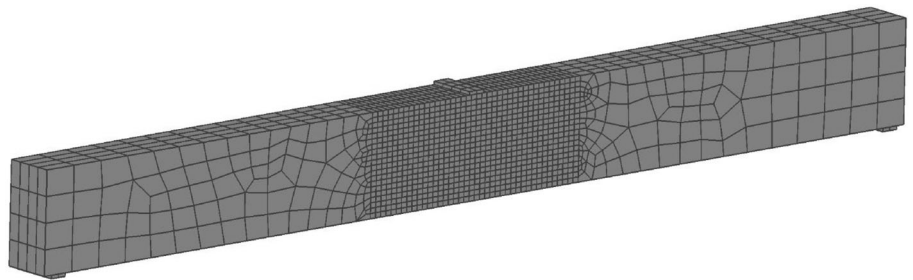


FIGURE 4 Typical mesh for the beam specimens with the mesh size of 30 mm in the middle



concrete. It should be noted that this assumption is not limiting since the bond failure can be captured quite well by the cracking of the finite elements next to the elements with the embedded reinforcement. The typical finite element mesh, shown in Figure 4, leads to an optimal number of finite elements (affecting computational costs) assuring the desired accuracy in the area of interest where the cracks develop.

The characteristic values considered to determine the mean values of strengths are $f_{y,k} = 500$ MPa and $f_{c,k} = 25$ MPa. Note that the effective depth of the T8 and T11 beams is $d = 0.565$ m, and thus $v_d = 0.05(200/d)^{2/3} = 0.025$ and $\mu_d = 1 - 0.05(200/d)^{2/3} = 0.975$. The important quantiles for the application of the presented safety formats and semi-probabilistic methods for the first two numerical examples are summarized in Table 6, and the results of NLFEA can be found in Table 7.

3.2.1 | The T8-A1 beam failing in bending

The obtained design values are further divided by $\gamma_{R_d} = 1.01$ reflecting the model uncertainties in bending.²⁷

In this simple example, all the utilized semi-probabilistic methods lead to an identical design value of resistance, $R_d = 40$ kN. The experimental result from the original publication was 50 kN.

As can be seen from the reference solutions in Figure 5 (top), there is a significant influence of correlation among input random variables on the variance of the quantity of interest (QoI)—the ultimate resistance of the structural member represented by the peak of the Load-Deflection diagram (LD). The highest variance is associated with the case of full correlation among all input random variables. One can see that the QoI of the highlighted LD realizations clearly corresponds to their rank since the correlation assures that all material characteristics increase proportionally. The second extreme case is the assumption of uncorrelated input random variables, which leads to the lowest variance of QoI, and the rank of realizations is not related to their ultimate resistance. The realistic correlation matrix leads to the variance close to the uncorrelated case due to a low influence of the concrete material characteristics on failure of this structural member, and thus it can be expected that ECoV methods and PSF will be conservative in comparison to this realistic design resistance.

Random variable	$X_{i,d}$	$X_{i,m}$	$X_{i\Delta} = X_{ik}$	$X_{i\frac{\Delta}{2}}$
Yield strength, f_y [MPa]	449	525	482	504
Compressive strength, f_c [MPa]	17.2	28	21.5	24.7
Fracture energy, G_f [MN/m]	9.02 e-5	1.76 e-4	1.23 e-4	1.49 e-4
Tensile strength, f_{ct} [MPa]	1.2	2.3	1.6	1.9

TABLE 6 Input random variables and the defined values for safety formats and ECoV methods

Abbreviation: ECoV, estimation of coefficient of variation.

TABLE 7 Results of NLFEA utilized in the presented safety formats and ECoV methods

	PSF	ECoV MC 2010		Eigen ECoV		
		R_m	R_k	R_m	$R_{\theta\frac{\Delta}{2}}$	$R_{\theta\Delta}$
T8 beam [kN]	40.0	50.1	45.4	50.1	47.8	45.4
T11 beam [kN]	308	376	340	376	358	340

Abbreviations: ECoV, estimation of coefficient of variation; NLFEA, non-linear finite element analysis.

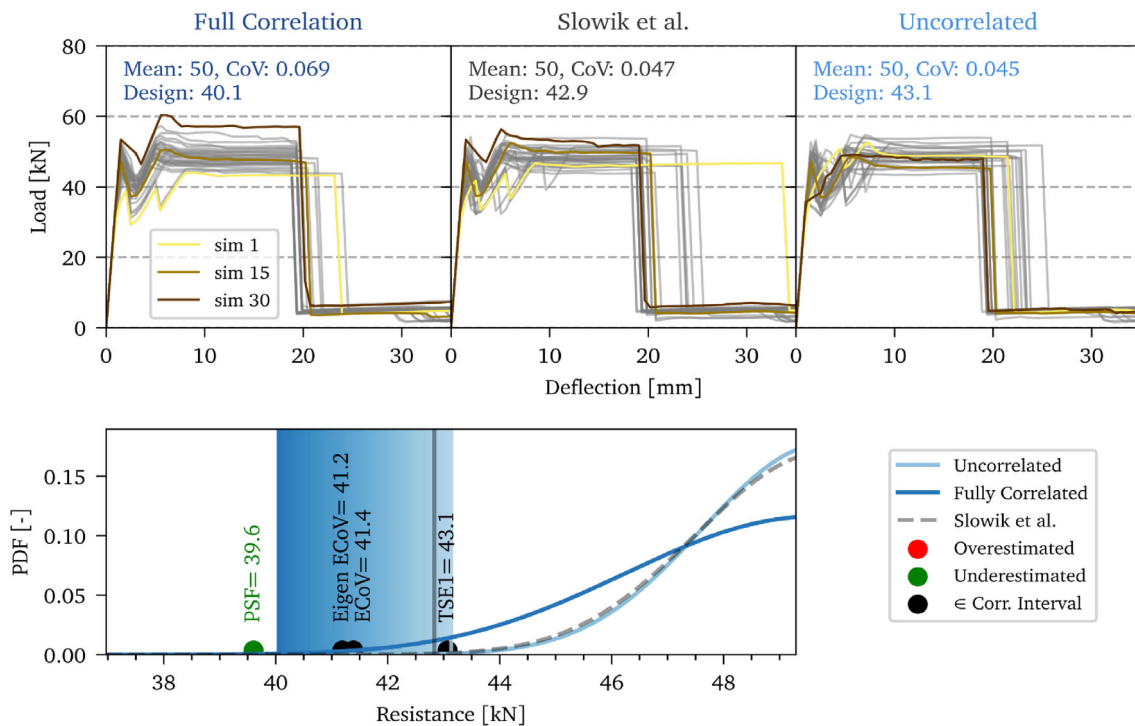


FIGURE 5 T8 beam—Three reference solutions obtained by LHS assuming different correlations (top) and comparison of design resistances (bottom). LHS, Latin Hypercube Sampling

The comparison of design values determined by ECoV methods and PSF can be seen in Figure 5 (bottom). The non-linearity of this example (steel failure in bending) is insignificant, and thus all the employed methods lead to similar design values, and the absolute differences are less than 10%. The highest design resistance is determined by TSE, which is almost identical to the reference uncorrelated solution. With respect to the solution assuming a realistic correlation among

input random variables (the vertical gray line in Figure 5), TSE is the most accurate, but also the most expensive method. The ECoV methods estimated almost identical design values inside of the correlation interval. Finally, the most conservative method is PSF, though the results of all the used methods are in close agreement. This example shows the typical results of semi-probabilistic methods in simple, almost linear mathematical models.

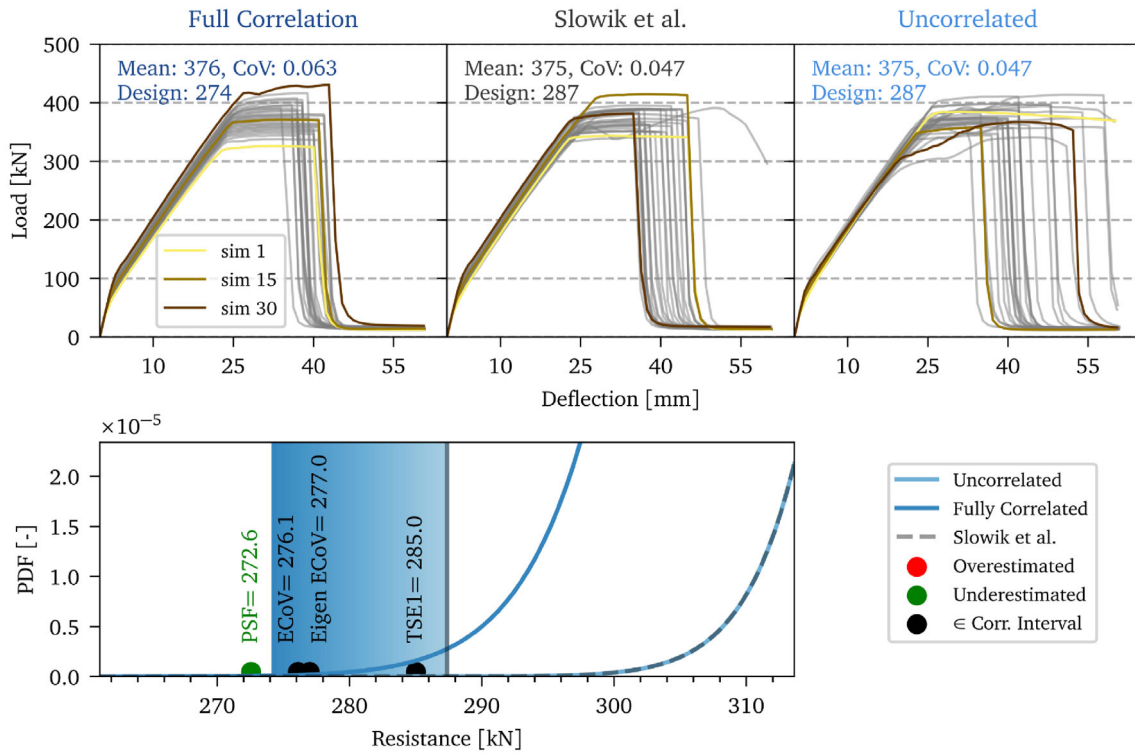


FIGURE 6 T11 beam—Three reference solutions obtained by LHS assuming different correlations (top) and comparison of design values (bottom). LHS, Latin Hypercube Sampling

3.2.2 | The T11-A1 beam failing in bending

The T11-A1 beam exhibits combined compressive crushing and shear failure by the yielding of stirrups. Therefore, the obtained design values are further reduced by $\gamma_{R_d} = 1.13$ reflecting the model uncertainties of shear failure.²⁷ Design resistances are depicted in Figure 6 together with the reference solution (the distribution and design quantile obtained by LHS). The experimental result from the original publication was 380 kN.

In this example, the yield strength of the reinforcement has the dominant influence on the ultimate structural resistance, since the realistic correlation matrix defining a strong correlation among concrete material characteristics leads to the identical variance as in an uncorrelated case. The comparison of the design values determined by ECoV methods and PSF can be seen in Figure 6 (bottom). Note that all the methods are in good agreement with the LHS and their results are according to expectations: ECoV methods lead to design values near the fully correlated boundary and TSE1 leads to the uncorrelated boundary of the correlation interval. The most conservative design value is obtained by PSF from a single simulation, though it is very close to the defined reference interval, and thus it may be seen as a very efficient method.

3.3 | Experimental program by Anderson and Ramirez

The third example is based on the experimental program by Anderson and Ramirez.³⁰ In this experiment, a beam with the cross-section of 406×406 mm was subjected to a four-point bending test with a shear span $a = 0.91$ m. The beam was designed to fail in shear, i.e. to comply with the condition for shear stress $V_{test}/(b_w d) > 6$. The beam geometry and reinforcement are shown in Figure 7 and the finite element model in Figure 8. The shear reinforcement is composed of double stirrups no. 3 with \emptyset 9.525 mm. The top longitudinal reinforcement consists of 5 bars no. 6 (\emptyset 19.05 mm) and the bottom reinforcement of 5 bars no. 9 (\emptyset 28.65 mm).

The characteristic values used for the determination of the mean values of material parameters are $f_{y,k} = 422$ MPa and $f_{c,k} = 25$ MPa. The effective depth of this beam is $d = 0.344$ m, and thus $v_d = 0.05(200/d)^{2/3} = 0.034$ and $\mu_d = 1 - 0.05(200/d)^{2/3} = 0.97$. The important quantiles for the application of the presented semi-probabilistic methods are summarized in Table 8.

From the obtained results of NLFEA, the beam is failing in shear (see Figure 9), and thus the obtained design values are further reduced by $\gamma_{R_d} = 1.13$ reflecting the model uncertainties of the shear failure.²⁷ Numerical results of

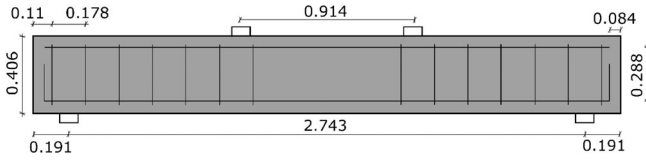


FIGURE 7 Geometry of the W1 beam of Anderson and Ramirez.³⁰ Note that the dimensions are in [m], but the original experiment was in imperial units

NLFEA for the ECoV methods and PSF are summarized in Table 9. The experimental result was 460 kN.

The results of the semi-probabilistic methods are presented in Figure 10 (bottom). From the reference solutions by LHS, one can see that there is a moderate influence of correlation among concrete material characteristics since the case with the realistic correlation matrix leads to a slightly higher variance in comparison to the uncorrelated case. The design values determined by the simplified ECoV methods are in good agreement with the reference solution. The estimate of the standard ECoV is close to the uncorrelated boundary while it should be closer to the fully correlated reference solution; the difference, however, is insignificant for practical applications. This can be attributed to a significant non-linearity in this example. Note that in contrast to the standard ECoV, an almost identical estimate by TSE1 is in agreement with the theoretical expectations and the high accuracy of this method is demonstrated by all the case studies. The additional simulation in Eigen ECoV significantly improves the estimation of standard ECoV, and its result is close to its reference solution (fully correlated). This is in agreement with the previous theoretical results,¹¹ since it should achieve a higher accuracy in comparison to the standard ECoV in the case of moderate non-linearity of the investigated mathematical models. Note that PSF leads to a very accurate estimation of design resistance, though it is not typical in the case of shear failure.²⁻⁴

4 | DISCUSSION

4.1 | Effect of correlation

Statistical correlation among material characteristics might play a crucial role, since it has a significant influence on the variance of QoI, particularly for concrete structures. Nonetheless, the exact information about the correlation matrix is usually unavailable and the recommendations in scientific literature widely differ depending on the concrete mixture, strength class, etc.^{2,31,32} For practical analyses of concrete structures, two extremes may be important: fully correlated random variables and uncorrelated random variables, which together define the

variance interval caused by insufficient information about the correlation. The fully correlated random input variables usually lead to a larger variance of QoI and conservative estimates of design values. The lower boundary of the variance corresponding to the uncorrelated case needs to be estimated by advanced probabilistic methods such as LHS, or approximated by TSE. In both cases, the number of calculations is significantly higher in comparison to the ECoV methods.

Based on the obtained results for the three structural members failing in different modes, it can be concluded that all the presented ECoV methods are well bounded by the correlation intervals. One should keep in mind that, Eigen ECoV and ECoV according to *fib* Model Code 2010 are based on the fully correlated case, and TSE correspond to the uncorrelated case. The examples indicate that Eigen ECoV provides better estimates for the fully correlated case and one additional simulation may significantly improve the estimate by ECoV (according to *fib* Model Code 2010). However, if only a single input random variable fully describes the variance of QoI, the standard ECoV has a superior efficiency as shown by the second example. In contrast, Eigen ECoV might be more suitable for shear failures with a higher non-linearity. The analysis of the obtained results and of the underlying assumptions reveals that the accuracy of the ECoV methods depends on a specific failure mode and assumed correlation matrix. Moreover, the ECoV methods may provide crude estimates in the case of multiple failure modes as briefly discussed in the following subsection.

4.2 | Limitation of simplified ECoV methods for multiple failure modes

Simplified safety formats are commonly devised to yield adequate estimates of design resistances in most practically relevant applications, while, inevitably in some cases, overconservative or unsafe approximations might be obtained. Though a detailed study of such errors is beyond the scope of this contribution, a fundamental example is analyzed here to provide the first insights. It is often argued that simplified safety formats may fail in the cases with several local extrema as typically caused by multiple failure modes. To verify this, two columns exposed to compression without eccentricity, acting as a series system, are analyzed (Figure 11).

The example is focused on a simple series system whose resistance R is obtained as a minimum of resistances of two identical RC columns:

$$R/A_c = \min(f_{c1} + \rho f_{y1}, f_{c2} + \rho f_{y2}) \quad (15)$$

FIGURE 8 Finite element mesh for the nonlinear analysis of the W1 beam of Anderson and Ramirez

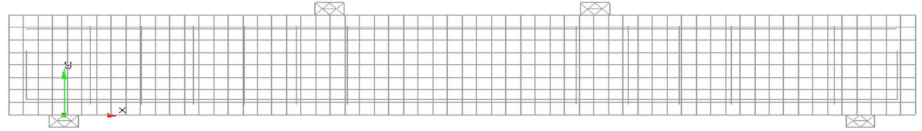


TABLE 8 Input random variables and the defined values for safety formats and ECoV methods

Random variable	$X_{i,d}$	$X_{i,m}$	$X_{i,\Delta} = X_{i,k}$	$X_{i,\frac{\Delta}{2}}$
Yield strength, f_y [MPa]	369	439	399	419
Compressive strength, f_c [MPa]	17.1	28	21.5	24.7
Fracture energy, G_f [MN/m]	9.02 e-5	1.76 e-4	1.23 e-4	1.49 e-4
Tensile strength, f_{ct} [MPa]	1.2	2.3	1.6	1.9

Abbreviation: ECoV, estimation of coefficient of variation.

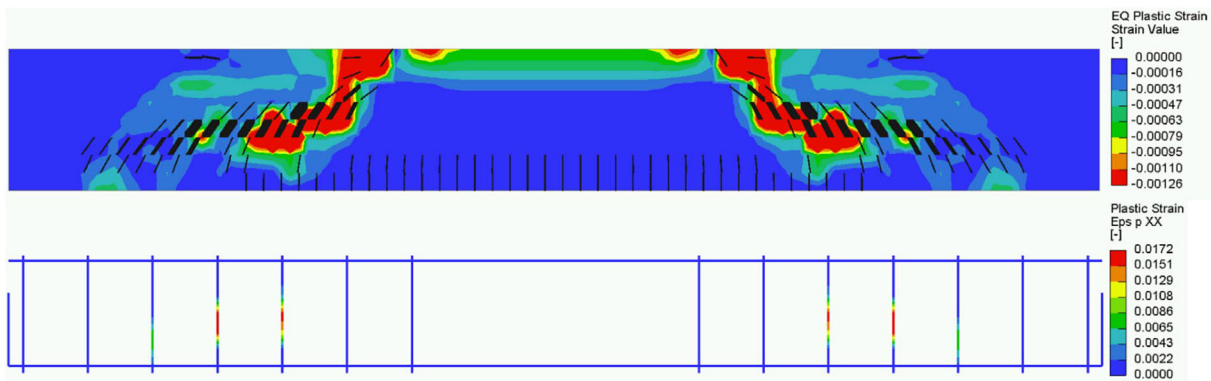


FIGURE 9 Results from the nonlinear analysis of the W1 beam failing in shear

TABLE 9 Results of NLFEA utilized in the presented safety formats and ECoV methods

	PSF	ECOV MC 2010		Eigen ECoV		
		R_m	R_k	R_m	$R_{\phi_{\frac{\Delta}{2}}}$	$R_{\phi_{\Delta}}$
W1 beam	307	387	351	387	367	351

Abbreviations: ECoV, estimation of coefficient of variation; NLFEA, non-linear finite element analysis.

where A_c denotes a concrete section area; and ρ is the reinforcement ratio common to both columns. The reinforcement ratio is a study parameter, arbitrarily varied disregarding practical constraints.

Mutually statistically independent strengths f_{ci} and f_{yi} are described by lognormal distributions with the following characteristics:

- $f_{cm} = 29.1$ MPa, $v_{fc} = 15\%$, and $f_{c0.05} = f_{ck} = 22.6$ MPa
- $f_{ym} = 455$ MPa, $v_{fy} = 5.8\%$, and $f_{y0.05} = f_{yk} = 414$ MPa

These assumptions are based on a more detailed study focused on the performance of safety formats for series systems.³³ Uncertainty in geometrical variables is

ignored here to keep focus on the key aspects affecting the performance of the simplified safety formats.

All the obtained design values are normalized to those obtained by the probabilistic approach using the numerical integration ($R_{d,prob}$). Besides $R_{d,PSF}$, all design values are obtained as a fractile of the system resistance corresponding to the probability of 1.12%, resulting from $\alpha_R = 0.8$ and $\beta = 3.8$. Model uncertainty is not considered in this section as it is typically treated separately, beyond the application of a particular safety format. Note that the justification of $\gamma_C = 1.5$ and $\gamma_S = 1.15$ according to Eurocode 2 Commentary³⁴ indicates that the model uncertainty factors related to the recommended values in EN 1992-2:2005¹ are very close to unity, and thus $\gamma_C = 1.5$ and $\gamma_S = 1.15$ are adopted without any adjustment in the following analysis, where model uncertainty is ignored.

Figure 12 displays a variability of $R_{d,safety\ format}/R_{d,prob}$ with a reinforcement ratio. For low ρ -values, the resistance of a column is governed by the concrete contribution, while the reinforcement contribution becomes important with the increasing ρ , and the distribution of column resistance attains a bimodal character. Figure 12

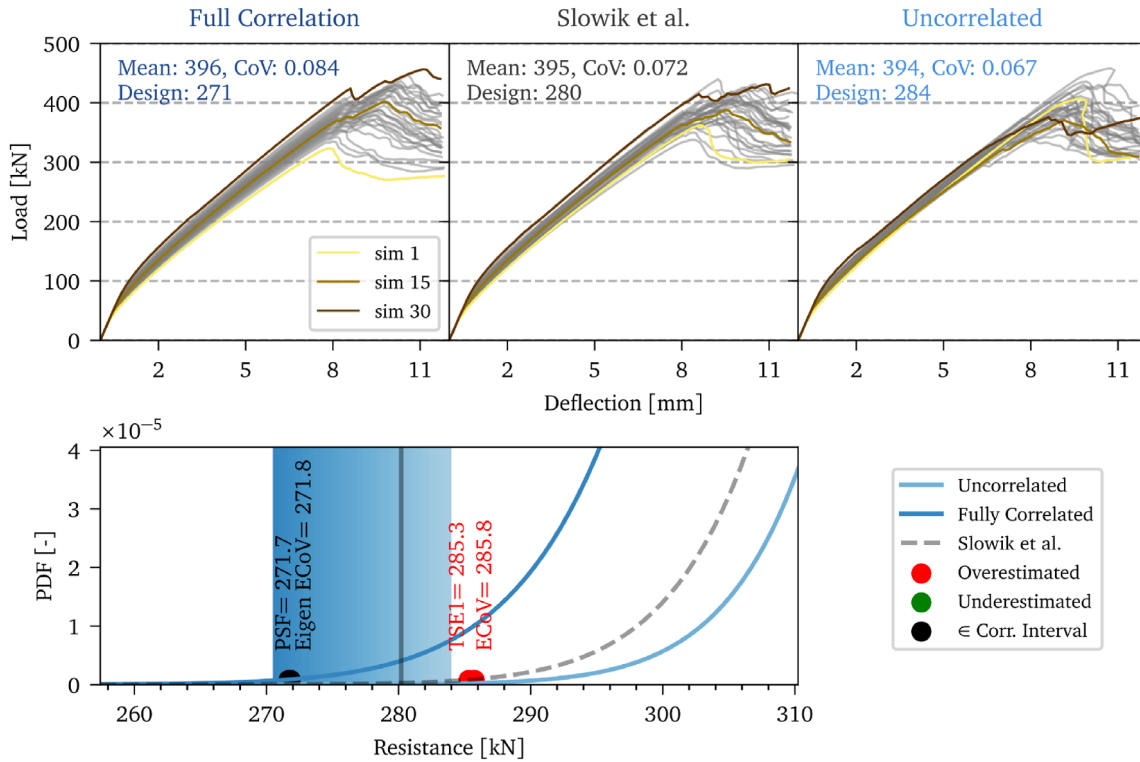


FIGURE 10 W1 beam—Three reference solutions obtained by LHS assuming different correlations (top) and comparison of design resistances (bottom). LHS, Latin Hypercube Sampling

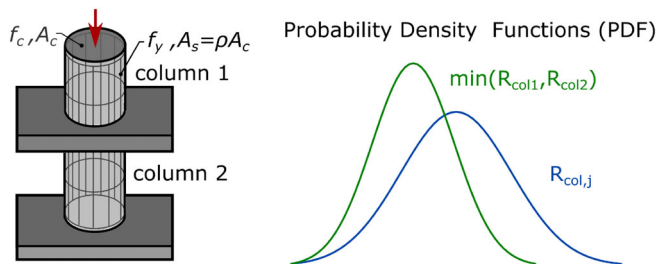


FIGURE 11 Illustration of the analyzed series system and probability density functions (PDFs) of component and system resistances

also shows that ECoV based on TSE leads to nearly the same design resistance as the probabilistic approach. All the other simplified formats provide reasonably conservative estimates, with errors mostly between 2% and 6%. Eigen ECoV (EE in the figure) performs slightly better than PSF and ECoV, but the differences are negligible in this case.

All the ECoV methods lead to conservative estimates of v_R as can be seen in Figure 13. As expected, the best estimates are obtained by TSE based on the largest number of limit state function evaluations, then following with Eigen ECoV and the standard ECoV.

What is interesting to observe is that TSE yields $R_{d,TSE} \approx R_{d,probab}$ for any $\rho > \rho_m$ while it systematically

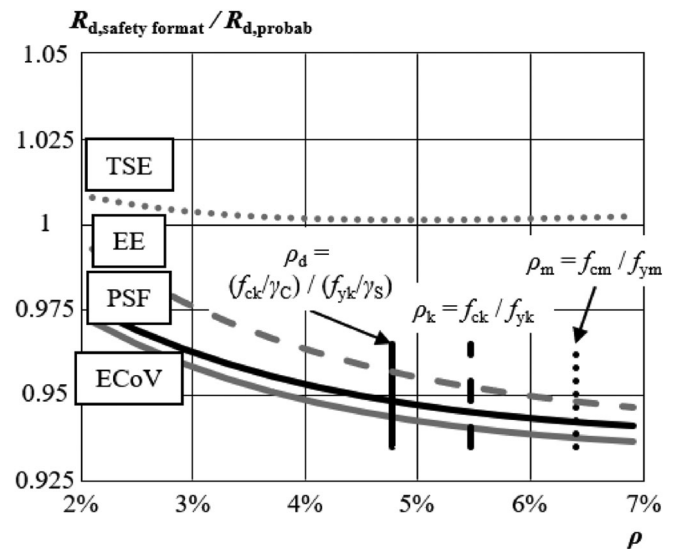


FIGURE 12 Variability of $R_{d,safety\ format} / R_{d,probab}$ with ρ

overestimates v_R (which should have led to $R_{d,TSE} < R_{d,probab}$). A detailed analysis indicates that this safe-sided error is nearly exactly outweighed by the failure in identifying the type of distribution of system resistance, ignoring the bimodal character of the distribution by TSE; this is common to all ECoV methods. The skewness of the bimodal distribution (reflected by the probabilistic

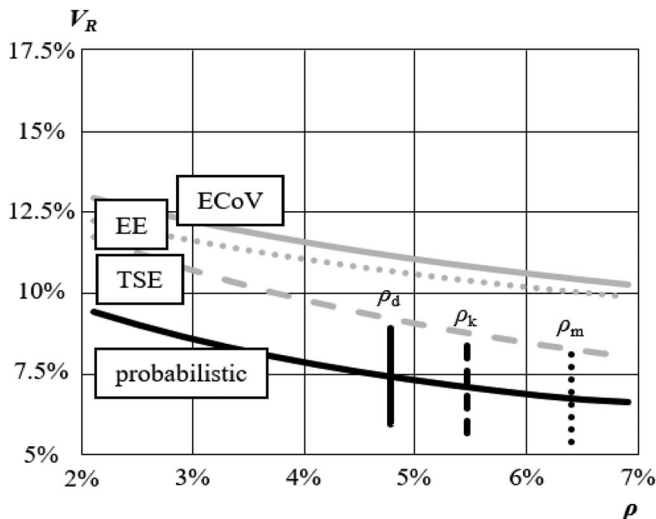


FIGURE 13 Variability of coefficients of variation estimated by various methods with ρ

approach) tends to be lower than that of a lognormal distribution, and the ECoV methods thus make unsafe errors here.

The presented limited analysis of the series system with two failure modes indicates a number of directions for further research:

- Most concrete structural systems are deemed to have properties closer to parallel systems as they are often indeterminate, providing for multiple load paths. The preliminary results for parallel systems, already partly presented,³⁵ indicate that the ECoV methods perform in a similar way as those observed here; the partial factor method tends to be conservative for parallel systems.
- Positive correlations between failure modes are expected to reduce the ECoV error for both types of systems. In contrast, the errors may amplify with an increasing number of failure modes of a similar importance. These counteracting effects should be investigated further.

5 | CONCLUSIONS

The comparison of the selected advanced semi-probabilistic methods is presented in three numerical examples failing in different modes. The case studies demonstrate how uncertainties in geometry can be combined with those in material properties and considered in NLFEA applications. The obtained results show that all the employed methods lead to design values close to the reference solution. The numerical differences become more significant with an increasing non-linearity of the failure mode. The theoretical behavior of the recently

proposed modification of Taylor Series Expansion (TSE) and its adaptation Eigen ECoV is successfully verified by realistic case studies. The correlation among input random variables might play a crucial role in determining the design values, and thus it might be beneficial to verify two limit situations: a fully correlated case by ECoV methods, and an uncorrelated case by TSE. For practical applications, recommendations should be provided as to when the examination of the two limiting situations is needed and how to proceed when a large difference between the design values is obtained.

ACKNOWLEDGMENTS

The authors would also like to express their gratitude for the support of project No. 20-01781S by the Czech Science Foundation. Preliminary theoretical results and the approach to NLFEA considering the uncertainties in basic variables have been discussed in *fib* COM 3 TG 3.1 and AG 8.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Lukáš Novák <https://orcid.org/0000-0001-9387-2745>

Jan Červenka <https://orcid.org/0000-0003-4945-1163>

Vladimír Červenka <https://orcid.org/0000-0002-1150-8171>

Drahomír Novák <https://orcid.org/0000-0003-0744-8265>

Miroslav Sýkora <https://orcid.org/0000-0001-9346-3204>

REFERENCES

1. CEN: EN 1992-2:2005 Eurocode 2: design of concrete structures – Part 2: concrete bridges – design and detailing rules, CEN; 2005.
2. Slowik O, Novák D, Novák L, Strauss A. Stochastic modelling and assessment of long-span precast prestressed concrete elements failing in shear. *Eng Struct*. 2021;228:111500. <https://doi.org/10.1016/j.engstruct.2020.111500>
3. Novák L, Novák D, Pukl R. Probabilistic and semi-probabilistic design of large concrete beams failing in shear. *Advances in engineering materials, structures and systems: innovations, mechanics and applications*. London: CRC Press; 2019. p. 1339–44. ISBN:978-1-138-38696-9.
4. Novák L, Pan L, Slowik O, Novák D. Advanced reliability and sensitivity analysis of prestressed concrete girders failing in shear. In: *Proc. of 12th fib International PhD Symposium in Civil Engineering*; 2018. p. 617–624. 978-80-01-06401-6.
5. Slobbe A, Rózsás Á, Allaix DL, Bigaj-van Vliet A. On the value of a reliability-based nonlinear finite element analysis approach in the assessment of concrete structures. *Struct Concr*. 2020;21:32–47. <https://doi.org/10.1002/suco.201800344>

6. McKay MD. Latin hypercube sampling as a tool in uncertainty analysis of computer models. In: Proceedings of the 24th Conference on Winter Simulation, WSC'92; 1992. p. 557–564.
7. Vořechovský M, Mašek J, Eliáš J. Distance-based optimal sampling in a hypercube: analogies to N-body systems. *Adv Eng Softw.* 2019;137:102709. <https://doi.org/10.1016/j.advengsoft.2019.102709>
8. Kennedy AC, Lennox CW. Moment operations on random variables, with applications for probabilistic analysis. *Probabilistic Eng Mech.* 2001;16(3):253–9. [https://doi.org/10.1016/S0266-8920\(01\)00007-8](https://doi.org/10.1016/S0266-8920(01)00007-8)
9. Zhang X, Pandey DM. Structural reliability analysis based on the concepts of entropy, fractional moment and dimensional reduction method. *Struct Saf.* 2013;43:28–40. <https://doi.org/10.1016/j.strusafe.2013.03.001>
10. Červenka V. Reliability-based non-linear analysis according to fib model code 2010. *Struct Concr.* 2013;14:19–28. <https://doi.org/10.1002/suco.201200022>
11. Novák L, Novák D. Estimation of coefficient of variation for structural analysis: the correlation interval approach. *Struct Saf.* 2021;92:102101. <https://doi.org/10.1016/j.strusafe.2021.102101>
12. Schlune H, Plos M, Gylltoft K. Safety formats for nonlinear analysis tested on concrete beams subjected to shear forces and bending moments. *Eng Struct.* 2011;33:2350–6.
13. Novák L, Novák D. On Taylor series expansion for statistical moments of functions of correlated random variables. *Symmetry.* 2020;12:1379. <https://doi.org/10.3390/sym12081379>
14. Novák L, Novák D. Shear strength design by stochastic simulation and semi-probabilistic methods. *AIP Conference Proceedings.* 2018;1978:500–4.
15. Novák D, Novák L, Slowik O, Strauss A. Prestressed concrete roof girders: Part III – Semi-probabilistic design. *Proceedings of the sixth international symposium on life-cycle civil engineering.* London: CRC Press, Taylor and Francis Group; 2018.
16. Allaix DL, Carbone VI, Mancini G. Global resistance factor for reinforced concrete beams. In: *Proceedings of the 5th international probabilistic workshop.* Ghent, Belgium: ACCO; 2007. 195–208.
17. Allaix DL, Carbone VI, Mancini G. Global safety format for nonlinear analysis of reinforced concrete structures. *Struct Concr.* 2013;14(1):29–42. <https://doi.org/10.1002/suco.201200017>
18. Rosenblueth E. Point estimates for probability moments. *Proc Natl Acad Sci USA.* 1975;72:3812–4.
19. Castaldo P, Gino D, Mancini G. Safety formats for non-linear finite element analysis of reinforced concrete structures: discussion, comparison and proposals. *Eng Struct.* 2019;193:136–53. <https://doi.org/10.1016/j.engstruct.2019.05.029>
20. CEN. prEN 1992-1-1 Eurocode 2: design of concrete structures – Part 1-1: general rules – rules for buildings, bridges and civil engineering structures, CEN, September 2021
21. Gino D, Castaldo P, Bertagnoli G, Giordano L, Mancini G. Partial factor methods for existing structures according to fib bulletin 80: assessment of an existing prestressed concrete bridge. *Struct Concr.* 2020;21:15–31. <https://doi.org/10.1002/suco.201900231>
22. fib. *fib model code for concrete structures.* Vol 2013. Lausanne: Ernst & Sohn; 2010. ISBN:9783433030615.
23. prEN 1990 Eurocode – Basis of structural and geotechnical design, CEN, September 2021
24. Červenka V, Červenka J, Jendele L. ATENA Program Documentation, Part 1: Theory. 2022, Cervenka Consulting s.r.o
25. Červenka J, Papanikolaou V. Three dimensional combined fracture-plastic material model for concrete. *Int J Plast.* 2008; 24:2192–220.
26. Červenka J, Červenka V, Eligehausen R. Fracture-plastic material model for concrete. Application to analysis of powder actuated anchors. *Proc FRAMCOS.* Volume 3. Germany: Aedificatio Publishers; 1998. p. 1107–16.
27. Červenka V, Červenka J, Kadlec L. Model uncertainties in numerical simulations of reinforced concrete structures. *Struct Concr.* 2018;19:2004–16. <https://doi.org/10.1002/suco.201700287>
28. Engen M, Hendriks MAN, Monti G, Allaix DL. Treatment of modelling uncertainty of NLFEA in fib model code 2020. *Struct Concr.* 2021;22:1–11. <https://doi.org/10.1002/suco.202100420>
29. Bosco C, Debernardi PG. Influence of some basic parameters on the plastic rotation of reinforced concrete elements. *CEB Bull Inf.* 1993;218:25–44.
30. Anderson N, Ramirez J. Detailing of stirrup reinforcement. *ACI Struct J.* 1989;86(5):507–15.
31. Zimmermann T, Lehký D, Strauss A. Correlation among selected fracture-mechanical parameters of concrete obtained from experiments and inverse analyses. *Struct Concr.* 2016;17: 1094–103. <https://doi.org/10.1002/suco.201500147>
32. Osman NO, Harun M, Mehmet AT. Factors determining the correlations between concrete properties. *Cem Concr Res.* 1996; 26(11):1629–37. [https://doi.org/10.1016/S0008-8846\(96\)00167-6](https://doi.org/10.1016/S0008-8846(96)00167-6)
33. Červenka J, Červenka V, Sykora M, Mlcoch J. Evaluation of safety formats for structural assessment based on nonlinear analysis. *Proc. EURO-C. Bad Hofgastein, Austria, London: CRC Press; 2018.* p. 669–78.
34. European concrete platform [Internet] Eurocode 2 Commentary. 2008 [cited 2022 Feb 21]. Available from: www.europeanconcrete.eu
35. Sykora M, Červenka J, Červenka V, Mlcoch J, Novak D, Novak L. Pilot comparison of safety formats for reliability assessment of RC structures. *Proc. fib Symposium 2019, 27–29 May 2019, Krakow. Lausanne: fib; 2019.* p. 2076–83.

AUTHOR BIOGRAPHIES



Lukáš Novák Assistant Professor, Faculty of Civil Engineering, Brno University of Technology, Brno, Czech Republic Email: novak.l@fce.vutbr.cz



Jan Červenka Cervenka Consulting, Prague, Czech Republic Email: jan.cervenka@cervenka.cz



Vladimír Červenka Cervenka Consulting, Prague, Czech Republic
Email: vladimir.cervenka@cervenka.cz



Drahomír Novák Full Professor, Faculty of Civil Engineering, Brno University of Technology, Brno, Czech Republic
Email: novak.d@fce.vutbr.cz



Miroslav Sýkora Associate Professor, Klokner Institute, Czech Technical University in Prague, Prague, Czech Republic
Email: miroslav.sykora@cvut.cz

How to cite this article: Novák L, Červenka J, Červenka V, Novák D, Sýkora M. Comparison of advanced semi-probabilistic methods for design and assessment of concrete structures. *Structural Concrete*. 2022. <https://doi.org/10.1002/suco.202200179>

3 Polynomial Chaos Expansion in Structural Mechanics

Truth is much too complicated to allow anything but approximations.

— John von Neumann

The propagation of uncertainties through complex mathematical models representing general physical systems is a very challenging task. Similarly to the previous chapter, such models are computationally expensive, and thus it is not feasible to employ standard sampling techniques. Moreover, UQ in a general setting contains various tasks including many types of sensitivity analyses and higher-order statistical analyses. Therefore, it is beneficial to create a computationally efficient surrogate model approximating the original mathematical model. Once the surrogate model is available, it is possible to employ various existing sampling-based methods for a statistical and sensitivity analysis [21, 11, 22]. Although there are many different types of surrogate models varying from methods as simple as TSE [18] to very complex Artificial Neural Networks [23], the analysis of physical systems is a very specific task requiring on the one hand accurate and numerically efficient surrogates, but on the other hand a certain level of interpretability of these surrogate models. This thesis is focused on non-intrusive PCE combining computational efficiency of polynomial basis functions, a well-developed theoretical background of regression methods, and interpretability stemming from carefully selected basis functions allowing for analytical UQ. Although a general theoretical background of PCE was proposed by a brilliant mathematician, Norbert Wiener, in 1938 [4], it took more than 50 years to adapt the method for UQ in engineering [5].

Assuming that QoI has a finite variance, PCE represents the $Y = \mathcal{M}(X)$ as a function of another random variable ξ called a germ with a known probability distribution function p_ξ . The function is in the form of infinite series of the polynomial chaos consisting of deterministic coefficients β and basis polynomials $\Psi(\xi)$ orthonormal with respect to p_ξ . In the case of \mathbf{X} and $\boldsymbol{\xi}$ being vectors containing M independent random variables, the polynomial $\Psi(\boldsymbol{\xi})$ is multivariate and it is built up as a tensor product of univariate orthonormal polynomials:

$$Y = \mathcal{M}(\mathbf{X}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} \beta_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \quad (3.1)$$

where $\boldsymbol{\alpha} \in \mathcal{A}$ is a multi-index containing integers reflecting polynomial degrees associated to each ξ_i . Coefficients β can be obtained in intrusive or non-intrusive frameworks. The intrusive approach requires a modification of a finite element

solver, and thus it is not commonly employed in industrial applications due to its implementation difficulties. This thesis is focused on a non-intrusive approach using any third-party solver as a black box representing an input-output transformation. From a statistical point of view, PCE is a simple linear regression model with intercept. Therefore, it is possible to use *ordinary least squares* (OLS) regression to minimize the error of the approximation. However, for the practical estimation of β , it is necessary to truncate PCE to a final number of terms.

The selection of a truncation scheme, and thus cardinality of $\#\mathcal{A}$, is a crucial step in the construction of an efficient approximation, since PCE suffers from the curse-of-dimensionality. This can be clearly seen in Fig. 3.1 presenting selected truncation schemes with 2 input random variables and maximum polynomial order of basis functions $p = 4$. The selected common truncation schemes are [24]: a simple tensor product of multivariate polynomials, a total polynomial order scheme, and a hyperbolic scheme. As can be seen from the comparison, the tensor product constructs a complete set of basis functions including all possible combinations of polynomials leading to the highest possible $\#\mathcal{A}$, the total polynomial order scheme neglects all interaction terms which have a total polynomial order (the sum of individual polynomial orders p_i) $|p| = \sum p_i \leq p$, and hyperbolic truncation further reduces the number of interaction terms by the parameter $q \in [0, 1]$. This can be justified by the sparsity-of-effects principle, which states that most models describing physical phenomena are dominated by the main effects and low-order interactions [25].

The obtained \mathcal{A} can be further reduced by sparse solvers utilizing the best model selection algorithms identifying the most suitable basis for a given experimental design (ED) containing realizations of the input random vector and corresponding model outputs, e.g. Least Angle Regression [26, 27], orthogonal matching pursuit [28] or Bayesian compressive sensing [29]. Note that, contrary to the truncation schemes, sparse solvers are highly dependent on the quality of the given ED, and thus it is important to carefully select the realizations of the input random vector.

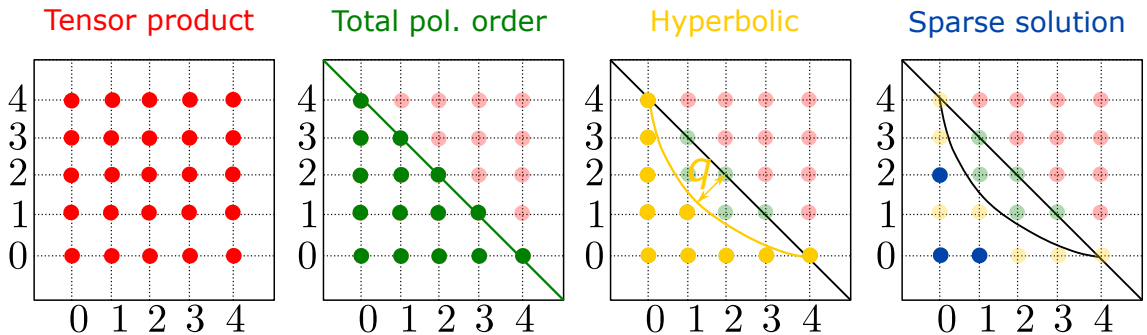


Fig. 3.1: 2D Graphical interpretation of truncation schemes utilized for the construction of PCE with maximum polynomial order $p = 4$.

The quality of ED is generally influenced by the positions and the number of samples, i.e. sampling technique. One of the most widely used techniques is LHS [30, 22, 31]. Another popular strategy for DoE is to uniformly fill the design domain according to some *space-filling* criteria [32] or to decrease the discrepancy of the samples [33, 34, 35, 36, 37]. These techniques can be used without any knowledge about the specific mathematical model or the type of surrogate model. Nonetheless, it is possible to increase the quality of ED by incorporating additional information into a sampling technique. In the context of regression-based surrogate models, the quality of ED can be measured by the optimality of the information matrix. There are various techniques developed for the selection of the best samples from a large candidate set improving the given optimality criterion of information matrix, i.e. alphabetical optimality leads to an optimal ED with respect to specific basis functions [38, 39, 40]. Furthermore, it is often beneficial to include additional information stemming from the specific mathematical model at hand and to sequentially improve the optimality of ED. The concept of adaptive sequential sampling for learning surrogate models is often termed *active learning*.

Naturally, it is ideal to use both *exploitation* (leveraging model behavior) criteria and *exploration* (space filling) criteria for the definition of an optimally balanced criterion [41]. Therefore, the first journal paper attached to this chapter entitled ***Variance-Based Adaptive Sequential Sampling for Polynomial Chaos Expansion*** [42] presents a novel approach for an adaptive sequential sampling developed specifically for PCE. The core idea of this approach is based on the general definition of the m th statistical moment and properties of PCE:

$$\begin{aligned} \langle y^m \rangle &= \int [\mathcal{M}(X)]^m p_X(x) dx = \int \left[\sum_{\alpha \in \mathbb{N}^M} \beta_\alpha \Psi_\alpha(\xi) \right]^m p_\xi(\xi) d\xi = & (3.2) \\ &= \sum_{\alpha_1 \in \mathbb{N}^M} \dots \sum_{\alpha_m \in \mathbb{N}^M} \beta_{\alpha_1} \dots \beta_{\alpha_m} \int \Psi_{\alpha_1}(\xi) \dots \Psi_{\alpha_m}(\xi) p_\xi(\xi) d\xi. \end{aligned}$$

It is evident that in the case of PCE, it is possible to get statistical moments by the integration of basis functions Ψ (orthonormal polynomials). Moreover, it is possible to analytically derive a variance of QoI directly from the orthonormality properties of basis functions as a sum of squared deterministic coefficients β . However, one can also see this formula as the integration of local contributions to the total variance, which is a function referenced in the paper as variance density. Variance density shows local deviations of the investigated mathematical model from its mean value. Motivated by the rationale of the Koksma-Hlawka inequality and the associated active learning method [41], the variance density was incorporated into the proposed Θ criterion defining the best possible locations of statistical samples for the surrogate model at hand. The Θ criterion consists of two parts: variance

density, and a geometrical term assuring uniform coverage of the whole design domain. Such an approach can be easily coupled with any existing sampling scheme for the construction of a candidate set.

The versatility and efficiency of the Θ criterion can be seen from published comparisons and applications, e.g. [43, 44]. Moreover, several extensions have already been proposed by several research teams (e.g. for multi-variate outputs [45] or reliability analysis [46]). Unfortunately, highly non-linear models or models with discontinuities can be approximated only by high-order PCEs inevitably leading to the curse-of-dimensionality despite the efficiency of sparse solvers and optimality of ED. This significant limitation can be alleviated by the decomposition of an input random space to several sub-domains approximated by low-order PCEs as presented in the second paper attached to this chapter – ***Active Learning-based Domain Adaptive Localized Polynomial Chaos Expansion*** [47]. Although the concept of domain decomposition and construction of local surrogate models is not entirely new [48, 49], the proposed method combines the decomposition with the active-learning methodology governed by the Θ criterion in the non-intrusive framework. Such a combination represents a very general approach, which can be combined with various sampling techniques and sparse PCE solvers. Moreover, the whole process is governed by the variance density, and thus it leads to the optimal decomposition with respect to UQ. Fig. 3.2 shows the three crucial steps of the iterative adaptive domain decomposition.

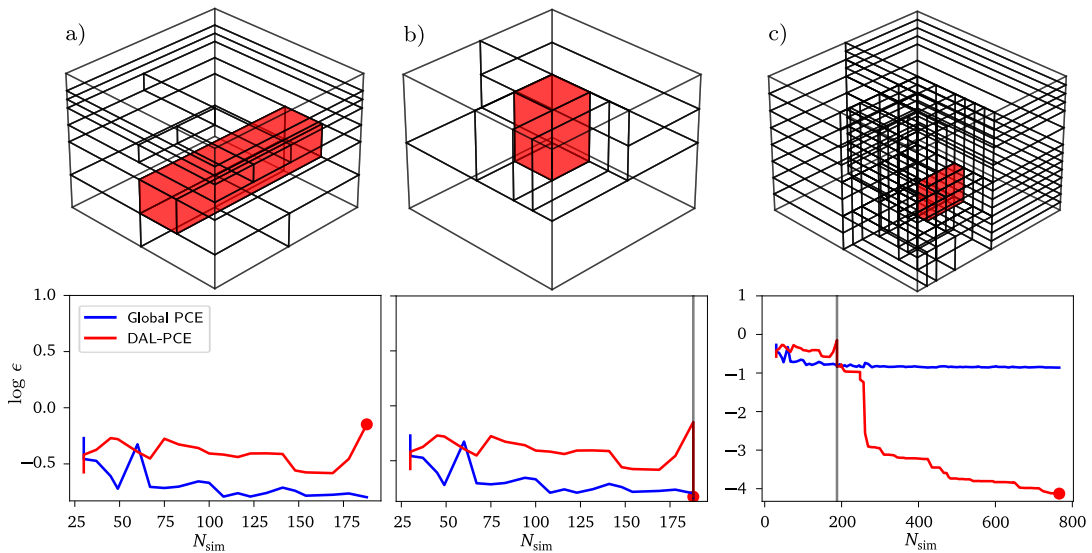


Fig. 3.2: Iterative decomposition of the input space governed by active learning. Top row shows the decomposed input domains, and bottom row presents associated mean squared errors. (Reprinted from [47])

In the first step of the algorithm referenced as DAL-PCE in the attached paper, the input space is decomposed into smaller sub-domains in a way that in each iteration there is divided a sub-domain containing a candidate sample associated to the highest Θ criterion. Furthermore, PCE is either reconstructed in the new sub-domain if possible, or new samples are added to the new sub-domain (Fig. 3.2a). During the process of decomposition and construction of localized PCEs, it is also possible to measure both the cumulative error of localized PCEs and the global PCE error. If the global PCE error is lower than the cumulative error of localized PCEs, the decomposition is discarded and the process starts again (Fig. 3.2b). The final decomposition thus contains the best possible decomposition of the input space for the given ED and maximum p of localized PCEs (Fig. 3.2c) with respect to the Θ criterion. Localized approximations are represented by low-order PCEs (e.g. $p = 2$), and thus it significantly reduces the size of ED in comparison to a single global high-order PCE. Moreover, the proposed algorithm combined with sparse solvers is extremely efficient for functions with localized non-linearity or discontinuity, especially for high-dimensional examples.

The active learning, various sparse solvers, and the domain decomposition lead to a construction of accurate PCEs for general mathematical models. Once the PCE is available, its specific form and basis functions can be exploited for an efficient UQ. It is well known that PCE is in the identical form as Hoeffding-Sobol decomposition of a function [50], and thus it is possible to analytically derive Sobol' indices and the first two statistical moments directly from the deterministic coefficients [51]. However, UQ in civil engineering is often focused on the estimation of a complete probability distribution of QoI, and so it is also necessary to estimate higher statistical moments as proposed in the third paper attached to this chapter – ***On Distribution-based Global Sensitivity Analysis by Polynomial Chaos Expansion*** [52]. Although the first two statistical moments can be obtained directly from PCE coefficients, getting higher moments is not such a straightforward process. The complication lies in the evaluation of triple and quad products of the basis functions as can be seen in Eq. 3.2. However, triple products can be analytically obtained for several polynomials using the standard linearization problem as shown in the attached paper for Hermite and Jacobi polynomials often used in PCE. The fourth-order product can be further obtained from the third-order product by a mathematical induction. The first four statistical moments can be used for the approximation of a probability distribution of QoI by the commonly known Gram-Charlier expansion or Edgeworth expansion. Finally, motivated by Hoeffding-Sobol decomposition and the recently proposed distribution-based sensitivity methods [53, 54], a novel distribution-based sensitivity analysis derived directly from PCE was proposed. The main concept is depicted in Fig. 3.3. Sensitivity indices are ob-

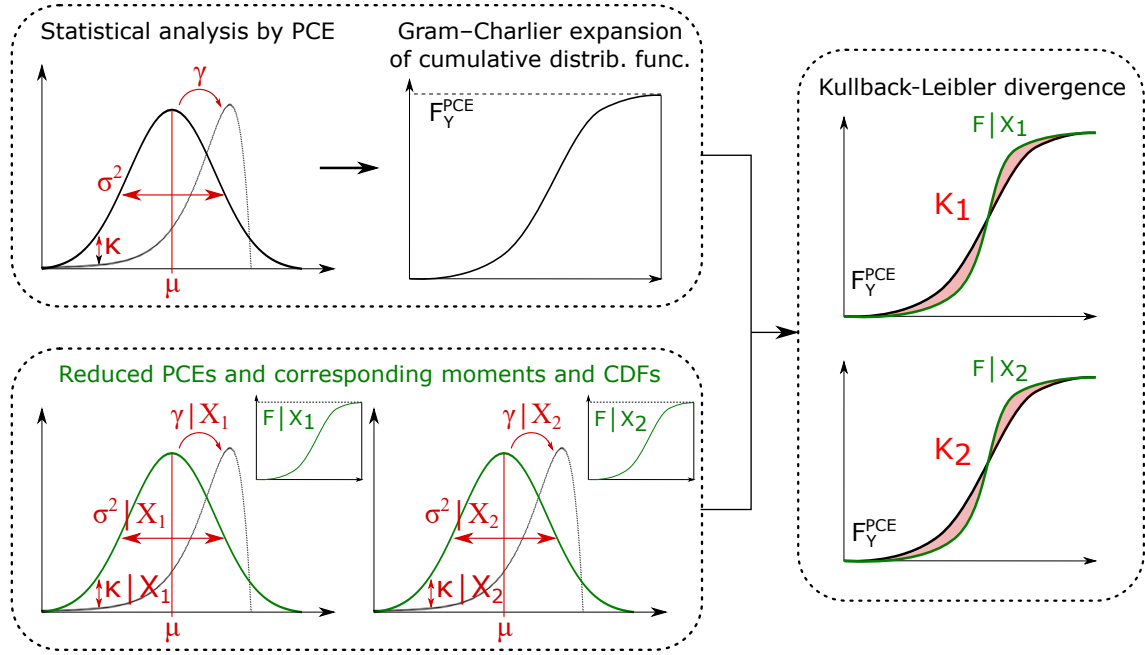


Fig. 3.3: The concept of the distribution-based sensitivity analysis. (Reprinted from [52])

tained as Kullback-Leibler divergence between the original cumulative distribution function (CDF) of QoI and conditional CDF neglecting PCE terms containing a selected input variable. CDFs are approximated by the Gram-Charlier expansion based on the first four statistical moments obtained analytically from the PCE neglecting selected terms – referenced as reduced PCE.

The reduced PCE is a general methodology for the selection of terms in PCE associated with analyzed input variables. Besides its application in distribution-based sensitivity, it can be utilized for sensitivity analysis in the space-time domain. In this case, some input variables are deterministic, though typically represented by uniform distributions and associated Legendre polynomials in the context of PCE. Therefore, it is necessary to filter-out their influence in statistical and sensitivity analysis. Using reduced PCE, it is possible to perform localized UQ for given space-time coordinates. A typical application of this approach is UQ of stochastic partial differential equations (PDE) as shown in the fourth paper attached to this chapter entitled *Physics-Informed Polynomial Chaos Expansions* [55]. Although surrogate models lead to generally accurate approximations, their applications in physics are still limited due to their data-driven nature, i.e. their quality is dependent on the quality of the given ED. The solution of this problem can be seen in the recently proposed physics-informed machine learning [56], improving data-driven approach by known physical principles, to ensure a realistic and physically

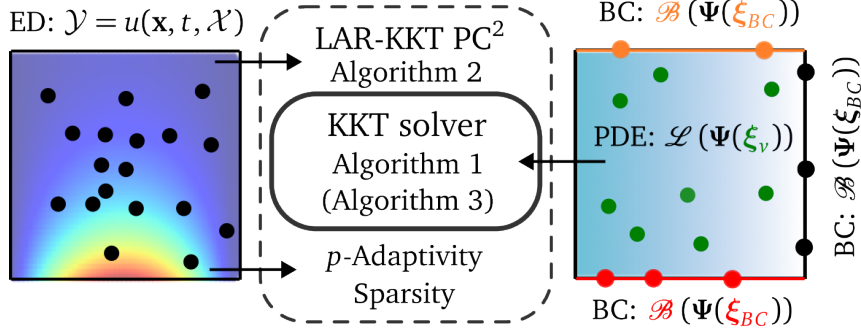


Fig. 3.4: Graphical representation of PC² combining p -adaptivity, LAR algorithm and KKT solver with virtual samples. (Reprinted from [55])

meaningful surrogate model behavior and to reduce training data demands. Following the framework of physics-informed machine learning, the paper presents a novel method referenced as physically constrained PCE (PC²) combining the developed non-intrusive PCE sparse solvers with constrained least squares based on Karush-Kuhn-Tucker (KKT) matrix. The proposed method represents a computationally very efficient solution of PDEs containing random parameters, or stochastic PDEs containing random fields. The methodology is schematically depicted in Fig. 3.4. The PCE is constructed from three separate sets of data points: boundary samples, standard ED containing ground-truth solutions, and virtual samples. While standard ED, containing solutions of the original mathematical model or measurements, is associated with a high computational burden, virtual samples represent only point-wise locations at which the PCE approximation is constrained to satisfy the PDE, and thus they, together with boundary samples, do not bring any additional significant computational cost. Surrogate models constructed by PC² satisfy the given physical constraints and do not lead to physically unrealistic results.

Finally, PC² in combination with reduced PCE is an efficient tool for UQ as evident in the following simple example of stochastic Euler beam with a constant stiffness and distributed load $q(x)$ represented by a Gaussian stochastic process:

$$\frac{\partial^2}{\partial x^2} \left(EI \frac{\partial^2 w}{\partial x^2} \right) = q(x). \quad (3.3)$$

The input random vector contains the first four components of the $q(x)$ obtained by Karhunen-Loève Expansion, and QoI is the deflection of the beam $w(x)$. Since the PCE is constrained by the given differential equation, it is also possible to obtain physically meaningful derivatives corresponding to bending moments, shear forces, and finally also the applied load directly from the PCE function. Reduced PCE methodology is utilized for a localized UQ as can be seen in Fig. 3.5 depicting the mean values and $\pm 3\sigma$ intervals of QoI and its derivatives. Similarly, one can

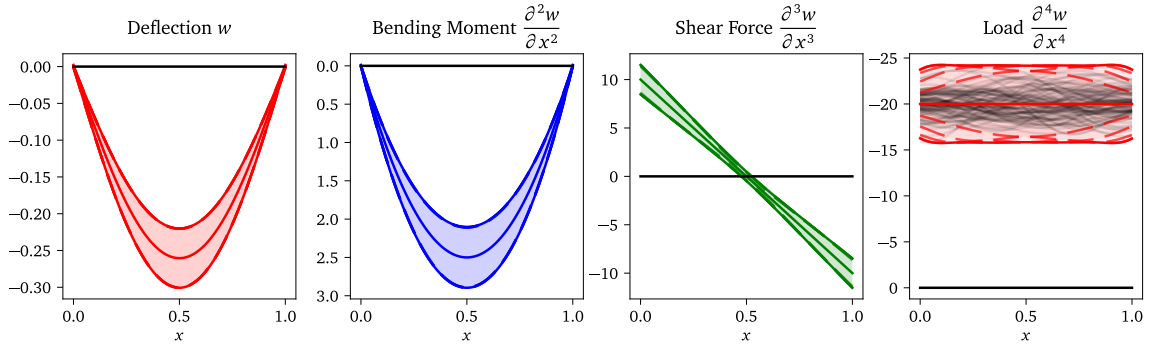


Fig. 3.5: PC² approximation of stochastic Euler beam equation and UQ.

obtain localized sensitivity measures (e.g. Sobol' indices) or higher-order statistical moments.

In order to provide the developed methods to a broad UQ community, all presented methods for adaptive sequential construction of PCE, advanced statistical and sensitivity analysis, and physics-informed PCE are implemented in a general Python package UQPy. The development team consists of researchers from top-tier institutions from the USA and Europe together with professional software developers. The UQPy package contains several modules for general uncertainty quantification, surrogate modeling and machine learning as described thoroughly in the last paper attached to this chapter *UQpy v4.1: Uncertainty Quantification with Python* [57]. The author of this thesis is the main developer of the PCE module containing various techniques for non-intrusive PCE implemented in a user-friendly framework. The UQPy can be found on GitHub repository and PyPI package manager, and it contains extensive documentation with various examples (the links are in the QR codes depicted in Fig. 3.6).



Fig. 3.6: UQPy package repository and documentation.

3.1 Variance-based Adaptive Sequential Sampling for Polynomial Chaos Expansion

NOVÁK, L.; VOŘECHOVSKÝ, M.; SADÍLEK, V.; SHIELDS, M. Variance-based adaptive sequential sampling for Polynomial Chaos Expansion. *COMPUTER METHODS IN APPLIED MECHANICS AND ENGINEERING*, 2021, vol. 386, p. 114105. ISSN: 0045-7825. (WoS-AIS: D1)
DOI: 10.1016/j.cma.2021.114105

Description

The paper presents a novel active learning method for PCE. The proposed technique enables a one-by-one extension of an ED while trying to obtain an optimal sample at each stage of the iterative surrogate model construction process. The proposed criterion for the sample selection balances both exploitation of the surrogate model and exploration of the design domain. The original idea comes from Koksma-Hlawka inequality and its utilization for sequential Monte Carlo sampling. In this paper, the proposed criterion consists of two parts: a local contribution to variance (directly derived from PCE basis functions), and a geometrical term assuring uniform coverage of the whole design domain. It can be seen from the numerical results that the proposed sequential sampling leads to a higher accuracy of the PCE in all the tested examples, including a study in high dimensions. Additionally, it was shown in the paper that the proposed adaptive sequential sampling technique can be used in tandem with any user-defined sampling method and any non-intrusive PCE algorithms.

These obtained results build the foundation for further research in various areas including active learning for multi-variate outputs, domain decompositions etc. The paper has already had a significant impact on the UQ community – 24 citations, according to Scopus.

Role of the author

Percentage of contribution: 33%

Lukáš Novák is the main author of this paper responsible for the concept, the methodology and the numerical results of the presented research. He created the theoretical background of the proposed method in collaboration with Miroslav Vořechovský and Michael D. Shields. The theoretical algorithm was transformed into a Python algorithm by Lukáš Novák with a significant help provided by Václav Sadílek in order to perform the extensive numerical investigation presented.



ELSEVIER



Available online at www.sciencedirect.com

ScienceDirect

Comput. Methods Appl. Mech. Engrg. 386 (2021) 114105

Computer methods
in applied
mechanics and
engineering

www.elsevier.com/locate/cma

Variance-based adaptive sequential sampling for Polynomial Chaos Expansion

Lukáš Novák^{a,*}, Miroslav Vořechovský^a, Václav Sadílek^a, Michael D. Shields^b

^a Brno University of Technology, Brno, Czech Republic

^b Johns Hopkins University, Baltimore, USA

Received 27 March 2021; received in revised form 5 August 2021; accepted 8 August 2021

Available online xxxx

Abstract

This paper presents a novel adaptive sequential sampling method for building Polynomial Chaos Expansion surrogate models. The technique enables one-by-one extension of an experimental design while trying to obtain an optimal sample at each stage of the adaptive sequential surrogate model construction process. The proposed sequential sampling strategy selects from a pool of candidate points by trying to cover the design domain proportionally to their local variance contribution. The proposed criterion for the sample selection balances both *exploitation* of the surrogate model and *exploration* of the design domain. The adaptive sequential sampling technique can be used in tandem with any user-defined sampling method, and here was coupled with commonly used Latin Hypercube Sampling and advanced Coherence D-optimal sampling in order to present its general performance. The obtained numerical results confirm its superiority over standard non-sequential approaches in terms of surrogate model accuracy and estimation of the output variance.

© 2021 Elsevier B.V. All rights reserved.

Keywords: Polynomial Chaos Expansion; Adaptive sampling; Sequential sampling; Coherence optimal sampling

1. Introduction

The Polynomial Chaos Expansion (PCE), originally proposed by Norbert Wiener [1] and further investigated in the context of engineering problems by many researchers, e.g. [2,3], represents a spectral expansion of the original *stochastic problem* in a polynomial basis. PCE approximation represents very efficient method for sensitivity analysis, uncertainty quantification or reliability analysis [4]. Moreover, once the PCE is available, it is possible to investigate the constructed explicit function in order to estimate additional information about the original problem including its statistical moments, output probability distribution or sensitivity indices without additional sampling [5], which is especially beneficial in industrial applications [6,7]. PCE can be generally formulated in intrusive or non-intrusive form. Despite the recent progress in research on the intrusive approach [8], it is still rarely employed in practical applications since it requires redesign of the mathematical model solver.

On the other hand, the non-intrusive approach offers a convenient way to perform probabilistic analysis of any black-box model. There are generally two types of non-intrusive methods for calculation of deterministic

* Corresponding author.

E-mail addresses: novak.l@fce.vutbr.cz (L. Novák), vorechovsky.m@vut.cz (M. Vořechovský), sadilek.v@fce.vutbr.cz (V. Sadílek), michael.shields@jhu.edu (M.D. Shields).

<https://doi.org/10.1016/j.cma.2021.114105>

0045-7825/© 2021 Elsevier B.V. All rights reserved.

coefficients: spectral projection and linear regression. The spectral projection approach utilizes the orthogonality of multivariate polynomials and calculates the coefficients using inner products. Although the integrals in spectral projection can be calculated by traditional tensor-product quadrature rules, the number of collocation points grows exponentially with the number of input random variables which is called *curse of dimensionality* and thus computationally far more efficient sparse grids [9,10] should be employed. The second type of the non-intrusive approach is based on linear regression. Although it is typically less expensive than spectral projection (the number of samples should be at least $\mathcal{O}(P \ln(P))$, where P is the number of terms in PCE [11,12]) it suffers from the curse of dimensionality as well, since number of PCE terms is extremely large for high dimensions and high polynomial orders. Therefore, it is necessary to employ advanced adaptive techniques for construction of sparse PCE in order to obtain efficient solutions for real-life physical systems. Moreover, the regression based PCE can be significantly affected by sampling schemes as was recently shown in extensive review paper [13]. Therefore, this paper is focused on the combination of PCE adaptivity with a sequential sampling strategy designed for the non-intrusive approach based on linear regression.

Since each evaluation of a computer model representing the engineering problem is typically highly time-consuming (e.g. non-linear finite element method), it is necessary to reduce the number of model evaluations as much as possible in the process of training the surrogate model, while maintaining the accuracy of the approximation. The balance between accuracy and computational requirements is strongly connected to the selection of the support points in the *design domain* of input variables — a computational design of experiments (DoE). Besides the commonly used crude Monte Carlo sampling, there are several advanced techniques developed in the fields of statistical/numerical estimation of integrals which improve the efficiency and accuracy of the DoE. One of the most widely used techniques is Latin Hypercube Sampling (LHS) [14], a variance reduction technique that uses stratified selection of sampling points. Another popular strategy for DoE is to uniformly fill the design domain according to some *space-filling* criteria such as miniMax, Maximin [15] or generalized versions of distance-based criteria [16,17], or to decrease the *discrepancy* of the point set. Low discrepancy designs can be obtained either by direct algorithmic minimization of selected discrepancy measure or as Quasi Monte Carlo sequences (known also as low-discrepancy sequences, or number-theoretical designs; see e.g. the sequence due to Halton [18,19], Sobol' [20,21], Niederreiter [22–24], Faure [25], the generalization of the Faure sequences by Tezuka [26], and others). These techniques for DoE can be used for general probabilistic analysis (=numerical integration) without any knowledge about the specific mathematical model or surrogate model.

Further it is often beneficial to include additional information into the DoE stemming from the specific type of the surrogate model at hand. The coherence-based sampling was proposed specifically for PCE constructed by ordinary least squares regression (OLS) [27] and it leads to higher stability in estimation of PCE coefficients in comparison to general sampling methods such as LHS, which is commonly used in combination with PCE. Another method developed specifically for OLS is induced sampling [28], which has been proved to be optimal for weighted least-squares methods.

Methods for DoE construction usually need to specify the number of simulations *a priori*. It is however much more efficient and practical to sample additional points one-by-one until desired accuracy of the approximation is reached. Such methodology for sample size extension is referred to as sequential sampling and it is especially beneficial in practical engineering applications. Sequential sampling schemes are often driven by a defined criterion to compare candidates for sample size extension. The concept of adaptive experimental design for learning surrogate models is often termed *active learning*. This approach is a common approach when the goal is reliability analysis with a surrogate: an initial experimental design is iteratively updated based on the current estimation of the limit-state surface in an active learning algorithm [29–31]. Active learning approach involving PCE in the context of reliability analysis was used e.g. in [32–34].

Although there are recent studies focused on general sequential sampling based on space-filling criteria or alphabetical optimality used for PCE [35,36], it is beneficial to use both *exploitation* (leveraging model behavior) criteria and *exploration* (space filling) criteria for definition of an optimally balanced criterion [37]. Such sequential sampling for sparse Bayesian learning PCE combining both aspects — epistemic uncertainty of the statistical inference (exploration) together with quadratic loss function (local exploitation) was recently proposed [38]. However, its application is limited to PCE build by sparse Bayesian learning only. This paper presents a novel adaptive sequential sampling technique with such a balanced criterion. The technique presented in this paper can be coupled with any sparse regression solver and common methods for DoE such as LHS and thus can be easily

implemented into the existing software solutions for PCE construction (e.g. [39–43]). Additionally, in order to increase the efficiency of the proposed scheme, the developed technique is coupled with coherence D-optimal sampling created specifically for non-intrusive PCE solved by OLS and thus all parts of the method are designed in order to increase the efficiency and accuracy of this particular type of surrogate model.

2. Polynomial Chaos expansion

Assume a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where Ω is an event space, \mathcal{F} is a σ -algebra on Ω (collection of subsets closed under complementation and countable unions) and \mathcal{P} is a probability measure on \mathcal{F} . If the input variable of a mathematical model, $Y = g(X)$, is a random variable $X(\omega)$, $\omega \in \Omega$, the model response $Y(\omega)$ is also a random variable. Assuming that Y has a finite variance, PCE represents the output variable Y as a function of an another random variable ξ called the *germ* with given distribution

$$Y = g(X) = g^{\text{PCE}}(\xi), \quad (1)$$

and representing the function $g(X)$ via polynomial expansion in a manner similar to the Fourier series of a periodic signal. A set of polynomials, orthogonal with respect to the distribution of the germ, is used as a basis of the Hilbert space $L^2(\Omega, \mathcal{F}, \mathcal{P})$ of all real-valued random variables of finite variance, where \mathcal{P} takes over the meaning of the probability distribution. The orthogonality condition for all $j \neq k$ is given by the inner product of $L^2(\Omega, \mathcal{F}, \mathcal{P})$ defined for any two functions ψ_j and ψ_k with respect to the weight function p_ξ (probability density function of ξ) as:

$$\langle \psi_j, \psi_k \rangle = \int \psi_j(\xi) \psi_k(\xi) p_\xi(\xi) d\xi = 0. \quad (2)$$

This means that there are specific orthogonal polynomials associated with the corresponding distribution of the germ via its weighting function. For example, Hermite polynomials orthogonal to the Gaussian measure are associated with normally distributed germs. Orthogonal polynomials corresponding to other distributions can be chosen according to Wiener–Askey scheme [44]. For further processing, it is beneficial to use normalized polynomials (orthonormal), where the inner product is equal to the Kronecker delta δ_{jk} , i.e. $\delta_{jk} = 1$ if and only if $j = k$, and $\delta_{jk} = 0$ otherwise

$$\langle \psi_j, \psi_k \rangle = \delta_{jk}. \quad (3)$$

In the case of \mathbf{X} and $\boldsymbol{\xi}$ being vectors containing M independent random variables, the polynomial $\Psi(\boldsymbol{\xi})$ is multivariate and it is built up as a tensor product of univariate orthogonal polynomials as

$$\Psi_\alpha(\boldsymbol{\xi}) = \prod_{i=1}^M \psi_{\alpha_i}(\xi_i), \quad (4)$$

where $\boldsymbol{\alpha} \in \mathbb{N}^M$ is a set of integers called the *multi-index*. The quantity of interest (QoI), i.e. the response of the mathematical model $Y = g(\mathbf{X})$, can then be represented, according to Ghanem and Spanos [3], as

$$Y = g(\mathbf{X}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} \beta_\alpha \Psi_\alpha(\boldsymbol{\xi}), \quad (5)$$

where β_α are deterministic coefficients and Ψ_α are multivariate orthogonal polynomials.

2.1. Non-intrusive computation of PCE coefficients

For practical computation, PCE expressed in Eq. (5) must be truncated to a finite number of terms P . The truncation is commonly achieved by retaining only terms whose total degree $|\boldsymbol{\alpha}|$ is less than or equal to a given p . Therefore, the truncated set of PCE terms is then defined as

$$\mathcal{A}^{M,p} = \left\{ \boldsymbol{\alpha} \in \mathbb{N}^M : |\boldsymbol{\alpha}| = \sum_{i=1}^M \alpha_i \leq p \right\}. \quad (6)$$

The cardinality of the truncated *index set* $\mathcal{A}^{M,p}$ is given by

$$\text{card } \mathcal{A}^{M,p} = \frac{(M+p)!}{M! p!}. \quad (7)$$

Moreover, in practical applications, it is beneficial to prefer only basis functions with lower number of interaction terms. Therefore, it was proposed by Blatman and Sudret [2] to create a PCE basis by a “hyperbolic” truncation scheme:

$$\mathcal{A}^{M,p,q} = \left\{ \boldsymbol{\alpha} \in \mathbb{N}^M : \|\boldsymbol{\alpha}\|_q \equiv \left(\sum_{i=1}^M \alpha_i^q \right)^{1/q} \leq p \right\}. \quad (8)$$

Note that selection of $q = 1$ corresponds to the standard truncation scheme according to Eq. (6) and, for $q < 1$, terms representing higher-order interactions are eliminated. Such an approach leads to a dramatic reduction in the cardinality of the truncated set for high total polynomial orders p and high dimensions M .

When PCE is truncated to a finite number of terms, there is an error ε of the approximation such that

$$Y = g(\mathbf{X}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \beta_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) + \varepsilon.$$

From a statistical point of view, PCE is a simple linear regression model with intercept. Therefore, it is possible to use *ordinary least square* (OLS) regression to minimize the error ε

$$\boldsymbol{\beta} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^P} \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} [\boldsymbol{\beta}^T \boldsymbol{\Psi}(\boldsymbol{\xi}^{(i)}) - g(\mathbf{x}^{(i)})]^2. \quad (9)$$

Knowledge of vector $\boldsymbol{\beta}$ fully characterizes the approximation via PCE. To solve for $\boldsymbol{\beta}$, first it is necessary to create n_{sim} realizations of the input random vector \mathbf{X} and the corresponding results of the original mathematical model \mathcal{Y} , together called the experimental design (ED). Then, the vector of P deterministic coefficients $\boldsymbol{\beta}$ is calculated as

$$\boldsymbol{\beta} = (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathcal{Y}, \quad (10)$$

where $\boldsymbol{\Psi}$ is the data matrix

$$\boldsymbol{\Psi} = \{ \Psi_{ij} = \Psi_j(\boldsymbol{\xi}^{(i)}), i = 1, \dots, n_{\text{sim}}, j = 0, \dots, P - 1 \}. \quad (11)$$

Note that the number of terms P is highly dependent on the number of input random variables M and the maximum total degree of polynomials p . Estimation of $\boldsymbol{\beta}$ by regression then needs at least the number of samples $\mathcal{O}(P \ln(P))$ for stable solution [11,12]. Therefore, in case of a large stochastic model, the problem can become computationally highly demanding. However, one can utilize advanced model selection algorithms such as Least Angle Regression (LAR) [45] to find an optimal set of PCE terms and thus reduce the number of samples needed to compute the unknown coefficients if the true coefficient vector is sparse or compressible as proposed by Blatman and Sudret [2]. Note that beside LAR, there are other best model selection algorithms such as orthogonal matching pursuit [46] or Bayesian compressive sensing [47] with comparable numerical results. The sparse set of basis functions obtained by any adaptive algorithm is further denoted for the sake of clarity as \mathcal{A} .

2.2. Estimation of approximation error

Once the PCE is constructed, it is crucial to estimate its accuracy. Further, the accuracy of PCE can be used for the direct comparison among several PCEs in order to choose the best surrogate model. Therefore it is beneficial to use methods which do not need any additional sampling of the original mathematical model. A common choice is the coefficient of determination R^2 , which is well known from machine learning. However, R^2 may lead to overfitting and thus advanced methods should be used. One of the most utilized methods for measuring the performance of the learning algorithm in recent years is the leave-one-out cross validation error Q^2 . This statistic is based on residuals between the original surrogate model and the surrogate model built with the ED while excluding one realization. This approach is repeated for all realizations in the ED and the average error is estimated. Although the calculation of Q^2 is typically highly time-consuming, it is possible to obtain results analytically from a single PCE as follows [48]

$$Q^2 = 1 - \frac{\frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \left[\frac{g(\mathbf{x}^{(i)}) - g^{\text{PCE}}(\mathbf{x}^{(i)})}{1 - h_i} \right]^2}{\sigma_{Y, \text{ED}}^2}, \quad (12)$$

where $\sigma_{Y,ED}^2$ is a variance of experimental design calculated using the original mathematical model and h_i represents the i th diagonal term of matrix $\mathbf{H} = \Psi (\Psi^T \Psi)^{-1} \Psi^T$.

2.3. Statistical moments derived from PCE

The specific form of PCE together with the orthogonality of the polynomials allows for a powerful and efficient post-processing. Once a PCE approximation is created, it is possible to obtain statistical moments of the QoI. Generally, its raw statistical moment of the m th order is defined as

$$\begin{aligned} \langle Y^m \rangle &= \int [g(X)]^m p_X(X) dX = \int \left[\sum_{\alpha \in \mathbb{N}^M} \beta_\alpha \Psi_\alpha(\xi) \right]^m p_\xi(\xi) d\xi \\ &= \int \sum_{\alpha_1 \in \mathbb{N}^M} \dots \sum_{\alpha_m \in \mathbb{N}^M} \beta_{\alpha_1} \dots \beta_{\alpha_m} \Psi_{\alpha_1}(\xi) \dots \Psi_{\alpha_m}(\xi) p_\xi(\xi) d\xi \\ &= \sum_{\alpha_1 \in \mathbb{N}^M} \dots \sum_{\alpha_m \in \mathbb{N}^M} \beta_{\alpha_1} \dots \beta_{\alpha_m} \int \Psi_{\alpha_1}(\xi) \dots \Psi_{\alpha_m}(\xi) p_\xi(\xi) d\xi. \end{aligned} \quad (13)$$

As can be seen from the final part of the formula, in case of PCE, it is necessary to integrate only over the basis functions (orthonormal polynomials), which leads to a dramatic simplification in comparison to the integration of the original mathematical function. Moreover, it is also possible to write an analytical expression of the integral in several cases. Specifically, the first statistical moment (mean value) is obtained as

$$\mu_Y = \langle Y^1 \rangle = \int \left[\sum_{\alpha \in \mathbb{N}^M} \beta_\alpha \Psi_\alpha(\xi) \right] p_\xi(\xi) d\xi = \sum_{\alpha \in \mathbb{N}^M} \beta_\alpha \int \Psi_\alpha(\xi) p_\xi(\xi) d\xi. \quad (14)$$

Considering the orthonormality of the polynomials

$$\int \Psi_\alpha(\xi) p_\xi(\xi) d\xi = 0 \quad \forall \alpha \neq 0, \quad \Psi_0 \equiv 1,$$

it is possible to obtain the mean value directly from the PCE deterministic coefficients. Namely, the mean value is equal to the first deterministic coefficient of the expansion

$$\mu_Y = \langle Y^1 \rangle = \beta_0. \quad (15)$$

The second raw statistical moment, $\langle Y^2 \rangle$, is written as

$$\begin{aligned} \langle Y^2 \rangle &= \int \left[\sum_{\alpha \in \mathcal{A}} \beta_\alpha \Psi_\alpha(\xi) \right]^2 p_\xi(\xi) d\xi = \sum_{\alpha_1 \in \mathcal{A}} \sum_{\alpha_2 \in \mathcal{A}} \beta_{\alpha_1} \beta_{\alpha_2} \int \Psi_{\alpha_1}(\xi) \Psi_{\alpha_2}(\xi) p_\xi(\xi) d\xi \\ &= \sum_{\alpha \in \mathcal{A}} \beta_\alpha^2 \int \Psi_\alpha(\xi)^2 p_\xi(\xi) d\xi = \sum_{\alpha \in \mathcal{A}} \beta_\alpha^2 \langle \Psi_\alpha, \Psi_\alpha \rangle. \end{aligned} \quad (16)$$

Considering again the orthonormality of the polynomials, defined by the inner product in Eq. (3), it is possible to obtain the variance $\sigma_Y^2 = \langle Y^2 \rangle - \mu_Y^2$ as the sum of all squared deterministic coefficients except the intercept (which represents the mean value), i.e.

$$\sigma_Y^2 = \sum_{\substack{\alpha \in \mathcal{A} \\ \alpha \neq 0}} \beta_\alpha^2. \quad (17)$$

Note that, higher statistical central moments skewness γ_Y (3rd moment) and kurtosis κ_Y (4th moment) need precomputing of the triple and quadruple products.

In the following, we select the variance of the output variable, i.e. σ_Y^2 in Eq. (17), as the *target characteristic* of Y and we focus on development of the sequential sampling strategy in order to estimate this variance as accurately as possible at any stage of PCE build. The reason for selection of variance is that we expect a monotonic relationship between function *variation* in the sense of Hardy and Krause [49] and its *variance*. As will be shown below, Eq. (16) can be modified in order to obtain the local contribution to variance. We conjecture that it is therefore

important to place samples densely in regions of large local variance and sparsely in regions of small local variance, in order to obtain a near-optimal sample [37]. This arrangement of samples may decrease the error of the function approximation and its integral.

3. Sampling methods

Assuming a non-intrusive approach for calculation of the PCE coefficients using OLS defined in Eq. (9), it is necessary to create an ED containing n_{sim} realizations of the input random vector and the vector of corresponding results of the original mathematical model. Typically we consider sampling of ξ using its density f_{ξ} , which represents the distribution according to the Wiener–Askey scheme. For the M -dimensional Legendre polynomials this means sampling uniformly from the M -dimensional hypercube $[-1, 1]^M$ and for M -dimensional Hermite polynomials, it corresponds to sampling from the M -dimensional Gaussian distribution with independent standardized Gaussian marginal distributions. Naturally, it is crucial to use an efficient sampling scheme for the DoE of ξ in order to obtain accurate results for a given computational budget.

DoE has been an area of interest for many researchers since the beginning of uncertainty quantification and structural reliability. The most simple but generally applicable method is crude Monte Carlo Sampling (MC), i.e. a method associated especially with robust (and inefficient) numerical integration. In standard Monte Carlo integration, the important condition of integration being unbiased is that the sample is selected *uniformly* and *independently* with respect to the target density. In MC the sampling points are selected independently of each other and therefore, clusters of points emerge randomly as well as regions which are not covered by any point.

There has been considerable effort spent on improving the spatial arrangement of points in a sample. The Koksma–Hlawka inequality [49], which was developed to predict an upper bound of integration error, motivates the decrease in *discrepancy* of the sample set (ED). Discrepancy in a way measures *uniformity* of a point set, i.e. the difference between the desired uniform distribution and the empirical distribution of the point set. Such uniformity of point distribution may be useful also for initial screening or building a surrogate model. Low discrepancy can be achieved either by direct minimization of a suitable discrepancy measure [17,50–52] or simply by using various Quasi Monte Carlo sequences mentioned in the Introduction section. Quasi Monte Carlo sequences are deterministic point sets and they allow for sequential addition of points one-by-one while retaining an optimal rate of star discrepancy decrease with increasing sample size [53].

Another branch of research focuses on variance reduction techniques such as *importance sampling* which place points according to a predefined or adaptively adjusted *sampling* density which can be different from the target density or methods of stratified selection of sampling points that improve spatial arrangement of the sampling points. One of the most widely used techniques is Latin Hypercube Sampling (LHS) first suggested by Conover [14]; see also [54]. LHS specifically has the effect of reducing variance associated with the additive components of a transformation. Hence, for functions that are dominated by the main effects of the individual variables, LHS will significantly reduce variance. For functions with significant variable interactions, it is less effective. [55–58].

Another important aspect of sample selection for DOE is the uniformity of filling of the design domain. This aspect is important for *function approximation* and resulted in the seminal works on *space-filling* criteria based on mutual distances among points: the Maximin and miniMax criteria [15]. These criteria prefer designs without point clusters or without large empty areas. A generalization of the Maximin criterion to the phi-criterion [16] was presented along with a heuristic construction algorithm that can be combined with the LHS strategy. Similarly, the miniMax criterion can be generalized [59]. Moreover, the recently developed *periodic* versions of the whole class of distance-based criteria [17,60] guarantee *statistically uniform* distribution of points along with *even point distribution* within each single design. These distance-based criteria can be employed for direct design optimization.

Uniformity of ED and space-filling criteria are important characteristics to obtain a quality ED for PCE. The LHS thus typically leads to more accurate estimation of β in comparison to crude MC [61,62]. Sampling for PCE, however, might be motivated by additional criteria of optimality than the space-filling property, discrepancy or statistical uniformity with respect to the target distribution. The optimality criterion should consider aspects related to the particular method of identification of the polynomial modes and their coefficients.

This naturally leads to sampling from a probability measure different from the input distributions, which minimize a selected characteristic of the ED. Moreover, it is also necessary to modify the basis functions in order to preserve orthonormality of the data matrix. Minimizing the coherence parameter of a PCE basis functions [63] leads to the coherence optimal and related asymptotic sampling [27], which is theoretically more efficient in comparison

to standard sampling based on the input distribution. Similarly the Christoffel sparse approximation [64] was derived using a different definition of the coherence parameter. Note that these sampling strategies are derived for a specific purpose (non-intrusive PCE) and thus their efficiency in general probabilistic analysis might be unsatisfying. Although coherence optimal sampling is more computationally demanding, it is not a crucial problem in engineering, where the calculation of mathematical models takes the largest part of the whole process.

Also, it would be optimal if the PCE that has been set at any stage of sampling based on the available information about the samples and the corresponding function values $g(\mathbf{x})$, can propose the new sampling point (sample size extension). With such an algorithm at hand, one can build the PCE approximation incrementally while exploiting all the knowledge available so far. For efficient accurate exploitation, it is necessary to use an adaptive algorithm for PCE construction (by *adaptive* we mean selection of the most important combinations of modes in the index set \mathcal{A}). There are several methods for adaptive selection of the optimal PCE basis functions such as LAR [45] employed in numerical experiments. The selected basis functions are further used for the process of exploitation and thus it is crucial to identify new basis functions in every iteration of the sequential sampling in order to obtain the best possible location for a new sampling point. Moreover, the adaptivity of the basis functions is important for the accurate coherence optimal sampling briefly described in the following paragraphs.

3.1. Coherence-optimal sampling

Generally, it is beneficial to take all pieces of information about the given mathematical task into account in order to choose a correct methodology. Although standard sampling based on the input distribution is suitable for any probabilistic analysis, there are more efficient methods developed specifically for PCE solved by over-determined OLS [27], which is employed in this paper.

Coherence-optimal sampling constructs a new sampling measure minimizing a coherence parameter associated with stability and convergence of the PCE solved by OLS. The coherence parameter $\mu(Y)$ is defined as [65]

$$\mu(Y) := \sup_{\xi} \sum_{j=1}^P |w(\xi)\psi_j(\xi)|^2, \quad (18)$$

where the weight function $w(\xi)$ is

$$w(\xi) := \frac{P}{B(\xi)}. \quad (19)$$

The analytical expression of $B(\xi)$ is generally not available, but it is possible to evaluate its value for arbitrary ξ as

$$B(\xi) = \sqrt{\sum_{j=1}^P |\psi_j(\xi)|^2}. \quad (20)$$

Finally, the coherence-optimal probability measure is defined as

$$f_{\text{coh}}(\xi) := P^{-1} f(\xi) B^2(\xi). \quad (21)$$

In order to sample from $f_{\text{coh}}(\xi)$ one may use Markov Chain Monte Carlo (MCMC) [65]. The proposal distribution for MCMC is suggested for case when $p \leq M$ where one should use standard sampling from distributions naturally orthogonal to the employed polynomial basis as already described for standard sampling. If $p \geq M$, samples should be independently drawn from a uniform distribution on a M -dimensional ball of radius $\sqrt{2}\sqrt{2p+1}$ for Hermite polynomials, and a M -dimensional Chebyshev distribution for Legendre polynomials.

Note that the coherence-optimal sampling does not generally lead to the orthonormal columns of the data matrix and the deviation $\langle \Psi_{\alpha_1}, \Psi_{\alpha_2} \rangle - \delta_{\alpha_1\alpha_2}$ may be significant. Therefore, one has to apply a weight function $w(\xi)$ to the basis functions and use weighted least squares in order to estimate β using coherence-optimal sampling as described in previous paragraphs. Weighted least squares method is a special case of generalized least squares containing only diagonal members of the correlation matrix of residuals. Optimal values of the deterministic coefficients is then obtained by solution of the following system of equations:

$$\mathbf{W} \Psi \beta \approx \mathbf{W} \mathcal{Y}, \quad (22)$$

where matrix \mathbf{W} is a diagonal matrix with $W_{i,i} = w(\xi^{(i)})$.

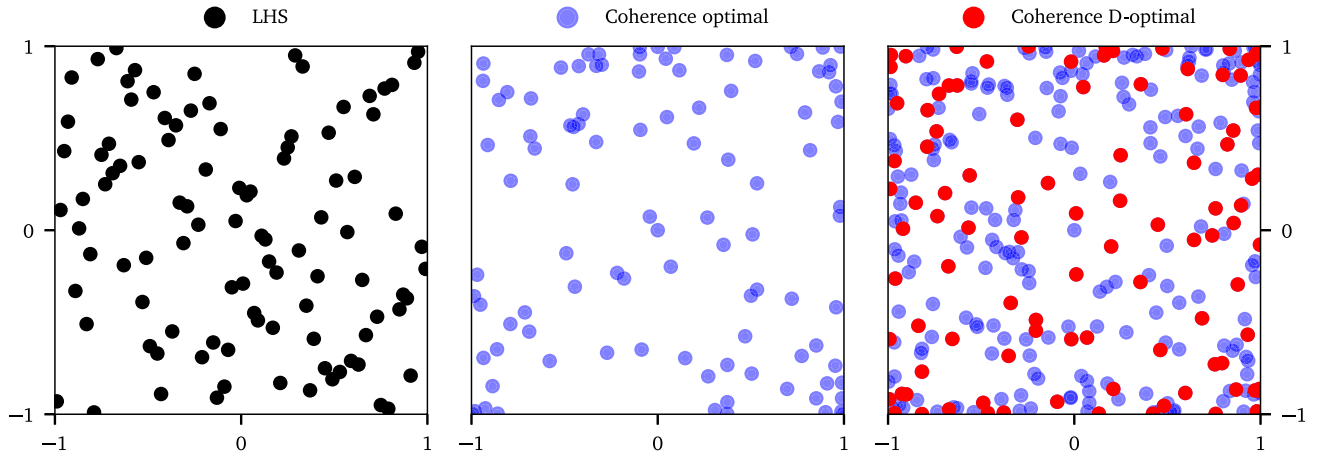


Fig. 1. An ED containing 100 samples generated by LHS, coherence-optimal sampling and coherence D-optimal sampling for Legendre polynomials.

3.2. D-optimal experimental design

Optimality of the ED for OLS can also be measured by the so-called alphabetic criteria of the information matrix $\mathcal{I} := \frac{1}{n_{\text{sim}}} \Psi^T \Psi$, which is crucial for stability of OLS. There are several types of criteria focused on various characteristics of the information matrix; see [62] for a review of criteria in the context of PCE. Although one of the most promising is the S-optimal criterion [35], it is also highly computationally demanding especially for large M . Therefore, in this paper we use the following cheaper and well known estimation-oriented criterion D-optimal design, which is focused on accurate estimation of β .

Since the PCE basis functions are orthonormal, \mathcal{I} is on average identity, but for finite sample size n_{sim} there is a deviation from the identity matrix

$$\|\mathcal{I} - \mathbf{I}\| > 0. \quad (23)$$

The D-optimal design, obtained by maximizing the determinant of the information matrix, leads to small deviation and thus stable estimation of β . For the practical construction of D-optimal ED, it is possible to employ Fedorov exchange algorithm [66], greedy algorithm [67] or rank revealing QR decomposition [68] to find an optimal information matrix containing n_{sim} rows out of a candidate pool containing $n_{\text{pool}} \gg n_{\text{sim}}$ rows.

3.3. Coherence D-optimal experimental design

As originally proposed together with coherence optimal sampling [68], it is beneficial to merge the previous two techniques in order to obtain a *Coherence D-optimal* (Coh D-opt) ED for the stable and accurate solution of WLS. Although Coh D-opt was originally proposed for compressed sensing in the context of PCE [68], D-optimality was already employed in combination with standard sampling (LHS and MC) for OLS and LAR in [35]. Therefore, Coh D-opt can similarly be used also for the non-intrusive approach based on WLS and LAR as employed in this paper. The pool of candidates is generated by coherence optimal sampling and further reduced by D-optimality criterion which leads to uniform ED without clusters of samples as can be seen in Fig. 1.

4. Adaptive sequential sampling

In industrial applications, it is often not feasible to perform a large number of evaluations of the original mathematical model (e.g. FEM) and thus it is important to reduce the number of simulations as much as possible. An efficient approach therefore is *adaptive sequential sampling*, which uses iterative selection of the new sampling points according to specific criteria while exploiting the already available information. Although general sequential sampling is an area of interest for many researchers [69–72], there is still a lack of studies focused specifically on PCE. The recent study [35] compared several simple sequential sampling methods based on D/S-optimality of samples or maximin criterion of samples.

Note that there are two different strategies for sequential sampling. The first is to enrich the initial ED according to a space-filling criterion (exploration) without assuming any knowledge of the mathematical model or PCE form; see e.g. [71,73]. The motivation is clear: we do not want to locate an augmented point very close to an existing point to avoid getting redundant information in the nearby region. However, by obtaining data sequentially, it is possible to learn from the early stages to inform subsequent data collection, minimize wasted resources, and provide answers for various objectives (exploitation). Therefore, the second strategy works with the structure of the PCE (basis functions) in order to identify an optimal sample. Unfortunately, in situations when the initial screening overlooks a globally important region, the exploitation criterion may continue refinement of some other, locally important region that was detected, and there is a risk of never discovering a globally important region. Therefore, it is beneficial to include a balance between both criteria in search for a suitable candidate. Note that, such approach was employed e.g. in [37] in a different context: a criterion motivated by the Koksma–Hlawka inequality [49] was proposed and coupled with stratified sampling in order to improve the efficiency of statistical integration.

As discussed above, the *adaptivity* feature of the PCE surrogate model can be ensured by any model selection algorithm. Moreover, it can be combined also with hyperbolic truncation according to Eq. (8), which is efficient for high P . A general adaptive sequential algorithm thus should adaptively reconstruct the PCE using model selection algorithms in order to identify a sparse set of basis functions \mathcal{A} in each iteration.

The *sequential* feature can be added by using a comparison criterion for selection of the best candidate from a pool of candidates while balancing between exploration and exploitation. *Exploitation* of the local areas of the design domain is focused on identification of sub-domains associated with a defined characteristic of the mathematical model such as high gradient, local maxima etc. The candidates from the identified sub-domains are further preferred. Another typical example can be identification of sub-domains associated with high variation of the mathematical model. Although exploitation is a powerful technique for identification of the best candidate, it is typically based on a built surrogate model and thus it is highly dependent on the quality of a given ED. On the other hand, *exploration* assures uniform coverage of the whole design domain, possibly with respect to specific characteristic as in case of alphabetical optimality [74], and it assures that the algorithm does not get stuck in local minima. It opens the door to detection and exploration of important areas in design domain, where the behavior of the studied function $g(\mathbf{x})$ might be significantly different from possibly incorrect expectation based on the surrogate.

4.1. The proposed Θ criterion for sequential sampling

We propose an adaptive sequential sampling strategy accompanied by a criterion designed for efficient and accurate estimation of β using least squares. Consider a pool of candidates containing n_{pool} realizations of the random vector ξ generated by an arbitrary sampling technique. Once the pool of candidates conditioned by the selected PCE basis is generated, it is necessary to construct a criterion for the selection of the best candidate balancing between the exploitation and exploration of the design domain. Such criterion, called the Θ criterion from here on, is proposed as follows

$$\Theta(\xi^{(c)}) \equiv \Theta_c = \underbrace{\sqrt{\sigma_{\mathcal{A}}^2(\xi^{(c)}) \cdot \sigma_{\mathcal{A}}^2(\xi^{(s)})}}_{\text{ave variance density}} \underbrace{l_{c,s}^M}_{\text{vol.}} \equiv \sqrt{\sigma_c^2 \cdot \sigma_s^2} l_{c,s}^M. \quad (24)$$

where we introduce an abbreviated notation by dropping the point designation $\xi^{(c)}$ and using simply the lower index instead and also by dropping the set of basic functions, \mathcal{A} , which is selected for every instant of the algorithm and thus may differ as the sample size increases. The criterion has an intuitive meaning and also has units of *variance* and is a product of two parts: the exploitation part (denoted as “ave variance density”) and the exploration part (the distance term $l_{c,s}$ raised to the domain dimension). Multiplication of these two independent contributions maintains the optimal balance between exploration and exploitation.

The *exploration* aspect is maintained by accounting for the distance $l_{c,s}$ between a candidate $\xi^{(c)}$ and its nearest neighboring point from the existing ED, $\xi^{(s)}$. For the distance term a suitable metric must be selected. In this paper, we select the Euclidean distance between the candidate and its nearest neighbor as

$$l_{c,s} = \sqrt{\sum_{i=1}^M |\xi_i^{(c)} - \xi_i^{(s)}|^2}, \quad (25)$$

although other distances can be considered, particularly in high dimension.

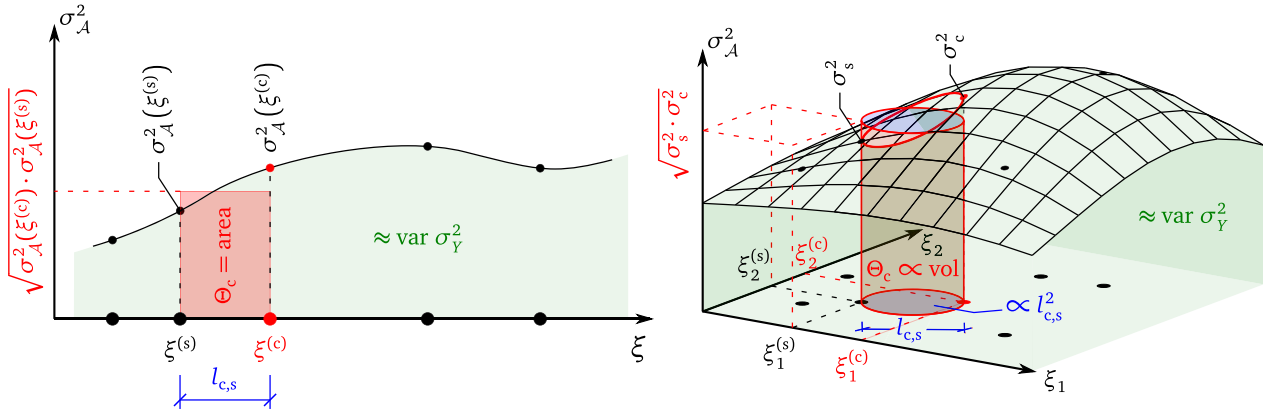


Fig. 2. Geometrical meaning of the proposed Θ criterion for a candidate “c” in two and three dimensions. Black solid circles are existing points and point “s” is the nearest neighbor to candidate “c”. The solid thin curve/surface represents the current estimation of variance density over the design domain.

The *exploitation* in candidate selection is motivated by our desire to *uniformly cover local contributions to the total variance*, σ_Y^2 . By recalling Eq. (17), we know that σ_Y^2 can be estimated simply as the sum of all squared deterministic coefficients except the intercept. The mean square can be obtained as the integral featuring the selected polynomial basis over the design domain; see Eq. (16). This means that the variance can be thought of as an integral of local contributions over the design domain indexed by coordinates ξ . In other words, we need to integrate a local *variance density* $\sigma_A^2(\xi)$. Once the PCE has been established at any given stage of the algorithm, the variance density is computationally cheap to evaluate for any location ξ as

$$\sigma_A^2(\xi) = \left[\sum_{\substack{\alpha \in \mathcal{A} \\ \alpha \neq 0}} \beta_\alpha w(\xi) \Psi_\alpha(\xi) \right]^2 p_\xi(\xi). \tag{26}$$

The local variance is therefore estimated based on the basis functions and coefficients β of the PCE. Depending on technique utilized for sampling of candidates, one should apply the weight $w(\xi)$ (also used in weighted least squares) to basis functions in order to reflect influence of sampling from a probability measure different from the input distribution. Specifically in this paper, $w(\xi)$ is defined for Coh D-opt according to Eq. (19) and $w(\xi) = 1$ for LHS. When considering a candidate “c”, one might think about the variance contribution of the region between the candidate and its nearest neighbor. A rough estimation may be obtained by considering an average of local variance densities between the candidate and its nearest neighbor, “s”. This average is represented by the *geometric mean* between the two numbers. The geometric mean between n numbers x_i is defined as $(\prod_{i=1}^n x_i)^{1/n}$ and therefore, we take the geometric mean of two local variance densities simply as the square root of the local variance densities of the candidate and its nearest neighbor, see the first term in Eq. (24). When this geometric mean is multiplied by the M th power of the distance between the two points, $l_{c,s}^M$, the volume (variance contribution) of a neighborhood between them is estimated; see the sketch in Fig. 2. In other words, the criterion estimates the amount of variance in a “bite” by the candidate. Note that Eq. (24) defines the bite as a hypercube of side-length $l_{c,s}$. However, other geometric entities may be considered without practical impact on the algorithm. The reason is that all the geometric volumes for various candidates under comparison would have the same positive multiplier of $l_{c,s}^M$ which can be dropped as it does not change the ranking of the compared candidates. Therefore, we can say that the proposed criterion helps to select a candidate with roughly the largest amount variance being refined. The balance between exploration and exploitation is maintained: a candidate which is close to an existing point can only be selected if the corresponding variance density is large. Similarly, when a region with low contribution is being detected by the PCE, candidates from such regions are ignored.

In situations when the variance density is a constant function, the criterion collapses to a simple space-filling criterion (a form of miniMax criterion); [15,60]. Such criterion ensures the preference of candidates filling the largest empty regions in the design domain and thus leading to uniform distribution of points in the sense of miniMax design criterion. We remark that miniMax design is a preferable choice for the construction of emulators because it minimizes the worst case prediction variance.

A question may arise: why do we propose to use the geometrical mean instead of arithmetical mean? The reason is that the criterion may also be used for infinite design domains (for example in conjunction with Gaussian germ). In such situation, the pool of candidates may contain points $\xi^{(c)}$ that are very far from the mean value and such a point may have (almost) zero variance density. Yet, the criterion would prioritize it due to the large distance from the nearest neighbor $\xi^{(s)}$ as the arithmetical average with its variance density would equal one half of the two local densities. In other words, infinitely distant candidate points would always win the comparison, despite a vanishing contribution of one of them. Using the geometrical mean prevents unimportant distant points with zero density from being selected.

Maximization of the criterion leads to the best candidate, which is added to active ED. As can be seen, the proposed criterion prefers candidate points in parts of multidimensional space associated with higher contribution to the variance of the mathematical model. This idea is similar to the sequential sampling proposed in [37] based on Koksma–Hlawka inequality respecting both variation of the function and discrepancy of realizations, however the proposed criterion is constructed specifically for PCE and thus it can use the PCE basis functions in order to increase the efficiency of the computation. The significant advantage of the proposed method is the ability to add candidates into existing ED one-by-one and thus it can be employed at any moment of the PCE construction process and it can be combined with any sampling algorithm for construction of initial ED marked with subscript as ξ_{ED} , \mathbf{W}_{ED} , \mathcal{Y}_{ED} .

4.2. Adaptive sequential sampling with LHS-based candidates

LHS represents perhaps the most common sampling technique in surrogate modeling and it can be easily coupled with the proposed sequential sampling in a simple manner. The pool of candidates is generated by LHS and the proposed criterion is employed for the selection of the best candidate. This process is repeated at every iteration of the sequential sampling (with the pool being either regenerated or reused from the preceding step). In order to illustrate the proposed sequential adaptive algorithm, we selected five iterations and depict the corresponding states in Fig. 3. The initial ED is represented by solid black circles and the sequentially selected realizations are plotted using solid red circles. The color maps represent the value of the proposed criterion (right column) and also its individual components (the preceding columns). Since the generation of the pool of candidates by LHS is simple and fast, LHS-based adaptive sequential sampling represents an efficient extension to non-sequential LHS and it could be easily implemented into existing software tools.

4.3. Adaptive sequential sampling with coherence-based candidates

The proposed criterion can generally be coupled with any sampling technique. However, since coherence-based sampling is highly affected by the set \mathcal{A} , it might be ideal if the pool of candidates is generated by coherence-based sampling. Further, in order to obtain stable estimates of β , the pool of candidates should be reduced using D-optimality criterion calculated by QR factorization with column pivoting [68] also called rank revealing QR factorization (RRQR). Note that it is necessary to evaluate the proposed criterion for every candidate and thus it might be computationally demanding for large n_{pool} . Therefore, we propose to generate the pool of candidates by coherence D-optimal sampling in every loop of sequential sampling instead of a single large pool generated before the start of the iteration process. Smaller pools for every iteration are not only computationally efficient but such approach reflects the actual sparse set of basis functions of PCE obtained by LAR in each iteration. This might be crucial for the candidate set generated by coherence-based sampling, since it is optimized for the selected basis functions in each step. Moreover, the proposed criterion gives higher importance to basis functions associated to higher β , which is important for identification of functional extremes.

Algorithm 1 thus reflects all pieces of information about the stochastic model (probabilistic distribution of input variables), investigated mathematical model (sparse set of basis functions) and even type of solver for PCE construction (weighted least squares) in order to obtain accurate and stable estimates of the deterministic PCE coefficients β . Combination of all techniques used in the algorithm thus leads to superior performance as will be shown in the numerical examples. However, generation of the pool of candidates is much more computationally demanding in comparison to the LHS-based adaptive sequential sampling.

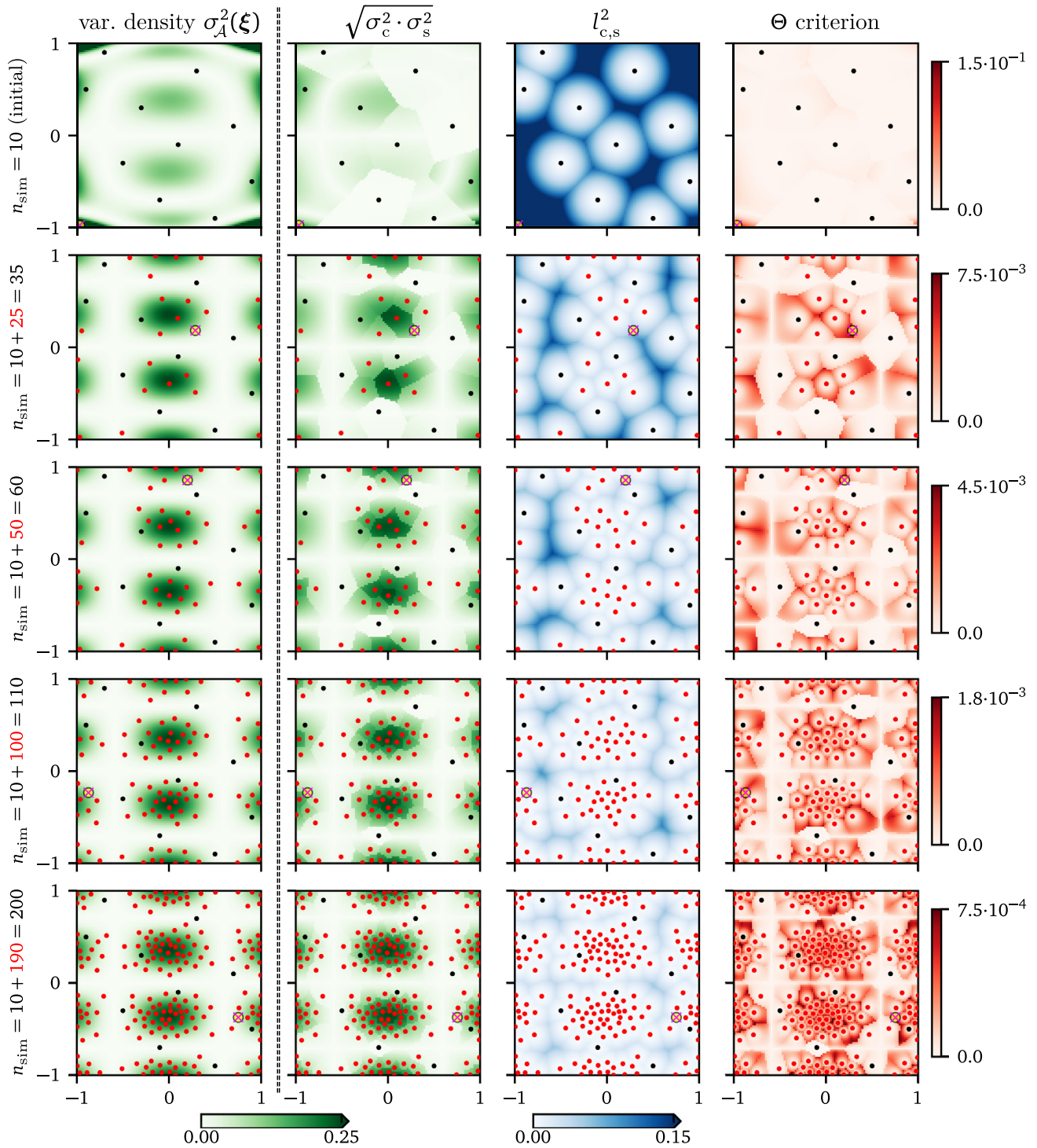


Fig. 3. Illustration of five stages during the proposed sequential sampling. Black solid circles: initial design. Red circles: extended sample. Crossed empty circle: the best candidate. The value of the proposed criterion and its components are depicted using the underlying color maps. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Numerical experiments

The proposed algorithm was numerically tested on several examples of increasing complexity. The setup common to all examples was as follows: PCE is solved by non-intrusive OLS (LHS) or WLS (Coh D-opt), a sparse set of the basis functions \mathcal{A} is obtained by LAR with maximum total polynomial order $p = 10$, if not said otherwise.

Algorithm 1 Coherence-Based Adaptive Sequential Sampling — one iteration

Input: $\xi_{ED}, \mathbf{W}_{ED}, \mathcal{Y}_{ED}, \mathcal{A}, \beta, (n_{sim})$	current Experimental Design and the corresponding PCE
1: $n = 5P, n_{pool} = 3P$	set the pool sizes based on the p and M
2: $\xi_{coh} \leftarrow n$ samples from $f_{coh}(\xi)$ using MCMC (Eq. (21))	draw points from the coherence density
3: $\mathbf{W}_{coh} \leftarrow$ weights corresponding to ξ_{coh} (Eq. (22))	calculate the corresponding weights
4: $\xi_{pool}, \mathbf{W}_{pool} \leftarrow n_{pool}$ D-optimal samples & weights	preselect the final pool of candidates
5: for all $\xi^{(c)} \in \xi_{pool}$ do	loop thru all candidates
6: $\xi^{(s)} \leftarrow \arg \min_{\xi \in \xi_{ED}} l_{c,s}(\xi^{(c)}, \xi)$	find the nearest neighbor
7: $\theta_c = \theta(\xi^{(c)}, \xi^{(s)})$	compute the criterion (Eq. (24))
8: end for	
9: $\xi^{(new)} \leftarrow \arg \max_{\xi^{(c)}} \theta_c$	select the best candidate
Output: $\xi^{(new)}, w(\xi^{(new)})$	return the best candidate and the corresponding weight

Identical p for all examples simulates a possible engineering situation with a black-box function (e.g. finite element analysis) where it is not possible to select the best p a priori. The initial ED for the PCE construction before the first step of the proposed iterative algorithm is generated by LHS and it contains an initial screening design with $n_{sim} = 10$ realizations of the input random vector for the first three examples and $n_{sim} = 20$ for the last example.

Although the proposed criterion can be coupled with any sampling technique for ED generation, only two selected techniques (LHS and Coh D-opt) were employed and compared in the numerical examples. LHS was selected for this study as it is the most common sampling technique for surrogate modeling due to its efficiency and simplicity. Existing software applications and packages for PCE construction (e.g. [39–43]) usually contain implementation of LHS and thus the process can be easily extended by proposed selection criterion. On the other hand, Coh D-opt is not a common approach, thus it is representative of advanced sampling methods suited specifically for least-squares PCE. Coh D-opt EDs usually achieve higher accuracy but their implementation is not straightforward.

Each example is solved by three types of strategies:

- non-sequential approach (non-seq) with ED generated via LHS for each sample size at once — this represents the most common approach employed in surrogate modeling,
- the proposed sequential sampling with candidates generated by LHS, and
- the proposed sequential sampling with candidates generated by Coh D-opt.

The sample size in the first (LHS) strategy is fixed and we study the PCE behavior for a range of these sample sizes selected a priori.

The two sequential sampling strategies differ in the way the pool of candidates is proposed. For the sake of clarity, the pool of candidates obtained by Coh D-opt for sequential sampling contains $3P$ D-optimal samples, which are selected from a greater pool of $5P$ Coh-optimal samples; for LHS sequential sampling, it contains $3P$ samples generated by LHS.

The results are compared in terms of the (i) *relative error in variance* of QoI

$$\epsilon = \frac{|\sigma^2 - \sigma_Y^2|}{\sigma_Y^2}, \quad (27)$$

defined as the absolute deviation of the estimated variance σ^2 from the exact value σ_Y^2 divided by the exact variance, and the (ii) *leave-one-out error* of PCE approximation Q^2 according to Eq. (12). In order to get a picture about reproducibility of the results, the calculations were repeated 100 times for each set of settings. The averages of $\log_{10}(\epsilon)$ (the order of relative error in variance) and $\log_{10}(1 - Q^2)$ are depicted by solid-lines and the scatters represent $\pm \sigma$ confidence intervals.

5.1. Toy 2D function

Consider a simple 2D function in which the two independent input random variables are uniformly distributed, $\mathbf{X} \sim \mathcal{U}$, with the mean values $\boldsymbol{\mu} = \{0, 0\}$ and variances $\boldsymbol{\sigma}^2 = \{6, 6\}$. The design domain is thus a square $[-\sqrt{18}; \sqrt{18}]^2$. The output variable (symmetrical and highly non-normal with zero mean value and skewness) is the following transformation

$$Y = \cos\left(\frac{X_1\pi}{5}\right) \sin\left(\frac{X_2\pi}{3}\right), \quad [\sigma_Y^2 \approx 0.199\,575\,434] \quad (28)$$

which should be easily approximated by all employed methods and the whole process is easily tractable.

Non-sequential LHS was calculated for 10 increasing sample sizes in range (24, 150). The supposedly uniform distribution of LH-samples is compared with the proposed algorithm for sequential selection of 140 candidates (to reach the same final sample size of 150 points). The rows in Fig. 3 show selected iterations of the sequential algorithm. The color maps show the local variance density (left column), and then spatial distribution of the two components of the proposed criterion: the average variance computed using the geometrical mean with the nearest neighbor and the squared distance to the nearest neighbor. The maps on the right hand side finally present the spatial distribution of the proposed Θ criterion. A relatively large pool of LHS candidates was generated and the color maps were computed on a fine grid of coordinates. The scale of Θ in the rightmost column is proportional to the amount of variance associated with individual candidates. It can be seen that the value Θ of the best candidate also indicates how much gain in the variance accuracy can be expected by adding one point. This information can be incorporated as a kind of “stopping criterion”. The scales in Fig. 3 show that refining from $n_{\text{sim}} = 10$ to $n_{\text{sim}} = 35$ decreases the variance bites from 10^{-1} to roughly $7.5 \cdot 10^{-3}$. However, further increase in sample size (and the associated time spent on evaluation of the $g(\mathbf{x})$ function) leads only to minor decrease in Θ criterion for the best candidate and this improvement may not be deemed as worth the expense.

This reasoning is well supported by the plots of variance error in Fig. 4 (left column). One can see that from the ED with $n_{\text{sim}} \approx 25$ points, further decreases in variance error are obtained at low rates. This is emphasized by a thin vertical dashed line showing that the available polynomial basis is saturated as the maximum polynomial order is exhausted. The decision to consider is either (i) stop the algorithm: the accuracy is acceptable or computing resources are too expensive (ii) increase the maximum polynomial order p with a chance to improve the convergence rate for further size extensions. Indeed, one can see in Fig. 4 (left column, bottom plot) that after adding about 15 points to the initial design with 10 points, the maximum polynomial degree of $p = 10$ gets almost always fully exploited and the error almost stabilizes (both the error in variance estimation and also the Q^2). When the experiment is repeated with $p = 20$ (see the middle column in Fig. 4), the saturation is postponed and both errors quickly decrease until the design reaches about $n_{\text{sim}} = 120$ points. This fact documents that the adaptivity feature should also increase the polynomial degree if higher accuracy is requested.

To conclude, the proposed sequential sampling clearly outperforms standard non-sequential sampling in error in estimated variance. Although difference in the approximation error Q^2 between non-seq and sequential LHS is not significant for $p = 10$, its sequential variant leads to the most accurate estimation of variance. In case of $p = 20$, the superiority of the sequential sampling is even more significant.

For the sake of completeness, the adaptivity of p was coupled with the proposed sequential sampling (see the right column in Fig. 4). Although there exist advanced adaptive algorithms such as adaptive Coh-D opt [75], the simplest algorithm with iterative increment of p was employed for both sampling schemes, i.e. p was iteratively increased from $p = 5$ to $p = 20$ and PCE was built for a given ED and finally p yielding the lowest approximation error measured by Q^2 was selected. The obtained results show the similar behavior of the sequential technique as in case of $p = 20$, as can be expected. Note that, techniques for p -adaptivity and many other advanced sampling techniques are beyond the scope of this paper. This research is focused on the proposed sequential sampling and for the possibility of direct comparison, such type of adaptivity is excluded in the following examples in order to show purely the role of Θ criterion, though they can generally lead to more accurate approximations.

5.2. Ishigami function

Consider now a three-dimensional Ishigami function [76]. The function is strongly nonlinear, non-monotonic and presents strong interactions. We set the coefficients as in [77]. Let $\mathbf{X} \sim \mathcal{U}[-\pi, \pi]^3$ and the mathematical model

$$Y = \sin(X_1) + 7 \sin^2(X_2) + 0.1X_3^4 \sin(X_1). \quad [\sigma_Y^2 \approx 13.844\,587\,940] \quad (29)$$

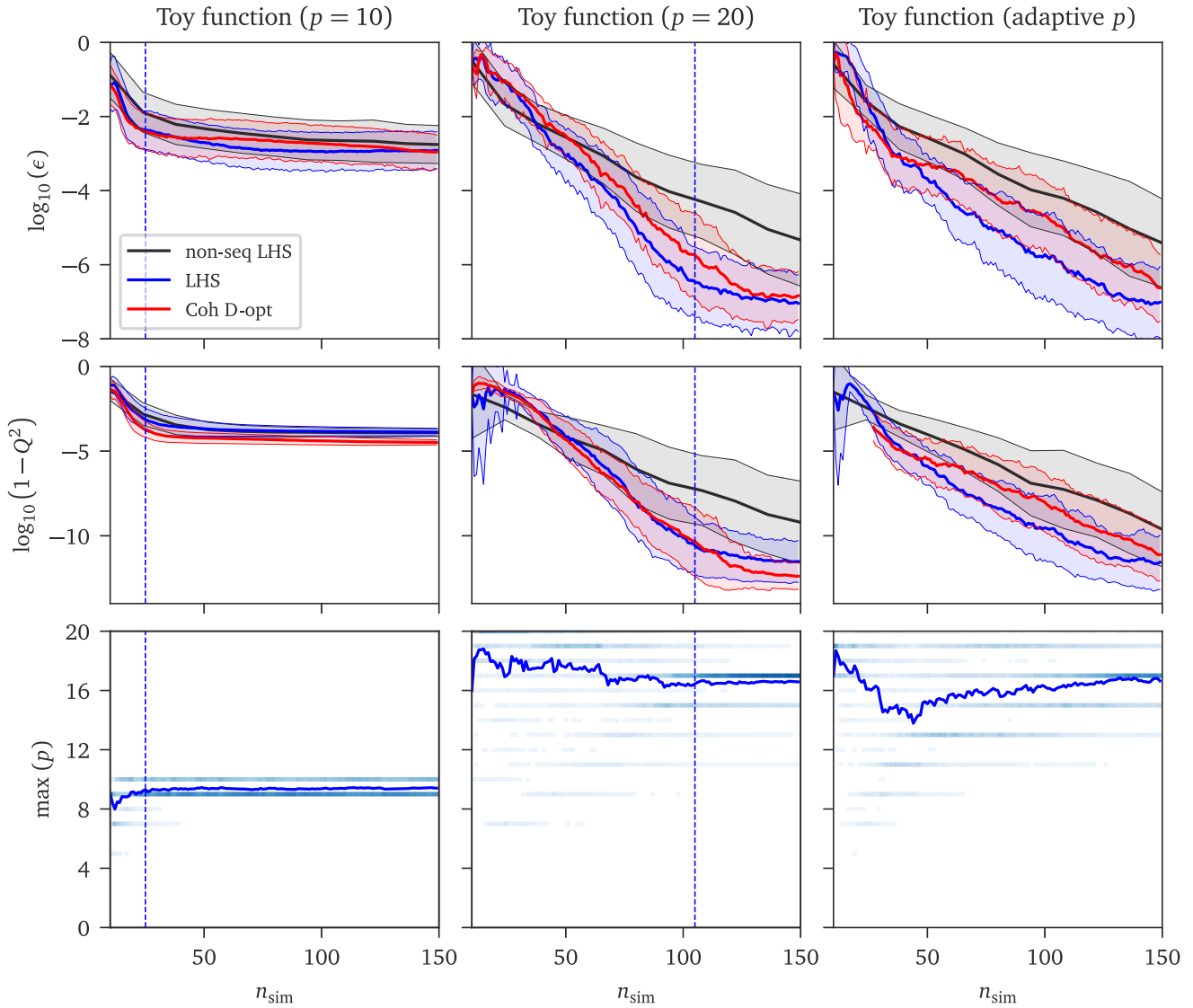


Fig. 4. Results for Toy 2D function obtained with maximum polynomial order $p = 10$ (left), $p = 20$ (middle) and adaptive p (right). The first two rows represent the accuracy measured by ϵ and Q^2 . The last row shows the mean value (solid blue line) and the empirical probability mass function (blue points) of the maximum order p in active \mathcal{A} for LHS. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The Ishigami function represents a well-known benchmark function for surrogate modeling and sensitivity analysis and thus additional analysis with a very large pool size was performed in order to show additional results discussed extensively in Section 6. First of all, non-sequential PCE based on Coh-D optimal sampling was created for the direct comparison. The obtained results for standard setting are summarized in Fig. 5 (left column).

The non-sequential Coherence-D optimal ED leads to unsatisfactory results both in variance estimation and Q^2 . Although the convergence rate is lower for our purpose (low number of samples), note that it is significantly more efficient with increasing number of samples. As can be seen, the proposed sequential sampling clearly outperforms the non-seq standard approach. Moreover, the convergence rate is significantly higher, until the polynomial chaos gets saturated by reaching the maximum order $p = 10$ as can be seen in Fig. 5 (left column, bottom plot). Naturally, non-sequential technique converges to identical accuracy with increasing number of samples.

Since the Ishigami function is a low-dimensional and inexpensive example, it was also possible to perform calculation with ED pool containing a large number $n_{\text{pool}} = 5000$ of simulations and 100 repetitions in order to

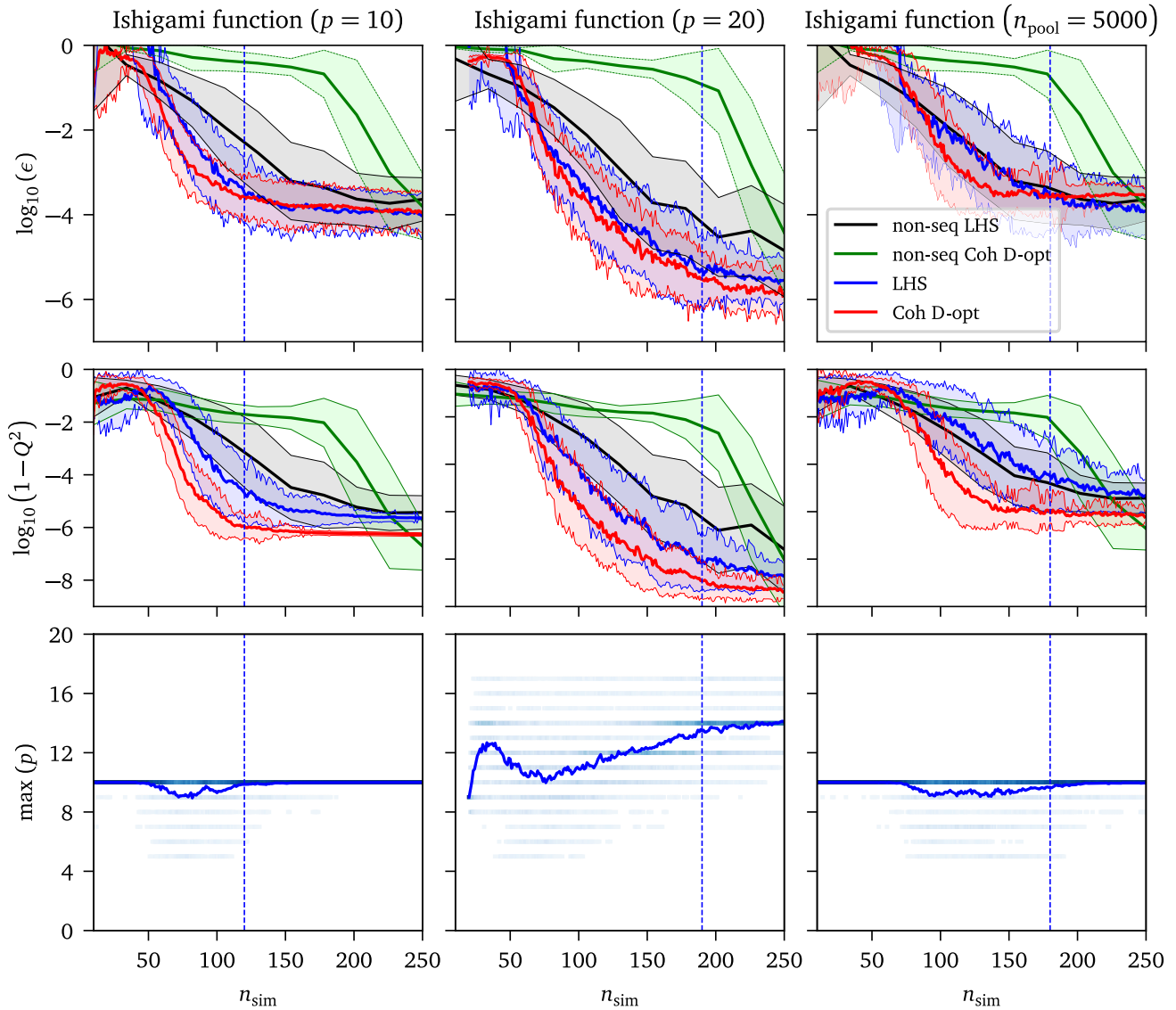


Fig. 5. Results for the Ishigami function. The first two rows represent the accuracy measured by ϵ and Q^2 . The last row shows the mean value (solid blue line) and the empirical probability mass function (blue points) of the maximum order p in active \mathcal{A} for LHS. The maximum polynomial order is $p = 10$ for the left ($n_{\text{pool}} = 858$) and right columns. The pool size for the middle column is $n_{\text{pool}} = 3P = 5313$ candidate points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

obtain statistical estimates. Note that the pool for Coh-D opt was obtained by D-optimal reduction from larger pool generated by coherence optimal sampling as in the previous examples.

The obtained results, for the case of an extremely large pool of candidates, depicted in Fig. 5 (right column) are slightly worse in comparison to the moderate size of the pool. This phenomenon is extensively discussed in Section 6. Additionally, note that the presented results are obtained for the fixed maximum polynomial order $p = 10$, which imposes a strong limit on achievable accuracy for $n_{\text{sim}} \gtrsim 120$. Similarly to the previous numerical example, allowing a higher p leads to higher accuracy of the PCE, which can be seen in Fig. 5 (middle column) using $p = 20$ (the saturation is postponed to about $n_{\text{sim}} \approx 200$).

5.3. 2D mirror line singularities

The 2D line singularities function is a mirrored version of a function used in [37]. The true surface of this function is studied for three different values of the parameter δ ; see Fig. 6. In order to document the effectiveness

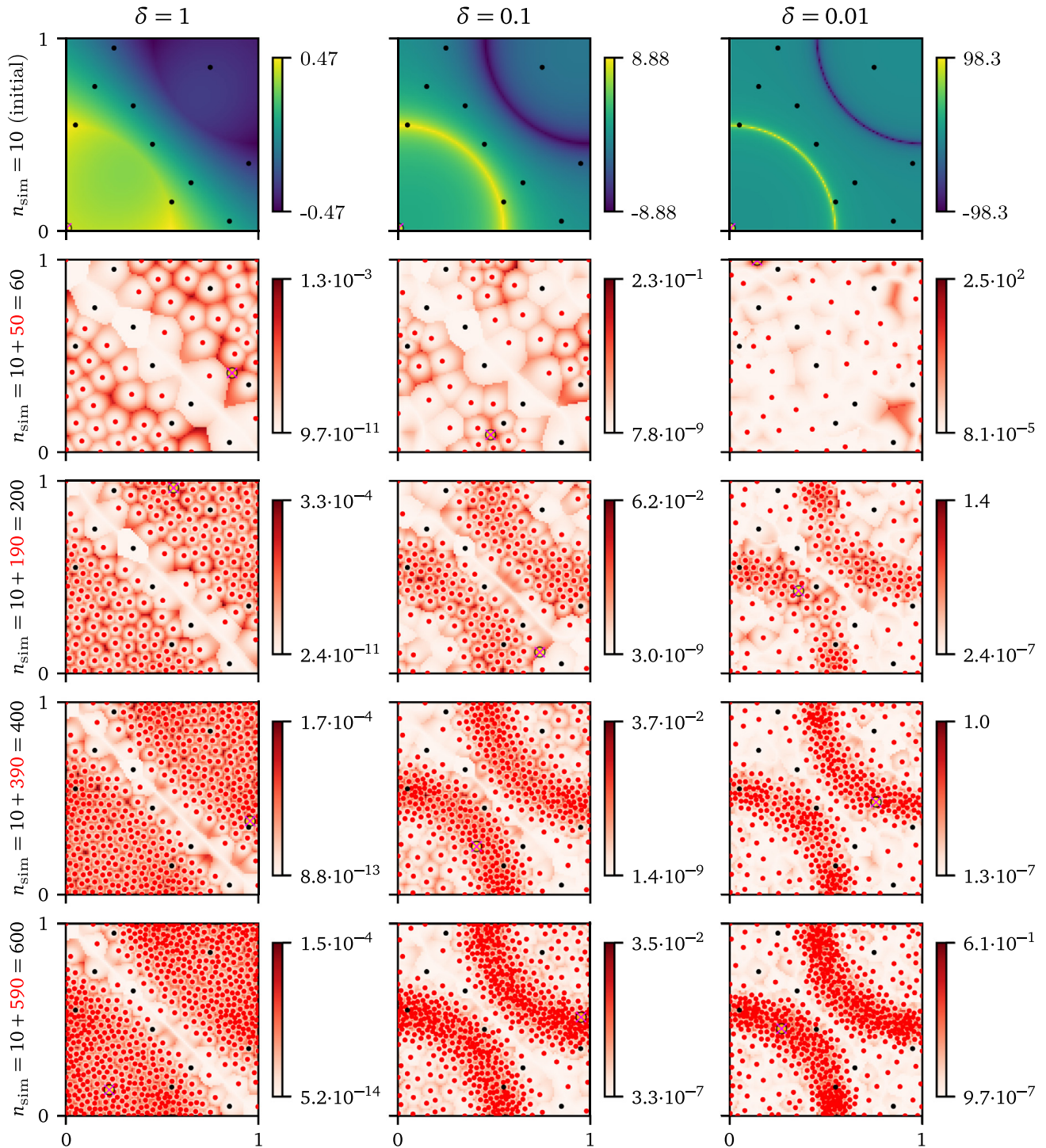


Fig. 6. 2D mirror line singularities: underlying map shows function values (the top row) and Θ criterion (others). Points represent the initial design ED (black points) and selected four iterations of algorithm adding candidates to the existing ED. Mathematical model for different parameters δ : $\delta = 1$ (left), $\delta = 0.1$ (middle), $\delta = 0.01$ (right). The maximum polynomial degree was $p = 12$ and therefore the ability of PCE to mimic a sharp singularity was limited.

of the proposed sequential sampling in the convergence plots, we selected the case of $\delta = 0.1$. Let $\mathbf{X} \sim \mathcal{U}[0, 1]^2$ and the mathematical model be in the following form

$$Y = \frac{1}{|0.3 - X_1^2 - X_2^2| + \delta} - \frac{1}{|0.3 - (1 - X_1)^2 - (1 - X_2)^2| + \delta}. \quad [\delta = 0.1 : \sigma_Y^2 \approx 13.070477042] \quad (30)$$

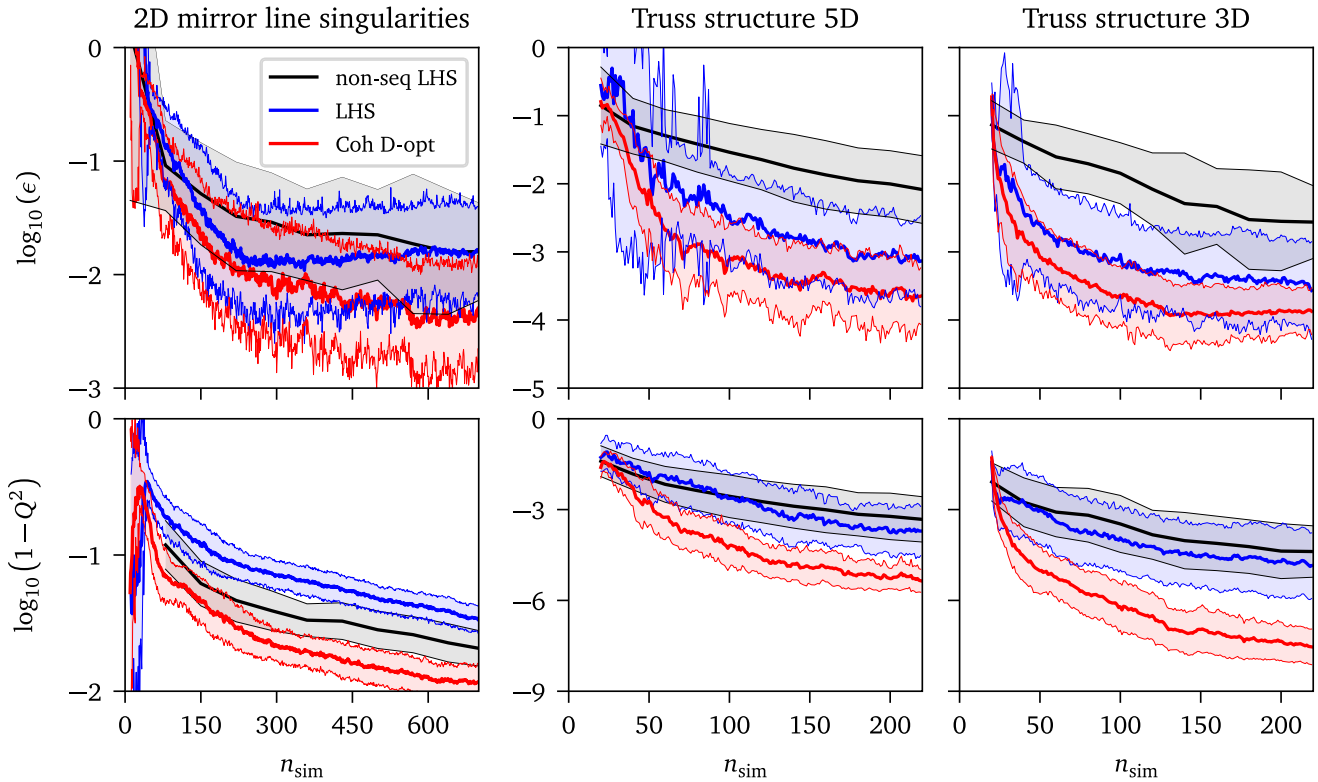


Fig. 7. Numerical results for Example 3 (left), Example 4 (middle) and Example 5 (right). The rows represent the accuracy measured via ϵ and Q^2 . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Note that, this example represents a challenging task for PCE especially for low values of δ . There are curved singularities located in a narrow vicinity of two circular arcs and thus it is crucial to identify the location of the singularity and use high-degree interacting polynomials for approximation. The values of the function range between its extremes of roughly $\pm\delta^{-1}$. We remark that the results of non-seq Coh D-opt are out of the graph range and thus this technique is not depicted in Fig. 7. Fig. 7 (top-left) shows a typical comparison of sequential and non-sequential LHS techniques. Sequential sampling is significantly better for mid-size ED, while the differences are reduced for large sample sizes when the polynomial chaos gets saturated. We remark that the high convergence rate for small to medium sample size is the practical range for which the proposed method is developed. Coh-D opt sequential strategy shows the best accuracy in Q^2 as well as in estimated variance. On the other hand, as can be seen in Fig. 7 (bottom-left) sequential LHS leads to low accuracy measured by Q^2 .

5.4. Truss structure (Hermite polynomials)

This problem involves nonuniform input variables and thus the selected set of polynomials is composed of Hermite polynomials. The design domain becomes open: \mathbb{R}^M . The mathematical model represents deflection of the truss structure depicted in Fig. 8. The deflection can be computed using the method of virtual work (unit load method). This method results in the following expression for the mid-span deflection

$$Y = F \left(\frac{552}{A_h E_h} + \frac{50.9117}{A_d E_d} \right). \quad [\sigma_Y^2 \approx 0.0002373] \quad (31)$$

The input random vector of the model, \mathbf{X} , consists of five independent random variables: the properties of the horizontal bars (Young's modulus E_h and cross-section area A_h), the properties of the diagonal bars (Young's modulus E_d and cross-section area A_d), and the magnitude of the loading forces F on the top joints. The properties of the input random variables are summarized in Table 1.

First of all, in order to construct PCE, the input random vector is transformed to standardized Gaussian random space by Nataf transformation [78,79] corresponding to Hermite germs. Further full set of polynomial basis functions

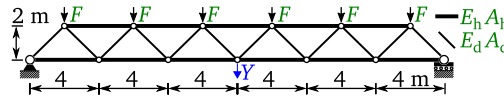


Fig. 8. Truss structure constructed from two types of bars.

Table 1

Five random variables featured in the truss example.

Variable	Distribution	Mean	Units	CoV
E_h, E_d	Log-normal	210	GPa	0.10
A_h	Log-normal	2000	mm ²	0.10
A_d	Log-normal	1000	mm ²	0.10
F	Gumbel-Max	50	kN	0.15

P is reduced by hyperbolic truncation with $q = 0.5$ according to Eq. (8). The initial ED contains $n_{sim} = 20$ realizations of X generated by LHS. We remark that Nataf transformation might increase the polynomial degree required to compute the PCE approximation of a function. To avoid the increase in nonlinearity of the problem by using the Nataf transformation, one could also compute the polynomials numerically in practical applications [80].

Obtained results can be seen in Fig. 7 (middle column). Similar to the previous example, sequential LHS leads to higher accuracy in estimated variance but the approximation error Q^2 is comparable to non-sequential LHS. The proposed sequential Coh D-opt method is clearly the most accurate and convergence rate is significantly faster both in Q^2 and variance estimation. The Coh-D sequential sampling converges to errors several orders of magnitude smaller than commonly used non-sequential LHS.

5.5. Truss structure (Hermite polynomials) — Reduced dimension

Adaptivity of the proposed algorithm is provided by the best model selection algorithm (LAR in this paper), which should be able to select the best possible set of basis functions \mathcal{A} . In order to examine this feature in the context of the proposed adaptive sequential sampling, the results of the previous example are compared to a manually reduced stochastic model of the previous example (Truss structure) preserving identical mathematical model as follows [81]:

$$Y = F \left(\frac{552}{\mathcal{A}_h} + \frac{50.9117}{\mathcal{A}_d} \right), \quad [\sigma_Y^2 \approx 0.0002373] \tag{32}$$

where \mathcal{A}_h is a Log-normal random variable with mean of 420 MN and \mathcal{A}_d is a Log-normal random variable with mean of 210 MN, CoV of both random variables equals 0.14177. The results obtained from the sequential sampling are summarized in Fig. 7 (right). As can be seen, the final accuracy of this example measured by relative error in variance ϵ is similar to the 5D formulation of the $g(X)$ when the size of ED reaches the final $n_{sim} = 220$, although the convergence of the reduced model to this value is significantly faster for lower n_{sim} . This is in compliance with the theoretical behavior of the model selection algorithm, which becomes more efficient with greater samples size. The faster convergence also affects the final accuracy of Coh D-opt measured by Q^2 , which is able to converge to lower values for the final n_{sim} .

6. Discussion

6.1. Optimal pool size and the maximum polynomial order p

As already briefly mentioned above, the selection of fixed maximum polynomial order p may impose a lower bound on achievable accuracy. This phenomenon was visible in the first “Toy” example (see Fig. 4) as well as for the “Ishigami function”, see Fig. 5. In these examples, the initial guess on $p = 10$ was found insufficient and selection of $p = 20$ allowed for much higher accuracy of the surrogate. This is documented by examination of the set \mathcal{A} , specifically maximum p in the active set of basis functions \mathcal{A} chosen by LAR depicted by blue color in the bottom rows of Figs. 4 and 5. It can be seen, that convergence rate is very low once the maximum order p in

active \mathcal{A} achieves its prescribed maximum of $p = 10$ since there are no additional possibilities for improvement of the approximation (the space of basis functions is out of suitable candidates). Therefore, if one extends the space of basis functions by sufficiently incrementing p , the decrease in error does not stop and ϵ for maximum n_{sim} is significantly lower. Unfortunately, the ideal p is not known a priori and thus one should adaptively increase it with growing active ED. However, as discussed already in [2], it is possible to employ an adaptive selection of the best p for each iteration. Such a feature might play a significant role in some cases. Such approach was deliberately not employed in this paper because the influence of the proposed sampling scheme might not be clearly separable from the model selection.

The proposed sequential technique achieved higher accuracy of estimated variance for lower number of samples in comparison to non-sequential sampling in all presented numerical examples. However, as can be seen from numerical results of the Ishigami function, increasing the pool size does not generally lead to higher accuracy or faster convergence rate. In fact, a very large number of candidates might cause slower convergence (comparable to non-sequential sampling). After detailed examination of this example, one can see the different structure of \mathcal{A} for the large set of candidates depicted in Fig. 5 (rightmost column). During the initial phase of the adaptive sequential sampling, LAR algorithm typically selects high-order polynomials since there is not enough information about a mathematical model, which leads to *overfitting*. Further increasing the sample size enables the adaptive algorithm to identify more appropriate low-order basis functions ($n_{\text{sim}} \approx 50$ for Ishigami function). Sequential sampling in the following steps of algorithm selects the candidates with respect to the current \mathcal{A} and exploration aspect. Note that high-order basis functions are selected by adaptive algorithm once the new regions associated to high local variance are discovered (peaks of the given mathematical model). However, such regions are not preferred by the selection criterion since low-order basis functions ignore the currently unknown extremes. Therefore, the convergence rate of the adaptive sequential sampling can be significantly affected by the number of candidates, since if there are no candidates in regions favored by low-order basis functions, the exploration part of the proposed criterion could investigate new functional extremes and adapt \mathcal{A} to a set of high-order basis functions. This phenomenon is illustrated by blue color in numerical results in Fig. 5 (left). In the first case with $n_{\text{pool}} = 3P$ (and $p = 10$), one can see a fast convergence for n_{sim} ranging between 50 and 150 (associated to lower maximum p in active \mathcal{A}) until the space of basis functions is out of suitable candidates as described in the paragraph above. On the other hand, for the second case with $n_{\text{pool}} = 5\,000$ (and $p = 10$ again; see Fig. 5 right), an active \mathcal{A} contains low-order basis functions for a range of significantly greater n_{sim} , which leads to a slower convergence of accuracy ultimately leading to the identical final error.

In summary, the achievable accuracy of the adaptive sequential sampling is limited by the maximum polynomial order p used. Simply, the flexibility of the polynomial approximation may not be sufficient to approximate the original function at a given precision level. In practical applications, a sufficient polynomial order is not known a priori and therefore p should be adaptively increased with increasing n_{sim} in order to achieve the best performance of the proposed algorithm. On the other hand, the convergence rate is significantly affected by n_{pool} and extremely large pool leads to slow convergence, since a selection of high-order basis functions is postponed to higher n_{sim} . Therefore the heuristic rule $n_{\text{pool}} = 3P$ is recommended for practical applications.

6.2. High dimensions

What remains an open question is the behavior of the proposed criterion in high dimensions. In particular, we need to understand the effect of the exploration part $l_{c,s}^M$ in Eq. (24). Generally, in high-dimensional space with independent and identically distributed (iid) components, the Minkowski distance of order $\mathcal{P} > 0$ (sometimes referred to as the \mathcal{P} -norm) concentrates, i.e. the coefficient of variation of the norm decreases with increasing dimension, M . This effect on \mathcal{P} -norm of letting M go large is well known in the computational learning literature. A discussion in the context of iid Gaussian distribution can be found e.g. in [82]. The asymptotic behavior of \mathcal{P} -norm and its convergence rates have been studied in [83].

The role of the distance term must be discussed especially in the case of uniform distribution of the germ which is defined over a hypercube. It is known that Euclidean distances (a special case of Minkowski distance with $\mathcal{P} = 2$) of any pair of points inside a hypercube tend to concentrate around its mean value when the dimension is high. It is known that the standard deviation of the Euclidean distance between any two randomly picked points stays approximately constant with increasing dimension while the mean value keeps growing proportionally to

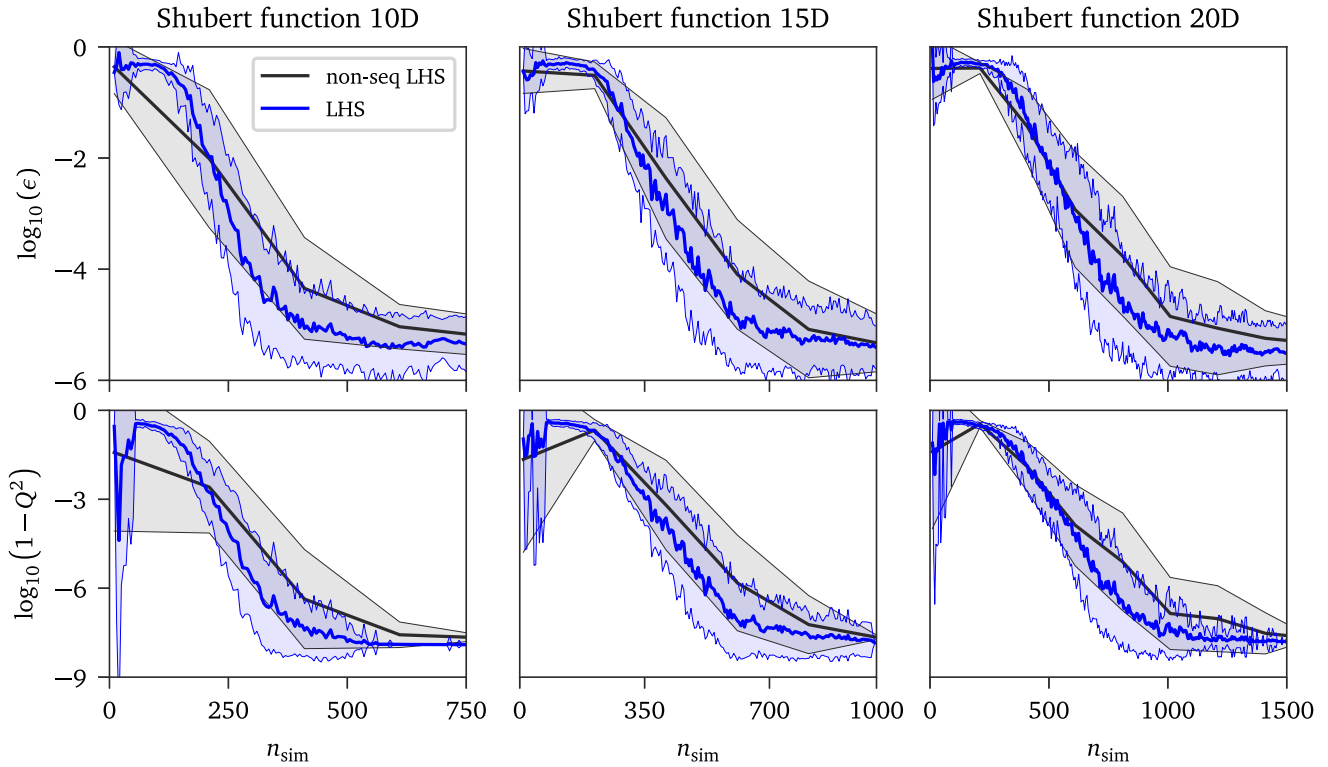


Fig. 9. Results for the Shubert function in 10 (left), 15 (middle) and 20 (right) dimensions. The first row represents the accuracy of estimated variance measured by ϵ and the second row shows the approximation accuracy measured by Q^2 .

\sqrt{M} [84,85]. Therefore, the coefficient of variation of a random distance is asymptotically proportional to $1/\sqrt{M}$. In such a case the distance contrast decreases and it is said that the distances concentrate [86,87]. The same holds also for the squared Euclidean distance l^2 between points picked at random: the squared distance has Gaussian distribution with the mean value of $M/6$ and variance $7M/180$. In order to obtain the hypervolume of a region in between a pair of points, the squared distance must be raised to $M/2$. We have found that the coefficient of variation of such volume quickly increases as the problem dimension grows high. Therefore, there is no problem with insufficient contrast as the proposed Θ criterion sufficiently varies even in domains of high dimension.

Although it will be necessary to perform extensive study with advanced sampling schemes in order to investigate the efficiency of the proposed criterion in high dimensions, the general behavior of the Θ criterion is demonstrated in the following simple example with input random vector $X \sim \mathcal{U}[-1, 1]^M$ — Shubert function No. 4 [88]:

$$Y = \sum_{i=1}^M \sum_{j=1}^5 j \cos((j+1)X_i + j). \quad [\sigma_Y^2 \approx M \cdot 45.266\,621\,664] \quad (33)$$

This function can be easily utilized as a benchmark for an arbitrary number of input random variables; in our study we use $M = \{10, 15, 20\}$. The maximum polynomial order $p = 12$ and hyperbolic truncation parameter $q = 0.3$ were used for construction of surrogate model, and the initial ED contains $n_{\text{sim}} = 10$ realizations of X generated by LHS. Since the hyperbolic truncation significantly reduces the P in high dimensions, we increased the pool size to $n_{\text{pool}} = 5P$. Note that, this study is focused only on stability of the proposed Θ criterion in higher dimensions and thus only the sequential and the non-sequential LHS were employed for direct comparison. From the obtained results depicted in Fig. 9, it can be seen that the proposed criterion enabling one-by-one extension of ED achieves higher accuracy in comparison to non-sequential approach independently of dimension M and thus the theoretical discussion in the previous paragraph was presented also numerically. Naturally, the absolute value of accuracy is highly dependent on the investigated function and on the particular sampling scheme coupled with Θ criterion, however this study demonstrates the applicability of the proposed criterion in high dimensions.

6.3. Further work

Since the DoE is usually not a bottleneck in probabilistic analysis and sequential sampling offers several advantages in comparison to non-sequential sampling, it can be recommended to employ the proposed algorithm for the construction of PCE despite higher computational requirements of repetitive DoE in each step. Although the choice of the sampling technique coupled with the proposed criterion for the sequential selection of the best candidate is arbitrary, it can be seen from the numerical examples, that LHS is efficient for low dimensional mathematical models. On the other hand, coherence D-optimal sampling achieved significantly higher accuracy in more complex examples. Therefore further work will be focused on an improvement of effectiveness by using advanced optimized space-filling designs for the generating of a pool of candidates and a comparative study of existing advanced sampling techniques coupled with the proposed criterion will be performed. Employment of designs generated from the target distribution that additionally avoids clustering [17,81,89] or empty regions [60] while maintaining true statistical homogeneity via periodic distance-based criteria has the potential to further improve the effectiveness of the proposed method, especially in high-dimensional space. Moreover, it was shown that fixed p represents a significant limitation for the proposed method and thus further work will be also focused on adaptive basis strategies [75,90], which have the potential to dramatically improve the final accuracy of PCE and solve the problem with LAR and with a large size of the pool of candidates.

7. Conclusion

A novel adaptive sequential sampling technique for accurate and efficient construction of a non-intrusive PCE was proposed in this paper and its performance was validated on several numerical examples of increasing complexity and dimensionality. The proposed technique selects the best candidate sample from a large pool maintaining the balance between exploration of the design domain and exploitation of the current characteristics of the PCE. The criterion driving the selection of the best candidate was successfully coupled with LHS and Coh-D optimal sampling and both variants were used in numerical examples. From the obtained results, it can be concluded that the proposed technique leads to higher accuracy of the constructed PCE in comparison to non-sequential sampling. The difference in accuracy between sequential and non-sequential sampling is especially significant for a low-size ED. However, it can be expected that the accuracy of sequential sampling converges to identical results as a non-sequential sampling for a very large ED in which the PCE is saturated. Comparing both sequential sampling techniques, superior performance was achieved by sequential adaptive Coh-D optimal sampling due to its adaptivity of the candidate sample in each iteration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge financial support provided by the Ministry of Education, Youth and Sports of the Czech Republic under Project No. LTAUSA19058. Additionally, the first author is also supported by the Ministry of Education, Youth and Sports of the Czech Republic under Project No. FAST-J-21-7209.

References

- [1] N. Wiener, The homogeneous chaos, *Amer. J. Math.* 60 (4) (1938) 897–936, <http://dx.doi.org/10.2307/2371268>.
- [2] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, *J. Comput. Phys.* 230 (6) (2011) 2345–2367, <http://dx.doi.org/10.1016/j.jcp.2010.12.021>.
- [3] R.G. Ghanem, P.D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer, New York, 1991, <http://dx.doi.org/10.1007/978-1-4612-3094-6>.
- [4] T. Crestaux, O.L. Maître, J.-M. Martinez, Polynomial chaos expansion for sensitivity analysis, *Reliab. Eng. Syst. Saf.* 94 (7) (2009) 1161–1172, <http://dx.doi.org/10.1016/j.res.2008.10.008>.
- [5] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, *Reliab. Eng. Syst. Saf.* 93 (7) (2008) 964–979, <http://dx.doi.org/10.1016/j.res.2007.04.002>.

- [6] N.-Z. Chen, C. Guedes Soares, Spectral stochastic finite element analysis for laminated composite plates, *Comput. Methods Appl. Mech. Eng.* 197 (51) (2008) 4830–4839, <http://dx.doi.org/10.1016/j.cma.2008.07.003>.
- [7] L. Novak, D. Novak, Surrogate modelling in the stochastic analysis of concrete girders failing in shear, in: *Proc. of the Fib Symposium 2019: Concrete - Innovations in Materials, Design and Structures*, 2019, pp. 1741–1747.
- [8] M. Chatzimanolakis, K.-D. Kantarakias, V. Asouti, K. Giannakoglou, A painless intrusive polynomial chaos method with RANS-based applications, *Comput. Methods Appl. Mech. Eng.* 348 (2019) 207–221, <http://dx.doi.org/10.1016/j.cma.2019.01.018>.
- [9] G.T. Buzzard, Global sensitivity analysis using sparse grid interpolation and polynomial chaos, *Reliab. Eng. Syst. Saf.* 107 (2012) 82–89, <http://dx.doi.org/10.1016/j.ress.2011.07.011>, SAMO 2010.
- [10] Z. Perko, L. Gilli, D. Lathouwers, J.L. Kloosterman, Grid and basis adaptive polynomial chaos techniques for sensitivity and uncertainty analysis, *J. Comput. Phys.* 260 (2014) 54–84, <http://dx.doi.org/10.1016/j.jcp.2013.12.025>.
- [11] A. Cohen, G. Migliorati, Optimal weighted least-squares methods, *SMAI J. Comput. Math.* 3 (2017) 181–203, <http://dx.doi.org/10.5802/smai-jcm.24>.
- [12] A.C. Narayan, J. Jakeman, T. Zhou, A christoffel function weighted least squares algorithm for collocation approximations, *Math. Comp.* 86 (2017) 1913–1947, <http://dx.doi.org/10.1090/mcom/3192>.
- [13] N. Lüthen, S. Marelli, B. Sudret, Sparse polynomial chaos expansions: Literature survey and benchmark, *SIAM/ASA J. Uncertain. Quantif.* 9 (2) (2021) 593–649, <http://dx.doi.org/10.1137/20M1315774>.
- [14] W. Conover, On a better method for selecting input variables, 1975, unpublished Los Alamos National Laboratories manuscript, reproduced as Appendix A of “Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems” by J.C. Helton and F.J. Davis, Sandia National Laboratories report SAND2001-0417, printed November 2002. URL <https://prod-ng.sandia.gov/techlib-noauth/access-control.cgi/2001/010417.pdf>.
- [15] M. Johnson, L. Moore, D. Ylvisaker, Minimax and maximin distance designs, *J. Statist. Plann. Inference* 2 (26) (1990) 131–148, [http://dx.doi.org/10.1016/0378-3758\(90\)90122-B](http://dx.doi.org/10.1016/0378-3758(90)90122-B).
- [16] M.D. Morris, T.J. Mitchell, Exploratory designs for computational experiments, *J. Statist. Plann. Inference* 43 (3) (1995) 381–402, [http://dx.doi.org/10.1016/0378-3758\(94\)00035-T](http://dx.doi.org/10.1016/0378-3758(94)00035-T).
- [17] M. Vořechovský, J. Eliáš, Modification of the Maximin and ϕ_p (phi) criteria to achieve statistically uniform distribution of sampling points, *Technometrics* 62 (3) (2020) 371–386, <http://dx.doi.org/10.1080/00401706.2019.1639550>.
- [18] J. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numer. Math.* 2 (1) (1960) 84–90, <http://dx.doi.org/10.1007/BF01386213>.
- [19] H. Faure, C. Lemieux, Generalized Halton sequences in 2008: A comparative study, *ACM Trans. Model. Comput. Simul.* 19 (4) (2009) 15:1–15:31, <http://dx.doi.org/10.1145/1596519.1596520>.
- [20] I. Sobol’, On the distribution of points in a cube and the approximate evaluation of integrals, *USSR Comput. Math. Math. Phys.* 7 (1967) [http://dx.doi.org/10.1016/0041-5553\(67\)90144-9](http://dx.doi.org/10.1016/0041-5553(67)90144-9).
- [21] I.M. Sobol’, Uniformly distributed sequences with an additional uniform property, *USSR Comput. Math. Math. Phys.* 16 (5) (1976) 236–242, [http://dx.doi.org/10.1016/0041-5553\(76\)90154-3](http://dx.doi.org/10.1016/0041-5553(76)90154-3), Short communication.
- [22] H. Niederreiter, Point sets and sequences with small discrepancy, *Monatsh. Math.* 104 (4) (1987) 273–337, <http://dx.doi.org/10.1007/BF01294651>.
- [23] H. Niederreiter, Low-discrepancy and low-dispersion sequences, *J. Number Theory* 30 (1) (1988) 51–70, [http://dx.doi.org/10.1016/0022-314X\(88\)90025-X](http://dx.doi.org/10.1016/0022-314X(88)90025-X).
- [24] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, in: *CBMS-NSF Regional Conference Series in Applied Mathematics*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, 1992, <http://dx.doi.org/10.1137/1.9781611970081>.
- [25] H. Faure, *Discrèpances de suites associées à un système de numération (en dimension un)* [Discrepancy of sequences associated with a number system (in dimension one)], *Bull. Soc. Math. France* 109 (1981).
- [26] S. Tezuka, Uniform Random Numbers: Theory and Practice, in: *The Springer International Series in Engineering and Computer Science* 315, Springer, Boston, MA, 1995, <http://dx.doi.org/10.1007/978-1-4615-2317-8>.
- [27] J. Hampton, A. Doostan, Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies, *J. Comput. Phys.* 280 (2015) 363–386, <http://dx.doi.org/10.1016/j.jcp.2014.09.019>.
- [28] A. Cohen, G. Migliorati, Optimal weighted least-squares methods, *SMAI J. Comput. Math.* 3 (2017) 181–203, <http://dx.doi.org/10.5802/smai-jcm.24>.
- [29] B. Echard, N. Gayton, M. Lemaire, AK-MCS: An active learning reliability method combining Kriging and Monte Carlo simulation, *Struct. Saf.* 33 (2) (2011) 145–154, <http://dx.doi.org/10.1016/j.strusafe.2011.01.002>.
- [30] L. Shi, B. Sun, D.S. Ibrahim, An active learning reliability method with multiple kernel functions based on radial basis function, *Struct. Multidiscip. Optim.* 60 (1) (2019) 211–229, <http://dx.doi.org/10.1007/s00158-019-02210-0>.
- [31] X. Yang, X. Cheng, Active learning method combining Kriging model and multimodal-optimization-based importance sampling for the estimation of small failure probability, *Internat. J. Numer. Methods Engrg.* 121 (21) (2020) 4843–4864, <http://dx.doi.org/10.1002/nme.6495>.
- [32] S. Marelli, B. Sudret, An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis, *Struct. Saf.* 75 (2018) 67–74, <http://dx.doi.org/10.1016/j.strusafe.2018.06.003>.
- [33] Y. Zhou, Z. Lu, W. Yun, Active sparse polynomial chaos expansion for system reliability analysis, *Reliab. Eng. Syst. Saf.* 202 (2020) 107025, <http://dx.doi.org/10.1016/j.ress.2020.107025>.
- [34] K. Cheng, Z. Lu, Active learning polynomial chaos expansion for reliability analysis by maximizing expected indicator function prediction error, *Internat. J. Numer. Methods Engrg.* 121 (14) (2020) 3159–3177, <http://dx.doi.org/10.1002/nme.6351>.

- [35] N. Fajraoui, S. Marelli, B. Sudret, Sequential design of experiment for sparse polynomial chaos expansions, *SIAM/ASA J. Uncertain. Quantif.* 5 (1) (2017) 1061–1085, <http://dx.doi.org/10.1137/16m1103488>.
- [36] M. Thapa, S.B. Mulani, R.W. Walters, Adaptive weighted least-squares polynomial chaos expansion with basis adaptivity and sequential adaptive sampling, *Comput. Methods Appl. Mech. Eng.* 360 (2020) 112759, <http://dx.doi.org/10.1016/j.cma.2019.112759>.
- [37] M.D. Shields, Adaptive Monte Carlo analysis for strongly nonlinear stochastic systems, *Reliab. Eng. Syst. Saf.* 175 (2018) 207–224, <http://dx.doi.org/10.1016/j.ress.2018.03.018>.
- [38] Y. Zhou, Z. Lu, K. Cheng, C. Ling, An efficient and robust adaptive sampling method for polynomial chaos expansion in sparse Bayesian learning framework, *Comput. Methods Appl. Mech. Engrg.* 352 (2019) 654–674, <http://dx.doi.org/10.1016/j.cma.2019.04.046>.
- [39] S. Marelli, B. Sudret, UQLab: A framework for uncertainty quantification in matlab, in: *Vulnerability, Uncertainty, and Risk*, 2014, pp. 2554–2563, <http://dx.doi.org/10.1061/9780784413609.257>.
- [40] E. Patelli, S. Tolo, H. George-Williams, J. Sadeghi, R. Rocchetta, M. de Angelis, M. Broggi, OpenCossan 2.0: an efficient computational toolbox for risk, reliability and resilience analysis, in: *Proceedings of the Joint ICVRAM ISUMA UNCERTAINTIES Conference*, 2018, pp. 1–8, URL <http://icvramisuma2018.org/cd/web/PDF/ICVRAMISUMA2018-0022.PDF>.
- [41] L. Novak, D. Novak, Polynomial chaos expansion for surrogate modelling: Theory and software, *Beton- Stahlbetonbau* 113 (2018) 27–32, <http://dx.doi.org/10.1002/best.201800048>.
- [42] J. Feinberg, H.P. Langtangen, Chaospy: An open source tool for designing methods of uncertainty quantification, *J. Comput. Sci.* 11 (2015) 46–57, <http://dx.doi.org/10.1016/j.jocs.2015.08.008>.
- [43] A. Olivier, D. Giovanis, B. Aakash, M. Chauhan, L. Vandanapu, M.D. Shields, UQpy: A general purpose Python package and development environment for uncertainty quantification, *J. Comput. Sci.* 47 (2020) 101204.
- [44] D. Xiu, G.E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, *SIAM J. Sci. Comput.* 24 (2) (2002) 619–644, <http://dx.doi.org/10.1137/s1064827501387826>.
- [45] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, *Ann. Statist.* 32 (2) (2004) 407–451, <http://dx.doi.org/10.2307/3448465>.
- [46] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inf. Theory* 53 (12) (2007) 4655–4666, <http://dx.doi.org/10.1109/tit.2007.909108>.
- [47] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, *IEEE Trans. Signal Process.* 56 (6) (2008) 2346–2356, <http://dx.doi.org/10.1109/TSP.2007.914345>.
- [48] G. Blatman, B. Sudret, An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis, *Probab. Eng. Mech.* 25 (2) (2010) 183–197, <http://dx.doi.org/10.1016/j.probengmech.2009.10.003>.
- [49] J.F. Koksma, Een algemeene stelling uit de theorie der gelijkmatige verdeling modulo 1, *Math. B* 11 (1942/1943) 7–11.
- [50] K.-T. Fang, Y. Wang, *Number-Theoretic Methods in Statistics*, first ed., Chapman and Hall/CRC, London; New York, 1993.
- [51] F.J. Hickernell, A generalized discrepancy and quadrature error bound, *Math. Comp.* 67 (221) (1998) 299–322, <http://dx.doi.org/10.1090/S0025-5718-98-00894-1>.
- [52] F.J. Hickernell, Lattice rules: How well do they measure up? in: P. Hellekalek, G. Larcher (Eds.), *Random and Quasi-Random Point Sets*, Springer New York, New York, NY, 1998, pp. 109–166, http://dx.doi.org/10.1007/978-1-4612-1702-2_3, Ch. Lattice Rules: How Well Do They Measure Up?.
- [53] J. Dick, F. Pillichshammer, *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, Cambridge University Press, 2010, <http://dx.doi.org/10.1017/cbo9780511761188>.
- [54] M.D. McKay, W.J. Conover, R.J. Beckman, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* 21 (1979) 239–245, <http://dx.doi.org/10.1080/00401706.1979.10489755>.
- [55] M. Stein, Large sample properties of simulations using Latin hypercube sampling, *Technometrics* 29 (2) (1987) 143–151, <http://dx.doi.org/10.1080/00401706.1987.10488205>.
- [56] A.B. Owen, A central limit theorem for Latin hypercube sampling, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 54 (2) (1992) 541–551, <http://dx.doi.org/10.2307/2346140>.
- [57] A.B. Owen, Controlling correlations in latin hypercube samples, *J. Amer. Statist. Assoc. (Theory Methods)* 89 (428) (1994) 1517–1522.
- [58] M.D. Shields, J. Zhang, The generalization of Latin hypercube sampling, *Reliab. Eng. Syst. Saf.* 148 (2016) 96–108, <http://dx.doi.org/10.1016/j.ress.2015.12.002>.
- [59] L. Pronzato, Minimax and maximin space-filling designs: some properties and methods for construction, *J. Soc. Fr. Stat.* 158 (1) (2017) 7–36.
- [60] J. Eliáš, M. Vořechovský, V. Sadílek, Periodic version of the minimax distance criterion for Monte Carlo integration, *Adv. Eng. Softw.* 149 (2020) 102900, <http://dx.doi.org/10.1016/j.advengsoft.2020.102900>.
- [61] S.-K. Choi, R.V. Grandhi, R.A. Canfield, C.L. Pettit, Polynomial chaos expansion with latin hypercube sampling for estimating response variability, *AIAA J.* 42 (6) (2004) 1191–1198, <http://dx.doi.org/10.2514/1.2220>.
- [62] M. Hadigol, A. Doostan, Least squares polynomial chaos expansion: A review of sampling strategies, *Comput. Methods Appl. Mech. Eng.* 332 (2018) 382–407, <http://dx.doi.org/10.1016/j.cma.2017.12.019>.
- [63] E.J. Candes, Y. Plan, A probabilistic and RIPless theory of compressed sensing, *IEEE Trans. Inf. Theory* 57 (11) (2011) 7235–7254, <http://dx.doi.org/10.1109/tit.2011.2161794>.
- [64] J.D. Jakeman, A. Narayan, T. Zhou, A generalized sampling and preconditioning scheme for sparse approximation of polynomial chaos expansions, *SIAM J. Sci. Comput.* 39 (3) (2017) A1114–A1144, <http://dx.doi.org/10.1137/16M1063885>.
- [65] J. Hampton, A. Doostan, Coherence motivated sampling and convergence analysis of least squares polynomial chaos regression, *Comput. Methods Appl. Mech. Eng.* 290 (2015) 73–97, <http://dx.doi.org/10.1016/j.cma.2015.02.006>.
- [66] A.J. Miller, N.-K. Nguyen, Algorithm AS 295: A fedorov exchange algorithm for D-optimal design, *J. R. Stat. Soc. Ser. C (Appl. Stat.)* 43 (4) (1994) 669–677, <http://dx.doi.org/10.2307/2986264>.

- [67] O. Dykstra, The augmentation of experimental data to maximize [X'X], *Technometrics* 13 (3) (1971) 682–688, <http://dx.doi.org/10.1080/00401706.1971.10488830>.
- [68] P. Diaz, A. Doostan, J. Hampton, Sparse polynomial chaos expansions via compressed sensing and D-optimal design, *Comput. Methods Appl. Mech. Eng.* 336 (2018) 640–666, <http://dx.doi.org/10.1016/j.cma.2018.03.020>.
- [69] Z. Wu, D. Wang, P.N. Okolo, K. Zhao, W. Zhang, Efficient space-filling and near-orthogonality sequential latin hypercube for computer experiments, *Comput. Methods Appl. Mech. Eng.* 324 (2017) 348–365, <http://dx.doi.org/10.1016/j.cma.2017.05.020>.
- [70] C. Tong, Refinement strategies for stratified sampling methods, *Reliab. Eng. Syst. Saf.* 91 (10) (2006) 1257–1265, <http://dx.doi.org/10.1016/j.ress.2005.11.027>.
- [71] M. Vořechovský, Hierarchical refinement of Latin Hypercube Samples, *Comput.-Aided Civ. Infrastruct. Eng.* 30 (5) (2015) 394–411, <http://dx.doi.org/10.1111/mice.12088>.
- [72] M. Shields, Refined latinized stratified sampling: a robust sequential sample size extension methodology for high-dimensional latin hypercube and stratified designs, *Int. J. Uncertain. Quantif.* 6 (2016) 79–97.
- [73] M.D. Shields, K. Teferra, A. Hapij, R.P. Daddazio, Refined stratified sampling for efficient Monte Carlo based uncertainty quantification, *Reliab. Eng. Syst. Saf.* 142 (2015) 310–325.
- [74] J.J. Borkowski, E.S. Valeroso, Comparison of design optimality criteria of reduced models for response surface designs in the hypercube, *Technometrics* 43 (4) (2001) 468–477, <http://dx.doi.org/10.1198/00401700152672564>.
- [75] J. Hampton, A. Doostan, Basis adaptive sample efficient polynomial chaos (BASE-PC), *J. Comput. Phys.* 371 (2018) 20–49, <http://dx.doi.org/10.1016/j.jcp.2018.03.035>.
- [76] T. Ishigami, T. Homma, An importance quantification technique in uncertainty analysis for computer models, in: *Proceedings. First International Symposium on Uncertainty Modeling and Analysis*, IEEE Comput. Soc. Press, 1990, pp. 398–403, <http://dx.doi.org/10.1109/isuma.1990.151285>.
- [77] A. Marrel, B. Iooss, B. Laurent, O. Roustant, Calculations of Sobol indices for the Gaussian process metamodel, *Reliab. Eng. Syst. Saf.* 94 (3) (2009) 742–751, <http://dx.doi.org/10.1016/j.ress.2008.07.008>.
- [78] A. Nataf, Détermination des distributions de probabilité dont les marges sont données, *C. R. Acad. Sci.* 225 (1962) 42–43.
- [79] R. Lebrun, A. Dutfoy, An innovating analysis of the Nataf transformation from the copula viewpoint, *Probab. Eng. Mech.* 24 (3) (2009) 312–320, <http://dx.doi.org/10.1016/j.probengmech.2008.08.001>.
- [80] J.D. Jakeman, F. Franzelin, A. Narayan, M. Eldred, D. Pflüger, Polynomial chaos expansions for dependent random variables, *Comput. Methods Appl. Mech. Engrg.* 351 (2019) 643–666, <http://dx.doi.org/10.1016/j.cma.2019.03.049>.
- [81] M. Vořechovský, J. Mašek, J. Eliáš, Distance-based optimal sampling in a hypercube: Analogies to N-body systems, *Adv. Eng. Softw.* 137 (2019) 102709, <http://dx.doi.org/10.1016/j.advengsoft.2019.102709>.
- [82] L. Katafygiotis, K. Zuev, Geometric insight into the challenges of solving high-dimensional reliability problems, *Probab. Eng. Mech.* 23 (2–3) (2008) 208–218, <http://dx.doi.org/10.1016/j.probengmech.2007.12.026>.
- [83] G. Biau, D.M. Mason, High-dimensional p -norms, in: *Mathematical Statistics and Limit Theorems*, Springer International Publishing, 2015, pp. 21–40, http://dx.doi.org/10.1007/978-3-319-12442-1_3.
- [84] R.S. Anderssen, R.P. Brent, D.J. Daley, P.A.P. Moran, Concerning $\int_0^1 \dots \int_0^1 (x_1^2 + \dots + x_k^2)^{1/2} dx_1 \dots dx_k$ and a Taylor series method, *SIAM J. Appl. Math.* 30 (1) (1976) 22–30, <http://dx.doi.org/10.1137/0130003>.
- [85] V. Sadílek, M. Vořechovský, Evaluation of pairwise distances among points forming a regular orthogonal grid in a hypercube, *J. Civ. Eng. Manage.* 24 (5) (2018) 410–523, <http://dx.doi.org/10.3846/jcem.2018.5189>.
- [86] C.C. Aggarwal, A. Hinneburg, D.A. Keim, On the surprising behavior of distance metrics in high dimensional space, in: *Database Theory — ICDT 2001*, Springer Berlin Heidelberg, 2001, pp. 420–434, http://dx.doi.org/10.1007/3-540-44503-x_27.
- [87] A. Flexer, D. Schnitzer, Choosing ℓ^p norms in high-dimensional spaces based on hub analysis, *Neurocomputing* 169 (2015) 281–287, <http://dx.doi.org/10.1016/j.neucom.2014.11.084>.
- [88] M. Jamil, X. Yang, A literature survey of benchmark functions for global optimisation problems, *Int. J. Math. Model. Numer. Optim.* 4 (2013) 150–194, <http://dx.doi.org/10.1504/IJMMNO.2013.055204>.
- [89] M. Vořechovský, J. Mašek, Distance-based optimal sampling in a hypercube: Energy potentials for high-dimensional and low-saturation designs, *Adv. Eng. Softw.* 149 (2020) 102880, <http://dx.doi.org/10.1016/j.advengsoft.2020.102880>.
- [90] G. Migliorati, Adaptive approximation by optimal weighted least-squares methods, *SIAM J. Numer. Anal.* 57 (5) (2019) 2217–2245, <http://dx.doi.org/10.1137/18M1198387>.

3.2 Active Learning-based Domain Adaptive Localized Polynomial Chaos Expansion

NOVÁK, L.; SHIELDS, M.; SADÍLEK, V.; VOŘECHOVSKÝ, M. Active learning-based domain adaptive localized polynomial chaos expansion. MECHANICAL SYSTEMS AND SIGNAL PROCESSING, 2023, vol. 204, ISSN: 0888-3270. (WoS-AIS: D1)

DOI: 10.1016/j.ymssp.2023.110728

Description

The paper extends the active learning methodology based on the Θ criterion to highly non-linear applications. The active learning is used for the decomposition of the input space to small sub-domains associated to the high variance density. The proposed algorithm leads to iterative decomposition of the input space and construction of local low-order PCEs. The numerical examples represent various types of functions: functions with high local non-linearity, functions with discontinuity, and high-dimensional functions with non-linear terms in low-dimensional subspace. The obtained numerical results show the superiority of the proposed method in comparison with the state-of-the-art technique of stochastic spectral embedding based on a similar concept. The discussion presents a detailed insight into the behavior of the algorithm, and explains the role of each of its components in the process of active learning. Additionally, the final part of the discussion is focused on the possible modifications and extensions of the proposed method.

Role of the author

Percentage of contribution: 33%

Lukáš Novák is the main author of this paper responsible for the concept, the methodology, and the numerical results of the presented research. Theoretical concept was created in collaboration with Michael D. Shields and Miroslav Vořechovský. The numerical algorithm was generalized for n -dimensional input space by Václav Sadílek.



Active learning-based domain adaptive localized polynomial chaos expansion

Lukáš Novák^{a,*}, Michael D. Shields^b, Václav Sadílek^a, Miroslav Vořechovský^a

^a Brno University of Technology, Brno, Czech Republic

^b Johns Hopkins University, Baltimore, USA

ARTICLE INFO

Communicated by M. Beer

Keywords:

Polynomial chaos expansion
Adaptive sampling
Sequential sampling
Local approximations
Active learning
Stochastic spectral embedding

ABSTRACT

The paper presents a novel methodology to build surrogate models of complicated functions by an active learning-based sequential decomposition of the input random space and construction of localized polynomial chaos expansions, referred to as domain adaptive localized polynomial chaos expansion (DAL-PCE). The approach utilizes sequential decomposition of the input random space into smaller sub-domains approximated by low-order polynomial expansions. This allows the approximation of functions with strong nonlinearities, discontinuities, and/or singularities that often appear in dynamical systems. Decomposition of the input random space and local approximations alleviates the Gibbs phenomenon for these types of problems and confines error to a very small vicinity near the non-linearity. The global behavior of the surrogate model is therefore significantly better than existing methods, as shown in numerical examples, including an engineering dynamical system exhibiting discontinuous response. The whole process is driven by an active learning routine that uses the recently proposed θ criterion to assess local variance contributions (Novák et al., 2021). The proposed approach balances both *exploitation* of the surrogate model and *exploration* of the input random space and thus leads to efficient and accurate approximation of the original mathematical model. The numerical results show the superiority of the DAL-PCE in comparison to (i) a single global polynomial chaos expansion and (ii) the recently proposed stochastic spectral embedding (SSE) method (Marelli et al., 2021) developed as an accurate surrogate model and which is based on a similar domain decomposition process. This method represents a general framework upon which further extensions and refinements can be based and which can be combined with any technique for non-intrusive polynomial chaos expansion construction.

1. Introduction

The Polynomial Chaos Expansion (PCE), originally proposed by Norbert Wiener [1] and further investigated in the context of engineering problems by many researchers, e.g. [2,3], is a preferred method for uncertainty quantification (UQ) and surrogate modeling in industrial applications [4,5] thanks to its efficiency and powerful post-processing. Once a PCE is available for a given problem, the constructed explicit function can be exploited to directly estimate important properties of the original problem including its statistical moments, response probability distribution or sensitivity indices (without additional sampling [6]), which brings significant efficiency for surrogate modeling, sensitivity analysis, uncertainty quantification and reliability analysis [7].

* Corresponding author.

E-mail addresses: novak.l@fce.vutbr.cz (L. Novák), michael.shields@jhu.edu (M.D. Shields), sadilek.v@fce.vutbr.cz (V. Sadílek), vorechovsky.m@vut.cz (M. Vořechovský).

<https://doi.org/10.1016/j.ymssp.2023.110728>

Received 2 May 2023; Received in revised form 23 July 2023; Accepted 26 August 2023

0888-3270/© 2023 Elsevier Ltd. All rights reserved.

The PCE, in its non-intrusive form, offers a convenient way to perform probabilistic analysis of any black-box model, e.g. finite element models representing complex physical systems in engineering. There are generally two types of non-intrusive methods to calculate the deterministic PCE coefficients: spectral projection and linear regression. The spectral projection approach utilizes the orthogonality of the multivariate polynomials and calculates the coefficients using inner products. The spectral projection leads to an explosion of computational complexity referred to as the *curse of dimensionality*. Therefore, the non-intrusive approach based on linear regression is often preferred. Although it is typically less expensive than the spectral projection (the number of samples should be at least $\mathcal{O}(P \ln(P))$, where P is the number of terms in the PCE [8,9]), it suffers from the *curse of dimensionality* as well, since the number of PCE terms grows rapidly with both dimension and maximum polynomial order. Therefore, it becomes necessary to employ advanced adaptive techniques to construct sparse PCEs that yield efficient solutions for real-world physical systems.

Regression-based PCE can be significantly affected by the selected sampling scheme, as was recently shown in an extensive review paper [10] comparing several general statistical sampling techniques. However, PCE construction as a linear regression model is a very problem specific task and it can be highly beneficial to use methods that exploit information from the given mathematical model and sequentially update the surrogate model — referred to as *active learning*. Active learning is a common approach for surrogate-based reliability analysis, wherein an initial experimental design is iteratively updated based on the current estimate of the limit-state surface [11–16]. Several active learning techniques for Kriging were recently developed [17,18] and used in multi-fidelity framework [19,20]. Active learning for reliability analysis with PCE was used e.g. in [21–23]. However, there is significantly less development in active learning for general UQ, some recent studies have focused on general sequential sampling for PCE based on space-filling criteria or alphabetical optimality [24,25]. However, it is beneficial to use both *exploitation* (leveraging model behavior) criteria and *exploration* (space filling) criteria to define an optimally balanced criterion [26]. Such sequential sampling for sparse Bayesian learning PCE combining both aspects — epistemic uncertainty of the statistical inference (exploration) together with quadratic loss function (local exploitation) – was recently proposed in [27]. However, its application is limited to PCE built by sparse Bayesian learning only.

The authors of this paper recently proposed a general active learning method based on sequential adaptive variance-based sampling [28], which is an efficient tool for accurate surrogate modeling that is sufficiently general for further extension [29]. Although this approach leads to superior results in comparison to standard approaches without active learning, it is limited by the inherently smooth nature of the PCE. More specifically, polynomial basis functions are not able to approximate functions with discontinuities or singularities. Moreover, it is necessary to use high-order polynomials to approximate functions with local non-linearities, even when the rest of the input random space could be easily approximated by a low-order PCE. This can lead to spurious oscillations in the approximation and over-fitting. To overcome this limitation, we propose a method to construct localized PCEs based on the concept of *divide-and-conquer*, i.e. decomposition of the input random space to sub-domains approximated by many low-order PCEs instead of a single high-order global PCE. Although this concept is not entirely new in stochastic finite elements [30] and stochastic collocation [31,32], there is no such approach for non-intrusive PCE. However there are two primary techniques based on similar concepts as described in the following section.

1.1. Related developments

Stochastic Spectral Embedding (SSE) [33] is a general approximation technique based on a decomposition of the input random space and the construction of embedded local approximations. Although it is generally possible to use any spectral approximation technique, it is beneficially coupled with PCE. SSE is based on a novel idea of *embedding* – instead of constructing local approximations of the original mathematical model, local surrogates are constructed to approximate the *residuals* between the model and approximation from the previous level of the decomposed space. Although such an approach can lead to significant improvement in comparison to a single global approximation [33], it is not a sequential approach based on active learning and thus it does not iteratively reflect new information obtained from the previous steps of the algorithm. Active learning is crucial in analysis of functions with discontinuity or singularity because it allows for the aforementioned exploration and exploitation necessary to find and resolve these features. For the sake of completeness, active learning for SSE has been proposed for reliability analysis [34], but it does not lead to an accurate approximation over the entire input random space. Its accuracy is limited to regions around the limit surface, which are important for an estimation of failure probability.

The second related technique is Multi-element generalized Polynomial Chaos Expansion (ME-gPC) [35] and its modification based on ANOVA decomposition [36], which improves its computational efficiency for high dimensions. ME-gPC was developed as an extension of generalized PCE based on Wiener-Askey scheme [37] allowing analysis of models with arbitrary distribution of input random vector. The ME-gPC method consists of three main parts: decomposition of the input random space, numerical construction of locally orthogonal polynomials and an adaptive procedure based on the decay rate of local error in estimated variance derived from local PCE. ME-PCE applies an h -type mesh refinement procedure akin to mesh refinement in finite element methods. By doing so, they introduce a structured grid of uniform points in each new element and solve for the PCE coefficients. This can be cumbersome and does not afford the flexibility to adaptively select sparse and near-optimal training points. Although ME-gPC was recently improved by p -adaptivity in anisotropic ME-gPC [38], we note that the ME-gPC was created mainly for uncertainty propagation in models with arbitrary input distributions, and thus in contrast to SSE, its objective is not necessarily to construct the best possible surrogate model using adaptive algorithms, but rather to minimize errors in response statistics. This is a subtle, but important difference that distinguishes its use as a predictive tool from that of a tool for statistical estimation.

Active learning has also been investigated in the context of collocation methods. Jakeman et al. [39] developed an efficient method for the detection of discontinuities followed by a decomposition of the input random space to several sub-domains

separated by discontinuities. The adaptive simplex stochastic collocation method [31,40] extends the applicability of ME-gPC to high-dimensional models on non-hypercube spaces. Further improvement of computational efficiency was achieved by hp -adaptivity in ME-gPC with re-use of existing samples in adapted weighted Leja collocation [41]. Although these recent developments in stochastic collocation show very promising results, they are specific to collocation schemes (usually on sparse grids) whose aim is integration and not surrogate model construction. Unlike these collocation methods, this paper focuses on the development of domain decomposition and active learning-based sequential sampling for constructing surrogate models with generally arbitrary sampling schemes.

1.2. Contributions of this paper

This paper describes a novel method, termed Domain Adaptive Localized PCE (DAL-PCE), that applies adaptive sequential decomposition of the input random space and adaptive sequential sampling within the sub-domains. Both of these features are based on the recently proposed criterion for variance-based sequential statistical sampling, developed specifically for PCE in [42]. In the context of previously described methods SSE and ME-gPC, the proposed novel approach can be thought to lie between them. Like SSE, it is developed specifically for the construction of accurate surrogate models, especially for functions with high non-linearity or discontinuity. But the decomposition of the input random space is rather similar to ME-gPC. The uniqueness of our proposal lies in the combination of active learning, sequential sampling, sequential decomposition of the input space and regression-based PCE using sparse solvers such as Least Angle Regression (LARS), allowing adaptivity and learning in each iteration of the proposed algorithm.

2. Polynomial chaos expansion

Consider a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where Ω is an event space, \mathcal{F} is a σ -algebra on Ω and \mathcal{P} is a probability measure on \mathcal{F} . If the input variable of a mathematical model, $Y = f(X)$, is a random variable $X(\omega)$, $\omega \in \Omega$, the model response $Y(\omega)$ is also a random variable. Assuming that Y has finite variance, PCE represents the output variable Y as a function of another random variable ξ called the *germ* with a known distribution

$$Y = f(X) = f^{\text{PCE}}(\xi), \quad (1)$$

and represents the function $f(X)$ via infinite polynomial expansion. A set of polynomials, orthogonal with respect to the distribution of the germ, are used as a basis of the Hilbert space $L^2(\Omega, \mathcal{F}, \mathcal{P})$ of all real-valued random variables of finite variance, where \mathcal{P} takes over the meaning of the probability distribution. The orthogonality condition is given by the inner product of $L^2(\Omega, \mathcal{F}, \mathcal{P})$ defined for any two functions ψ_j and ψ_k for all $j \neq k$ with respect to the weight function p_ξ (probability density function of ξ) as

$$\langle \psi_j, \psi_k \rangle = \int \psi_j(\xi) \psi_k(\xi) p_\xi(\xi) d\xi = 0. \quad (2)$$

This means that there are specific orthogonal polynomials associated with the corresponding distribution of the germ via its weighting function. For example, Hermite polynomials orthogonal to the Gaussian measure are associated with normally distributed germs. Orthogonal polynomials corresponding to other distributions can be chosen according to Wiener-Askey scheme [37] or constructed numerically [43]. For further processing, it is beneficial to use normalized polynomials (orthonormal), where the inner product of i th and j th polynomials is equal to the Kronecker delta δ_{jk} , i.e., $\delta_{jk} = 1$ if and only if $j = k$, and $\delta_{jk} = 0$ otherwise.

In the case of X and ξ being vectors containing M independent random variables, the polynomial $\Psi(\xi)$ is multivariate and it is built up as a tensor product of univariate orthonormal polynomials, i.e.,

$$\Psi_\alpha(\xi) = \prod_{i=1}^M \psi_{\alpha_i}(\xi_i), \quad (3)$$

where $\alpha \in \mathbb{N}^M$ is a set of integers called the *multi-index* reflecting polynomial degrees associated to each ξ_i . The quantity of interest (QoI), i.e. the response of the mathematical model $Y = f(X)$, can then be represented as [3]

$$Y = f(X) = \sum_{\alpha \in \mathbb{N}^M} \beta_\alpha \Psi_\alpha(\xi), \quad (4)$$

where β_α are deterministic coefficients and Ψ_α are multivariate orthonormal polynomials.

2.1. Non-intrusive computation of PCE coefficients

For practical computation, the PCE expressed in Eq. (4) must be truncated to a finite number of terms P . One can generally choose any truncation rule (e.g., the tensor product of polynomials up to the selected order p), but the most common truncation is achieved by retaining only terms whose total degree $|\alpha|$ is less than or equal to a given p , in which case the truncated set of PCE terms is then defined as

$$\mathcal{A}^{M,p} = \left\{ \alpha \in \mathbb{N}^M : |\alpha| = \sum_{i=1}^M \alpha_i \leq p \right\}. \quad (5)$$

The cardinality of the truncated *index set* $\mathcal{A}^{M,p}$ is given by

$$\text{card } \mathcal{A}^{M,p} = \frac{(M+p)!}{M! p!} \equiv P. \quad (6)$$

When the PCE is truncated to a finite number of terms, there is an error ε in the approximation such that

$$Y = f(\mathbf{X}) = \sum_{\alpha \in \mathcal{A}} \beta_{\alpha} \Psi_{\alpha}(\xi) + \varepsilon.$$

From a statistical point of view, PCE is a simple linear regression model with intercept. Therefore, it is possible to use *ordinary least squares* (OLS) regression to minimize the error ε . Knowledge of vector β fully characterizes the approximation via PCE. To solve for β , first it is necessary to create N_{sim} realizations of the input random vector \mathbf{X} and the corresponding results of the original mathematical model \mathcal{Y} , together called the experimental design (ED). Then, the vector of P deterministic coefficients β can be determined by OLS as

$$\beta = (\Psi^T \Psi)^{-1} \Psi^T \mathcal{Y}, \quad (7)$$

where Ψ is the data matrix

$$\Psi = \{ \Psi_{ij} = \Psi_j(\xi^{(i)}), i = 1, \dots, N_{\text{sim}}, j = 0, \dots, P-1 \}. \quad (8)$$

A well-known problem, the *curse of dimensionality*, states that P is highly dependent on the number of input random variables M and the maximum total degree of polynomials p , which is clear from Eq. (6). Considering that estimation of β by regression requires at least $\mathcal{O}(P \ln(P))$ number of samples for stable solution [8,9], the problem can become computationally highly demanding in case of a large or strongly non-linear stochastic models. Although one can use advanced model selection algorithms such as Least Angle Regression (LAR) [2,44], orthogonal matching pursuit [45] or Bayesian compressive sensing [46] to find an optimal set of PCE terms, and thus reduce the number of samples needed to compute the unknown coefficients, the benefit of these techniques is significant only if the true coefficient vector is sparse or compressible. The sparse set of basis functions obtained by any adaptive algorithm is further denoted by \mathcal{A} for the sake of clarity.

2.2. Approximation error estimation

Once the PCE is constructed, it is crucial to estimate its accuracy. Further, the PCE accuracy can be used to directly compare several PCEs to choose the best surrogate model. Ideally, the ED should be divided into validation and training sets, but this might be extremely computationally demanding in engineering applications with complex numerical models. Therefore in the field of uncertainty quantification (UQ) of engineering models, it is preferred to estimate the approximation error directly from the training set without any additional sampling of the original model. A common choice is the coefficient of determination R^2 , which is well-known from machine learning or statistics. However, R^2 may lead to over-fitting, and thus, advanced methods should be used. One of the most widely-used methods is the leave-one-out cross-validation (LOO-CV) error Q^2 . The LOO-CV is based on residuals between the original surrogate model and the surrogate model built with the ED while excluding one realization. This approach is repeated for all realizations in the ED and the average error is estimated. Although the calculation of Q^2 is typically highly time-consuming, it is possible to obtain results analytically from a single PCE as follows [47]

$$Q^2 = \frac{\frac{1}{N_{\text{sim}}} \sum_{i=1}^{N_{\text{sim}}} \left[\frac{g(\mathbf{x}^{(i)}) - g^{\text{PCE}}(\mathbf{x}^{(i)})}{1 - h_i} \right]^2}{\sigma_{Y,\text{ED}}^2}, \quad (9)$$

where $\sigma_{Y,\text{ED}}^2$ is the variance of the ED calculated using the original mathematical model and h_i represents the i th diagonal term of matrix $\mathbf{H} = \Psi (\Psi^T \Psi)^{-1} \Psi^T$.

2.3. Statistical moments derived from PCE

The form of PCE as a linear summation over orthonormal polynomials allows for powerful and efficient post-processing. In particular, once a PCE approximation is created, it is possible to directly estimate statistical moments of the output from the expansion.

The first statistical moment (the mean value) is simply the first deterministic coefficient of the expansion $\mu_Y = \langle Y^1 \rangle = \beta_0$. The second raw statistical moment, $\langle Y^2 \rangle$, can be estimated by

$$\begin{aligned} \langle Y^2 \rangle &= \int \left[\sum_{\alpha \in \mathcal{A}} \beta_{\alpha} \Psi_{\alpha}(\xi) \right]^2 p_{\xi}(\xi) d\xi = \sum_{\alpha_1 \in \mathcal{A}} \sum_{\alpha_2 \in \mathcal{A}} \beta_{\alpha_1} \beta_{\alpha_2} \int \Psi_{\alpha_1}(\xi) \Psi_{\alpha_2}(\xi) p_{\xi}(\xi) d\xi \\ &= \sum_{\alpha \in \mathcal{A}} \beta_{\alpha}^2 \int \Psi_{\alpha}(\xi)^2 p_{\xi}(\xi) d\xi = \sum_{\alpha \in \mathcal{A}} \beta_{\alpha}^2 \langle \Psi_{\alpha}, \Psi_{\alpha} \rangle. \end{aligned} \quad (10)$$

Considering the orthonormality of the polynomials, it is possible to obtain the variance $\sigma_Y^2 = \langle Y^2 \rangle - \mu_Y^2$ as the sum of all squared deterministic coefficients except the intercept (which represents the mean value), i.e.

$$\sigma_Y^2 = \sum_{\substack{\alpha \in \mathcal{A} \\ \alpha \neq 0}} \beta_\alpha^2. \quad (11)$$

Note that the computation of higher statistical central moments, specifically skewness γ_Y (3rd moment) and kurtosis κ_Y (4th moment), are more complicated since they require triple and quad products. These can be obtained analytically only for certain polynomial families, e.g., formulas for Hermite and Legendre polynomials (and their combination) can be found in [42].

3. Active learning-based domain adaptive localized PCE (DAL-PCE)

In this section, we propose a novel methodology to construct localized PCEs designed for highly non-linear functions, termed Domain Adaptive Localized PCE (DAL-PCE). Instead of increasing the maximum polynomial order p (p -adaptivity), which brings high computational requirements due to the *curse of dimensionality*, we propose to decompose the input random space into several sub-domains approximated by low-order PCEs (h -adaptivity). Although this idea is not entirely new, we use this approach in combination with novel active learning methods to identify domains for refinement and for sequential sample selection and regression-based PCEs. This allows us to use any sparse adaptive solver (e.g. LAR), and thus it can be easily implemented into the existing software packages [48,49]. In the following sections, we define the requisite components of the proposed method and provide an algorithm (Algorithm 1) for its implementation.

3.1. Variance-based adaptive sequential sampling

The decomposition of the input random space is a sequential process coupled with adaptive sampling, assuring optimal coverage of the sub-domains of interest. The whole process thus consists of two steps: (i) identification of an important sub-domain, that is, a domain that is either large compared to other sub-domains or that is associated with a high local variance; and (ii) identification of the best positions for additional samples extending the current ED in the selected sub-domain. Each of these steps must be based on a criterion that balances *exploration* of the input random space with *exploitation* of the surrogate model, which in our case is in the form of a PCE. The Θ -criterion for adaptive sequential sampling, which is driven by the output variance and its approximation via local variance using PCE [28], is employed for both steps. We will first discuss the process for adaptive sequential sampling within a specified sub-domain in this section. This will be followed by the process for refinement of the domain in the subsequent sections.

Consider a pool of candidate samples containing realizations of the random vector ξ generated by an arbitrary sampling technique, e.g., Latin Hypercube Sampling (LHS) [50,51] or Coherence sampling [8,52,53]. From this pool of candidates, we select the best sample using a method inspired by the sequential sampling proposed in [26] and based on Koksma-Hlawka inequality [54]. The Θ -criterion for PCE, which accounts for both variation of the function and discrepancy of the samples, was proposed as follows [28]

$$\Theta(\xi^{(c)}) \equiv \Theta^c = \underbrace{\sqrt{\sigma_A^2(\xi^{(c)}) \cdot \sigma_A^2(\xi^{(s)})}}_{\text{ave variance density}} \underbrace{l_{c,s}^M}_{\text{vol.}} \equiv \sqrt{\sigma_c^2 \cdot \sigma_s^2} l_{c,s}^M. \quad (12)$$

The criterion is a product of two terms — the *exploitation* term (denoted as “ave variance density”) and the *exploration* part (the distance term $l_{c,s}$ raised to the domain dimension) – which are multiplied to maintain an optimal balance between exploration and exploitation [28].

The *exploration* aspect is maintained by accounting for the distance $l_{c,s}$ between a candidate $\xi^{(c)}$ and its nearest neighboring realization from the existing ED, $\xi^{(s)}$ as

$$l_{c,s} = \sqrt{\sum_{i=1}^M |\xi_i^{(c)} - \xi_i^{(s)}|^2}. \quad (13)$$

If the criterion was reduced to this term only, sequential filling of the greatest empty regions would occur, converging to uniform space coverage in the spirit of the space-filling “miniMax criterion” [55–57].

The *exploitation* component is motivated by the desire to sample points in regions with the greatest contributions to the total variance of the QoI σ_Y^2 , i.e. at points with the highest *variance density*. Once the PCE has been established at any given stage of the algorithm, the *variance density* is computationally cheap to evaluate for any location ξ as

$$\sigma_A^2(\xi) = \left[\sum_{\substack{\alpha \in \mathcal{A} \\ \alpha \neq 0}} \beta_\alpha \Psi_\alpha(\xi) \right]^2 p_\xi(\xi). \quad (14)$$

The local variance is therefore estimated directly using the basis functions and coefficients β of the PCE. When considering a candidate “c”, an estimate of the variance contribution of the region between the candidate and its nearest neighbor “s” may be obtained by averaging the local variance densities between the two. Therefore, we can say that the candidate with the greatest Θ^c criterion is the one that represents the largest amount of total variance to be refined by its selection.

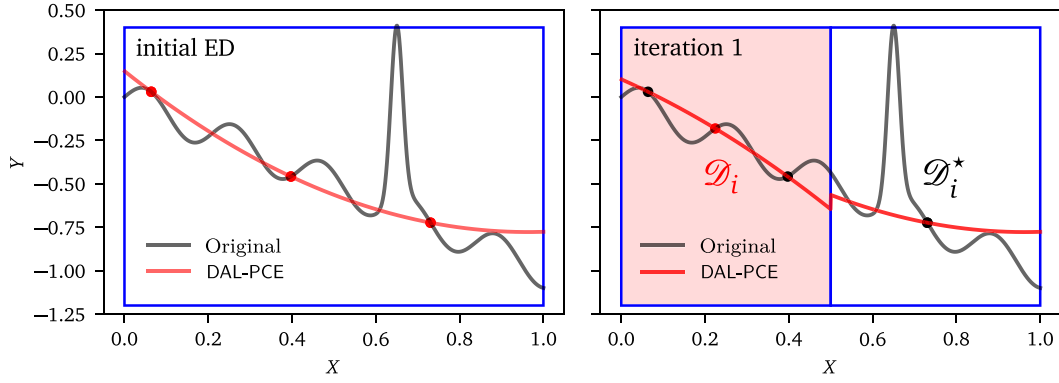


Fig. 1. The first iteration of the algorithm: the original sub-domain is split and the new local PCE is constructed in \mathcal{D}_i (red background), while the second part in \mathcal{D}_i^* inherits the PCE approximation from the original domain.

A significant advantage of this method is the ability to add candidates to an existing ED one-by-one. Thus, it can be employed at any moment of the PCE construction process. Moreover, this learning function can be combined with any sampling algorithm for the construction of the initial ED and candidates for extension. The ideas behind the θ criterion will now be used in the proposed domain decomposition and ED extension algorithm.

3.2. Decomposition of input random space

The core of the proposed approach is a sequential decomposition of the input random space \mathcal{D} for the construction of local approximations. This approach assumes that the original mathematical model can be approximated by piecewise low-order PCEs that are valid only in individual sub-domains of \mathcal{D} . Therefore, in the proposed approach, the input random space is sequentially decomposed into $n_{\mathcal{D}}$ smaller non-overlapping sub-domains $\mathcal{D}_i \subset \mathcal{D}$ that collectively fill the full input random space \mathcal{D} , i.e.

$$\bigcup_{i=1}^{n_{\mathcal{D}}} \mathcal{D}_i = \mathcal{D} \quad \text{such that} \quad \mathcal{D}_i \cap \mathcal{D}_j = \emptyset \quad \forall i, j. \quad (15)$$

In each iteration of the algorithm, a single sub-domain \mathcal{D}_i (referred to as the parent) is identified for refinement and divided by a plane perpendicular to the direction of one selected input random variable. Specifically, \mathcal{D}_i is divided into a refinement-child \mathcal{D}_i , which is further processed, and an inheriting-child \mathcal{D}_i^* adopting the PCE from the parent as illustrated for a one-dimensional function in Fig. 1. In this case, we see that the space is divided into two subdomains. In the left (refinement child) a new PCE is constructed. In the right (inheriting child), the original PCE is retained. Such process assures an exhaustive decomposition into disjoint subsets i.e. $\mathcal{D}_i = \mathcal{D}_i \oplus \mathcal{D}_i^*$. This sequential domain decomposition is illustrated in Fig. 2, which depicts the original input random space and the first four iterations of the decomposition process.

In contrast to SSE [33], the selection of a single sub-domain for refinement in each iteration is based on an active learning approach, the details of which are provided in subsequent sections. Importantly, actively integrating information from the original mathematical model leads to a significantly more effective decomposition of the space and thus assures accurate approximations, even for small-size EDs. On the other hand, the identified decomposition and the associated ED are directly connected to the given mathematical model and, therefore, might be inefficient for general statistical analysis.

The complete surrogate model is assembled from the $n_{\mathcal{D}}$ local PCEs associated with all sub-domains \mathcal{D}_i as

$$Y \approx \sum_{i=0}^{n_{\mathcal{D}}} \sum_{\alpha_i \in \mathcal{A}_i} \beta_{\alpha_i} \Psi_{\alpha_i}(\xi) \mathbb{1}_{\mathcal{D}_i}(\xi), \quad (16)$$

where $\mathbb{1}_{\mathcal{D}_i}(\xi)$ represents indicator function, i.e. $\mathbb{1}_{\mathcal{D}_i}(\xi) = 1$ only if $\xi \in \mathcal{D}_i$ and $\mathbb{1}_{\mathcal{D}_i}(\xi) = 0$ otherwise. In other words, to approximate the original model at any point, it suffices to determine the one relevant sub-domain and use the corresponding local PCE. Each such local PCE has its own set of basis functions \mathcal{A}_i and corresponding coefficients β_{α_i} , which can be obtained by any model-selection algorithm. In this paper, the OLS and LAR algorithms are employed, but generally, any non-intrusive technique can be used.

3.3. Domain selection via modified variance-based criterion

The selection process to identify the “best” subdomain for possible division is governed by extending the θ -criterion from Eq. (12) as follows

$$\theta_i = \underbrace{\mathcal{W}_i \cdot \exp(Q_i^2)}_{\text{weight of subdomain}} \cdot \underbrace{\sqrt{\sigma_{\mathcal{A}_i}^2(\xi^{(c)}) \cdot \sigma_{\mathcal{A}_i}^2(\xi^{(s)})}}_{\theta^c \text{ in } i\text{th subdomain}} I_{c,s}^M. \quad (17)$$

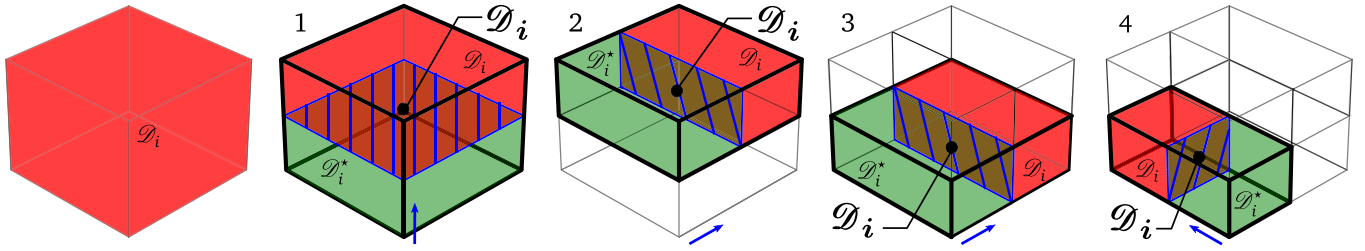


Fig. 2. The first four steps of the decomposition of a 3D space of input random variables. The thick black lines outline the parent domain selected for division. The red and green boxes inside it represent the two newly created refinement-child \mathcal{D}_i (red) and inheriting-child \mathcal{D}_i^* (green) sub-domains created by splitting the parent domain \mathcal{D}_i (bold boundaries), selected via Eq. (17), by the cutting plane (blue). The cutting plane is perpendicular to the variable selected for splitting (blue arrow).

This extended criterion aims to identify sub-domains of the input random space associated with the maximum value of θ^c , while simultaneously accounting for the size of each subdomain and the accuracy of the existing local PCE. The former is calculated using Eq. (12) calculated for a rich pool of screening global candidates, while the latter are measured by incorporating the volume of each sub-domain \mathcal{W}_i and the LOO-CV error Q_i^2 , respectively. The LOO-CV term, $\exp(Q_i^2)$, can be thought to artificially inflate the domain volume as a penalization for inaccurate approximation. When the approximation is perfect ($Q_i^2 = 0$), the true volume of the sub-domain is used. Meanwhile, a poor approximation with $Q_i^2 = 1$ leads to roughly 2.72 times increased volume.

The three terms featured in Eq. (17) aim at different aspects affecting the accuracy of the final surrogate model: large sub-domains are preferred by \mathcal{W}_i , sub-domains containing poor PCE approximation are promoted via $\exp(Q_i^2)$ and finally, θ^c prefers sub-domains with a high concentration of variance. Note that θ^c is calculated for a rich pool of screening candidates, and \mathcal{W}_i and $\exp(Q_i^2)$ are calculated directly from the geometry of existing sub-domain and the local PCE model, respectively. The product of all three terms in the extended criterion, therefore, maintains the desired balance and assures the selection of the sub-domain, \mathcal{D}_i , that currently seems to be the most important for increasing the accuracy of the PCE surrogate model.

Sub-domain \mathcal{D} with the greatest θ_i is selected and one of the operations described in detail in Section 3.6 is performed, depending on whether \mathcal{D}_i contains a critical number of ED points. Two scenarios can occur:

- \mathcal{D}_i contains a sufficient number of ED points ($n_i \geq n_{\text{sim}}$) to ensure accuracy of a PCE on the domain. Therefore, it becomes a parent \mathcal{D}_i (bold boundaries in Fig. 2) and is divided into two parts by a selected rule. The child domain containing the decisive candidate with the greatest θ^c becomes the refinement-child \mathcal{D}_i (see the red subdomains in steps 1 – 4 in Fig. 2). The remaining volume becomes an inheriting-child denoted \mathcal{D}_i^* (see the green subdomains in Fig. 2), which retains the PCE from the parent. Division occurs by a cutting plane, oriented perpendicular to the selected direction (blue arrows in Fig. 2), and naturally, the coordinates of the cutting plane are restricted to the bounding box of the selected parent \mathcal{D}_i , see Section 3.6. If needed, the refinement-child domain \mathcal{D}_i is sequentially filled with additional ED points (according to θ^c) to reach $n_i = n_{\text{sim}}$ needed to construct a new PCE approximation.
- \mathcal{D}_i does *not* contain a sufficient number of ED points ($n_i < n_{\text{sim}}$). The domain is not divided because the suggestion for division is based on insufficient information. Instead, new ED points are sequentially added to \mathcal{D}_i , again using the θ^c criterion. Note that this scenario practically arises when the selected domain was an inheriting-child in the previous iteration. In this case, the selected domain has inherited a PCE model that was constructed over a larger domain. When that domain was divided, it was left with an insufficient number of points from which to construct a new PCE.

3.4. PCE basis functions

Without loss of generality, the proposed method operates on the M -dimensional unit hypercube with uniform distributions of input random variables, i.e., $\mathbf{X} \sim \mathcal{U}[0, 1]^M$. In the case of a general joint probability distribution of \mathbf{X} , it is always possible to transform input random vector to the unit hypercube by Rosenblatt transformation [58], Nataf transformation [59] or various methods based on copulas [60]. Standard normalized Legendre polynomials, orthonormal to the uniform distribution, can thus be used as basis functions for the PCE. However, due to the decomposition of the input random space to smaller sub-domains, each with lower bound a_i and upper bound b_i , it is necessary to use univariate scaled orthonormal Legendre polynomials of n th order $\tilde{\psi}_n(\xi)$ defined as follows

$$\tilde{\psi}_n(\xi) = \psi_n\left(\frac{2\xi - a_i - b_i}{b_i - a_i}\right), \quad (18)$$

where ψ_n represents standard orthonormal Legendre polynomials. Naturally, the transformation of the original input random vector to the unit hypercube might bring additional non-linearity, and thus one might prefer the direct construction of polynomials locally orthonormal to the given original probability measure as proposed in the Me-gPC [35]. While certainly possible, this brings additional computational demands and thus it is not employed here.

3.5. Local and global statistical estimates from DAL-PCE

The significant advantage of PCE is that analytically post-processing of the expansion yields highly efficient estimates of statistical moments [42], sensitivity indices [6] and LOO-CV [2]. In the proposed DAL-PCE, since the original domain \mathcal{D} is decomposed into a set of sub-domains (see Eq. (15)), standard analytical post-processing can be applied locally, and global characteristics can be obtained by simple weighted summations that converge to the true values as $n_{\mathcal{D}}$ increases. Specifically, the global mean value and variance of a QoI are obtained from localized PCEs (denoted by subscript \mathcal{D}_i) as follows

$$\mu_Y = \sum_{i=1}^{n_{\mathcal{D}}} \mathcal{W}_i \beta_{0_i} = \sum_{i=1}^{n_{\mathcal{D}}} \mathcal{W}_i \mu_{\mathcal{D}_i}, \quad (19)$$

$$\sigma_Y^2 = \sum_{i=1}^{n_{\mathcal{D}}} \mathcal{W}_i \sum_{\substack{\alpha_i \in \mathcal{A}_i \\ \alpha_i \neq \emptyset}} \beta_{\alpha_i}^2 = \sum_{i=1}^{n_{\mathcal{D}}} \mathcal{W}_i \sigma_{\mathcal{D}_i}^2, \quad (20)$$

where the local mean $\mu_{\mathcal{D}_i}$ and variance $\sigma_{\mathcal{D}_i}^2$ are obtained as described in Section 2.3.

Local Sobol' indices, $S_{\mathcal{D}_i}$, of any order, can be derived directly from localized PCEs and their first-order (main effect) estimates are given by

$$S_{\mathcal{D}_i}^{X_j} = \frac{1}{\sigma_{\mathcal{D}_i}^2} \sum_{\alpha_i \in \mathcal{A}_i^{X_j}} \beta_{\alpha_i}^2 \quad \mathcal{A}_i^{X_j} = \left\{ \alpha_i \in \mathcal{A}_i : \alpha_i^j > 0, \alpha_i^{k \neq j} = 0 \right\}. \quad (21)$$

These local Sobol' indices are used in the DAL-PCE to determine the cut direction (see Section 3.6). Likewise, global Sobol' indices can be obtained easily from the weighted summation of local contributions to partial variances normalized by σ_Y^2 as follows

$$S_{X_j} = \frac{\sum_{i=1}^{n_{\mathcal{D}}} \mathcal{W}_i \sum_{\alpha_i \in \mathcal{A}_i^{X_j}} \beta_{\alpha_i}^2}{\sigma_Y^2}. \quad (22)$$

Similarly, global LOO-CV, Q^2 , of a QoI can be approximated by the weighted summation of the local contributions as

$$Q^2 = \sum_{i=1}^{n_{\mathcal{D}}} \mathcal{W}_i Q_{\mathcal{D}_i}^2, \quad (23)$$

where $Q_{\mathcal{D}_i}^2$ are obtained from each local PCE using Eq. (9).

These estimates are used throughout the proposed DAL-PCE, as described in detail next.

3.6. Numerical algorithm

Based on the presented theoretical background, we now present the numerical algorithm for the domain adaptive localized PCE. As mentioned above, the whole process can be divided to two iterative tasks: (i) decomposition of the input random space and (ii) construction of localized PCEs. Both of these tasks are described in the following paragraphs with specific reference to the steps in Algorithm 1.

The first task identifies the important sub-domain \mathcal{D}_i that should be divided and over which low-order local PCE should be constructed. The sub-domain \mathcal{D}_i is specifically identified using the θ_i criterion from Eq. (17), which again incorporates three important characteristics for accurate surrogate modeling — the size of the sub-domain \mathcal{W}_i , the accuracy of the existing local PCE measured by $Q_{\mathcal{D}_i}^2$, and the original θ^c criterion measuring the variance contribution in \mathcal{D}_i . While \mathcal{W}_i and $Q_{\mathcal{D}_i}^2$ are computed for the whole sub-domain, θ^c is computed at specific realizations of input random vector. Therefore, it is necessary to cover the sub-domains with a sufficiently large number of screening candidates, such that the total global number of screening candidates is given by $n_{c,g}$. Based on numerical experiments, we recommend $n_{c,g} \geq 1000 M$ to ensure that each sub-domain contains a sufficient number of screening candidates. Note that the screening candidates are used only to identify \mathcal{D}_i [step 5]. They are not used for the ED, and thus even high $n_{c,g}$ does not bring any additional computational demand.

Once \mathcal{D}_i is identified, it is necessary to check whether there are enough samples to construct a PCE inside the sub-domain. We start with finding out how many points belong to the selected domain \mathcal{D}_i [step 6]. If the number of samples in the identified sub-domain, n_i , is greater than (or equal to) n_{sim} [step 7], a local PCE already exists for \mathcal{D}_i . The subdomain is then assigned as a parent \mathcal{D}_i for division [step 8] and the first-order Sobol' indices are estimated by Eq. (22) [step 9]. This identified parent \mathcal{D}_i is divided in the direction of the highest first-order Sobol' index $S_{\mathcal{D}_i}^{X_j}$. The new restricted coordinates of refinement-child \mathcal{D}_i are identified and the inheriting-child \mathcal{D}_i^* is created [step 10]. Further, the number of ED samples n_i in the refinement-child \mathcal{D}_i is determined [step 11]. On the other hand, if the identified sub-domain \mathcal{D}_i does not contain enough samples (i.e. $n_i < n_{\text{sim}}$), the inherited PCE from the previous iteration is not sufficiently local (it was trained over a domain that has since been divided) and it is necessary to add new samples to \mathcal{D}_i before constructing a new local PCE.

The second task of the proposed algorithm is sequential sampling and adaptive PCE construction in sub-domain \mathcal{D}_i . Recall that this domain may be either

Algorithm 1 DAL-PCE: Active Domain Decomposition and Construction of Localized PCEs

Input: maximum local polynomial order p , number of screening global candidates $n_{c,g}$, number of local candidates $n_{c,l}$, number of iterations n_{iter}

- 1: set the minimum number of realizations for local PCE construction $n_{sim} \in \langle P, 2P \rangle$
- 2: generate a rich pool of $n_{c,g}$ screening candidates
- 3: generate the initial ED (size n_{sim}) and construct the initial global PCE
- 4: **for** 1 to n_{iter} **do**
- 5: identify the sub-domain \mathcal{D}_i with the highest θ_i based on screening candidates
- 6: $n_i \leftarrow$ number of ED samples existing in \mathcal{D}_i
- 7: **if** $n_i \geq n_{sim}$ **then**
- 8: the identified sub-domain \mathcal{D}_i becomes a parent \mathcal{D}_i
- 9: identify the direction of the highest first-order Sobol' index $S_{\mathcal{D}_i}$ of the parent \mathcal{D}_i
- 10: restrict coordinates of $\mathcal{D}_i \rightarrow \mathcal{D}_i^*$ and create \mathcal{D}_i^*
- 11: $n_i \leftarrow$ number of ED samples existing in \mathcal{D}_i
- 12: **end if**
- 13: generate $n_{c,l}$ local candidates in \mathcal{D}_i
- 14: **while** $n_i < n_{sim}$ **do**
- 15: extend size of local ED n_i using the local θ^c criterion
- 16: **end while**
- 17: reconstruct local PCEs in the \mathcal{D}_i
- 18: **end for**

Output: list of subdomains and corresponding PCEs

- (i) a refinement-child that was just divided but does not contain a sufficient number of points ($n_i < n_{sim}$) or,
- (ii) an inheriting-child that now does not contain at least n_{sim} ED samples.

Next, a set of local candidates is generated in region \mathcal{D}_i [step 13]. To ensure sufficient assessment of the coverage of the domain, the number of local candidates is empirically recommended as $n_{c,l} \in \langle 3P, 5P \rangle$ [28]. From these candidates, the standard θ^c criterion in Eq. (12) is used to iteratively select the best candidates until there are n_{sim} samples in \mathcal{D}_i [step 14-16]. This sequential extension of the sample in \mathcal{D}_i is adaptive in the sense that the pairwise distances in Eq. (12) between candidates and existing ED points are updated after the addition of each new point. However, because $n_i < n_{sim}$ the local variance densities are estimated from the previously existing PCE, which cannot be updated until a sufficient number of samples are available in \mathcal{D}_i .

The last step of each iteration is to construct the local PCE using scaled Legendre polynomials as basis functions (see Eq. (18)) [step 17]. Any non-intrusive technique can be used to estimate the coefficients β ; we use LARS and OLS for an adaptive construction of the local PCEs in this paper. At the end of the iteration, all sub-domains are re-numbered and a list of sub-domains with corresponding PCEs can be exported, or the next iteration can be started. Note that the final surrogate model according to Eq. (16) is discontinuous, which might be problematic for continuous functions. However, this problem is less significant with an increasing number of iterations because the sub-domains are smaller, and the θ criterion prefers samples near boundaries.

3.7. Adaptivity in PCE construction and domain decomposition

Adaptivity is central to the proposed DAL-PCE. In the proposed algorithm, there are two types of adaptivity employed:

- (i) adaptivity in PCE construction (selection of the optimal set of basis functions), and
- (ii) adaptivity in domain decomposition

Since the PCE can be constructed by any regression technique in each sub-domain, PCE adaptivity is incorporated by sparse solvers and best model selection algorithms, e.g. Least Angle Regression [44], orthogonal matching pursuit [45] or Bayesian compressive sensing [46]. Although sparse solvers are often used for PCE with high p , this adaptivity is also important for reducing the number of basis functions (and thus the minimum number of ED samples) for high-dimensional examples or, in our case, for very low-size ED in each \mathcal{D}_i approximated by low- p local PCE.

The second type of adaptivity is the proposed adaptivity in the domain decomposition. At any point in the iterative process, the existing ED samples can be used to construct local PCEs or a single global PCE. The DAL-PCE is not guaranteed to provide a better approximation than the global PCE. This can be measured via Q^2 , specifically by computing Q_{local}^2 from Eq. (23) and Q_{global}^2 from a single global PCE according to Eq. (9). If $Q_{local}^2 > Q_{global}^2$ at a given iteration, the domain decomposition is deemed to be poor and the whole decomposition process is *re-started*. That is, the complete geometrical decomposition is forgotten and all existing ED points are taken as an initial ED for a brand new run of the algorithm. This is illustrated in Fig. 3, which shows the decomposition (top) and the associated error (bottom) right before the restart (a) at $N_{sim} = 181$, (b) the new decomposition and error right after the restart, and (c) the final decomposition/error, which shows significant improvement over the global PCE. These histories show the standard R^2 error defined in Eq. (24). It is not necessary to check this criterion at every iteration, but it is suggested to check it periodically, every n_r steps, to ensure adequate local refinement.

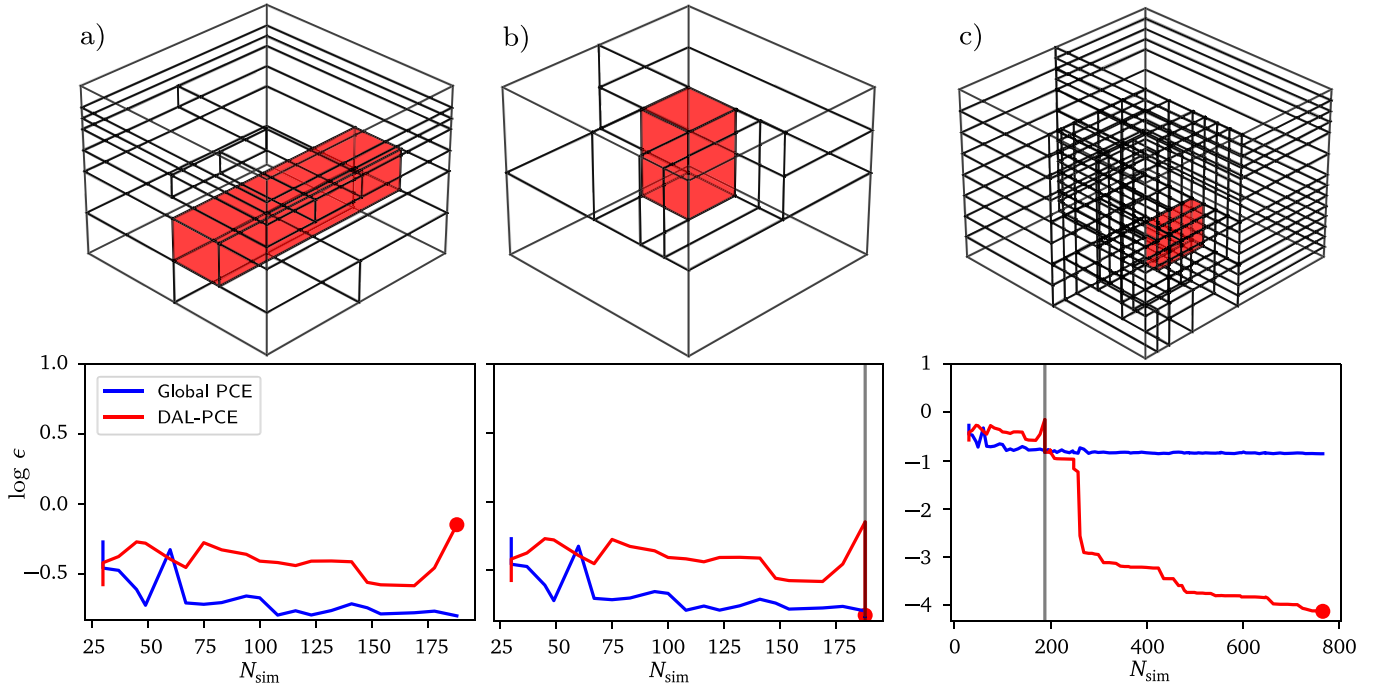


Fig. 3. Illustration of domain decomposition restart. (a) decomposition and error evolution prior to restart, (b) rebuilt decomposition and error drop right after the restart, (c) final decomposition and error showing that the restart unlocks a dramatic decrease in error.

3.8. Stopping criteria

The proposed DAL-PCE algorithm can be fully automated by adding an adequate stopping criterion. A simple but practical stopping criterion is based on computational budget, i.e. once the total number of model evaluations N_{sim} or number of iterations n_{iter} have reached a critical level/budget. One may also use a stopping criterion based on decomposition pattern, e.g. the smallest or the largest volumes of any subdomain, to ensure a desired resolution. Valuable stopping criterion can be also obtained directly from Q^2 , corresponding to a target/threshold level of achieved accuracy. Regardless of the selected stopping criteria, it can easily be applied before *step 5* of the proposed algorithm (start of each iteration).

4. Numerical experiments

The proposed DAL-PCE is presented on four numerical examples of increasing complexity and which illustrated different aspects of the approach. The obtained results are compared (a) to the standard global PCE approach with adaptive maximum order $p \in [5, 25]$ and (b) to SSE [33], as current state-of-the-art non-intrusive surrogate modeling technique based on the domain decomposition. The PCE is constructed using the UQPy package [48] and the original implementation of SSE is used from the UQLab package [49]. To compare methods, the relative mean squared errors ϵ are calculated for all three approximations \tilde{f} on a validation set containing a large pool of 10^6 integration points generated by crude Monte Carlo according to

$$\epsilon(\mathbf{X}) := \frac{\mathbb{E}[(f(\mathbf{X}) - \tilde{f}(\mathbf{X}))^2]}{\mathbb{D}[f(\mathbf{X})]}, \quad (24)$$

where $\mathbb{E}[\cdot]$ and $\mathbb{D}[\cdot]$ are the mean value and variance operators, respectively.

To show representative results of the proposed DAL-PCE algorithm, the calculations were repeated 100 times, and the same settings of the algorithm for all examples were selected as follows: maximum local polynomial degree $p = 2$, number of global candidates $n_{c,g} = 1000 M$, number of local candidates $n_{c,l} = 5P$, the minimum number of samples for local PCE construction $n_{\text{sim}} = 1.5P$, the minimum number of iterations before checking for restart $n_r = 20$, and β are obtained by LARS and OLS algorithm. The minimum number of samples in sub-domains required to justify an expansion for SSE was set identically to DAL-PCE, and polynomial order is adaptively selected in the range $p \in [2, 6]$. Since the SSE is not a sequential approach, the presented results were obtained for ten discrete sample sets of increasing size to compare the convergence of the method. Note that all samples and candidates are generated by LHS for all compared approaches, though it was shown [28] that for the variance-based sequential sampling, it is significantly better to use advanced techniques such as Coherence D-optimal sampling [53].

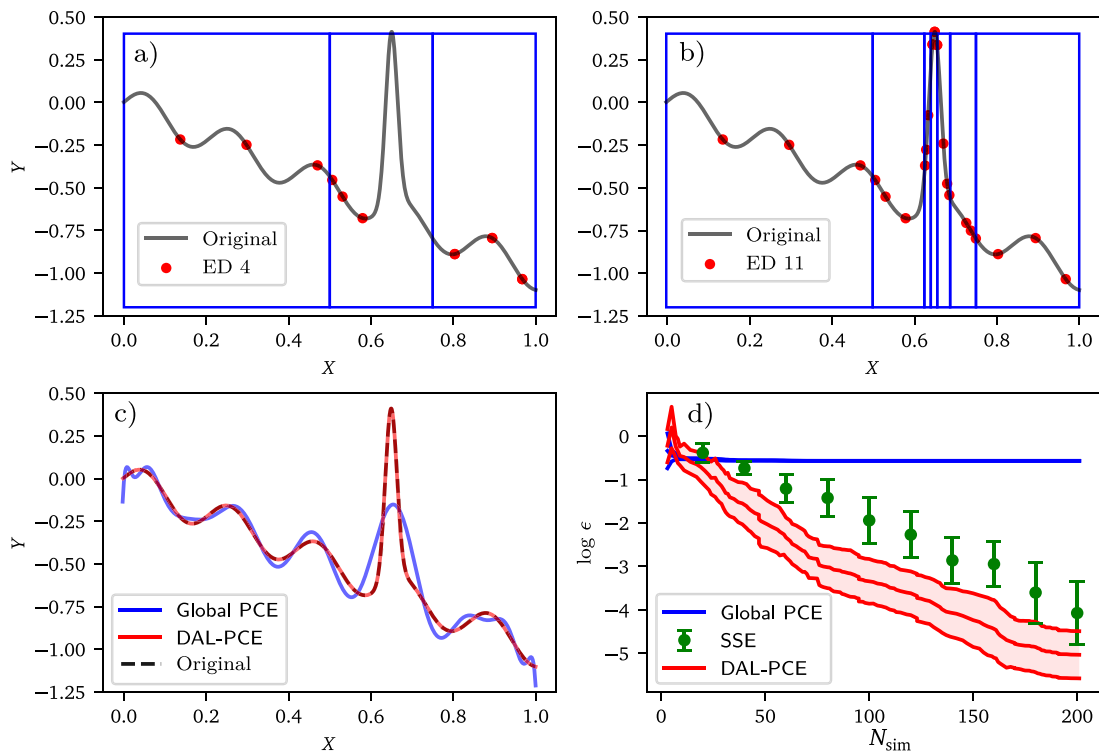


Fig. 4. (a), (b) The adapted domain and ED before (iteration 4) and after (iteration 11) exploration and discovery of the exponential part of the mathematical model. (c) Final surrogate models from global PCE and DAL-PCE. (d) Convergence plot comparing the mean square error for global PCE, SSE, and DAL-PCE. The convergence plots for Global PCE and DAL-PCE show continuous mean value $\pm\sigma$ intervals from 100 repeated trials, while those for SSE are plotted for several discrete ED sizes.

4.1. One-dimensional toy example

The first example involves a simple 1D function [33] that is extremely difficult to approximate with PCE due to the third, highly nonlinear “exp” term

$$f(X) = -X + 0.1 \sin(30X) + \exp(-(50(X - 0.65))^2), \quad X \sim \mathcal{U}[0, 1]. \quad (25)$$

The poor performance of a single global PCE learned from 200 samples is depicted by the blue line in Fig. 4c where it is clear that a single global PCE is not able to accurately approximate the function even for a high number of samples and high maximum polynomial order $p \in [5, 25]$. This function was originally developed to demonstrate the efficiency of SSE based on domain decomposition and thus, it is a natural choice for comparison of the proposed DAL-PCE and SSE.

Fig. 4a–b show a typical realization of the DAL-PCE where the algorithm sequentially decomposes the domain and adds additional samples to the ED. Specifically shown are the 4th and 11th iterations. The boundaries of sub-domains are represented by blue vertical lines and red dots show the positions of samples in the ED. Once the algorithm discovers the highly nonlinear region (the steep peak caused by exp), it progressively refines this region and adds more samples there as a result of the high variance density. Of course, these figures show only one realization of the algorithm, and the decomposition is dependent on the initial ED. Therefore, it is necessary to repeat the algorithm many times with random initial ED to assess convergence.

Fig. 4d shows convergence of the error ϵ from 100 repeated trials. The single global PCE is unable to accurately approximate the original function even when using high p , and thus, ϵ does not converge, as expected. Both methods based on domain decomposition (DAL-PCE and SSE) achieve great accuracy already for 200 samples. However, the DAL-PCE consistently has 1–2 orders of magnitude higher accuracy than SSE for the given number of samples. Moreover, increase in variance of ϵ is, in general, slower in DAL-PCE than in SSE. Fast increment in the variance of SSE can also be seen in the original paper [33]. Finally, we again observe that convergence is continuous with DAL-PCE, where convergence can only be assessed at discrete sample sizes with SSE through a new analysis. All of these advantages of the DAL-PCE can be attributed to active learning, which both explores the space and exploits the behavior of the function to decompose the domain and add samples. Although active learning might lead to lower accuracy (higher ϵ) initially (for small $n_{sim} = 10\text{--}20$) as it is dominated by exploration, it rapidly improves once it identifies important features and begins to favor exploitation.

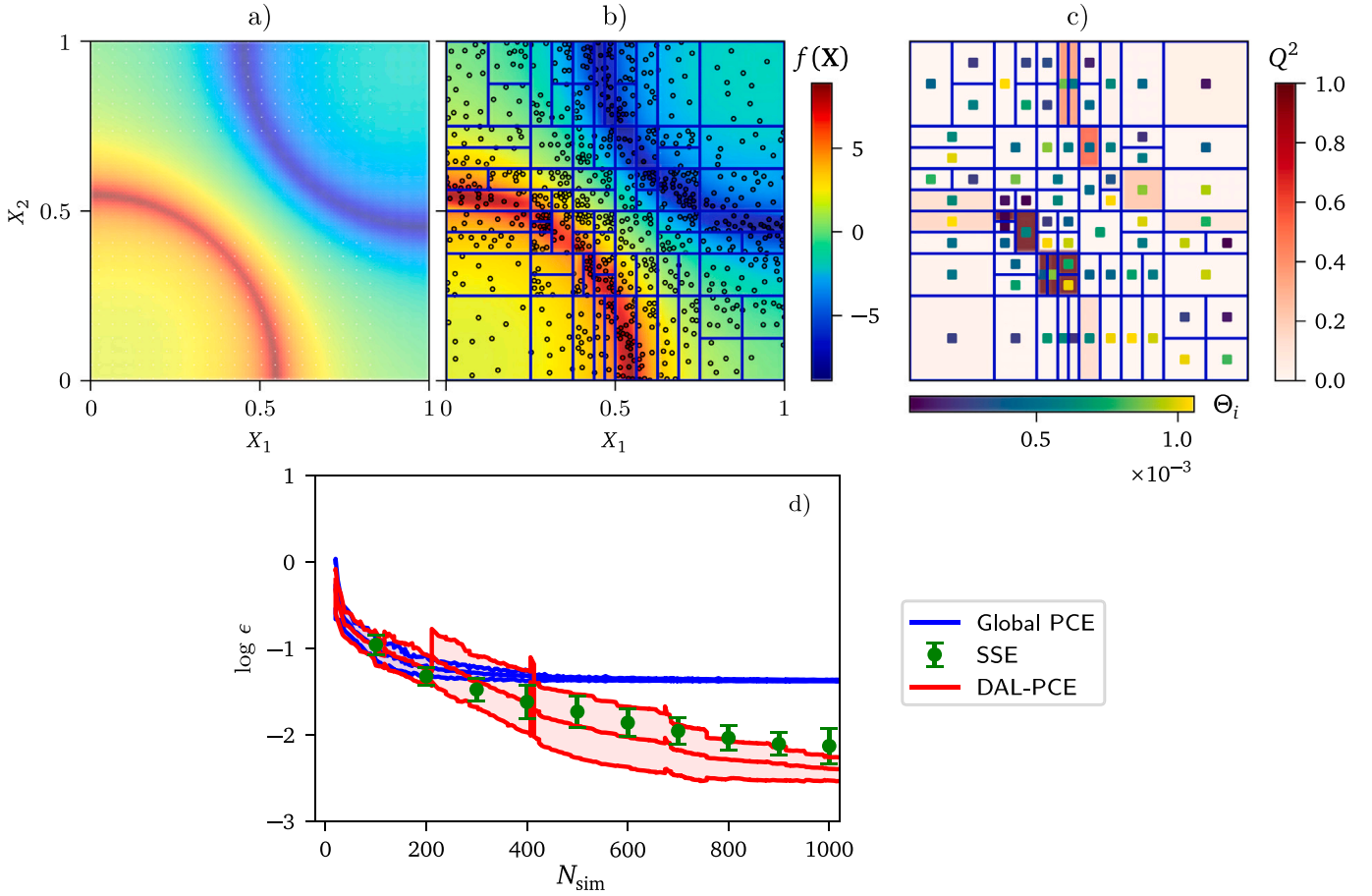


Fig. 5. Results for the 2-dimensional Singularity function: (a) original mathematical model, (b) approximation via DAL-PCE (background color), current domain division and the corresponding ED, (c) local LOO-CV $Q^2_{\mathcal{D}_i}$ and Θ_i value for each sub-domain, (d) convergence plots for DAL-PCE, Global PCE, and SSE showing the mean value and $\pm\sigma$ interval. Convergence plots for SSE show the mean $\pm\sigma$ at discrete sample sizes.

4.2. Two-dimensional singularity

The second example involves a 2D function with mirrored quarter-circle arc line singularities [28]. The form of the function is given by

$$f(\mathbf{X}) = \frac{1}{|0.3 - X_1^2 - X_2^2| + \delta} - \frac{1}{|0.3 - (1 - X_1)^2 - (1 - X_2)^2| + \delta}, \quad \mathbf{X} \sim \mathcal{U}[0, 1]^2, \quad (26)$$

where the strength of the singularities is controlled by the parameter δ , which we set as $\delta = 0.1$. The singularities in this example represent a challenging task for a global PCE even with high order, due to the well-known Gibbs phenomenon [61]. It is thus beneficial to identify the location of the singularity, locally decompose the domain, and construct low-order local PCEs.

Fig. 5 illustrates the decomposition and DAL-PCE approximation at a given stage of the computation. Panel (a) visualizes the true values of the function via a background color. The same coloring scheme is used in panel (b) for the pointwise information available in the current ED (small circles) and for the function approximation via DAL-PCE by the background color. Panels (b) and (c) show also the final domain decomposition. The symmetry in the decomposition documents the great convergence of the DAL-PCE thanks to an adaptive decomposition described in the previous section. Plot (c) shows the local $Q^2_{\mathcal{D}_i}$ error in each individual sub-domain (darker color corresponds to higher local error). These local errors clearly show localization of the prediction error to very small areas near singularities, which are continually being refined. The color of the small solid squares in the center of each sub-domains shows the Θ_i value for that sub-domain.

Finally, the convergence plot in Fig. 5(d) shows that both DAL-PCE and SSE outperform the global PCE, as expected. The SSE performs comparable to or slightly better than DAL-PCE for small N_{sim} , but the DAL-PCE begins to outperform SSE as N_{sim} grows thanks to the active learning approach that targets samples in the vicinity of the singularities. Note that the error converges for both SSE and DAL-PCE as we approach 1000 samples and does not seem to substantially reduce after this. This is due to the fundamental limitation of trying to approximate this singularity, even locally, with low-order polynomials.

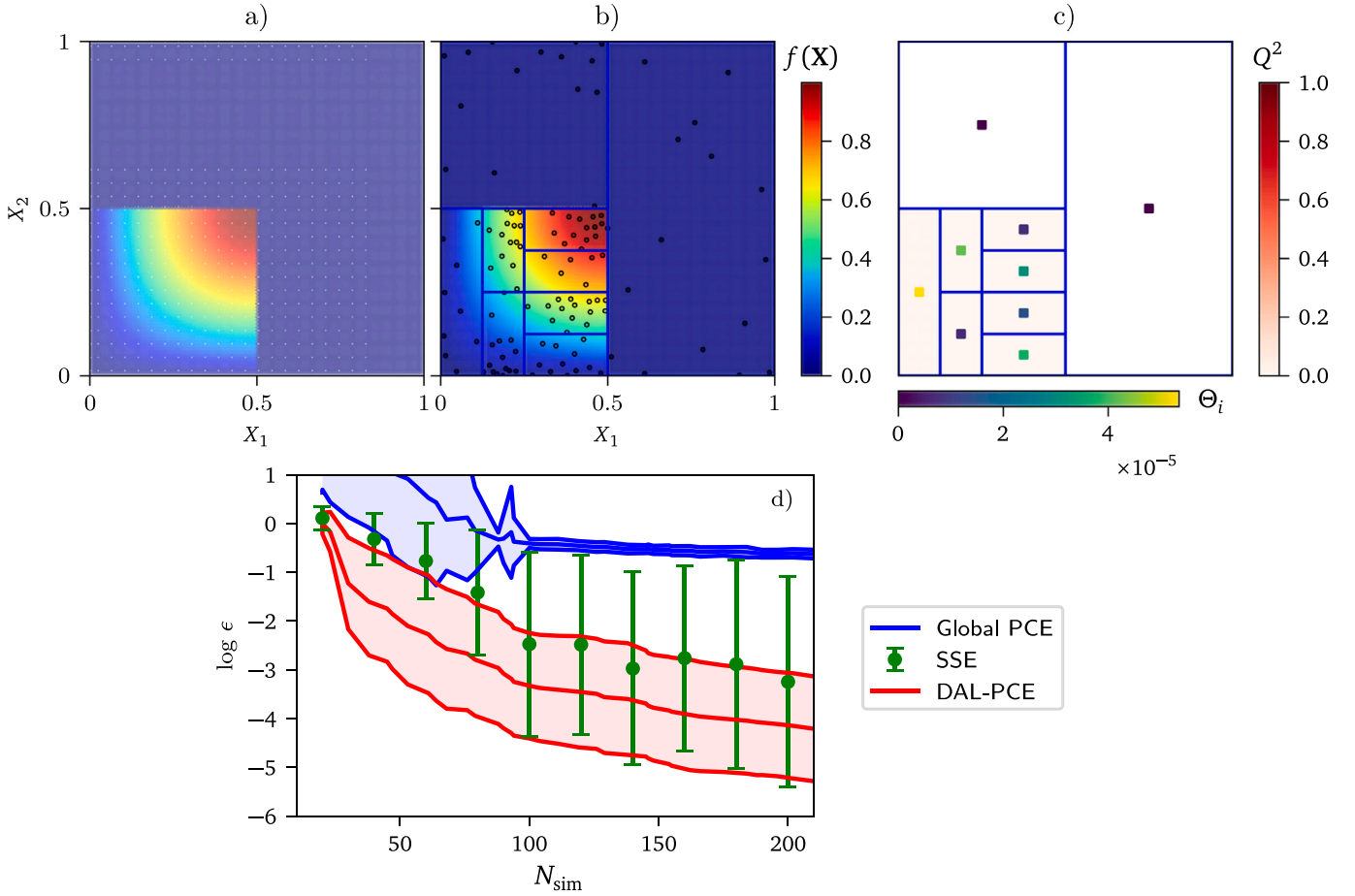


Fig. 6. Results for the 2-dimensional discontinuity function: (a) original mathematical model, (b) approximation via DAL-PCE and ED, (c) local LOO-CV $Q^2_{\Theta_i}$ and Θ_i value for each sub-domain, (d) convergence plots for DAL-PCE, Global PCE, and SSE showing the mean value and $\pm\sigma$ interval. Convergence plots for SSE show the mean $\pm\sigma$ at discrete sample sizes.

4.3. M -dimensional discontinuity

The third example investigates the role of dimensionality on the performance of the proposed DAL-PCE. The following discontinuous function is defined for an arbitrary number of input random variables M [32]

$$f(\mathbf{X}) = \begin{cases} \sin(X_1\pi)\sin(X_2\pi) & \text{if } x_1 \leq 0.5 \text{ and } x_2 \leq 0.5 \\ \sum_{i=3}^M X_i & \text{otherwise,} \end{cases} \quad \mathbf{X} \sim \mathcal{U}[0,1]^M. \quad (27)$$

This function has a discontinuity in the first two input random variables, which can be seen in Fig. 6a. A single global PCE cannot accurately approximate the function because of the discontinuity, although the function $f(\mathbf{X})$ can be easily approximated by two separate PCEs in the two regions for which the definitions differ. But, this requires *a priori* knowledge of the discontinuity location. Since the location of the discontinuity is assumed to be unknown, this function is a good example for domain adaptation using DAL-PCE.

The detailed results for a 2D version of this problem are depicted in Fig. 6 in identical form as in the previous example. Note that the local Q^2_i errors Fig. 6c show perfect accuracy in the part of the input random space where $f(\mathbf{X}) = 0$ and thus the associated sub-domains are not preferred for further decomposition. The convergence plot in Fig. 6d confirms that a single global PCE is not able to create an accurate approximation, and adding more points to ED does not lead to significant improvements in the approximation. The mean values of errors ϵ associated with the proposed DAL-PCE approach are significantly lower in comparison to SSE (1–2 orders of magnitude) – similarly as in the first example, though the convergence trend is similar for both methods. SSE, however, uses a random splitting routine. This can lead to very high variance of results since the accuracy is highly dependent on the pattern of the decomposed input random space. This clearly shows the advantage of an active learning approach.

The influence of dimensionality M on the convergence of the DAL-PCE, SSE, and global PCE is studied in Fig. 7 for (a) 3, (b) 5, (c) 6, and (d) 8 input random variables. As the domain dimension increases, the linear part of the function $f(\mathbf{X})$ occupies an increasing proportion of the domain while the discontinuity remains low-dimensional. The proposed DAL-PCE greatly improves the convergence because it is able to identify an ideal decomposition and local samples to resolve the discontinuity. For low-dimensions ($M = 2, 3$), SSE error ϵ shows a decreasing trend that is better than global PCE but has an extremely high variance.

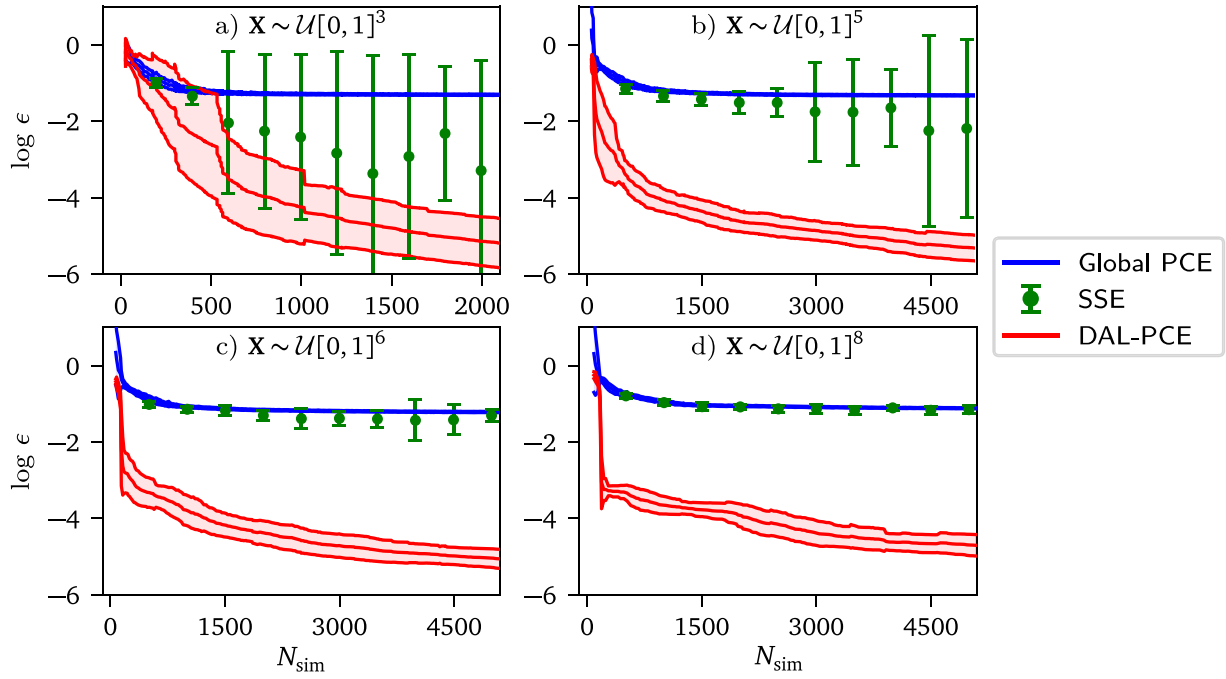


Fig. 7. Convergence plots for the M -dimensional function: (a) 3-dimensional version, (b) 5-dimensional version, (c) 6-dimensional version, and (d) 8-dimensional version. Convergence plots for the DAL-PCE and global PCE show the mean value $\pm\sigma$ interval. Convergence plots for SSE also show the mean $\pm\sigma$, but at discrete sample sizes.

This is caused by a lack of control in sample placement. The domain decomposition in SSE is a product of sample location, and without active learning to guide the sample placement, SSE will sometimes produce very good decomposition and sometimes very poor decomposition. Meanwhile, the proposed DAL-PCE errors have comparably low variance for low-dimensions and consistently have accuracy comparable to, or better than, the best SSE realizations.

As the dimension, M , increases, the DAL-PCE is able to maintain a very high level of accuracy, while the accuracy degrades completely for the SSE such that it is comparable to the global PCE. The DAL-PCE is able to maintain its low error because the discontinuity remains low-dimensional and the active learning process is able to target this region for domain refinement and sampling. This means that the DAL-PCE remains largely independent of the problem dimension and instead depends predominantly on the intrinsic dimension of the discontinuous/nonlinear features of the model. The performance of SSE, on the other hand, degrades with dimension because its domain decomposition depends only on a set of *a priori* specified points that are not selected in a way that is aware of the important features of the model. Consequently, as the dimension increases, the algorithm becomes less likely to refine the domain appropriately around an embedded low-dimensional feature. We remark that this desirable scalable convergence trend of the DAL-PCE is not likely a universal property, as the trend may break down in problems where the intrinsic dimension of the discontinuity/nonlinearity is high or where the discontinuity occupies a very small proportion of the domain — in which case exploration of the space to find the important feature may take a very large number of samples.

In the present example, the discontinuity in the function given in Eq. (27) lies at $x_1 = 0.5$ and $x_2 = 0.5$, which corresponds to the exact location where the domain will be split for both SSE and during the early iterations of the DAL-PCE. One might argue that this presents an unreasonable advantage for the proposed algorithm. We therefore modified the function such that the discontinuity lies at $x_1 = 0.61$ and $x_2 = 0.61$. Fig. 8 shows the convergence for the DAL-PCE and SSE for this modified function with varying dimensions, M . The absolute errors ϵ exhibit a slower decrease, especially for dimensions $M = 3$ and $M = 5$. However, the proposed active learning still leads to superior results (especially for higher dimensions, as in the previous case). Note that there are visible spikes in the DAL-PCE convergence graph for the 3-dimensional example. Although the results were statistically processed, these spikes are caused by the restart adaptivity occurring at the same N_{sim} in each replication. In this case, the optimal decomposition pattern is very complicated, and therefore, the algorithm activates the restart adaptivity frequently (after multiples of n_r steps) until it finds a suitable pattern to continue convergence. SSE in the 3- and 5-dimensional cases has a higher mean error and significantly lower variance in comparison to the previous example. This is caused by the fact that the modified discontinuity location no longer lies along the boundary of the domain decomposition. In the previous example, some SSE realizations achieved near-perfect accuracy because the domain was coincidentally divided along the discontinuity.

This phenomenon is investigated more closely in Fig. 9, which compares the number of outliers in both versions of 3D examples. In addition to the mean $\pm\sigma$ seen previously, the figure also shows standard boxplots for SSE (median along with lower and upper quartiles) and the corresponding number of “extreme” realizations producing very high accuracy (top axis) for (a) the original position of discontinuity; and (b) discontinuity at $x_1 = 0.61$ and $x_2 = 0.61$. As can be seen, in panel (a) there are many outliers producing $\epsilon < -7$, which effectively decreases μ relative to the median while also significantly increasing the variance. In contrast,

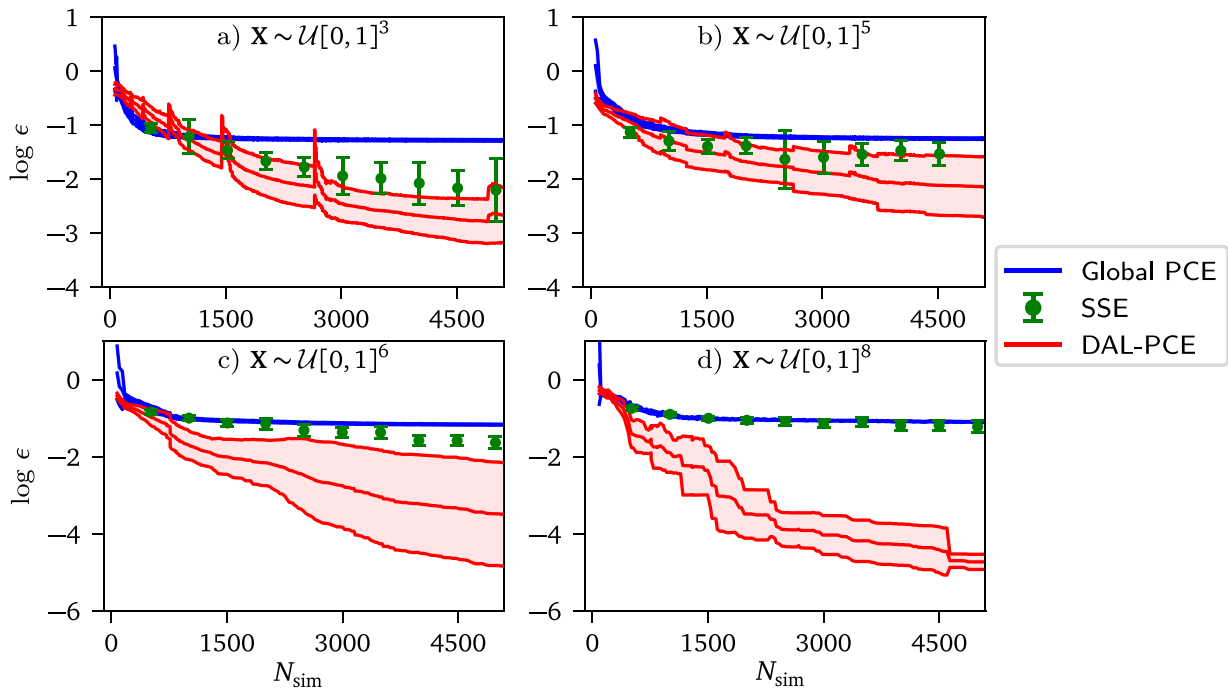


Fig. 8. Convergence plots for the modified M -dimensional function: (a) 3-dimensional version, (b) 5-dimensional version, (c) 6-dimensional version, and (d) 8-dimensional version. Convergence plots for the DAL-PCE and global PCE show the mean value $\pm\sigma$ interval. Convergence plots for SSE also show the mean $\pm\sigma$, but at discrete sample sizes.

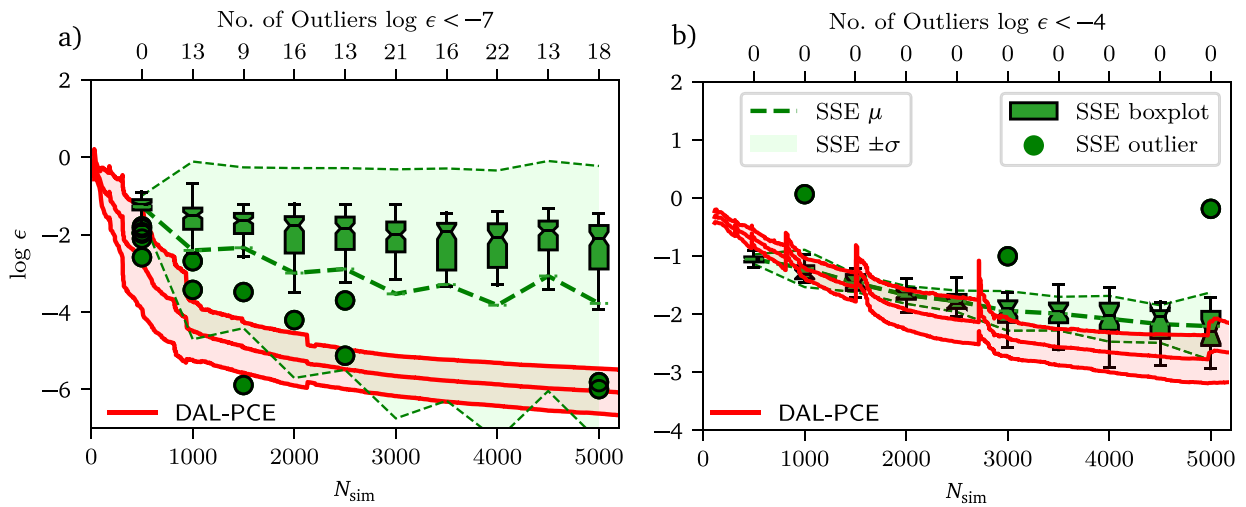


Fig. 9. Convergence plots for DAL-PCE and SSE with additional boxplots for SSE showing the median, lower and upper quartiles and outliers for: (a) the 3D example with a discontinuity at $x_1 = 0.5$ and $x_2 = 0.5$, (b) the 3D example with a discontinuity at $x_1 = 0.61$ and $x_2 = 0.61$.

DAL-PCE has no outliers, and it leads to very consistent results. In panel (b), there are no outliers for either SSE or DAL-PCE, and the results are thus consistent with low variance for both methods.

4.4. Dynamical response of asymmetric shallow von Mises truss

In this section, we demonstrate the relevance of the proposed method for a representative engineering example exhibiting discontinuous response. Consider the shallow two-bar planar truss subjected to a vertical load at its top joint, as presented in [15] and illustrated in Fig. 10a. The target is to approximate the outcome of a detailed dynamical analysis. The approximated quantity of interest is the displacement of the loading point. These types of engineering systems are prone to snap-through instability. The von Mises truss pictured in Fig. 10a is an example of bistable shallow structures, which has been used for many years as a benchmark in the numerical analysis of nonlinear structures, including the dynamic buckling of structures.

The truss is formed by two prismatic bars made of a hard wood (density 800 kg/m^3 , modulus of elasticity $E = 12 \text{ GPa}$). There are two variables in the studied von Mises truss: (i) the loading vertical force F , and (ii) a half sine-wave imperfection of the left

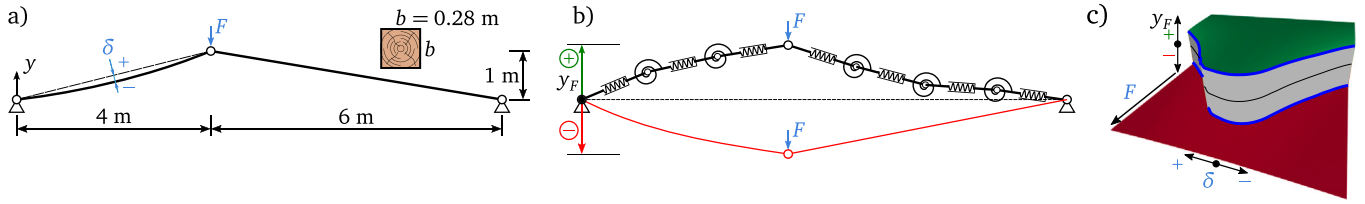


Fig. 10. Asymmetric shallow von Mises truss. (a) Initial geometry with two random variables F and δ ; (b) illustrative sketch of the discrete dynamical model and the meaning of output variable y_f (the course of dynamical response obtained by the FyDiK code is shown in [FyDiK animation](#), see the supplementary material); (c) illustration of the discontinuous response function of the two input variables.

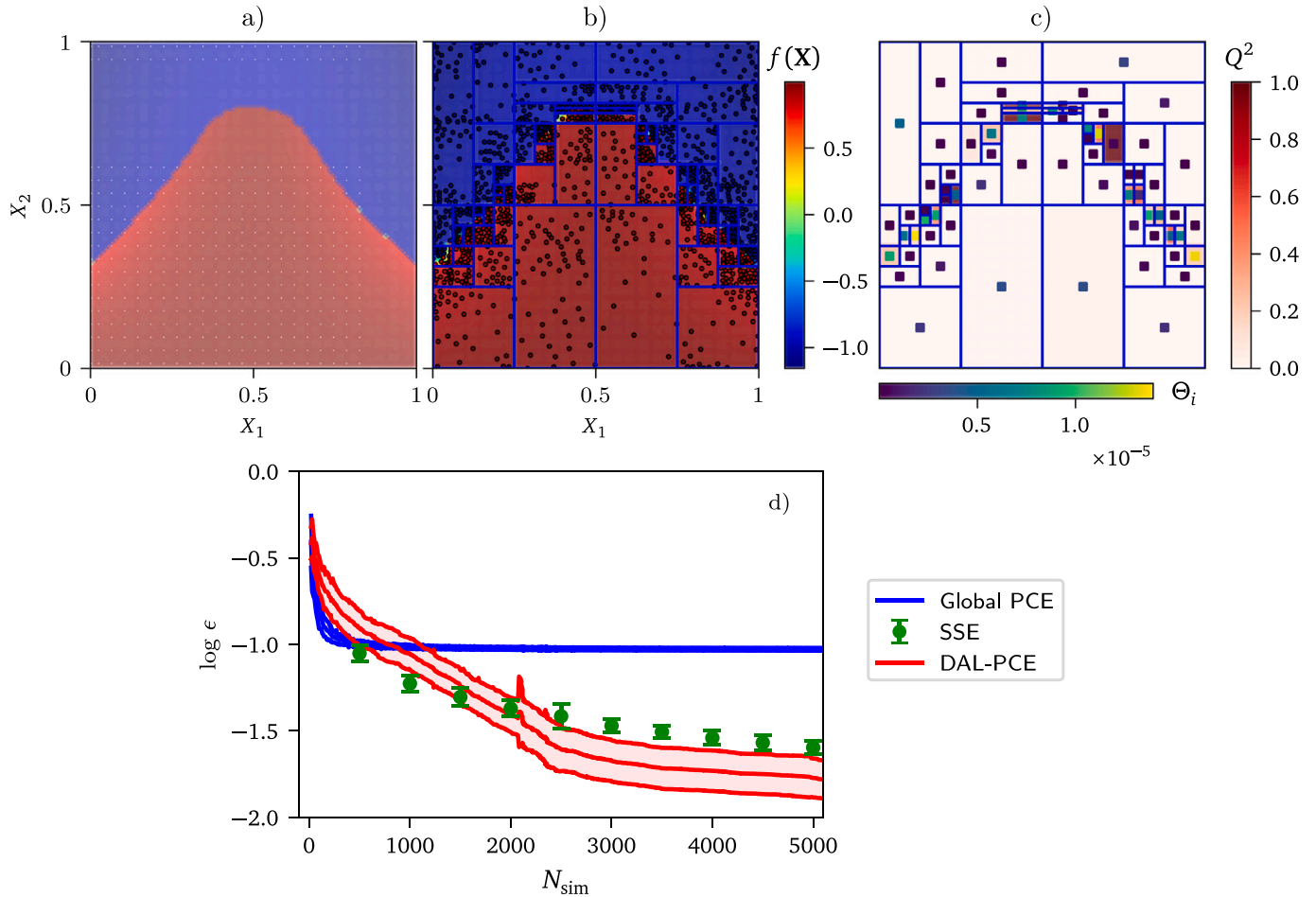


Fig. 11. Results for the von Mises truss example: (a) original mathematical model (numerical solution), (b) approximation via DAL-PCE and ED, (c) local LOO-CV Q^2 and Θ_i value for each sub-domain, (d) convergence plots for DAL-PCE, Global PCE, and SSE showing the mean value and $\pm\sigma$ interval; convergence plots for SSE show the mean $\pm\sigma$ at discrete sample sizes.

bar having magnitude δ , see the sketch in Fig. 10a. The load is applied dynamically as a step function at time zero for an unlimited duration. The structure is modeled, as illustrated in Fig. 10b. In particular, the mass of the bar is concentrated in 21 mass points, including the supports and the loading point. These mass points are connected via 10 + 10 translational springs representing the normal stiffness of the true bars. The pairs of the axial members are connected via a rotational spring having zero moment for a zero angle between adjacent bars. The only exceptions are the loading and support points where there are no rotational springs attached (hinges). The damping is associated with the mass points via linear viscous damping coefficient set to 11 N s/(kg m) approximating the relative damping of about 3%. Explicit dynamics solver FyDiK [62,63] was used to solve the equations of equilibrium at the mass points; see the attached [FyDiK animation](#) (see the supplementary material). The animation compares two solutions with a small difference in the loading force magnitude: the top truss keeps its upright configuration while the bottom truss is loaded by a larger force leading to snap-through. The changing colors of individual segments of the bars represent the pulsating normal forces. The numerical solution lasts up to two seconds, which is the time needed for almost complete stabilization of the solution (kinetic energy drops below a negligible threshold).

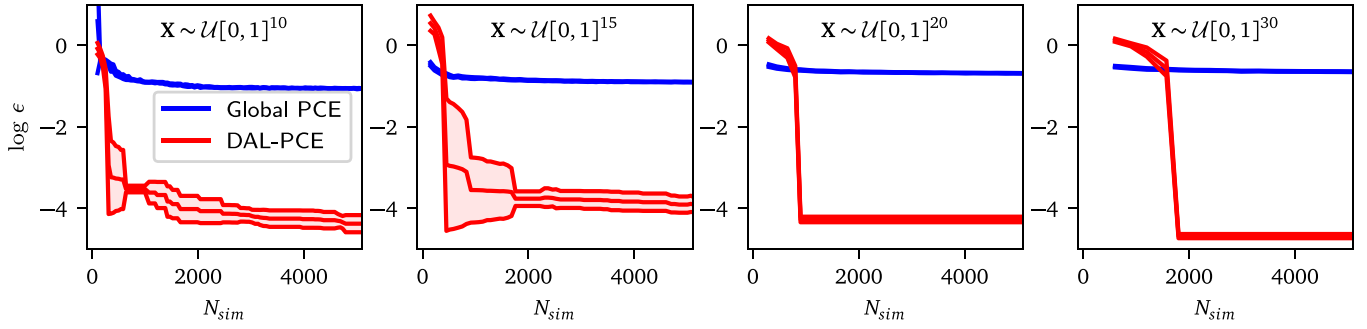


Fig. 12. Convergence plots of DAL-PCE in high-dimensional example: M dimensional function with localized low-dimensional discontinuity. The input random space contains $M = 10, 15, 20, 30$ random variables (from left to right).

Since the structure is very shallow, sudden application of the vertical force can cause snap-through buckling, wherein the loading point drops down between the supports and the members switch from a state of compression to tensile stresses in the final stable state. We specifically study the horizontal coordinate y_F of the loading point after the dynamic response stabilizes to the final deformed shape. The force $F \in (31.6, 772.6)$ kN and initial imperfection $\delta \in (-0.4, 0.4)$ m are treated as uniform random variables mapped to the unit square such that the model input $\mathbf{X} \sim \mathcal{U}[0, 1]^2$. Because of the potential snap-through buckling, the solution is discontinuous as illustrated in Fig. 10c. On each side of the discontinuity, the solution y_F is smooth and slowly varying, having values near +1 m and -1 m, respectively. Note that the output is *not symmetric* with respect to $\delta = 0$ because the dynamical response evolves differently for concave and convex initial displacements.

The sharp boundary between the buckled and unbuckled regions, shown in Fig. 11a cause global PCE to produce poor approximations that are vulnerable to the Gibbs phenomenon, similar to the example in . This is shown by the convergence plots in Fig. 11d comparing global PCE, DAL-PCE, and SSE. Clearly, the complexity of this example and the complicated shape of the discontinuity limits the accuracy of all the surrogate models. The proposed DAL-PCE achieves low accuracy for small sample sizes because the correspondingly small number of sub-domains and low-order PCEs are unable to sufficiently approximate the boundary. Therefore, the global PCE and SSE (with a low number of embedding levels) are initially better. With an increasing number of samples, the proposed DAL-PCE approach leads to superior results because the active learning is able to resolve the discontinuity as illustrated in Fig. 11b, which shows the domain decomposition and approximation after 2000 samples. Fig. 11c shows the corresponding LOO-CV errors for each subdomain, demonstrating the errors are confined to small, localized regions near the boundary.

5. Discussion & future work

The proposed DAL-PCE approach is a general methodology for the decomposition of the input random space and construction of localized PCEs using active learning. The proposed active learning is based on a novel Θ criterion that optimally balances global *exploration* with local *exploitation* of the model. However, an important topic of further research is to study the behavior of the proposed criterion in higher dimensions. In order to show the computational possibilities of the proposed algorithm for higher dimensions, the study of the third example with growing M can be seen in Fig. 12. As can be seen, the trend of the convergence is identical to in the previously presented lower-dimensional cases. The numerical construction of DAL-PCE for growing M is not significantly more costly because just a single sparse PCE is constructed in each iteration (i.e., there is only a computational cost of an ordinary least square regression). Therefore, it is possible to use DAL-PCE for arbitrary dimensions M under the assumption that it is possible to locally construct standard sparse PCE. The computational cost of the domain decomposition is negligible, and so is also the computational cost of evaluating the Θ criterion for all candidates. However, we would like to emphasize that this example has low dimensional discontinuity; thus, the DAL-PCE is extremely efficient in this case. However, the efficiency could be significantly lower for examples with high-dimensional discontinuities, where sparse solvers do not bring any benefits. The proposed algorithm can be easily applied to examples in low to medium-sized space dimensions. However, there are still some unresolved theoretical questions when it comes to high-dimensional space. In particular, the geometrical terms $I_{c,s}^M$ and \mathcal{W}_i likely cause poor convergence in high dimensions. Although some preliminary results focused on investigating the term $I_{c,s}^M$ in high dimensions was previously performed in the paper [28] proposing the original Θ criterion, it is still necessary to perform an extensive study of its behavior as well as investigating the influence of \mathcal{W}_i , which may need to be reformulated for high dimensions.

The roles of the individual components of the Θ criterion are investigated in Fig. 13, which reports their evolution during the convergence of the first 1D example. The exploration part (orange line) is composed of the subdomain weights (blue) and the powered distances between the global candidates and the existing ED samples (green). The exploitation part (red) accounts for the local variance density. The left column of the figure shows statistical results based on 50 repetitions (averages \pm one standard deviation). Based on the average trends of the red and orange lines, we can identify three stages with typical proportions between exploration and exploitation. In the first stage lasting about 20 samples, the sampling is slightly dominated by exploration. This corresponds to the situation in which only little is known about the approximated function. In the second stage (roughly between 20–50 samples), the exploitation term does not tend to decrease, which suggests the dominance of the exploitation aspect in the

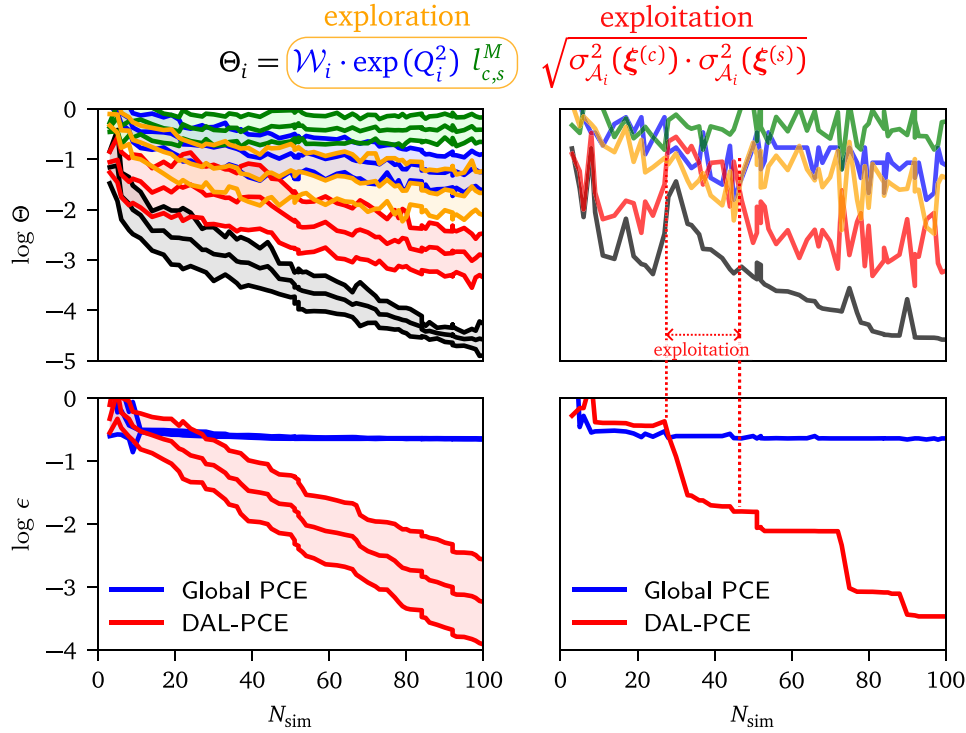


Fig. 13. Components of θ criterion in the 1D example (first row): statistical results based on 50 repetitions (left); a selected single realization (right). The convergence of the achieved accuracy can be seen in the second row similarly as in the previous examples.

candidate selection and corresponds to sampling near the explored regions with significant non-linearities. In the third stage (from about 50 samples), the algorithm promotes exploration in the search for new important regions as expected.

The behavior of the algorithm is better visible when focusing on just one realization (run). The right-hand side column of Fig. 13 shows one selected realization of the algorithm. The figure clearly identifies stages dominated by exploitation: the convergence plot shows peaks in the composite criterion caused by the sudden increases in the exploitation term (red). In these stages, the approximation errors rapidly decrease thanks to significant adjustments of the local PCEs, which are based on exploitative samples.

Besides already presented comparisons of DAL-PCE to SSE based on similar concepts (PCE combined with domain decomposition), the following Fig. 14 shows a comparison to adaptive Kriging. Kriging is typically preferred for an approximation of highly non-linear functions, and it shares some of the possibilities of the analytical post-processing for sensitivity and statistical analysis [64]. Therefore it is, besides PCE and SSE, a natural choice for comparison with the proposed DAL-PCE. Specifically, we use adaptive Kriging based on expected improvement for global fit learning function [65] implemented in UQPpy [48]. Kriging is constructed with linear trend, Matérn kernel, correlation lengths equal to $l_{GP} = 0.1$ and the process variance $\sigma_{GP} = 0.1$ in both examples. The learning function governing adaptivity is in the following form

$$\mathcal{L}(\mathbf{x}^{(c)}) = [\tilde{f}(\mathbf{x}^{(c)}) - f(\mathbf{x}^{(s)})]^2 + \sigma_{\tilde{f}}(\mathbf{x}^{(c)})^2, \quad (28)$$

where $\mathbf{x}^{(c)}$ is a candidate, $\mathbf{x}^{(s)}$ is a closest sample in the existing ED and $\sigma_{\tilde{f}}(\mathbf{x}^{(c)})^2$ is a Kriging variance in the location of candidate. Candidates were generated by LHS, similarly as in DAL-PCE. The numbers of initial samples and candidates are identical as for DAL-PCE. Although there are many learning functions for reliability analysis based on the Kriging prediction, these are not useful for the global approximation addressed in the present paper. However, as can be seen in Eq. (28), the selected learning function has a very similar rationale as the θ criterion: it aims to construct a space-filling design via the second term, but it prefers regions associated with high local non-linearities; therefore, it can be directly compared to DAL-PCE.

Adaptive Kriging was compared to DAL-PCE in two selected examples: (a) the first 1D example and (b) the third example with discontinuity in 2D. On the one hand, Kriging is significantly more efficient in the 1D highly non-linear example, as expected. The reason is that the approximated function was specifically proposed as a very challenging task for PCE due to the exponential term, which polynomials cannot easily approximate. However, the exponential term is not very challenging for Kriging or other surrogate models. On the other hand, Kriging clearly fails in example (b) containing discontinuity because continuous surrogate models cannot approximate such functions without domain decomposition as in DAL-PCE or SSE. Moreover, adaptive Kriging (light green in Fig. 14b) is even worse than standard Kriging without adaptive sampling (solid green) in this example because the discontinuity leads to ED consisting of black dots (initial sample by LHS) adaptively extended by an enormous number of samples very close to each other (green dots in the vicinity of function discontinuity), i.e., the first term (exploitation part) of the additive learning function is almost always dominating in Eq. (28).

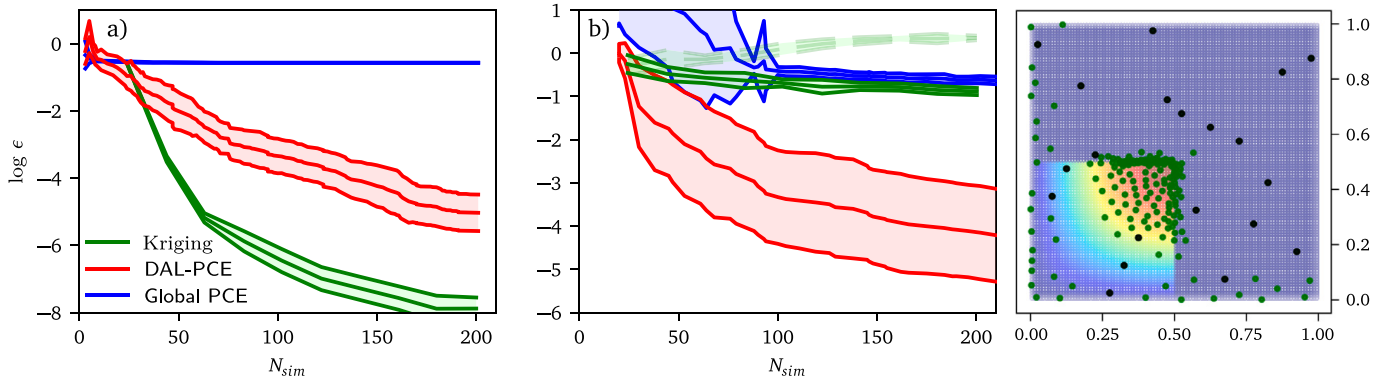


Fig. 14. Comparison of DAL-PCE and adaptive Kriging in two examples: (a) convergence plot of 1D non-linear example, (b) convergence plot of 2D example with discontinuity and final adapted ED.

From the obtained numerical results supporting the theoretical characteristics of both methods, it can be stated that the proposed DAL-PCE offer a general extension of standard global PCE, allowing an approximation of highly non-linear functions. However, it does not necessarily lead to superior performance in comparison to Kriging, which is an extremely powerful interpolator, especially for reliability analysis, though the PCE is often preferred as a global surrogate for its extensive analytical post-processing for UQ [42].

Although this paper presents one specific learning algorithm, the methodology is general and amenable to modifications to reflect the specific user's needs. The whole process can be divided into two tasks: (A) decomposition of the input random space and (B) construction of localized PCEs; and both can be easily modified as discussed further:

- (A) The most important sub-domain \mathcal{D}_i is identified by extended θ according to Eq. (17) evaluated for a large number of global candidates. In this paper, we use standard LHS for candidate generation, but it may be beneficial to use different sampling methods that produce more uniform coverage of the whole input random space (see e.g. [57,66,67]). Although generating a large number of candidates is generally possible, it might be challenging to uniformly cover the entire input random space, especially in high dimensions. Thus, one can use any sampling technique suitable for a specific example, e.g., [68]. Once the \mathcal{D}_i is identified via Eq. (17), it is either divided (providing it contains enough ED points) or the sample is extended inside it, to achieve a better PCE approximation. The simplest division occurs by splitting the volume into two parts of identical hypervolume in the direction of the highest first-order Sobol' index. However, the algorithm can accommodate various different approaches. For example, it is possible to divide the \mathcal{D}_i into a higher number of sub-domains, not just two. Moreover, other criteria can be used instead of splitting the domain into parts of equal hypervolume. For example, the cutting plane can be positioned so to split the domain variance into equal parts.
- (B) The user can choose to employ any existing method to construct the non-intrusive PCEs, including various sparse solvers or adaptive algorithms, which may be preferable for certain applications [10]. For example, we use LARS with OLS. However, it is generally more efficient to use active learning based on the θ criterion for PCE as shown in [28], which employs variance-based sequential sampling. This improvement can be integrated within the DAL-PCE to make local PCE more efficient in each subdomain, thereby improving the overall convergence. This can be compounded by the use of advanced sampling techniques within the subdomains, such as Coherence D-optimal sampling [52,53].

As seen from the previous paragraphs, the whole algorithm can be adapted for specific needs reflecting the characteristics of a given mathematical model, such as dimensionality, sparsity, non-linearity etc., by simply exchanging components of the proposed algorithm for suitable existing (or new) techniques. Note that even after the modification, the whole methodology based on θ criterion is still valid and can be used for uncertainty quantification and surrogate modeling as described in this paper. Moreover, in comparison to SSE, the DAL-PCE sequentially adds points and divides the sub-domains one-by-one based on information obtained from the previous iteration. Another significant advantage of the DAL-PCE is that it provides estimates of the local errors, $Q_{\mathcal{D}_i}$, associated with each sub-domain. Since localized PCEs are constructed independently, local errors estimate the local accuracy/uncertainty of the surrogate model directly and can be assembled to provide global error measures. Naturally, local accuracy is very important information that can be used for further probabilistic analysis and active learning. Although this paper does not propose any specific approach for further processing of this information, it could serve as a main ingredient for various active learning algorithms. For example, it could be directly used to quantify a prediction uncertainty in industrial applications and possibly extend the ED in a sub-domain of interest.

6. Conclusion

The paper presented a novel approach, domain adaptively localized PCE, for the adaptive sequential construction of localized PCEs based on active learning and decomposition of the input random space. It combines adaptive sequential sampling based on the recently proposed θ criterion to maintain the balance between exploration of the input random space and exploitation of the current

characteristics of the PCE together with the adaptive sequential decomposition of the input random space creating sub-domains approximated by local surrogate models. The methodology offers a general technique that can be easily adapted or modified for specific functions extending its applicability. The performance of the proposed methodology was validated on several numerical examples of increasing complexity investigating different aspects of the algorithm and leading to superior results in comparison to a single global PCE and the recently proposed SSE.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The first author acknowledges the financial support provided by the Czech Science Foundation under project number 22-00774S. Additionally, the major part of this research was conducted during the research stay of the first author at Johns Hopkins University, supported by the project International Mobility of Researchers of Brno University of Technology, Czechia, under project No. EF18_053/0016962.

Appendix A. Supplementary data

The supplementary material contains an animation of the studied von Mises truss, depicting its dynamical simulation in both stable and unstable configurations.

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.ymsp.2023.110728>.

References

- [1] N. Wiener, The homogeneous chaos, *Amer. J. Math.* 60 (4) (1938) 897–936, <http://dx.doi.org/10.2307/2371268>.
- [2] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, *J. Comput. Phys.* 230 (6) (2011) 2345–2367, <http://dx.doi.org/10.1016/j.jcp.2010.12.021>.
- [3] R.G. Ghanem, P.D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer, New York, 1991, <http://dx.doi.org/10.1007/978-1-4612-3094-6>.
- [4] N.-Z. Chen, C. Guedes Soares, Spectral stochastic finite element analysis for laminated composite plates, *Comput. Methods Appl. Mech. Engrg.* 197 (51) (2008) 4830–4839, <http://dx.doi.org/10.1016/j.cma.2008.07.003>.
- [5] L. Novák, D. Novák, Surrogate modelling in the stochastic analysis of concrete girders failing in shear, in: *Proc. of the Fib Symposium 2019: Concrete - Innovations in Materials, Design and Structures*, 2019, pp. 1741–1747.
- [6] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, *Reliab. Eng. Syst. Saf.* 93 (7) (2008) 964–979, <http://dx.doi.org/10.1016/j.res.2007.04.002>.
- [7] T. Crestaux, O.L. Maître, J.-M. Martinez, Polynomial chaos expansion for sensitivity analysis, *Reliab. Eng. Syst. Saf.* 94 (7) (2009) 1161–1172, <http://dx.doi.org/10.1016/j.res.2008.10.008>.
- [8] A. Cohen, G. Migliorati, Optimal weighted least-squares methods, *SMAI J. Comput. Math.* 3 (2017) 181–203, <http://dx.doi.org/10.5802/smai-jcm.24>.
- [9] A.C. Narayan, J. Jakeman, T. Zhou, A Christoffel function weighted least squares algorithm for collocation approximations, *Math. Comp.* 86 (2017) 1913–1947, <http://dx.doi.org/10.1090/mcom/3192>.
- [10] N. Lüthen, S. Marelli, B. Sudret, Sparse polynomial chaos expansions: Literature survey and benchmark, *SIAM/ASA J. Uncertain. Quantif.* 9 (2) (2021) 593–649, <http://dx.doi.org/10.1137/20M1315774>.
- [11] B. Echard, N. Gayton, M. Lemaire, AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation, *Struct. Saf.* 33 (2) (2011) 145–154, <http://dx.doi.org/10.1016/j.strusafe.2011.01.002>.
- [12] L. Shi, B. Sun, D.S. Ibrahim, An active learning reliability method with multiple kernel functions based on radial basis function, *Struct. Multidiscip. Optim.* 60 (1) (2019) 211–229, <http://dx.doi.org/10.1007/s00158-019-02210-0>.
- [13] X. Yang, X. Cheng, Active learning method combining Kriging model and multimodal-optimization-based importance sampling for the estimation of small failure probability, *Internat. J. Numer. Methods Engrg.* 121 (21) (2020) 4843–4864, <http://dx.doi.org/10.1002/nme.6495>.
- [14] M. Li, Z. Wang, Deep learning for high-dimensional reliability analysis, *Mech. Syst. Signal Process.* 139 (2020) 106399, <http://dx.doi.org/10.1016/j.ymsp.2019.106399>.
- [15] M. Vořechovský, Reliability analysis of discrete-state performance functions via adaptive sequential sampling with detection of failure surfaces, *Comput. Methods Appl. Mech. Engrg.* 401 (2022) 115606, <http://dx.doi.org/10.1016/j.cma.2022.115606>.
- [16] A. Gerasimov, M. Vořechovský, Failure probability estimation and detection of failure surfaces via adaptive sequential decomposition of the design domain, *Struct. Saf.* 104 (2023) 102364, <http://dx.doi.org/10.1016/j.strusafe.2023.102364>.
- [17] B. Gaspar, A. Teixeira, C. Guedes Soares, Adaptive surrogate model with active refinement combining Kriging and a trust region method, *Reliab. Eng. Syst. Saf.* 165 (2017) 277–291, <http://dx.doi.org/10.1016/j.res.2017.03.035>.
- [18] F. Cadini, F. Santos, E. Zio, An improved adaptive kriging-based importance technique for sampling multiple failure regions of low probability, *Reliab. Eng. Syst. Saf.* 131 (2014) 109–117, <http://dx.doi.org/10.1016/j.res.2014.06.023>.
- [19] M. Li, Z. Wang, Reliability-based multifidelity optimization using adaptive hybrid learning, *ASCE-ASME J. Risk Uncertain. Eng. Syst. B* 6 (2) (2020) 021005, <http://dx.doi.org/10.1115/1.4044773>.
- [20] M. Li, Z. Wang, Active resource allocation for reliability analysis with model bias correction, *J. Mech. Des.* 141 (5) (2019) 051403, <http://dx.doi.org/10.1115/1.4042344>.

- [21] S. Marelli, B. Sudret, An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis, *Struct. Saf.* 75 (2018) 67–74, <http://dx.doi.org/10.1016/j.strusafe.2018.06.003>.
- [22] Y. Zhou, Z. Lu, W. Yun, Active sparse polynomial chaos expansion for system reliability analysis, *Reliab. Eng. Syst. Saf.* 202 (2020) 107025, <http://dx.doi.org/10.1016/j.ress.2020.107025>.
- [23] K. Cheng, Z. Lu, Active learning polynomial chaos expansion for reliability analysis by maximizing expected indicator function prediction error, *Internat. J. Numer. Methods Engrg.* 121 (14) (2020) 3159–3177, <http://dx.doi.org/10.1002/nme.6351>.
- [24] N. Fajraoui, S. Marelli, B. Sudret, Sequential design of experiment for sparse polynomial chaos expansions, *SIAM/ASA J. Uncertain. Quantif.* 5 (1) (2017) 1061–1085, <http://dx.doi.org/10.1137/16m1103488>.
- [25] M. Thapa, S.B. Mulani, R.W. Walters, Adaptive weighted least-squares polynomial chaos expansion with basis adaptivity and sequential adaptive sampling, *Comput. Methods Appl. Mech. Engrg.* 360 (2020) 112759, <http://dx.doi.org/10.1016/j.cma.2019.112759>.
- [26] M.D. Shields, Adaptive Monte Carlo analysis for strongly nonlinear stochastic systems, *Reliab. Eng. Syst. Saf.* 175 (2018) 207–224, <http://dx.doi.org/10.1016/j.ress.2018.03.018>.
- [27] Y. Zhou, Z. Lu, K. Cheng, C. Ling, An efficient and robust adaptive sampling method for polynomial chaos expansion in sparse Bayesian learning framework, *Comput. Methods Appl. Mech. Engrg.* 352 (2019) 654–674, <http://dx.doi.org/10.1016/j.cma.2019.04.046>.
- [28] L. Novák, M. Vořechovský, V. Sadílek, M.D. Shields, Variance-based adaptive sequential sampling for Polynomial Chaos Expansion, *Comput. Methods Appl. Mech. Engrg.* 386 (2021) 114105, <http://dx.doi.org/10.1016/j.cma.2021.114105>.
- [29] J. Zhang, W. Gong, X. Yue, M. Shi, L. Chen, Efficient reliability analysis using prediction-oriented active sparse polynomial chaos expansion, *Reliab. Eng. Syst. Saf.* 228 (2022) 108749, <http://dx.doi.org/10.1016/j.ress.2022.108749>.
- [30] M.K. Deb, I.M. Babuška, J. Oden, Solution of stochastic partial differential equations using Galerkin finite element techniques, *Comput. Methods Appl. Mech. Engrg.* 190 (48) (2001) 6359–6372, [http://dx.doi.org/10.1016/S0045-7825\(01\)00237-7](http://dx.doi.org/10.1016/S0045-7825(01)00237-7).
- [31] J.A. Witteveen, G. Iaccarino, Simplex stochastic collocation with random sampling and extrapolation for nonhypercube probability spaces, *SIAM J. Sci. Comput.* 34 (2) (2012) A814–A838.
- [32] A. Bhaduri, Y. He, M.D. Shields, L. Graham-Brady, R.M. Kirby, Stochastic collocation approach with adaptive mesh refinement for parametric uncertainty analysis, *J. Comput. Phys.* 371 (2018) 732–750, <http://dx.doi.org/10.1016/j.jcp.2018.06.003>.
- [33] S. Marelli, P.-R. Wagner, C. Lataniotis, B. Sudret, Stochastic spectral embedding, *Int. J. Uncertain. Quantif.* 11 (2) (2021) 25–47, <http://dx.doi.org/10.1615/int.j.uncertaintyquantification.2020034395>.
- [34] P.-R. Wagner, S. Marelli, I. Papaioannou, D. Straub, B. Sudret, Rare event estimation using stochastic spectral embedding, *Struct. Saf.* 96 (2022) 102179, <http://dx.doi.org/10.1016/j.strusafe.2021.102179>.
- [35] X. Wan, G.E. Karniadakis, An adaptive multi-element generalized polynomial chaos method for stochastic differential equations, *J. Comput. Phys.* 209 (2) (2005) 617–642, <http://dx.doi.org/10.1016/j.jcp.2005.03.023>.
- [36] J. Foo, G.E. Karniadakis, Multi-element probabilistic collocation method in high dimensions, *J. Comput. Phys.* 229 (5) (2010) 1536–1557, <http://dx.doi.org/10.1016/j.jcp.2009.10.043>.
- [37] D. Xiu, G.E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, *SIAM J. Sci. Comput.* 24 (2) (2002) 619–644, <http://dx.doi.org/10.1137/s1064827501387826>.
- [38] A. Basmaji, A. Fau, J. Urrea-Quintero, M. Dannert, E. Voelsen, U. Nackenhorst, Anisotropic multi-element polynomial chaos expansion for high-dimensional non-linear structural problems, *Probab. Eng. Mech.* 70 (2022) 103366, <http://dx.doi.org/10.1016/j.probingmech.2022.103366>.
- [39] J.D. Jakeman, A. Narayan, D. Xiu, Minimal multi-element stochastic collocation for uncertainty quantification of discontinuous functions, *J. Comput. Phys.* 242 (2013) 790–808, <http://dx.doi.org/10.1016/j.jcp.2013.02.035>.
- [40] D. Giovanis, M. Shields, Variance-based simplex stochastic collocation with model order reduction for high-dimensional systems, *Internat. J. Numer. Methods Engrg.* 117 (11) (2019) 1079–1116, <http://dx.doi.org/10.1002/nme.5992>.
- [41] A. Galetzka, D. Loukrezis, N. Georg, H. De Gersem, U. Römer, An hp-adaptive multi-element stochastic collocation method for surrogate modeling with information re-use, *Internat. J. Numer. Methods Engrg.* <http://dx.doi.org/10.1002/nme.7234>.
- [42] L. Novák, On distribution-based global sensitivity analysis by polynomial chaos expansion, *Comput. Struct.* 267 (2022) 106808, <http://dx.doi.org/10.1016/j.compstruc.2022.106808>.
- [43] W. Gautschi, On generating orthogonal polynomials, *SIAM J. Sci. Stat. Comput.* 3 (3) (1982) 289–317, <http://dx.doi.org/10.1137/0903018>.
- [44] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, *Ann. Statist.* 32 (2) (2004) 407–451, <http://dx.doi.org/10.2307/3448465>.
- [45] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inform. Theory* 53 (12) (2007) 4655–4666, <http://dx.doi.org/10.1109/tit.2007.909108>.
- [46] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, *IEEE Trans. Signal Process.* 56 (6) (2008) 2346–2356, <http://dx.doi.org/10.1109/TSP.2007.914345>.
- [47] G. Blatman, B. Sudret, An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis, *Probab. Eng. Mech.* 25 (2) (2010) 183–197, <http://dx.doi.org/10.1016/j.probingmech.2009.10.003>.
- [48] A. Olivier, D. Giovanis, B. Aakash, M. Chauhan, L. Vandanapu, M.D. Shields, UQpy: A general purpose Python package and development environment for uncertainty quantification, *J. Comput. Sci.* 47 (2020) 101204.
- [49] S. Marelli, B. Sudret, UQLab: A framework for uncertainty quantification in Matlab, in: *Vulnerability, Uncertainty, and Risk*, 2014, pp. 2554–2563, <http://dx.doi.org/10.1061/9780784413609.257>.
- [50] M.D. McKay, W.J. Conover, R.J. Beckman, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* 21 (1979) 239–245, <http://dx.doi.org/10.1080/00401706.1979.10489755>.
- [51] W. Conover, On a better method for selecting input variables, 1975, unpublished Los Alamos National Laboratories manuscript, reproduced as Appendix A of “Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems” by J.C. Helton and F.J. Davis, Sandia National Laboratories report SAND2001-0417, printed November 2002. URL <https://prod-ng.sandia.gov/techlib-noauth/access-control.cgi/2001/010417.pdf>.
- [52] J. Hampton, A. Doostan, Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies, *J. Comput. Phys.* 280 (2015) 363–386, <http://dx.doi.org/10.1016/j.jcp.2014.09.019>.
- [53] P. Diaz, A. Doostan, J. Hampton, Sparse polynomial chaos expansions via compressed sensing and D-optimal design, *Comput. Methods Appl. Mech. Engrg.* 336 (2018) 640–666, <http://dx.doi.org/10.1016/j.cma.2018.03.020>.
- [54] J.F. Koksma, Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1, *Math. B* 11 (1942/1943) 7–11.
- [55] M. Johnson, L. Moore, D. Ylvisaker, Minimax and maximin distance designs, *J. Statist. Plann. Inference* 2 (26) (1990) 131–148, [http://dx.doi.org/10.1016/0378-3758\(90\)90122-B](http://dx.doi.org/10.1016/0378-3758(90)90122-B).
- [56] L. Pronzato, Minimax and maximin space-filling designs: some properties and methods for construction, *J. Soc. Fr. Stat.* 158 (1) (2017) 7–36.
- [57] J. Eliáš, M. Vořechovský, V. Sadílek, Periodic version of the minimax distance criterion for Monte Carlo integration, *Adv. Eng. Softw.* 149 (2020) 102900, <http://dx.doi.org/10.1016/j.advengsoft.2020.102900>.
- [58] M. Rosenblatt, Remarks on a multivariate transformation, *Ann. Math. Stat.* 23 (3) (1952) 470–472, <http://dx.doi.org/10.1214/aoms/1177729394>.
- [59] A. Nataf, Détermination des distributions de probabilité dont les marges sont données, *C. R. Acad. Sci.* 225 (1962) 42–43.
- [60] F. Wang, H. Li, System reliability under prescribed marginals and correlations: Are we correct about the effect of correlations? *Reliab. Eng. Syst. Saf.* 173 (2018) 94–104, <http://dx.doi.org/10.1016/j.ress.2017.12.018>.
- [61] J.M. Davis, P. Hagelstein, Gibbs phenomena for some classical orthogonal polynomials, *J. Math. Anal. Appl.* 505 (1) (2022) 125574.

- [62] P. Frantík, FyDik - A software for interactive simulations of dissipative nonlinear dynamical systems based on physical discretization, 2000–2022, <http://fydik.kitnarf.cz/>.
- [63] P. Frantík, Simulation of the stability loss of the von Mises truss in an unsymmetrical stress state, *Eng. Mech.* 14 (3) (2007) 155–161.
- [64] K. Cheng, Z. Lu, C. Ling, S. Zhou, Surrogate-assisted global sensitivity analysis: an overview, *Struct. Multidiscip. Optim.* 61 (3) (2020) 1187–1213, <http://dx.doi.org/10.1007/s00158-019-02413-5>.
- [65] C.Q. Lam, W.I. Notz, Sequential adaptive designs in computer experiments for response surface model fit, *Stat. Appl.* 6 (2008) 207–233.
- [66] M. Vořechovský, J. Eliáš, Modification of the Maximin and ϕ_p (phi) criteria to achieve statistically uniform distribution of sampling points, *Technometrics* 62 (3) (2020) 371–386, <http://dx.doi.org/10.1080/00401706.2019.1639550>.
- [67] M. Vořechovský, J. Mašek, J. Eliáš, Distance-based optimal sampling in a hypercube: Analogies to N-body systems, *Adv. Eng. Softw.* 137 (2019) 102709, <http://dx.doi.org/10.1016/j.advengsoft.2019.102709>.
- [68] M. Vořechovský, J. Mašek, Distance-based optimal sampling in a hypercube: Energy potentials for high-dimensional and low-saturation designs, *Adv. Eng. Softw.* 149 (2020) 102880, <http://dx.doi.org/10.1016/j.advengsoft.2020.102880>.

3.3 On Distribution-based Global Sensitivity Analysis by Polynomial Chaos Expansion

NOVÁK, L. On distribution-based global sensitivity analysis by polynomial chaos expansion. *COMPUTERS & STRUCTURES*, 2022, vol. 267. ISSN: 0045-7949. (WoS-AIS: Q1)
DOI: 10.1016/j.compstruc.2022.106808

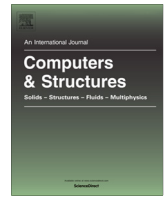
Description

The author proposed a novel technique for a distribution-based sensitivity analysis derived directly from PCE. First, the theory of PCE is deeply reviewed and extended by the results from a standard linearization problem of Hermite and Jacobi polynomials. Further, analytical formulas for PDF and CDF of Gram-Charlier expansion are provided, and the concept of reduced PCE for sensitivity analysis is presented. Finally, a complex methodology combining the presented theoretical results for sensitivity analysis is proposed. The proposed method is based on Kullback-Leibler divergence between probability distribution of QoI and conditional distributions of QoI derived directly from PCE using analytical estimations of the first four statistical moments. The main advantage of the proposed method is its numerical efficiency, since it is analytically derived from the PCE, and due to the fact that it takes into account the whole probability distribution in contrast to the well-known Sobol' indices. The proposed method is applied to several examples including analytical functions, finite element models, and also an application for the reliability-oriented sensitivity analysis.

Role of the author

Percentage of contribution: 100%

Lukáš Novák is the author of this paper responsible for the concept, the methodology, and the numerical results of the presented research.



On distribution-based global sensitivity analysis by polynomial chaos expansion



Lukáš Novák

Faculty of Civil Engineering, Brno University of Technology, Veveri 331/95, Brno 60200, Czech Republic

ARTICLE INFO

Article history:

Received 18 January 2022

Accepted 12 April 2022

Keywords:

Sensitivity analysis

Gram–Charlier expansion

Surrogate modelling

Kullback–Leibler divergence

Polynomial chaos expansion

ABSTRACT

This paper presents a novel distribution-based global sensitivity analysis based on the Kullback–Leibler divergence derived directly from generalized polynomial chaos expansion (PCE). The synergy between PCE and Gram–Charlier expansion is utilized for derivation of novel and computationally efficient sensitivity indices. In contrast to a standard procedure for estimation of higher statistical moments, this paper reviews standard linearization problem of Hermite and Jacobi polynomials in order to efficiently estimate skewness and kurtosis directly from PCE. Higher statistical moments are used for an estimation of probability distribution by Gram–Charlier expansion, which is represented by derived explicit cumulative distribution function. The proposed sensitivity indices taking the whole probability distribution into account are calculated for several numerical examples of increasing complexity in order to present their possibilities. It is shown, that the proposed sensitivity indices are obtained without any additional computational demands together with Sobol indices, and thus can be easily used as complementary information for a complex sensitivity analysis or any decision making in industrial applications. Application of the proposed approach on engineering structure is presented in case of prestressed concrete roof girders failing shear. Moreover, the potential of the proposed approach for reliability-oriented sensitivity analysis is investigated in pilot numerical example.

© 2022 Elsevier Ltd. All rights reserved.

1. Introduction

The mathematical model of a physical system can be seen as a function of an input vector $g(\mathbf{X})$. It is necessary to consider the uncertain input variables described by specific probability distribution in order to obtain realistic results. The task of an analyst is then an uncertainty quantification of the model response - quantity of interest (QoI). There are efficient probabilistic methods commonly used in stochastic analysis for this purpose, typically based on pseudo-random sampling of the input vector and repetitive calculation of deterministic $g(\mathbf{X})$ in order to obtain corresponding results, which are further statistically processed for moment and sensitivity analysis.

Unfortunately, the combination of sampling-based probabilistic methods and mathematical models in industrial applications is highly time-consuming or even not feasible, especially in the case of large mathematical models with many random input variables. Moreover, the mathematical function is often represented by non-linear finite element model (NLFEM) in industrial applications, which is an accurate, but highly time-consuming numerical algo-

rithm. The solution can be an approximation of the original model by an explicit function usually called the surrogate model. This paper is focused on Polynomial Chaos Expansion (PCE) as a powerful surrogate model with broad possibilities of post-processing due to its special form of basis functions, as will be shown further.

PCE originally proposed by Norbert Wiener [1] and further investigated in the context of engineering problems by many researchers, e.g. [2,3], represents spectral expansion of the original stochastic problem in polynomial basis. Such approach is often far more efficient in comparison to the sampling methods. Moreover, once the PCE is available, it is possible to investigate the explicit function in order to obtain additional information about a mathematical model including statistical moments, probability distribution of QoI or sensitivity indices without additional sampling, which is especially beneficial in industrial applications [4].

There are generally two types of sensitivity analyses. On the one hand, there is local sensitivity analysis focused on the behavior of a function around a point of interest (e.g. one-at-a-time and screening) [5]. Such information is typically obtained by a differentiation of the mathematical model in the area of interest. On the other hand, there is global sensitivity analysis investigating the whole design domain e.g. the well-known analysis of variance (ANOVA) [6,7]. Global sensitivity analysis is an area of interest for many

E-mail address: novak.l@fce.vutbr.cz

researchers, especially ANOVA represented by Sobol indices is used frequently nowadays. Nevertheless, ANOVA methods are still highly computationally demanding, and thus their computation in combination with time-consuming mathematical models is often not feasible. Fortunately, it was shown by Sudret [8] how to derive Sobol indices directly from PCE. It leads to a significant reduction of computational demands in comparison with the traditional pick and freeze Monte Carlo approach [7].

Although ANOVA represents a strong tool for global sensitivity analysis, it takes only the first two statistical moments into account. Therefore, recent theoretical research is focused on the so-called moment-independent or distribution-based sensitivity analysis. These methods generally take the whole distribution of random variables into account. The first distribution-based indices were derived by Borgonovo [9,10], and later Gamboa et al. [11] proposed indices based on the Hoeffding-Sobol decomposition and Cramer-von Misses distance. Both approaches are well suited for pseudo-random sampling to obtain sensitivity indices, which are computationally highly demanding, and their accuracy is highly dependent on the number of samples. Another approach is represented by the relative entropy measure called Kullback-Leibler divergence [12], which was proposed by Park [13] and has already been used in several engineering applications [14,15]. This approach is interpreted as the measure of the information loss when the original distribution is approximated by distribution neglecting the uncertainty of selected input variables. The use of relative entropy for sensitivity analysis makes perfect sense, since the obtained results are often used for reduction of stochastic models neglecting the uncertainty of selected variables. However, for this method, it is necessary to obtain the probability distribution of the original model, and the probability distribution of the model neglecting the uncertainty of selected input variables according to omission sensitivity [16], assuming it as a deterministic value (typically mean value). Unfortunately, such approach requires a high number of numerical calculations.

This paper is focused on the derivation of sensitivity indices based on Kullback-Leibler divergence directly from PCE without any additional computational demands together with Sobol indices. This novel approach is based on analytical expressions for higher statistical moments obtained from PCE and their utilization for CDF approximation by Gram-Charlier expansion, which is also derived in analytical form in this paper. Thanks to analytical expressions, the proposed approach can be employed without additional calculations of the original mathematical model and thus it leads to significant reduction of computational requirements.

The proposed distribution-based sensitivity measure might be especially important in sensitivity analysis with respect to reliability of structures. The reliability analysis represents a crucial task in a design and assessment of structures and thus it is important to identify input random variables, which are able to significantly affect the failure probability. However, it is not often trivial to understand the influence of uncertainty of input random variables to extremely low failure probabilities of structure. Therefore scientists are ultimately interested in Reliability-Oriented Sensitivity Analysis (ROSA) [17,18]. ROSA is generally able to identify the importance of input variables directly with respect to reliability of real structures [19]. Unfortunately important drawback of ROSA is extreme computational burden, since it is necessary to estimate conditional failure probabilities associated to each input random variable. Therefore scientists are focused on novel efficient derivation of sensitivity measure for ROSA, e.g. Contrast function [20] and Entropy [21,22]. The significantly higher efficiency of ROSA could be also obtained by surrogate modeling, such as recently proposed Krigging-MC strategy [23]. In this paper, a possibility of the proposed approach based on PCE for efficient derivation of ROSA

indices based on two selected sensitivity measures is presented in a pilot example.

2. Polynomial chaos expansion

2.1. Surrogate modelling with PCE

Assume a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where Ω is an event space, \mathcal{F} is a σ -algebra on Ω (collection of subsets closed under complementation and countable unions) and \mathcal{P} is a probability measure on \mathcal{F} . If the input variable of the mathematical model is a random variable $X(\omega)$, $\omega \in \Omega$ the model response $Y(\omega)$ is a random variable. PCE represents the variable Y as a function of another random variable ξ called germ with given distribution:

$$Y = g(X) = g^{PCE}(\xi), \quad (1)$$

and representing that function as a polynomial expansion in a manner similar to the Fourier series of a periodic signal. A set of polynomials orthogonal with respect to the distribution of the germ are used as a polynomial basis of Hilbert space $L^2(\Omega, \mathcal{P}, \mathbb{R})$ of all real-valued random variables of finite variance, where \mathcal{P} takes over the meaning of the probability distribution. The orthogonality condition for all $j \neq k$ is given by the inner product of $L^2(\Omega, \mathcal{P}, \mathbb{R})$ defined for any two functions ψ_j and ψ_k with respect to the weight function p_ξ (probability density function of ξ) as

$$\langle \psi_j, \psi_k \rangle = \int \psi_j(\xi) \psi_k(\xi) p_\xi d\xi = 0. \quad (2)$$

It means that there are specific orthogonal polynomials associated with the specific distribution of a germ via its weighting function p_ξ . For example, Hermite polynomials orthogonal to the Gaussian measure are associated to the normally distributed germs. Orthogonal polynomials corresponding to other distributions can be chosen according to Wiener-Askey scheme [24]. For further processing, it is beneficial to use normalized polynomials (orthonormal), where the inner product is equal to the Kronecker delta, i.e. $\delta_{jk} = 1$ if and only if $j = k$ and otherwise $\delta_{jk} = 0$:

$$\langle \psi_j, \psi_k \rangle = \delta_{jk}. \quad (3)$$

In the case of \mathbf{X} and ξ being vectors containing M random variables, the polynomial Ψ is multivariate and it is constructed as a tensor product of univariate orthogonal polynomials

$$\Psi_\alpha(\xi) = \prod_{i=1}^M \psi_{\alpha_i}(\xi_i), \quad (4)$$

where $\alpha \in \mathbb{N}^M$ is a set of integers called multi-index reflecting degrees of all univariate polynomials in the multivariate PCE term. The orthonormality of multivariate polynomials is defined as:

$$\langle \psi_\alpha, \psi_\beta \rangle = \int \psi_\alpha(\xi) \psi_\beta(\xi) p_\xi d\xi = \delta_{\alpha\beta}. \quad (5)$$

The QoI, the response of the mathematical model $Y = g(X)$, can be then represented according to Ghanem and Spanos [3] as:

$$Y = g(\mathbf{X}) = \sum_{\alpha \in \mathbb{N}^M} \beta_\alpha \Psi_\alpha(\xi) \quad (6)$$

where β_α are deterministic coefficients and Ψ_α are multivariate orthogonal polynomials.

Note that in case of arbitrary probability distribution of input random variables, it is not possible to use Wiener-Askey scheme of well-known polynomial families. However, it is possible to construct basis polynomials directly by Gram-Schmidt orthogonalization process [25] or transform all input random variables into standardized Gaussian space by Nataf transformation [26-28].

Unfortunately, for specific combinations of input parameters and correlation coefficients, there is not a guaranteed solution; more details about the limitations of Nataf transformation can be found in [29]. The solution of such specific cases can be a general framework based on the theory of copula [30].

2.2. Non-intrusive computation of PCE coefficients

PCE according to Eq. 6 must be truncated to a final number of terms P for practical computation. It is commonly achieved by retaining only terms whose total degree $|\alpha|$ is equal or less than the given p . Therefore, the truncated set of PCE terms is defined as

$$\mathcal{A}^{M,p} = \left\{ \alpha \in \mathbb{N}^M : |\alpha| = \sum_i \alpha_i \leq p \right\}. \quad (7)$$

The cardinality of the truncated set $\mathcal{A}^{M,p}$ is given by:

$$\text{card } \mathcal{A}^{M,p} = \frac{(M+p)!}{M! p!} \quad (8)$$

Moreover, in practical applications, it is beneficial to prefer only basis functions with lower number of interaction terms, e.g. Hyperbolic truncation [31].

From a statistical point of view, PCE is a simple linear regression model with intercept. Therefore, it is possible to use ordinary least squares (OLS) regression to minimize the error ε :

$$\beta = \arg \min \frac{1}{n} \sum_{i=1}^n \left(\beta^T \Psi(\xi^{(i)}) - g(x^{(i)}) \right)^2. \quad (9)$$

First, it is necessary to create n realizations of input random vector \mathbf{X} and corresponding results of the original mathematical model \mathcal{Y} , together called experimental design (ED). Then, deterministic coefficients β are calculated as

$$\beta = (\Psi^T \Psi)^{-1} \Psi^T \mathcal{Y}, \quad (10)$$

where Ψ is the data matrix:

$$\Psi = \left\{ \Psi_{ij} = \Psi_j(\xi^{(i)}), i = 1, \dots, n, \quad j = 0, \dots, P-1 \right\}. \quad (11)$$

Note that, the number of terms P is highly dependent on the number of input random variables M and the maximum total degree of polynomials p . Therefore, in case of a large stochastic model, it is easy to obtain a computationally highly demanding problem. The solution can be a utilization of advanced model selection algorithms e.g. Least Angle Regression (LAR) [32] to find an optimal set of PCE terms as proposed by Blatman and Sudret [31] and implemented in several software packages together with similar model selection algorithms [33] for practical construction of PCE [34–37].

3. Statistical moments derived from PCE

A specific form of PCE and orthogonality of polynomials allows for a powerful and efficient post-processing. Once a PCE approximation is created, it is possible to obtain statistical moments of QoI. Generally, a statistical moment of any order is defined as:

$$\begin{aligned} \langle Y^m \rangle &= \int [g(\mathbf{X})]^m p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int \left[\sum_{\alpha \in \mathbb{N}^M} \beta_{\alpha} \Psi_{\alpha}(\xi) \right]^m p_{\xi}(\xi) d\xi = \\ &= \int \sum_{\alpha_1 \in \mathbb{N}^M} \dots \sum_{\alpha_m \in \mathbb{N}^M} \beta_{\alpha_1} \dots \beta_{\alpha_m} \Psi_{\alpha_1}(\xi) \dots \Psi_{\alpha_m}(\xi) p_{\xi}(\xi) d\xi = \\ &= \sum_{\alpha_1 \in \mathbb{N}^M} \dots \sum_{\alpha_m \in \mathbb{N}^M} \beta_{\alpha_1} \dots \beta_{\alpha_m} \int \Psi_{\alpha_1}(\xi) \dots \Psi_{\alpha_m}(\xi) p_{\xi}(\xi) d\xi. \end{aligned}$$

As can be seen from the final part of the formula, in case of PCE, it is necessary to integrate over basis functions (orthonormal polynomials), which leads to dramatic simplification in comparison to the integration of the original mathematical function. Moreover, it is also possible to write an analytical expression of the integral in several cases. Specifically, the first statistical moment (mean value) is obtained as

$$\mu_Y = \langle Y^1 \rangle = \int \left[\sum_{\alpha \in \mathbb{N}^M} \beta_{\alpha} \Psi_{\alpha}(\xi) \right] p_{\xi}(\xi) d\xi = \sum_{\alpha \in \mathbb{N}^M} \beta_{\alpha} \int \Psi_{\alpha}(\xi) p_{\xi}(\xi) d\xi.$$

Considering the orthonormality of polynomials

$$\int \Psi_{\alpha}(\xi) p_{\xi}(\xi) d\xi = 0 \quad \forall \alpha \neq 0, \quad \Psi_0 \equiv 1,$$

it is possible to obtain the mean value directly from the PCE deterministic coefficients. Namely, the mean value is equal to the first deterministic coefficient of the expansion

$$\mu_Y = \langle Y^1 \rangle = \beta_0. \quad (12)$$

The second statistical moment $\langle Y^2 \rangle$ is written as

$$\begin{aligned} \langle Y^2 \rangle &= \int \left[\sum_{\alpha \in \mathcal{A}^{M,p}} \beta_{\alpha} \Psi_{\alpha}(\xi) \right]^2 p_{\xi}(\xi) d\xi = \\ &= \sum_{\alpha_1 \in \mathcal{A}^{M,p}} \sum_{\alpha_2 \in \mathcal{A}^{M,p}} \beta_{\alpha_1} \beta_{\alpha_2} \int \Psi_{\alpha_1}(\xi) \Psi_{\alpha_2}(\xi) p_{\xi}(\xi) d\xi = \\ &= \sum_{\alpha \in \mathcal{A}^{M,p}} \beta_{\alpha}^2 \int \Psi_{\alpha}(\xi)^2 p_{\xi}(\xi) d\xi = \sum_{\alpha \in \mathcal{A}^{M,p}} \beta_{\alpha}^2 \langle \Psi_{\alpha}, \Psi_{\alpha} \rangle. \end{aligned}$$

Considering again the orthonormality of polynomials, defined by the inner product in Eq. 3, it is possible to obtain the variance $\sigma_Y^2 = \langle Y^2 \rangle - \mu_Y^2$ as a sum of all squared deterministic coefficients except the intercept, which represents the mean value:

$$\sigma_Y^2 = \sum_{\substack{\alpha \in \mathcal{A} \\ \alpha \neq 0}} \beta_{\alpha}^2. \quad (13)$$

Higher statistical central moments, skewness γ_Y (3rd moment) and kurtosis κ_Y (4th moment), are generally obtained as follows:

$$\gamma_Y := \frac{1}{\sigma^3} \mathbb{E} \left[(Y - \mu_Y)^3 \right] = \frac{1}{\sigma^3} \sum_{\substack{\alpha \in \mathcal{A} \\ \alpha \neq 0}} \sum_{\substack{\beta \in \mathcal{A} \\ \beta \neq 0}} \sum_{\substack{\gamma \in \mathcal{A} \\ \gamma \neq 0}} \langle \Psi_{\alpha} \Psi_{\beta} \Psi_{\gamma} \rangle \beta_{\alpha} \beta_{\beta} \beta_{\gamma} \quad (14)$$

$$\begin{aligned} \kappa_Y &:= \frac{1}{\sigma^4} \mathbb{E} \left[(Y - \mu_Y)^4 \right] \\ &= \frac{1}{\sigma^4} \sum_{\substack{\alpha \in \mathcal{A} \\ \alpha \neq 0}} \sum_{\substack{\beta \in \mathcal{A} \\ \beta \neq 0}} \sum_{\substack{\gamma \in \mathcal{A} \\ \gamma \neq 0}} \sum_{\substack{\delta \in \mathcal{A} \\ \delta \neq 0}} \langle \Psi_{\alpha} \Psi_{\beta} \Psi_{\gamma} \Psi_{\delta} \rangle \beta_{\alpha} \beta_{\beta} \beta_{\gamma} \beta_{\delta} \end{aligned} \quad (15)$$

Note that the integration of higher products may be time consuming, especially in a high-dimensional space with high total polynomial order. Unfortunately, there is no closed analytical formula for these higher-order products applicable for orthonormal polynomials of any type. Nevertheless, there are known explicit expressions of the third-order products for several types of polynomials e.g. Hermite, Jacobi and generalized Laguerre. Finding the third-order product of orthogonal polynomials is a subject of the so-called standard linearization problem, which has been deeply investigated by mathematicians [38–41]. Although this topic is extremely broad and investigating various polynomial families, this paper is limited only to review of main results involving Hermite and Jacobi polynomials often used in PCE associated to Gaussian resp. Uniform distributions of germs.

The standard linearization formula for general orthogonal polynomials P_n is defined by the following expression:

$$\langle P_k P_l \rangle = \int P_k(\xi) P_l(\xi) p_\xi d\xi = \omega_l \delta_{kl}, \tag{16}$$

where δ_{kl} is Kronecker delta and $\omega_l > 0$ is the normalization constant. The expression for the third-order product is then as follows:

$$\langle \widehat{P}_k \widehat{P}_l \widehat{P}_m \rangle = \sqrt{\frac{\omega_m}{\omega_k \omega_l}} B(k, l, m) = \widehat{B}(k, l, m). \tag{17}$$

Note that $\widehat{B}(k, l, m)$ are associated to normalized polynomials $\widehat{P}_n = \omega_n^{-\frac{1}{2}} P_n$. The fourth-order product can be obtained from the third-order product by a mathematical induction:

$$\begin{aligned} \langle \widehat{P}_j \widehat{P}_k \widehat{P}_l \widehat{P}_m \rangle &= \sum_{n=0}^{j+k} \widehat{B}(j, k, n) \langle \widehat{P}_l \widehat{P}_m \widehat{P}_n \rangle \\ &= \sum_{n=0}^{j+k} \widehat{B}(j, k, n) \widehat{B}(l, m, n). \end{aligned} \tag{18}$$

Specifically the most important family of polynomials used in PCE is represented by Hermite polynomials associated to Gaussian distribution of a germ. The univariate probabilists' Hermite polynomial of n^{th} order normalized by $\omega_n = n!$ is:

$$H_n(\xi) = \frac{(-1)^n}{\sqrt{n!}} e^{\frac{\xi^2}{2}} \frac{d^n}{d\xi^n} e^{-\frac{\xi^2}{2}}, \quad n \in \mathbb{N} \tag{19}$$

The coefficients $\widehat{B}(k, l, m)$ for Hermite polynomials are

$$\widehat{B}(k, l, m) = \left[\binom{k}{g-m} \binom{l}{g-m} \binom{m}{g-k} \right]^{\frac{1}{2}}, \tag{20}$$

whenever $2g = k + l + m$ is even and $|k - l| \leq m \leq k + l$. Otherwise $\widehat{B}(k, l, m) = 0$. The proof can be found in [42,39].

The second important distribution of germ is uniform, which is associated to Legendre polynomials. The orthonormal Legendre polynomials are special case of Jacobi polynomials, deeply investigated in literature and thus only selected results are summarized further. The univariate Legendre polynomial of n^{th} order normalized by $\omega_n = \frac{1}{2n+1}$ is given by Rodrigues' formula:

$$L_n(\xi) = \frac{1}{(2n+1)2^n n!} \frac{d^n}{d\xi^n} (\xi^2 + 1)^n, \quad n \in \mathbb{N} \tag{21}$$

The coefficients $\widehat{B}(k, l, m)$ for Legendre polynomials are

$$\widehat{B}(k, l, m) = \sqrt{\frac{\omega_m}{\omega_k \omega_l}} (2m+1) \binom{k \quad l \quad m}{0 \quad 0 \quad 0}^2 \tag{22}$$

whenever $2g = k + l + m$ is even and l, m, n satisfy triangle inequality, otherwise $\widehat{B}(k, l, m) = 0$. Note that the last squared term represents Wigner (3j) coefficient function for compactness of formula. Proof can be simply obtained from Neumann-Adams formula [41] by proper normalization.

Finally, from Eq. 4 and Eq. 17 it follows for general multivariate case:

$$\langle \widehat{P}_\alpha \widehat{P}_\beta \widehat{P}_\gamma \rangle = \prod_{m=1}^M \widehat{B}(\alpha_m, \beta_m, \gamma_m) \tag{23}$$

and similarly from Eq. 18:

$$\langle \widehat{P}_\alpha \widehat{P}_\beta \widehat{P}_\gamma \widehat{P}_\delta \rangle = \prod_{m=1}^M \sum_{n=0}^{\alpha_m + \beta_m} \widehat{B}(\alpha_m, \beta_m, n) \widehat{B}(\gamma_m, \delta_m, n). \tag{24}$$

where α_m represents m th position of multindex α etc. Note that it is possible to combine Legendre and Hermite polynomials for each of

input variables and use corresponding formulas for $\widehat{B}(k, l, m)$, since the final statistical moments are obtained as a product of separate $\widehat{B}(k, l, m)$ in each dimension. Utilization of the reviewed analytical formulas leads to significant improvement of computational efficiency and thus it allows estimation of higher statistical moments even for high M and p .

4. Probability distribution function of QoI

It is important to remember that PCE is an expansion to construct a random variable with a probability distribution identical to the quantity of interest Y . It is not an expansion of its probability distribution function f_Y (PDF) or cumulative distribution function F_Y (CDF). If one is interested in the distribution of Y^{PCE} , it is possible to utilize one of the well-known classical distribution expansions such as Gram-Charlier expansion (G-C) or a similar Edgeworth expansion. The Edgeworth expansion contains one more Hermite polynomial compared to G-C, while keeping the number of parameters constant. However, it was shown [43], that the range of parameters, for which positivity of the approximation is guaranteed, is smaller than for G-C. For this reason, G-C is used here for the estimation of distribution of QoI.

Let us assume that it is possible to write probability distribution of Y as a perturbation of Gaussian PDF ϕ . Once the QoI is normalized to be zero-mean and unit-variance, the Gram-Charlier expansion of f_Y is as follows:

$$f_Y(y) = \sum_{n=0}^{\infty} c_n H_n(y) \phi(y). \tag{25}$$

After several mathematical operations, considering the characteristics of Hermite polynomials, it is possible to write the approximation of PDF in the terms of its central moments (e.g. the first four moments) in the form of G-C type A:

$$f_Y(y) = \left[1 + \frac{\gamma_Y}{\sqrt{3!}} H_3(y) + \frac{\kappa_Y - 3}{\sqrt{4!}} H_4(y) \right] \phi(y). \tag{26}$$

In practical computation, the moments are estimated from samples and the estimation is highly sensitive to outliers. Fortunately, it is possible to get statistical moments analytically in case of PCE as was shown in the previous section. Therefore, once the PCE is available, it is possible to analytically obtain $f_Y(y)$. For the further sections, it is necessary to derive an explicit expression for CDF as well.

Theorem 1. The Gram-Charlier expansion of CDF F_Y is in the following form:

$$F_Y = \Phi(y) - \left[\frac{\gamma_Y}{3\sqrt{2!}} H_2(y) + \frac{\kappa_Y - 3}{4\sqrt{3!}} H_3(y) \right] \phi(y). \tag{27}$$

Proof. The integration of orthogonal probabilists' Hermite polynomials $H_n(y)$ and standard Gaussian pdf $\phi(y)$ is:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^t H_n(y) e^{-\frac{y^2}{2}} dy.$$

Polynomials $H_n(y)$ are described by the Hermite differential equation

$$\frac{d^2 H_n(y)}{dy^2} - y \frac{dH_n(y)}{dy} + nH_n(y) = 0$$

and Appell sequence

$$\frac{dH_n(y)}{dy} = nH_{n-1}(y).$$

First both sides of the above equations are multiplied by $e^{-\frac{1}{2}y^2}$:

$$e^{-\frac{1}{2}y^2} \frac{d^2 H_n(y)}{dy^2} - e^{-\frac{1}{2}y^2} y \frac{dH_n(y)}{dy} + e^{-\frac{1}{2}y^2} nH_n(y) = 0 \quad (28)$$

and

$$e^{-\frac{1}{2}y^2} \frac{dH_n(y)}{dy} = e^{-\frac{1}{2}y^2} nH_{n-1}(y). \quad (29)$$

From Eq. 29, by differentiation with respect to the variable y , the following expression is obtained:

$$\begin{aligned} & -ye^{-\frac{1}{2}y^2} \frac{dH_n(y)}{dy} + e^{-\frac{1}{2}y^2} \frac{d^2 H_n(y)}{dy^2} + \\ & + ye^{-\frac{1}{2}y^2} nH_{n-1}(y) - e^{-\frac{1}{2}y^2} n \frac{dH_{n-1}(y)}{dy} = 0 \end{aligned}$$

Then, the Eq. 28 is subtracted from the above formula:

$$ye^{-\frac{1}{2}y^2} nH_{n-1}(y) - e^{-\frac{1}{2}y^2} n \frac{dH_{n-1}(y)}{dy} - e^{-\frac{1}{2}y^2} nH_n(y) = 0,$$

which can be rewritten as

$$-\left(-ye^{-\frac{1}{2}y^2} H_{n-1}(y) + e^{-\frac{1}{2}y^2} \frac{dH_{n-1}(y)}{dy}\right) = e^{-\frac{1}{2}y^2} H_n(y).$$

Finally, the following relationship is obtained:

$$-\frac{d}{dy} \left(e^{-\frac{1}{2}y^2} H_{n-1}(y) \right) = e^{-\frac{1}{2}y^2} H_n(y). \quad (30)$$

Using the obtained result, the solution of the integral is

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t H_n(y) e^{-\frac{1}{2}y^2} dy &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \left[-\frac{d}{dy} \left(e^{-\frac{1}{2}y^2} H_{n-1}(y) \right) \right] dy = \\ &= -\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} H_{n-1}(t). \end{aligned}$$

For orthonormal Hermite polynomials, it is necessary to divide both sides of Eq. 30 by a normalization constant $\sqrt{n!}$ which leads to:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^t H_n(y) e^{-\frac{1}{2}y^2} dy = -\frac{1}{\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} H_{n-1}(t). \quad (31)$$

The integration of Gram–Charlier expansion in order to obtain CDF is

$$\begin{aligned} F_Y &= \int_{-\infty}^t f_Y(y) dy = \Phi(t) + \frac{\gamma_Y}{\sqrt{3!}} \int_{-\infty}^t H_3(y) \phi(y) dy \\ &+ \frac{\kappa_Y}{\sqrt{4!}} \int_{-\infty}^t H_4(y) \phi(y) dy = \Phi(t) - \left[\frac{\gamma_Y}{3\sqrt{2!}} H_2(t) + \frac{\kappa_Y - 3}{4\sqrt{3!}} H_3(t) \right] \phi(t). \end{aligned}$$

where the last operation arises from the Eq. 31. \square

5. Hoeffding-sobol decomposition and reduced PCE

One of the advanced methods for global sensitivity analysis – ANOVA – is represented by Sobol indices derived from the Hoeffding-Sobol decomposition [7]. A brief overview of the theoretical background is given in this subsection (more details about the derivation of Sobol indices directly from PCE can be found in [8]). Let \mathbf{X} be a random vector with independent margins. For any $\mathbf{x} \in \mathbb{R}^M$ and any subset $\mathbf{u} \subseteq I = \{1, \dots, M\}$, $\mathbf{x}_{\mathbf{u}}$ concatenates the components of \mathbf{x} whose indices are included in \mathbf{u} . According to the Hoeffding-Sobol decomposition, any square integrable function $g(\mathbf{x})$ can be decomposed as:

$$\begin{aligned} g(\mathbf{x}) &= g_0 + \sum_{i=1}^M g_i(x_i) + \sum_{1 \leq i < j \leq M} g_{ij}(x_i, x_j) + \dots + g_{1,2,\dots,M}(\mathbf{x}) = \\ &= g_0 + \sum_{\substack{\mathbf{u} \subseteq \{1, \dots, M\} \\ \mathbf{u} \neq \emptyset}} g_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}). \end{aligned} \quad (32)$$

In consequence of the defined decomposition, the variance of Y can be decomposed as well as:

$$\sigma_Y^2 = \text{Var}[Y] = \sum_{\substack{\mathbf{u} \subseteq \{1, \dots, M\} \\ \mathbf{u} \neq \emptyset}} \text{Var}[g_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})], \quad (33)$$

where $\text{Var}[g_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})]$ are partial variances. The first Sobol indices are obtained if \mathbf{u} contains a single i -th input variable, the second-order indices correspond to two input variables, etc. Important information about the influence of input variables and all interactions can be expressed by total Sobol indices representing the first order influence and influence of all interactions, which are computed as

$$S_i^T = \sum_{\mathbf{u} \in \mathbf{u}} S_{\mathbf{u}}. \quad (34)$$

A numerical computation of Sobol indices by Monte Carlo is typically highly computationally demanding. Fortunately, there is a connection between PCE and the Hoeffding-Sobol decomposition as was already shown in [8]. PCE can be rewritten in the form of the Hoeffding-Sobol decomposition by a simple reordering of the terms:

$$g(\mathbf{x}) = g_0 + \sum_{\substack{\mathbf{u} \subseteq \{1, \dots, M\} \\ \mathbf{u} \neq \emptyset}} g_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}) \approx g^{PCE}(\mathbf{x}) = \beta_0 + \sum_{\alpha \in \mathcal{A}_{\mathbf{u}}} \beta_{\alpha} \Psi_{\alpha}(\xi), \quad (35)$$

where the set of basis multivariate polynomials dependent on $\mathbf{x}_{\mathbf{u}}$ is defined as:

$$\mathbf{A}_{\mathbf{u}} = \{ \alpha \in \mathcal{A}^{M,p} : \alpha_k \neq 0 \leftrightarrow k \in \mathbf{u} \}. \quad (36)$$

Therefore, Sobol indices can be analytically obtained directly from PCE in virtue of Eq. 13 as follows [8]:

$$S_i^{PCE} = \frac{\sum_{\alpha \in \mathcal{A}_i} \beta_{\alpha}^2}{\sigma_Y^2} \quad \mathcal{A}_i = \{ \alpha \in \mathcal{A}^{M,p} : \alpha_i > 0, \alpha_{j \neq i} = 0 \}, \quad (37)$$

$$S_i^{T,PCE} = \frac{\sum_{\alpha \in \mathcal{A}_i^T} \beta_{\alpha}^2}{\sigma_Y^2} \quad \mathcal{A}_i^T = \{ \alpha \in \mathcal{A}^{M,p} : \alpha_i > 0 \}. \quad (38)$$

The above expressions represent just a selection of the specific PCE coefficients associated to $\mathbf{x}_{\mathbf{u}}$. However, for the further sections, it is beneficial to understand this process as a reduction of the original function to $g_{\mathbf{u}}^{PCE}$ as the first step and a calculation of the variance of $g_{\mathbf{u}}^{PCE}$ using Eq. 13 as the second step. The obtained partial variance is then normalized by the original variance assuming all terms of PCE.

Formally written, for any set $\mathbf{u} \subseteq I = \{1, \dots, M\}$ let $\sim \mathbf{u}$ be the complement to \mathbf{u} , i.e. $\sim \mathbf{u} = I \setminus \mathbf{u}$. The reduced PCE approximation $g_{\mathbf{u}}^{PCE}$ (neglecting the influence of selected variables $\mathbf{x}_{\mathbf{u}}$, whose indices are included in \mathbf{u}) of the original model has the following form:

$$g_{\mathbf{u}}^{PCE}(\mathbf{x}) = \beta_0 + \sum_{\alpha \in \mathcal{A}_{\sim \mathbf{u}}} \beta_{\alpha} \Psi_{\alpha}(\xi), \quad (39)$$

where the reduced set of multivariate polynomials is defined as

$$\mathcal{A}_{\sim \mathbf{u}} = \{ \alpha \in \mathcal{A}^{M,P} : \alpha_k \neq 0 \leftrightarrow k \in \sim \mathbf{u} \}. \quad (40)$$

For the sake of clarity, the set of multivariate polynomials for n^{th} order indices, i.e. when \mathbf{u} contains n variables, is defined as

$$\mathcal{A}_{\sim \mathbf{u}} = \left\{ \alpha \in \mathcal{A}^{M,P} : \alpha_{\mathbf{u}} = 0 \cup \sum_{\alpha} \alpha_i \neq \sum_{\alpha_{\mathbf{u}}} \alpha_j \right\}, \quad (41)$$

and for total indices, i.e. when \mathbf{u} contains a single variable and the analyst is interested in the total influence of single variables together with all their interactions, it is defined as

$$\mathcal{A}_{\sim \mathbf{u}} = \{ \alpha \in \mathcal{A}^{M,P} : \alpha_{\mathbf{u}} = 0 \}. \quad (42)$$

The first order or total Sobol indices are then calculated using reduced PCE as

$$S_{\mathbf{u}}^{\text{PCE}} = 1 - \frac{\sigma_{g_{\mathbf{u}}^{\text{PCE}}}^2}{\sigma_{g^{\text{PCE}}}^2}. \quad (43)$$

As can be seen, the influence of variables $\mathbf{X}_{\mathbf{u}}$ is measured by reduced PCE, which neglects these variables, and by the proper normalization of the result. Note that it is possible to create reduced PCE in the first or total order sense, neglecting just the first order terms associated with the input variable (terms containing only the selected variable), in other words, all terms associated with the input variable (the first order and all terms with interactions). This approach for the construction of reduced PCE is the core of the proposed sensitivity analysis described in the following sections.

6. Distribution-based global sensitivity analysis using reduced PCE

6.1. Theoretical derivation of sensitivity indices

Although ANOVA represents an effective global sensitivity analysis, it is a second-order method in sense of statistical moments. It is not able to take higher statistical moments into account, and as a result, it may not be able to correctly identify the role of input variables. The solution is a sensitivity method working with a shape of the probability distribution function. The main difference between these methods is the measure of distance between given distributions [44]. In the following paragraphs, there is a brief description of selected measures, which were proposed for distribution-based global sensitivity analysis in literature.

The pilot work on this topic was proposed by Borgonovo [9]. The rationale of Borgonovo sensitivity indices is as follows: let f_Y be the density function of model response obtained with all input parameters \mathbf{X} free to vary according to their probability distribution. If one variable is frozen on specific value x_i , the conditional density function $f_{Y|x_i}$ of the model response given x_i is obtained. Generally, it is possible to assume a fixed vector $\mathbf{x}_{\mathbf{u}}$ at a specific value. The shift $s(\mathbf{X}_{\mathbf{u}})$ between f_Y and $f_{Y|x_{\mathbf{u}}}$ can be interpreted as the influence of $\mathbf{X}_{\mathbf{u}}$ on the shape of distribution of Y . Of course, $f_{Y|x_{\mathbf{u}}}$ is a random variable and thus it is necessary to estimate the expected shift $\mathbb{E}_{\mathbf{x}_{\mathbf{u}}}[s(\mathbf{X}_{\mathbf{u}})]$. Sensitivity indices based on $\mathbb{E}_{\mathbf{x}_{\mathbf{u}}}[s(\mathbf{X}_{\mathbf{u}})]$ take the whole distribution into account in contrast to Sobol indices. The shift $s(\mathbf{X}_{\mathbf{u}})$ can be generally measured by various metrics. Two of them are briefly described in following paragraphs.

According to Borgonovo [9], the expected shift is defined utilizing the following measure:

$$\mathbb{E}_{\mathbf{x}_{\mathbf{u}}}[s(\mathbf{X}_{\mathbf{u}})] = \int_{\mathbb{R}^M} f_{\mathbf{x}_{\mathbf{u}}}(\mathbf{x}_{\mathbf{u}}) \left[\int |f_Y(y) - f_{Y|x_{\mathbf{u}}}(y)| dy \right] d\mathbf{x}_{\mathbf{u}}. \quad (44)$$

In order to ensure that sensitivity indices are in interval $(0, 1)$, Borgonovo proposed the moment-independent sensitivity indices based on the normalized expected shift as $\delta_{\mathbf{u}} = \mathbb{E}_{\mathbf{x}_{\mathbf{u}}}[s(\mathbf{X}_{\mathbf{u}})]/2$. In a later study, Liu and Homma [45] proposed to use CDF instead of PDF.

Further, following the idea of Borgonovo, researchers Gamboa et al. [11] have recently proposed a sensitivity measure based on Cramér-von Mises distance between the CDF $F_Y(t) = P(Y \leq t)$, $t \in \mathbb{R}$ of the mathematical model response and the conditional cumulative distribution function (CCDF) $F_Y^{\mathbf{u}}(t) = P(Y \leq t | \mathbf{X}_{\mathbf{u}})$, $t \in \mathbb{R}$ of Y given $\mathbf{X}_{\mathbf{u}}$. Using the Hoeffding-Sobol decomposition of function, Gamboa et al. derived following Cramér-von Mises indices in a similar manner as Sobol indices:

$$C_{\mathbf{u}} = \frac{\int_{\mathbb{R}} \mathbb{E} \left[(F_Y^{\mathbf{u}}(t) - F_Y(t))^2 \right] dF_Y(t)}{\int_{\mathbb{R}} F_Y(t)(1 - F_Y(t)) dF_Y(t)}, \quad (45)$$

where $s(\mathbf{X}_{\mathbf{u}}) = (F_Y^{\mathbf{u}}(t) - F_Y(t))^2$ is a Cramér-von Mises distance between CDF and CCDF, which is normalized by a denominator analogous to Sobol indices. The obtained indices also have similar properties to Sobol indices, and thus can be easily interpreted.

Both methods presented above are designed for the estimation by a double-loop Monte Carlo approach, which may be computationally demanding or even not feasible for industrial applications. Therefore, the following section is focused on an efficient derivation of distribution-based sensitivity indices directly from PCE using results of the previous sections.

Generally, if one is interested in the quantification of influence of input variable to higher-order moments, it is natural to study the shape of probability distribution, which contains information from all statistical moments. Further, it is assumed that distributions of input variables and QoI can be analytically expressed, and thus Monte Carlo approach is not needed. Sensitivity analysis is then focused on the comparison between two distribution functions (PDF or CDF). There are several distance measures in mathematical literature, e.g. Hellinger distance, Kullback–Leibler distance or energy distance [46].

In this paper, the Gram–Charlier expansion is used as the analytical expression of probability distribution in order to obtain a fast and accurate calculation of sensitivity measure. Of course, one can use any type of measure to quantify the sensitivity indices. Here, the relative entropy of two distributions is compared by the Kullback–Leibler divergence [12]:

$$K_{f_Y}^{\mathbf{u}} = \int_{\mathbb{R}} f_Y^{\text{PCE}}(t) \ln \frac{f_Y^{\text{PCE}}(t)}{f_{\mathbf{u}}^{\text{PCE}}(t)} dt, \quad (46)$$

where the probability distribution function f_Y^{PCE} is obtained analytically by the Gram–Charlier expansion from PCE and $f_{\mathbf{u}}^{\text{PCE}}$ is similarly obtained from the reduced PCE defined in previous section. Note that the reduced PCE $g_{\mathbf{u}}^{\text{PCE}}$ has the identical mean value (the first deterministic coefficient) as the original PCE, and thus it neglects only the uncertainty caused by selected variable. Although the Kullback–Leibler divergence is not a true distance measure, because it is non-symmetric and it does not satisfy the triangular inequality, it is a non-negative measure and thus can be used for the quantification of the difference between any two distributions. Moreover $K_{f_Y}^{\mathbf{u}} = 0$ if and only if $f_Y^{\text{PCE}} = f_{\mathbf{u}}^{\text{PCE}}$ and thus neglecting the uncertainty of $X_{\mathbf{u}}$ has no effect on the probability distribution of the mathematical model.

The numerical stability of Gram–Charlier expansion is higher for CDF, thus the cumulative Kullback–Leibler divergence proposed by Park [47] is recommended for industrial applications:

$$K_{F_Y}^{\mathbf{u}} = \int_l^r F_Y^{\text{PCE}}(t) \ln \frac{F_Y^{\text{PCE}}(t)}{F_{\mathbf{u}}^{\text{PCE}}(t)} dt + \mathbb{E}[Y] - \mathbb{E}[g_{\mathbf{u}}^{\text{PCE}}]. \quad (47)$$

Note that reduced PCE satisfies $\mathbb{E}[Y] = \mathbb{E}[g_{\mathbf{u}}^{PCE}]$ and $\sigma_{g_{\mathbf{u}}^{PCE}}^2 > \sigma_{g_{\mathbf{u}}^{PCE}}^2$, thus it is possible to define the following global boundaries of the integral:

$$l = \inf\{t \in \mathbb{R} : F_{\mathbf{u}}^{PCE}(t) > 0\}, r = \sup\{t \in \mathbb{R} : F_{\mathbf{u}}^{PCE}(t) < 1\}. \quad (48)$$

Since the Kullback–Leibler divergence is a measure of the information lost when $F_{\mathbf{u}}^{PCE}$ is used to approximate F_Y^{PCE} , there is a problem with the proper normalization of indices. However, for the decision making, one might prefer relative values of indices, which provide an easy interpretation. Since $K_{F_Y}^{\mathbf{u}}$ for all possible subsets $\mathbf{u} \subset I = \{1, \dots, M\}$ can be calculated without any additional computational demands, it is proposed here to normalize the indices in order to express a relative measure in the interval $(0, 1)$ as:

$$K_{\mathbf{u}} = \frac{K_{F_Y}^{\mathbf{u}}}{\sum_{\substack{\Delta = (I) \\ \Delta \neq I}} K_{F_Y}^{\Delta}}, \quad (49)$$

where $\Delta = (I)$ is the power set of I , i.e. Δ contains all possible subsets of I of a given order, e.g. if one is interested in the relative first order influence of variables, $\Delta = (I)$ contains all single input variables; if one is interested only in second order influence, $\Delta = (I)$ contains all possible subsets of I containing two variables etc. Of course, it is possible to use this approach for total influence just by a different definition of $\mathcal{A}_{-\mathbf{u}}$ using Eq. 41 or Eq. 42 similarly as in the case of Sobol indices. Note that it is also possible to use directly $K_{F_Y}^{\mathbf{u}}$, if one is interested in influence of higher order interactions and thus the differences between first order and total order indices are important. Unfortunately, $K_{F_Y}^{\mathbf{u}}$ have no meaningful interpretation in contrast to Sobol indices and thus it is recommended to use direct values only for comparative study of higher-order interactions. It is obvious that in practical computation, it is possible to get $K_{\mathbf{u}}$ of the desired order if and only if the PCE contains the terms of the desired order.

6.2. Computational algorithm of the proposed sensitivity analysis

First of all, non-intrusive form of PCE approximating given QoI must be constructed. The algorithm for sparse PCE using LAR as

the best model selection algorithm is employed in numerical examples [35]. Once the PCE approximation of the QoI is available, it is possible to obtain $K_{\mathbf{u}}$ together with $S_{\mathbf{u}}$ by its simple post-processing summarized in Algorithm 1 and graphically represented in Fig. 1. The whole process consists of several tasks presented in previous sections: derivation of the higher statistical moments according to Section 3 (using Eq. 22 and Eq. 24); analytical approximation of CDF using the Gram–Charlier expansion according to Section 4; finally reduced PCEs $g_{\mathbf{u}}^{PCE}$ in the first order and total order sense are built for all input variables and used for the calculation of the proposed distribution-based indices $K_{\mathbf{u}}$ and Sobol indices $S_{\mathbf{u}}^{PCE}$, which are obtained from $g_{\mathbf{u}}^{PCE}$ without additional computational demands. Note that thanks to analytical equations for higher moments and CDF, it is possible to employ this algorithm even for high M and p .

Algorithm 1. Distribution-based sensitivity analysis using PCE

-
- Input:** \mathcal{A}, β
- 1: analytical derivation of μ, σ^2, γ and κ from g^{PCE}
 - 2: construction of F_Y^{PCE} by Gram–Charlier expansion
 - 3: **for all** $\mathbf{X}_{\mathbf{u}}$
 - 4: construction of reduced PCE $g_{\mathbf{u}}^{PCE}$ (neglecting selected terms)
 - 5: analytical derivation of σ^2, γ and κ from $g_{\mathbf{u}}^{PCE}$
 - 6: construction of $F_{\mathbf{u}}^{PCE}$ by Gram–Charlier expansion
 - 7: calculation of $K_{F_Y}^{\mathbf{u}}$ by numerical integration
 - 8: **end for**
 - 9: calculation of $K_{\mathbf{u}}$ by normalization of $K_{F_Y}^{\mathbf{u}}$
- Output:** $K_{\mathbf{u}}$
-

7. Numerical examples

7.1. Illustrative examples

The first, very simple example presents the process of the proposed distribution-based sensitivity analysis and its comparison to

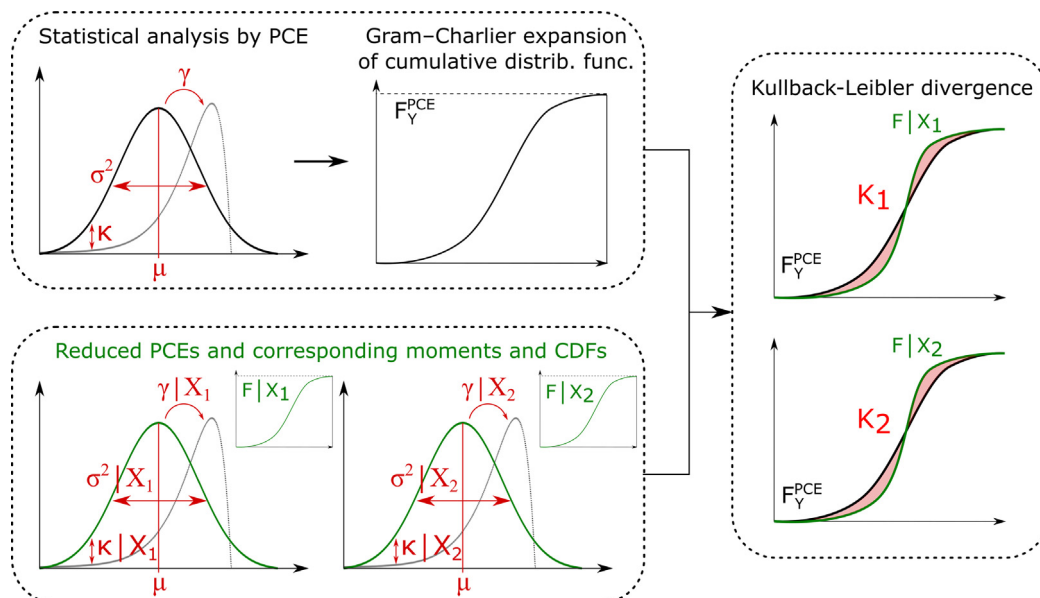


Fig. 1. Graphical summary of the proposed sensitivity algorithm.

Sobol indices. The general motivation of the distribution-based sensitivity measure is the following: let $X_1 \sim \mathcal{L}\mathcal{N}$ and $X_2 \sim \mathcal{U}$ with identical mean $\mu = 10$ and variance $\sigma^2 = 1$ but generally different shape of distribution. The QoI is described by the sum of random variables:

$$Y = X_1 + X_2. \tag{50}$$

Note that mean and variance of QoI are identical to the case with Gaussian input variables. The numerators of Sobol indices can be obtained very easily as

$$\text{Var}(\mathbb{E}[Y | X_1]) = \text{Var}(X_1) = \text{Var}(X_2) = \text{Var}(\mathbb{E}[Y | X_2])$$

In other words, Sobol indices are identical for both input random variables regardless of the distribution of variables. However, the input variables do not share the same probability distribution, and thus should not have generally the same importance. Therefore, it is necessary to use distribution-based sensitivity analysis in order to quantify the difference between X_1 and X_2 .

The numerical results obtained by PCE and the proposed distribution-based sensitivity method are summarized in Table 1, where K_u represents the proposed distribution-based Kullback–Leibler relative indices and S_u^{PCE} Sobol indices derived from PCE. Note that the first order and total indices (marked with the superscript T) are identical due to the simple form of the mathematical model without any interactions.

For the sake of clarity, two elementary examples using same QoI are described in this paragraph. In case of variables with identical type of distribution, the indices are generally identical for K_u as well as for S_u^{PCE} in a relative sense due to a different normalization, i.e. in case of the simple sum of two variables assumed above, with independent identically distributed variables (i.i.d.) random variables one gets $S_i^{\text{PCE}} = K_i = 0.5$. Also note that sensitivity indices are not affected by mean values (by definition of reduced PCE) and thus in case of the simple sum of two variables assumed above with $X_1 \sim \mathcal{N}(100, 0.001)$ and $X_2 \sim \mathcal{N}(0, 1)$ one gets $S_1^{\text{PCE}} = K_1 = 0$ and $S_2^{\text{PCE}} = K_2 = 1$.

Further, let us investigate a modification of the previous QoI. Now it is described by the following non-linear mathematical model

$$Y = X_1 + X_2 + X_1^2 X_2^2, \tag{51}$$

and $X_1 \sim \mathcal{L}\mathcal{N}$ and $X_2 \sim \mathcal{U}$ with identical mean $\mu = 10$ and variance $\sigma^2 = 4$. The whole process of construction of indices is identical to the one in the previous example and the obtained results are summarized in Table 2.

As can be seen, the indices are in compliance with Sobol indices. In this case, Sobol indices would not be able to recognize the difference between input variables if X_1 and X_2 had identical first four statistical moments, but generally different distribution [11].

Table 1
Numerical comparison of Sobol indices and proposed indices.

Variable	S_i^{PCE}	K_i	$S_i^{\text{T,PCE}}$	K_i^T
X_1	0.50	0.46	0.50	0.46
X_2	0.50	0.54	0.50	0.54

Table 2
Difference between Sobol indices and proposed indices for non-linear function.

Variable	S_i^{PCE}	K_i	$S_i^{\text{T,PCE}}$	K_i^T
X_1	0.49	0.59	0.57	0.58
X_2	0.43	0.41	0.51	0.42

The next example taken from [44] is focused on the difference between the total order and first order sense of sensitivity indices. The input variables are i.i.d. standard Gaussian random variables and the mathematical model is

$$Y = 5 + 2X_1 + 7X_2X_3 + 3X_2^2 + 2X_1X_2X_3. \tag{52}$$

Note that the random variable X_3 has only an interaction effect, and thus the first order sensitivity index should be zero. The distribution-based sensitivity analysis is performed identically as in the previous examples assuming the first order reduced PCE and the total order reduced PCE. The numerical results of the proposed indices and Sobol indices are summarized again in Table 3.

As can be seen, the proposed method correctly quantified the influence of the variable X_3 , i.e. the first order effect is zero but it has significant total order effect. Although the results are in compliance with Sobol indices, the absolute values are different due to the normalization of K_u providing an easy interpretation of the results in range $(0, 1)$, which might be beneficial for industrial applications and decision making. Specifically, it can be represented as a percentage expression of information lost due to the neglect of uncertainty of input variable, and such information may be crucial for further reduction of the stochastic model.

The next multiplicative model was selected as a benchmark with known analytical solution [10]:

$$Y = \prod_{i=1}^M X_i^{a_i}. \tag{53}$$

The reference solutions of Borgonovo distribution-based sensitivity indices δ_u are given for the following setting: weights are $\mathbf{a} = 1$, $M = 3$ and X_i are independent random variables having log-normal distribution $\mathcal{L}\mathcal{N}(\lambda_Y, \zeta_Y)$ with mean values $\boldsymbol{\mu} = 1$ and variances $\boldsymbol{\sigma}^2 = [16, 4, 1]$.

Surrogate model represented by PCE was created with maximal polynomial degree $p = 10$ and ED containing 1000 samples generated by LHS. The sparse PCE is created by the algorithm presented in [35] (based on LAR) but generally any algorithm might be employed for efficient construction of PCE. The accuracy of PCE measured by coefficient of determination is $R^2 = 0.98$. Obtained results of the proposed technique, reference values obtained analytically [10] and the reference values normalized by their sum (relative values) are compared in Table 4. As can be seen, the proposed indices lead to an accurate estimation of relative probability-based sensitivity indices. The small difference is caused by an accuracy of PCE. The normalization is necessary due to different sensitivity metrics of reference solution and the proposed technique.

Table 3
Results of sensitivity analysis on example focused on first and total order influence.

Variable	S_i^{PCE}	K_i	$S_i^{\text{T,PCE}}$	K_i^T
X_1	0.053	0.086	0.107	0.001
X_2	0.240	0.914	0.947	0.920
X_3	0.000	0.000	0.707	0.079

Table 4
Comparison of the proposed sensitivity indices to the reference analytical solution.

Variable	Reference δ	Normalized δ	K_i^T
X_1	0.472	0.68	0.68
X_2	0.155	0.22	0.23
X_3	0.071	0.10	0.09

Table 5
Stochastic model of mid-span deflection of a fixed beam.

Parameter	b	h	L	E	F
μ	0.15	0.3	5	30	100
σ	0.0075	0.015	0.05	4.5	20
Units	[m]	[m]	[m]	[GPa]	[kN]

Table 6
Comparison of first four statistical moments obtained by proposed method with analytical solution and experimental design.

Method	μ	σ^2	γ	κ
Analytical	6.69	4.09	0.91	4.46
PCE	6.69	4.08	0.88	4.41
ED (LHS)	6.65	3.41	0.78	4.10

Table 7
Global sensitivity analysis of deflection of a fixed beam.

Variable	S_i^{PCE}	K_i	$S_i^{T,PCE}$	K_i^T
b	0.026	0.004	0.029	0.004
h	0.251	0.185	0.266	0.175
L	0.01	0.001	0.011	0.001
E	0.250	0.177	0.263	0.166
F	0.438	0.633	0.460	0.654

7.2. Engineering analytical example

The next example is represented by a typical engineering problem with a known analytical solution, a maximum deflection of a fixed beam loaded by a single force in mid-span

$$Y = \frac{1}{16} \frac{FL^3}{Ebh^3}, \quad (54)$$

containing 5 lognormally distributed uncorrelated random variables according to Table 5, where b and h represent the width and height of the rectangular cross-section, E is the modulus of elasticity of concrete, F is the loading force and L is the length of the beam.

For this example, it is simple to obtain an analytically lognormal distribution of Y , and thus numerical results can be compared to

analytical results assumed as a reference solution. The product of lognormally distributed variables is a lognormal variable $Y \sim \mathcal{L.N}(\lambda_Y, \zeta_Y)$, where parameters of distribution are obtained as:

$$\lambda_Y = \ln\left(\frac{1}{16}\right) + \lambda_F + 3\lambda_L - \lambda_E - \lambda_b - 3\lambda_h,$$

$$\zeta_Y = \sqrt{\zeta_F^2 + (3\zeta_L)^2 + \zeta_E^2 + \zeta_b^2 + (3\zeta_h)^2}.$$

Using the adaptive algorithm, it was possible to build the PCE with ED containing 100 samples generated by Latin Hypercube Sampling (LHS) [48] and maximal polynomial order $p = 5$. Since the function is not in a polynomial form, the PCE is an approximation of the original model with an accuracy measured by Leave-one-out cross validation $Q^2 = 0.998$. The obtained statistical moments of PCE used for G-C expansion are compared with analytical reference solution and statistical moments of an initial experimental design in Table 6.

As can be seen, 100 samples are enough for the estimation of the first four statistical moments by PCE and it is far more accurate in comparison to simple LHS sampling used for ED. The results of sensitivity analysis derived from PCE are summarized in Table 7. The first order indices as well as total order indices have an identical trend as Sobol indices and even small differences between E and h are identified correctly.

7.3. Engineering applications

7.3.1. Concrete beam ARW1 failing in shear 2D

The first engineering application is a replication of experimental programme by Anderson & Ramirez [49]. In this experiment a beam ARW1 of cross-section 406×406 mm was subjected to a four-point bending test with a shear span $a = 0.91$ m. The beam was designed to fail in shear. The shear reinforcement is composed of double stirrups No. 3 with diameter 9.525 mm. Top longitudinal reinforcement consists of 5 bars No. 6 (diameter 19.05 mm) and bottom reinforcement of 5 bars No. 9 (diameter 28.65 mm). NLFEM was created in Atena Science software focused on non-linear fracture mechanics of concrete structures [50] and the 'Nonlinear Cementitious' material model was used for the concrete material. The geometry of the beam, the reinforcement and results of NLFEM are depicted in Fig. 2.

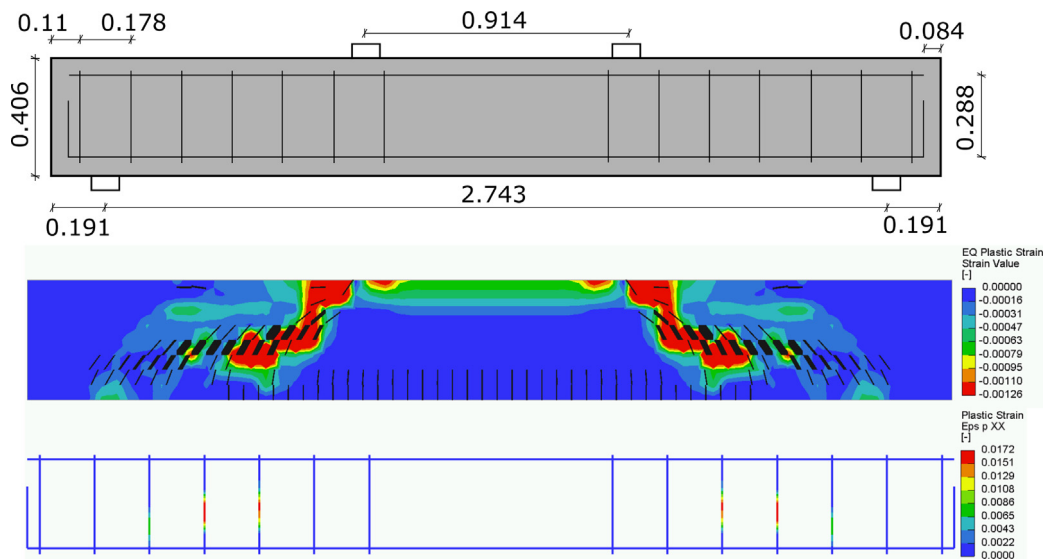


Fig. 2. Geometry in [m] of ARW1 beam and results of NLFEM.

Table 8
Stochastic model and obtained results for the beam ARW1.

Parameter	f_t	f_c	G_f	f_y
μ	2.3	28	176	439
CoV [%]	22	16	22	6.7
Units	[MPa]	[MPa]	[Jm ²]	[MPa]
K_i	0.03	0.84	0.12	0.01
K_i^T	0.05	0.74	0.20	0.01
S_i^{PCE}	0.13	0.52	0.25	0.02
$S_i^{T,PCE}$	0.18	0.52	0.35	0.02

From the obtained results of equivalent plastic strain in concrete and plastic strain in reinforcement (see Fig. 2), it is clear that the beam is failing in shear. Shear failure mechanism represents complex phenomenon and thus it is beneficial to perform sensitivity analysis.

PCE was created using the adaptive algorithm as in previous examples with ED containing 30 samples generated by LHS and maximal polynomial order $p = 8$. The obtained accuracy measured by Leave-one-out cross validation $Q^2 = 0.92$. Note that although such accuracy is not sufficient for a reliability analysis, it is acceptable for a sensitivity analysis.

The stochastic model contains 4 lognormal random variables (parameters of concrete and reinforcement): f_c - compressive strength of concrete, f_y - yield strength of the reinforcement, f_t - tensile strength of concrete, G_f - fracture energy of concrete. The statistical moments of input variables together with obtained results are summarized in Table 8.

From the obtained results, it is clear that a shear mechanism is dominantly affected by an uncertainty of f_c and G_f . Moreover, there are important interactions between f_t and G_f , while f_c has dominant first-order influence. As can be seen, results of the proposed sensitivity indices are supported by results of Sobol indices. These results are in compliance also with theory assuming that the shear mechanism is resisted by concrete struts acting in compression and a development of cracks is governed by G_f and f_t .

7.3.2. Prestressed concrete roof girder 3D

The second analyzed structure is a full-scale LDE7 roof girder produced by Franz Oberndorfer GmbH & Co KG in Austria. Details of NLFEM and stochastic model are described in the following paragraphs. Presented results are a part of the long-term practical research project focused on development of digital twin including laboratory experiments [51,52], stochastic modelling [53], semi-

probabilistic design [54] and surrogate modelling [55,4,56]. The girder is made from concrete C50/60 and is prestressed by 2x8 strands in each web (Cables - St 1570/1770 - F93). Strands are located in following distances from bottom: 70 and 7x40 mm. It has a TT-shaped cross-section and the total length is 30.00 m and the height is 0.50 m at ends and 0.90 in the middle. The reinforcement and geometry of the beam is symmetrical according to middle cross-sectional and longitudinal plane. The load was applied 4.125 m from support above both webs and the ultimate limit state is represented by the critical value of the force applied during the simulation (QoI is the peak of load-deflection diagram). The geometry of the girder, the cross-section and a place of applied load can be seen in Fig. 3.

The '3D Nonlinear Cementitious 2' material model was used for the concrete. The geometry of the beam, supports, and reinforcement was created exactly according to drawings provided by the manufacturer. The '3D Nonlinear Cementitious 2' material model was used for the concrete. The steel reinforcement and prestressing tendons were modeled using 1D elements with a multi-linear stress vs. strain diagram with hardening. Prestressing was applied in the form of initial strain in the tendons. Prestress losses (immediate and long-term) were taken into account according to the fib Model Code 2010. The evaluation of FEM is highly time-consuming, therefore using the surrogate model is necessary to perform the sensitivity analysis.

The stochastic model is based on laboratory experiments and it contains 5 lognormal random variables as can be seen in Table 9: material parameters of concrete as in previous example together with E -Young's modulus of concrete and $I.L.U$ - uncertainty of calculated immediate prestress losses. Moreover in order to obtain realistic results, the correlations among concrete characteristics (summarized in Table 10) were assumed according to results of the previous research [53].

PCE was created using the adaptive algorithm as in previous examples with ED containing 100 samples generated by LHS, maximal polynomial order $p = 8$. Since the function is not in a polynomial form, the PCE is an approximation of the original model with an accuracy measured by Leave-one-out cross validation $Q^2 = 0.98$. Note that input random vector was transformed into uncorrelated standardized space by Nataf transformation in order to use Hermite polynomials as PCE basis. The obtained results of global sensitivity analysis obtained directly from PCE are summarized in Table 11 and graphically interpreted in Fig. 4.

Note that obtained distribution-based sensitivity indices lead to same ranking of input random variables, though K_i amplify the dominance of f_c . This is caused by significant influence of f_c not

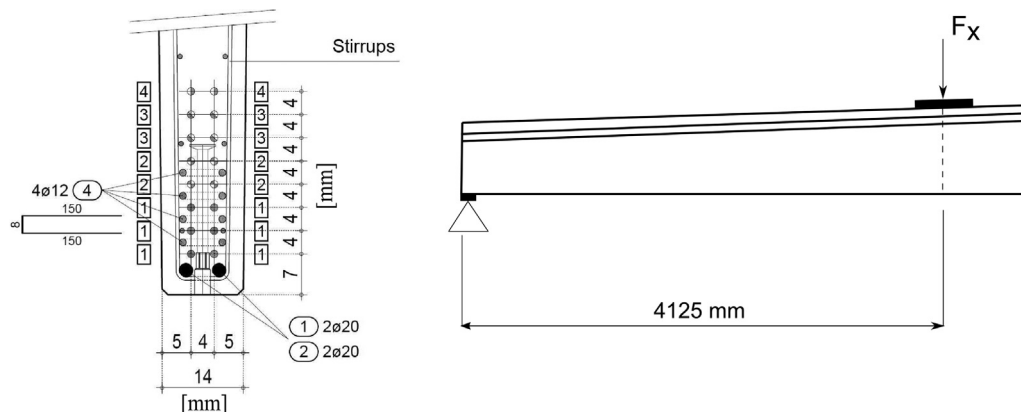


Fig. 3. Geometry of the prestressed roof girder LDE7.

Table 9
Stochastic model of the girder LDE7.

Parameter	$I.L.U$	f_c	E	f_t	G_f
μ	1	77	34.8	3.9	219.8
CoV [%]	10	6.4	10.6	10.6	12.8
Units	[-]	[MPa]	[GPa]	[MPa]	[Jm ²]

Table 10
Correlation matrix of concrete material parameters.

	f_c	E	f_t	G_f
f_c	1	0.8	0.7	0.6
E	0.8	1	0.5	0.5
f_t	0.7	0.5	1	0.8
G_f	0.6	0.5	0.8	1

Table 11
Global sensitivity analysis of the LDE7 girder failing in shear.

Variable	S_i^{PCE}	K_i	$S_i^{T,PCE}$	K_i^T
$I.L.U$	0.18	0.10	0.18	0.05
f_c	0.56	0.83	0.57	0.89
E	0.02	0.01	0.02	0.01
f_t	0.19	0.05	0.19	0.04
G_f	0.05	0.01	0.05	0.01

only to variance of QoI but also to skewness and kurtosis. In order to reduce the uncertainty of resistance in shear of LDE7 and thus increase the design value of resistance, it is crucial to reduce and/or identify the uncertainty of f_c . From obtained results it can be concluded that it is not necessary to include G_f and E into stochastic model, since its influence is negligible.

7.4. ROSA example

The last analytical example is focused on potential of the proposed approach in structural reliability – estimation of ROSA indices. Two of recently proposed ROSA methods were selected for numerical example performed in this paper: contrast indices [20] and sensitivity indices according to Xiao et al. [57]. The failure probability is the key quantity of interest in both cases. The first ROSA method is first-order Contrast index measuring the difference between P_f and conditional $P_f|X_i$. The index is calculated using a formula derived in [21]:

$$C_i = \frac{P_f(1 - P_f) - \mathbb{E}(P_f|X_i(1 - P_f|X_i))}{P_f(1 - P_f)} \quad (55)$$

The formula above gives a better insight into the essence of the calculation of C_i in comparison to its general form available in [20]. The second type is the first order index according to Xiao et al. defined as

$$K_i = \frac{\mathbb{E}(|P_f - P_f|X_i|)}{2P_f} \quad (56)$$

where $|P_f - P_f|X_i|$ measures the absolute difference between unconditional and conditional failure probability. Note that, both selected ROSA methods use mean value of difference $\mathbb{E}(\cdot)$ and thus the practical computations typically need double-loop approach, which can be numerically very demanding in industrial (numerical) applications.

The typical example of structural reliability used for numerical study is taken from [21], where it was calculated by double-loop numerical integration and thus reference numerical solution is available. The mathematical model of structural reliability is in the following fundamental form:

$$Z = R - F, \quad (57)$$

where the resistance of structure $R \sim \mathcal{N}(412.54, 34.132^2)$ is independent of the load action $F \sim \mathcal{N}(\mu_F, 34.132^2)$. The mean value of the load action $\mu_F \in \langle 92.54, 722.54 \rangle$ increases with step

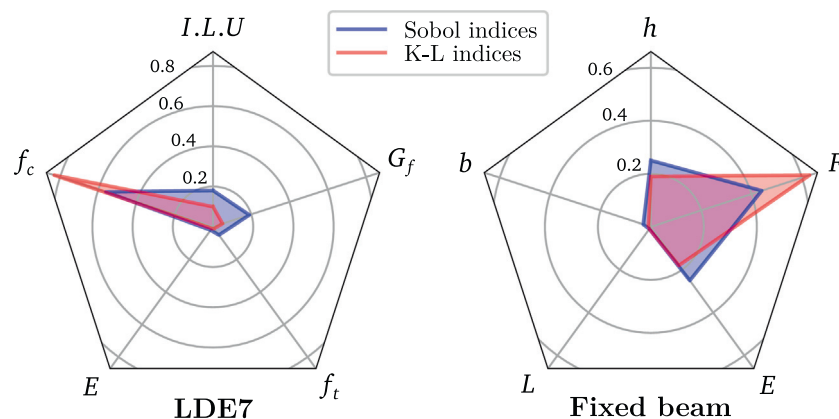


Fig. 4. Obtained Sobol indices (blue) and the proposed Kullback–Leibler indices derived from PCE (red) for fixed beam (right) and roof girder LDE7 (left).

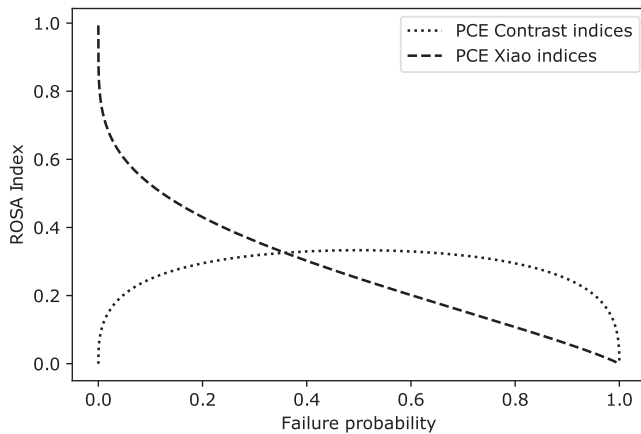


Fig. 5. PCE approximation of ROSA indices dependent on failure probability.

$\Delta_F = 10$. The increasing of μ_f leads to higher failure probability, since the mean value of safety margin μ_z decreases.

The PCE is based on experimental design containing 100 samples generated by LHS in order to uniformly cover the whole design domain. Once the PCE approximation is available, it is possible to derive conditional failure probabilities from reduced PCEs and Gram–Charlier expansions similarly as in previous examples focused on distribution-based sensitivity. Obtained results (depicted in Fig. 5) of first order Contrast indices and ROSA indices according to Xiao et al. are identical to reference solution by double-loop Monte Carlo.

Note that the ROSA indices estimated by PCE are in this case exact and thus reference solutions coincide with PCE estimators in Fig. 5. Although it is clear that the obtained PCE approximation leads to the exact solution due to a simple form of mathematical model and thus obtained perfect accuracy of PCE (which is not realistic in practical examples), it shows the methodology to be correct and it can be generally used as a fast approximation for time-consuming models and errors of estimated ROSA indices are given by accuracy of PCE approximation.

8. Discussion and further work

Numerical results presented in the previous section provide an explanation and an illustration of the proposed distribution-based sensitivity indices derived directly from PCE together with Sobol indices. The proposed method can thus be used as complementary information to the commonly known results of ANOVA, since the distribution-based sensitivity analysis takes the whole distribution into account, which might be crucial in some applications, and its derivation is without any additional computational demands. In this section, some possibilities of the proposed method and solutions of the selected problems beyond the scope of this paper are briefly discussed, in order to generalize the proposed approach for industrial applications.

First of all, there is an important property of the proposed method for a localized sensitivity analysis, i.e. it is possible to study the influence of input random variables in the area of interest (e.g. around the design point). Since the Kullback–Leibler divergence compares two probability distributions in a given interval, it is possible to calculate the sensitivity indices by the same procedure, just with different integral boundaries l_i and r_i covering the area of interest:

$$K_{F_Y}^u = \int_{l_i}^{r_i} F_Y^{PCE}(t) \ln \frac{F_Y^{PCE}(t)}{F_u^{PCE}(t)} dt. \quad (58)$$

Although, it would be possible to define $K_{F_Y}^u$ with switched position of $F_Y^{PCE}(t)$ and $F_u^{PCE}(t)$, it would lead to computational issues since $F_u^{PCE}(t)$ neglects the uncertainty of selected variables, and thus can be significantly lower than original CDF $F_Y^{PCE}(t)$ or even close to zero in the area of interest. Therefore, it is recommended to use expression above for sensitivity analysis.

In case of arbitrary polynomial basis which can not be used for analytical derivation of statistical moments, it is possible to utilize PCE as a response surface to obtain probability density function f_Y^{PCE} of approximated model response. For this purpose, it is necessary to evaluate the PCE with a sufficiently large sample set (around $n_{sim} \approx 10^6$) and estimate the probability density function using e.g. a kernel density estimator:

$$f_Y^{PCE}(y) = \frac{1}{n_{sim} h} \sum_{i=1}^{n_{sim}} K\left(\frac{y - g^{PCE}(\mathbf{x}_i)}{h}\right), \quad (59)$$

where g^{PCE} is a response surface (an approximation of the original mathematical model) represented by PCE, K is the kernel function (e.g. standard normal density) and h is the bandwidth. The cumulative distribution function can be estimated assuming CDF as a kernel function (e.g. standard normal cumulative distribution function). This solution may be highly computationally demanding, since it needs a sufficient number of simulations for the estimation of $F_u^{PCE}(t)$ for all subset of I , using the reduced PCE g_u^{PCE} as a response surface. The rest of the process is identical to the one described in this paper. This approach was already investigated in a previous study [58] of the author of this paper. Identical approach should be used if a distribution of QoI is highly non-Gaussian and thus it is not possible to employ Gram–Charlier expansion for its analytical approximation.

Naturally an accuracy of estimations of higher statistical moments is given by an accuracy of PCE, which is highly affected by statistical sampling. Further work will be focused on comparison of advanced sampling methods (e.g. [59–61]) in context of distribution-based sensitivity analysis. Moreover, recently proposed adaptive sequential sampling developed for correct estimation of variance of QoI [62] will be employed in further studies in order to increase the accuracy and stability of estimated moments.

9. Conclusion

PCE is usually constructed as a surrogate model for further Monte Carlo type analysis, nonetheless, it is beneficial to study the structure of PCE in order to obtain additional information on a given stochastic problem. The novel approach for the distribution-based sensitivity analysis derived directly from PCE, which offers important information about mathematical model, is presented here. It is shown how to derive the reduced PCE neglecting the influence of selected input variables, and how to construct the corresponding probability distribution using derived analytical formula for Gram–Charlier expansion. The influence of an input variable is measured by the Kullback–Leibler divergence, which takes the whole shape of distribution into account. The proposed sensitivity indices are obtained after a relative normalization in order to allow for a simple interpretation in industrial applications and decision making. The numerical results of the proposed method are compared to the commonly known and frequently used Sobol indices. Although the results are in compliance with Sobol indices in most cases, it is shown that Sobol indices are only a second-order method, and thus may fail to identify the correct influence of input variables. It can be seen that the proposed method can yield important complementary information in sensitivity analysis together with Sobol indices obtained without additional computational demands. The results presented herein

could be interesting for theoretical researchers in the area of sensitivity analysis and PCE as well as for engineers working in industry, as the proposed approach presents additional information for sensitivity analysis taking the whole distribution into account.

List of Abbreviations and Selected Symbols

polynomial chaos expansion (PCE); quantity of interest (QoI); non-linear finite element model (NLFEM); analysis of variance (ANOVA); reliability-oriented sensitivity analysis (ROSA); ordinary least squares (OLS); experimental design (ED); least angle regression (LAR); probability distribution function (PDF); cumulative distribution function (CDF); Gram–Charlier expansion (G-C); conditional cumulative distribution function (CCDF); independent identically distributed variables (i.i.d.); latin hypercube sampling (LHS).

input random vector \mathbf{X} ; QoI (model response) Y ; mathematical model $g(\mathbf{X})$; multi-index α ; deterministic coefficients of PCE β ; multivariate orthogonal polynomials in PCE Ψ ; number of terms in PCE P ; set of basis functions in PCE \mathcal{A} ; number of input random variables M ; maximum total degree of polynomials p ; mean value of QoI μ_Y ; variance of QoI σ_Y^2 ; skewness of QoI γ_Y ; kurtosis of QoI κ_Y ; PDF of QoI f_Y ; CDF of QoI F_Y ; conditional cumulative distribution function of QoI F_Y^u ; Kullback–Leibler divergence $K_{F_Y}^u$; normalized Kullback–Leibler indices K_u ; Sobol indices derived from PCE S_u^{PCE} ; Borgonovo sensitivity indices δ .

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The author would like to express his thanks for the support provided by the Czech Science Foundation Project No. 20-01734S.

References

- Wiener N. The Homogeneous Chaos. *Am J Mathe* 1938;60(4):897–936. <https://doi.org/10.2307/2371268>.
- Sudret B, Berveiller M, Lemaire M. A stochastic finite element procedure for moment and reliability analysis. *Eur J Comput Mech* 2006;15(7–8):825–66. <https://doi.org/10.3166/remn.15.825-866>.
- Ghanem RG, Spanos PD. *Stochastic Finite Elements: A Spectral Approach*. Berlin, Heidelberg: Springer-Verlag; 1991.
- Novak L, Novak D. Surrogate modelling in the stochastic analysis of concrete girders failing in shear. In: *Proc. of the Fib Symposium 2019: Concrete - Innovations in Materials, Design and Structures*; 2019. p. 1741–1747, ISBN: 9782940643004.
- Borgonovo E, Plischke E. Sensitivity analysis: A review of recent advances. *Eur J Oper Res* 2016;248(3):869–87. <https://doi.org/10.1016/j.ejor.2015.06.032>.
- Chastaing G, Gamboa F, Prieur C. Generalized Hoeffding–Sobol decomposition for dependent variables - application to sensitivity analysis. *Electronic J Stat* 2012;6:2420–48. <https://doi.org/10.1214/12-EJS749>.
- Sobol I. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathe Comput Simul* 2001;55(1):271–80. [https://doi.org/10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6).
- Sudret B. Global sensitivity analysis using polynomial chaos expansions. *Reliab Eng Syst Saf* 2008;93(7):964–79. <https://doi.org/10.1016/j.ress.2007.04.002>.
- Borgonovo E. A new uncertainty importance measure. *Reliab Eng Syst Saf* 2007;92(6):771–84. <https://doi.org/10.1016/j.ress.2006.04.015>.
- Borgonovo E, Castangs W, Tarantola S. Moment independent importance measures: New results and analytical test cases. *Risk Anal* 2011;31(3):404–28. <https://doi.org/10.1111/j.1539-6924.2010.01519.x>.
- Gamboa F, Klein T, Lagnoux A. Sensitivity analysis based on cramer–von mises distance. *SIAM/ASA J Uncertainty Quantif* 2018;6(2):522–48. <https://doi.org/10.1137/15M1025621>.
- Kullback S, Leibler RA. On Information and Sufficiency. *Ann Math Stat* 1951;22(1):79–86. <https://doi.org/10.1214/aoms/1177729694>.
- Park CK, Ahn K-I. A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment. *Reliab Eng Syst Saf* 1994;46(3):253–61. [https://doi.org/10.1016/0951-8320\(94\)90119-8](https://doi.org/10.1016/0951-8320(94)90119-8).
- Liu H, Chen W, Sudjianto A. Relative Entropy Based Method for Probabilistic Sensitivity Analysis in Engineering Design. *J Mech Des* 2005;128(2):326–36. <https://doi.org/10.1115/1.2159025>.
- Teixeira R, O'Connor A, Nogal M. Probabilistic sensitivity analysis of offshore wind turbines using a transformed kullback-leibler divergence. *Struct Saf* 2019;81:101860. <https://doi.org/10.1016/j.strusafe.2019.03.007>.
- Madsen HO. Omission sensitivity factors. *Struct Saf* 1988;5(1):35–45. [https://doi.org/10.1016/0167-4730\(88\)90004-5](https://doi.org/10.1016/0167-4730(88)90004-5).
- Luyi L, Zhenzhou L, Jun F, Bintuan W. Moment-independent importance measure of basic variable and its state dependent parameter solution. *Struct Saf* 2012;38:40–7. <https://doi.org/10.1016/j.strusafe.2012.04.001>.
- Wei P, Lu Z, Hao W, Feng J, Wang B. Efficient sampling methods for global reliability sensitivity analysis. *Comput Phys Commun* 2012;183(8):1728–43. <https://doi.org/10.1016/j.cpc.2012.03.014>.
- Kala Z. Global sensitivity analysis of reliability of structural bridge system. *Eng Struct* 2019;194:36–45. <https://doi.org/10.1016/j.engstruct.2019.05.045>.
- Fort J-C, Klein T, Rachdi N. New sensitivity analysis subordinated to a contrast. *Commun Stat - Theory Methods* 2016;45(15):4349–64. <https://doi.org/10.1080/03610926.2014.901369>.
- Kala Z. Sensitivity analysis in probabilistic structural design: A comparison of selected techniques. *Sustainability* 2020;12(11). <https://doi.org/10.3390/su12114788>.
- Kala Z. New importance measures based on failure probability in global sensitivity analysis of reliability. *Mathematics* 2021;9(19). <https://doi.org/10.3390/math9192425>.
- Ling C, Lu Z, Cheng K, Sun B. An efficient method for estimating global reliability sensitivity indices. *Probab Eng Mech* 2019;56:35–49. <https://doi.org/10.1016/j.probengmech.2019.04.003>.
- Xiu D, Karniadakis GE. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM J Sci Comput* 2002;24(2):619–44. <https://doi.org/10.1137/S1064827501387826>.
- Jakeman JD, Franzelin F, Narayan A, Eldred M, Plüfger D. Polynomial chaos expansions for dependent random variables. *Comput Methods Appl Mech Eng* 2019;351:643–66. <https://doi.org/10.1016/j.cma.2019.03.049>.
- Kiureghian AD, Liu P. Structural reliability under incomplete probability information. *J Eng Mech* 1986;112(1):85–104. [https://doi.org/10.1061//ASCE/0733-9399\(1986\)112:1\(85\)](https://doi.org/10.1061//ASCE/0733-9399(1986)112:1(85)).
- Lebrun R, Dufloy A. A generalization of the Nataf transformation to distributions with elliptical copula. *Probab Eng Mech* 2009;24(2):172–8. <https://doi.org/10.1016/j.probengmech.2008.05.001>.
- Lebrun R, Dufloy A. Do Rosenblatt and Nataf isoprobabilistic transformations really differ? *Probab Eng Mech* 2009;24(4):577–84. <https://doi.org/10.1016/j.probengmech.2009.04.006>.
- Lebrun R, Dufloy A. An innovating analysis of the Nataf transformation from the copula viewpoint. *Probab Eng Mech* 2009;24(3):312–20. <https://doi.org/10.1016/j.probengmech.2008.08.001>.
- Torre E, Marelli S, Embrechts P, Sudret B. A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas. *Probab Eng Mech* 2019;55:1–16. <https://doi.org/10.1016/j.probengmech.2018.08.001>.
- Blatman G, Sudret B. Adaptive sparse polynomial chaos expansion based on least angle regression. *J Comput Phys* 2011;230(6):2345–67. <https://doi.org/10.1016/j.jcp.2010.12.021>.
- Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. *Ann Stat* 2004;32(2):407–99. <https://doi.org/10.1214/009053604000000067>.
- Lüthen N, Marelli S, Sudret B. Sparse polynomial chaos expansions: Literature survey and benchmark. *SIAM/ASA J Uncert Quantif* 2021;9(2):593–649. <https://doi.org/10.1137/20M1315774>.
- Olivier A, Giovanis DG, Aakash B, Chauhan M, Vandanapu L, Shields MD. UQpy: A general purpose python package and development environment for uncertainty quantification. *J Comput Sci* 2020;47:101204. <https://doi.org/10.1016/j.jocs.2020.101204>.
- Novak L, Novak D. Polynomial chaos expansion for surrogate modelling: Theory and software. *Beton- und Stahlbetonbau* 2018;113(S2):27–32. <https://doi.org/10.1002/best.201800048>.
- Marelli S, Sudret B. UQLab: A framework for uncertainty quantification in Matlab. In: *Vulnerability, Uncertainty, and Risk*; 2014. p. 2554–2563. <https://doi.org/10.1061/9780784413609.257>.
- Patelli E, Tolo S, George-Williams H, Sadeghi J, Rocchetta R, de Angelis M, Broggi M. Opencossan 2.0: an efficient computational toolbox for risk, reliability and resilience analysis. In: *Proceedings of the joint ICVRAM ISUMA UNCERTAINTIES conference*; 2018. p. 1–8.
- Gasper G. Linearization of the product of jacobi polynomials. i. *Can J Mathe* 1970;22(1):171–5. <https://doi.org/10.4153/CJM-1970-020-2>.
- Savin E, Faverjon B. Computation of higher-order moments of generalized polynomial chaos expansions. *Int J Numer Meth Eng* 2017;111(12):1192–200. <https://doi.org/10.1002/nme.5505>.
- Chaggara H, Koepf W. On linearization and connection coefficients for generalized hermite polynomials. *J Comput Appl Math* 2011;236(1):65–73. <https://doi.org/10.1016/j.cam.2011.03.010>.
- Adams JC. On the expression of the product of any two legendre's coefficients by means of a series of legendre's coefficients. *Proc Roy Soc London* 1878;27(185–189):63–71. <https://doi.org/10.1098/rpsl.1878.0016>.

- [42] Sullivan TJ. Introduction to Uncertainty Quantification, Vol. 63 of Texts in Applied Mathematics, Springer; 2015. <https://doi.org/10.1007/978-3-319-23395-6>.
- [43] Barton DE, Dennis KE. The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal. *Biometrika* 1952;39(3/4):425–7. <https://doi.org/10.2307/2334037>.
- [44] Greegar G, Manohar C. Global response sensitivity analysis using probability distance measures and generalization of sobol's analysis. *Probab Eng Mech* 2015;41:21–33. <https://doi.org/10.1016/j.probenmech.2015.04.003>.
- [45] Liu Q, Homma T. A new importance measure for sensitivity analysis. *J Nucl Sci Technol* 2010;47(1):53–61. <https://doi.org/10.1080/18811248.2010.9711927>.
- [46] Ullah A. Entropy, divergence and distance measures with econometric applications. *J Stat Plann Inference* 1996;49(1):137–62. [https://doi.org/10.1016/0378-3758\(95\)00034-8](https://doi.org/10.1016/0378-3758(95)00034-8).
- [47] Park S, Rao M, Shin DW. On cumulative residual kullback–leibler information. *Stat Probab Lett* 2012;82(11):2025–32. <https://doi.org/10.1016/j.spl.2012.06.015>.
- [48] McKay MD, Conover WJ, Beckman RJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 1979;21:239–45. <https://doi.org/10.1080/00401706.1979.10489755>.
- [49] Anderson NS, Ramirez JA. Detailing of stirrup reinforcement. *Acı Struct J* 1989;86:507–15.
- [50] Cervenka J, Papanikolaou VK. Three dimensional combined fracture-plastic material model for concrete. *Int J Plast* 2008;24(12):2192–220. <https://doi.org/10.1016/j.ijplas.2008.01.004>.
- [51] Strauss A, Zimmermann T, Lehký D, Novák D, Keršner Z. Stochastic fracture-mechanical parameters for the performance-based design of concrete structures. *Struct Concr* 2014;15(3):380–94. <https://doi.org/10.1002/suco.201300077>.
- [52] Strauss A, Krug B, Slowik O, Novak D. Combined shear and flexure performance of prestressing concrete t-shaped beams: Experiment and deterministic modeling. *Struct Concr* 2018;19(1):16–35. <https://doi.org/10.1002/suco.201700079>.
- [53] Slowik O, Novák D, Novák L, Strauss A. Stochastic modelling and assessment of long-span precast prestressed concrete elements failing in shear. *Eng Struct* 2021;228:111500. <https://doi.org/10.1016/j.engstruct.2020.111500>.
- [54] Novák, L. Novák D, Slowik O, Strauss A. Prestressed concrete roof girders: Part III – semi-probabilistic design. In: Proceedings of the Sixth International Symposium on Life-Cycle Civil Engineering (IALCCE 2018), CRC Press, Taylor and Francis Group; 2018. p. 510–517, ISBN: 978-113862633-1.
- [55] Pan L, Novák L, Lehký D, Novák D, Cao M. Neural network ensemble-based sensitivity analysis in structural engineering: Comparison of selected methods and the influence of statistical correlation. *Comput Struct* 2021;242:106376. <https://doi.org/10.1016/j.compstruc.2020.106376>.
- [56] Lehký D, Novák D, Novák L, Šomodíková M. Prestressed concrete roof girders: Part II – surrogate modeling and sensitivity analysis. In: Proceedings of the Sixth International Symposium on Life-Cycle Civil Engineering (IALCCE 2018), CRC Press, Taylor and Francis Group; 2018. p. 2649–2698, ISBN: 978-113862633-1.
- [57] Xiao S, Lu Z. Structural reliability sensitivity analysis based on classification of model output. *Aerosp Sci Technol* 2017;71:52–61. <https://doi.org/10.1016/j.ast.2017.09.009>.
- [58] Novak L, Novak D. On the possibility of utilizing Wiener-Hermite polynomial chaos expansion for global sensitivity analysis based on Cramer-von Mises distance. In: Proceedings of 2019 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, QR2MSE 2019; 2019. p. 646–654. <https://doi.org/10.1109/QR2MSE46217.2019.9021206>.
- [59] Vořechovský M, Mašek J, Eliáš J. Distance-based optimal sampling in a hypercube: Analogies to N-body systems. *Adv Eng Softw* 2019;137:102709. <https://doi.org/10.1016/j.advengsoft.2019.102709>.
- [60] Cohen A, Migliorati G. Optimal weighted least-squares methods. *SMAI J Comput Mathe* 2017;3:181–203. <https://doi.org/10.5802/smai-icm.24>.
- [61] Migliorati G. Adaptive approximation by optimal weighted least-squares methods. *SIAM J Num Anal* 2019;57(5):2217–45. <https://doi.org/10.1137/18M1198387>.
- [62] Novák L, Vorechovský M, Sadílek V, Shields MD. Variance-based adaptive sequential sampling for polynomial chaos expansion. *Comput Methods Appl Mech Eng* 2021;386:114105. <https://doi.org/10.1016/j.cma.2021.114105>.

3.4 Physics-Informed Polynomial Chaos Expansions

L. Novák, H. Sharma, M. D. Shields, Physics-Informed Polynomial Chaos Expansions, [arXiv:2309.01697](https://arxiv.org/abs/2309.01697)

DOI: 10.48550/arXiv.2309.01697

[Note: Accepted for publication in the Journal of Computational Physics on 08.03.2024 (WoS-AIS: D1)]

Description

The paper presents a novel approach for the construction of non-intrusive PCE referenced as physics-informed PCE. The proposed approach is motivated by the recently proposed physics-informed neural networks and generally a framework of physics-informed machine learning. The proposed method employs constrained least squares with Karush-Kuhn-Tucker matrix in order to constrain a PCE approximation to obey the given physical constraints and boundary conditions. In this pilot work, constraints are limited to equality constraints in the form of stochastic differential equations. A significant benefit of the proposed method referenced as physics-constrained polynomial chaos (PC²) is its non-intrusive regression-based nature allowing for the utilization of various existing methods for an adaptive construction of sparse PCE (including the previous work of the author). The proposed method is presented in several numerical examples including deterministic and stochastic differential equations. The most significant benefit of the proposed PC² lies in its generality and simplicity allowing for an easy implementation into various existing algorithms and software packages. The paper is currently under review (29.2.2024) and is publicly available on arxiv in preprint form, which already has 3 citations.

Role of the author

Percentage of contribution: 60%

Lukáš Novák is the main author of this paper responsible for the concept, the methodology, and the numerical results of the presented research. The theoretical concept was created in collaboration with Michael D. Shields, and some numerical aspects of the optimization process were investigated by Himanshu Sharma.

Physics-Informed Polynomial Chaos Expansions

Lukáš Novák

Brno University of Technology, Brno, Czech Republic

Himanshu Sharma, Michael D. Shields*

Johns Hopkins University, Baltimore, USA

Abstract

Developing surrogate models for costly mathematical models representing physical systems is challenging since it is typically not possible to generate large training data sets, i.e. to create a large experimental design. In such cases, it can be beneficial to constrain the surrogate approximation to adhere to the known physics of the model. This paper presents a novel methodology for the construction of physics-informed polynomial chaos expansions (PCE) that combines the conventional experimental design with additional constraints from the physics of the model represented by a set of differential equations and specified boundary conditions. A computationally efficient means of constructing physically constrained PCEs, termed PC², is proposed and compared to the standard sparse PCE. Algorithms are presented for both full-order and sparse PC² expansions and an iterative approach is proposed for addressing nonlinear differential equations. It is shown that the proposed algorithms lead to superior approximation accuracy and do not add significant computational burden over conventional PCE. Although the main purpose of the proposed method lies in combining training data and physical constraints, we show that the PC² can also be constructed from differential equations and boundary conditions alone without requiring model evaluations. We further show that the constrained PCEs can be easily applied for uncertainty quantification through analytical post-processing of a reduced PCE by conditioning on the deterministic space-time variables. Several deterministic examples of increasing complexity are provided and the proposed method is demonstrated for uncertainty quantification.

Keywords: Polynomial Chaos Expansion, Physical Constraints, Surrogate modelling, Uncertainty Quantification, Physics-informed Machine Learning

1. Introduction

Mathematical models of real-life physical systems are typically highly computationally demanding and contain various uncertain variables. It is therefore necessary to develop surrogate models as computationally cheap approximations to perform *uncertainty quantification* (UQ), optimization, parametric studies, and other tasks. These surrogate models treat the original model as a black-box and are trained from several deterministic simulations at given data points in the design domain. Their practical use requires a sufficient number of data points covering the space of input variables, which can be prohibitively expensive. To reduce training data demands and improve approximation accuracy, it can be beneficial to incorporate additional constraints, e.g., known physical principles, in the training process to ensure realistic and physically meaningful surrogate model behavior. Therefore, in recent years there has been considerable interest in developing machine learned surrogate models capable of satisfying physical constraints – a field broadly referred to as *physics-informed machine learning* [1].

*Corresponding author

Email addresses: novak.l@fce.vutbr.cz (Lukáš Novák), {hsharm10, michael.shields}@jhu.edu (Himanshu Sharma, Michael D. Shields)

Much of the work in physics-informed machine learning has focused on neural network models. Although the first studies on physics-constrained neural networks emerged in the 1990s [2, 3, 4], this topic has garnered significant attention by the hugely popular physics-informed neural networks (PINNs) [5] and the related physics-informed neural operators [6]. Numerous variants of deep neural networks with physical constraints are now available, including physics-informed autoencoders [7], the Deep Ritz Method [8], and the Deep Galerkin Method [9] to name just a few. In fact, the recent developments and applications of PINNs and related neural networks are so numerous we cannot possibly cover them here. But neural networks are not the only machine learning models that benefit from physical constraints. Recent studies have also incorporated physical constraints into Gaussian process regression models [10, 11], which have the benefit of providing a natural measure of uncertainty. These physically constrained Gaussian processes have been applied for UQ of highly complex physical systems [12].

This paper focuses on imposing constraints on Polynomial Chaos Expansions (PCE) [13], which we treat here as a machine learning regression method [14, 15]. Our interest in PCEs stems from their usefulness for UQ tasks that include moment estimation and sensitivity index computation [16], which derive from their orthogonality properties with respect to the probability measures of the input variables, as well as their relative computational efficiency in training. Although PCE surrogate models have distinct benefits for UQ [17, 18], they suffer from the *curse of dimensionality*, since the number of PCE terms grows rapidly with both dimension and maximum polynomial order. Naturally, this significantly affects the number of data points needed for regression-based training. As a result, recent research has focused on exploiting the model to minimize the size of the basis set and maximize the information obtained from each data point. This involves identifying an optimally sparse basis using adaptive algorithms [19], determining the optimal positions of data points for regression [20, 21, 22, 23], and constructing various types of localized surrogates consisting of many low-order PCEs [24, 25, 26]. In addition to the information obtained directly from model evaluations, there are often specific physical characteristics and constraints derived from the nature of the model or quantity of interest (QoI) that should be incorporated into the PCE. We propose a novel and computationally efficient method to incorporate known physics in the form of ordinary differential equations (ODEs) and partial differential equations (PDEs) and their boundary conditions (Dirichlet, Neumann, or mixed) in the PCE training, referred to as a physics constrained polynomial chaos (PC²) expansion. We enforce the constraints at discrete points in the input domain, referred to as virtual points, to solve a constrained least squares optimization problem for the PC² coefficients. This approach can be generalized to incorporate any form of equality-type constraints, where we specifically explore equality constraints that derive from ODEs/PDEs. Extension of the PC² method to more general constraints (e.g. inequality constraints) requires the use of more advanced optimizers and is the topic of a future work.

Incorporating known physics and boundary conditions into the construction of a PCE approximation improves generalization and typically results in higher prediction accuracy, especially in regions of the input space containing an insufficient number of training points. In the proposed PC² framework, we can generally adopt any suitable optimization technique to solve the constrained optimization problem. However, as the number of PC² coefficients and virtual points increases, certain numerical optimization techniques may suffer from convergence issues, or their computational cost could be significant, which restricts their use to low or moderate dimensional problems. In this paper, we focus specifically on equality-type constraints with the objective of minimizing the sum of squared residuals (SSR), and thus it is possible to use the well-known method of Lagrange multipliers. The Karush-Kuhn-Tucker (KKT) stationarity condition yields a system of equations that can be solved directly for the PC² coefficients. This approach is a natural extension of the ordinary least squares (OLS) approach, as used in standard PCE [27], to incorporate known equality-type constraints in a computationally efficient manner. We further show that the approach can be combined with sparse regression methods, specifically Least Angle Regression (LAR) [28], to efficiently reduce the basis set. We then apply the method to several benchmark ODEs and PDEs to illustrate its performance. Finally, we demonstrate how it can be used for general UQ purposes.

2. Non-intrusive Polynomial Chaos Expansion

Assume a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where Ω is an event space, \mathcal{F} is a σ -algebra on Ω and \mathcal{P} is a probability measure on \mathcal{F} . If the Doob-Dynkin lemma is satisfied and the input variable of a mathematical model, $\mathcal{Y} = u(\mathcal{X})$, is a random variable $\mathcal{X}(\omega)$, $\omega \in \Omega$, the model response $\mathcal{Y}(\omega)$ is also a random variable. Assuming that \mathcal{Y} has

finite variance, PCE represents the output variable \mathcal{Y} as a function of another random variable, ξ , called the *germ* with a known distribution as

$$\mathcal{Y} = u(\mathcal{X}) = u^{\text{PCE}}(\xi), \quad (1)$$

and represents the function $u(\mathcal{X})$ by an infinite expansion of polynomials that are orthogonal with respect to the probability density, $p_\xi(\xi)$, of the germ. These polynomials form a basis of the Hilbert space $L^2(\Omega, \mathcal{F}, \mathcal{P})$ of all real-valued random variables of finite variance, where \mathcal{P} takes over the meaning of the probability distribution. The orthogonality condition is given by the inner product on $L^2(\Omega, \mathcal{F}, \mathcal{P})$ defined for any two functions ψ_j and ψ_k with respect to the weight function $p_\xi(\xi)$ as:

$$\langle \psi_j, \psi_k \rangle = \int \psi_j(\xi) \psi_k(\xi) p_\xi(\xi) d\xi = 0 \quad \forall j \neq k. \quad (2)$$

This means that there are specific orthogonal polynomials associated with the corresponding distribution of the germ. Orthogonal polynomials corresponding to common distributions can be chosen according to the Wiener-Askey scheme [29], or numerically constructed for non-standard distributions, e.g. for non-Wiener-Askey distributions, dependent random variables [30, 31, 32], or moment-based/data-driven arbitrary PCEs [33, 34]. For further processing, it is beneficial to use orthonormal polynomials, where the inner product of the polynomials is equal to the Kronecker delta δ_{jk} , i.e. $\delta_{jk} = 1$ if $j = k$, and $\delta_{jk} = 0$ otherwise.

In the case of \mathcal{X} and ξ being vectors containing M independent random variables, the polynomial basis $\Psi(\xi)$ is multivariate and it is built up as a tensor product of univariate orthonormal polynomials, i.e.

$$\Psi_{\alpha}(\xi) = \prod_{i=1}^M \psi_{\alpha_i}(\xi_i), \quad (3)$$

where $\alpha \in \mathbb{N}^M$ is a set of integers called the *multi-index* reflecting polynomial degrees associated to each ξ_i . The quantity of interest (QoI), i.e. the response of the model $\mathcal{Y} = u(\mathcal{X})$, can then be represented as [35]

$$\mathcal{Y} = u(\mathcal{X}) = \sum_{\alpha \in \mathbb{N}^M} \beta_{\alpha} \Psi_{\alpha}(\xi), \quad (4)$$

where β_{α} are deterministic coefficients and Ψ_{α} are multivariate orthonormal polynomials.

For practical computation, the PCE expressed in Eq. (4) must be truncated to a finite number of terms, P . One can generally choose any truncation rule (e.g. tensor product of polynomials up to the selected order p), but the most common truncation is achieved by retaining only terms whose total degree $|\alpha|$ is less than or equal to a given p , in which case the truncated set of PCE terms is defined as

$$\mathcal{A}^{M,p} = \left\{ \alpha \in \mathbb{N}^M : |\alpha| = \sum_{i=1}^M \alpha_i \leq p \right\}. \quad (5)$$

The cardinality of the truncated *index set* $\mathcal{A}^{M,p}$ is given by

$$\text{card } \mathcal{A}^{M,p} = \frac{(M+p)!}{M! p!} \equiv P. \quad (6)$$

When the PCE is truncated to a finite number of terms, there is an error ε in the approximation such that

$$\mathcal{Y} = u(\mathcal{X}) = \sum_{\alpha \in \mathcal{A}} \beta_{\alpha} \Psi_{\alpha}(\xi) + \varepsilon = \mathcal{Y}^{\text{PCE}} + \varepsilon.$$

From a statistical point of view, the PCE is a simple linear regression model with intercept. Therefore, it is possible to use *ordinary least squares* (OLS) regression to solve for the coefficients β that minimize the error ε . To solve this regression problem for β , we first generate n_{sim} realizations of the input random vector \mathcal{X} , denoted \mathbf{X} , and

compute the corresponding results of the model Y , together called the experimental design (ED) or training data set. Then, the vector of P deterministic coefficients $\boldsymbol{\beta}$ can be determined by OLS as

$$\boldsymbol{\beta} = (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T Y, \quad (7)$$

where $\boldsymbol{\Psi}$ is the data matrix:

$$\boldsymbol{\Psi} = \{\Psi_{ij} = \Psi_j(\boldsymbol{\xi}^{(i)}), i = 1, \dots, n_{\text{sim}}, j = 0, \dots, P-1\}. \quad (8)$$

Note, however, that Eq. (7) is not typically used in practice due to ill-conditioning. Instead, more stable numerical methods, such as singular value decomposition, will usually be employed. Moreover, it is clear from Eq. (6) that P is strongly dependent on the number of input random variables M and the maximum total degree of polynomials p . Considering that estimating $\boldsymbol{\beta}$ by OLS requires at least $\mathcal{O}(P \ln(P))$ samples for a stable solution [36, 37], the problem can become computationally demanding for large or strongly non-linear models. To reduce computational expense, one can use advanced model selection algorithms such as LAR [38, 28] to find an optimal set of PCE terms, and thus reduce the number of samples needed to compute the unknown coefficients. Further reduction can be obtained by incorporating additional physical constraints.

3. Physically Constrained Polynomial Chaos Expansion (PC²) – Deterministic Formulation

In this section, we propose a novel approach to perform PCE regression with known physical constraints. The approach, referred to as Physically Constrained Polynomial Chaos (PC²), expands the classes of regression models that can obey physical constraints to include PCE, which is widely used for UQ (Section 4). Yet, a physics-informed PCE may be an attractive option even for deterministic problems because, for sufficiently smooth problems, it will produce an accurate solution with fewer required training points. Combined with computationally efficient optimization strategies, this may make it less costly to train than other methods, such as PINNs. We therefore begin by formulating the constrained regression problem for the solution of deterministic PDEs.

Consider the general partial differential equation given by

$$\begin{aligned} \mathcal{L}(x, t; u(x, t)) &= f(x, t), \quad \forall x \in \mathcal{D}, t \in \mathcal{T} \\ \mathcal{B}(x, t; u(x, t)) &= g(x, t), \quad \forall x \in \partial\mathcal{D}, t \in \mathcal{T} \end{aligned} \quad (9)$$

where $\mathcal{T} \subset \mathbb{R}$, $\mathcal{D} \subset \mathbb{R}^3$ with boundary $\partial\mathcal{D}$, \mathcal{L} is a differential operator with boundary operator \mathcal{B} , $u(\cdot)$ is the response of the system, and f, g are external forces/source terms. We aim to solve the PCE regression problem described above as constrained by the general Eq. (9). That is, we define the objective function by

$$\begin{aligned} \mathcal{M}(\boldsymbol{\beta}) &= \min_{\boldsymbol{\beta}} \sum_{j=1}^{n_{\text{sim}}} [Y^j - u^{\text{PCE}}(\boldsymbol{x}^j, t^j)]^2 = \min_{\boldsymbol{\beta}} \|\boldsymbol{Y} - \boldsymbol{\Psi}\boldsymbol{\beta}\|^2 \\ \text{s.t. } \mathcal{L}(\boldsymbol{x}_V, t_V; u(\boldsymbol{x}_V, t_V)) &= f(\boldsymbol{x}_V, t_V), \\ \mathcal{B}(\boldsymbol{x}_{\text{BC}}, t_{\text{BC}}; u(\boldsymbol{x}_{\text{BC}}, t_{\text{BC}})) &= g(\boldsymbol{x}_{\text{BC}}, t_{\text{BC}}), \end{aligned} \quad (10)$$

where we define three discrete sets of samples of size n_{sim} , n_{BC} , and n_V . Respectively, these define the size of the experimental design (n_{sim}) corresponding to (in the deterministic setting) the number of points in space and time at which the PDE solution is obtained through model evaluation, the number of boundary points (n_{BC}) at which only the boundary conditions are enforced on the PCE model, and the number of virtual points (n_V) at which the PCE approximation is constrained to satisfy the PDE. Note that, in general, there is no relationship between n_{sim} , n_{BC} , and n_V – all of which may be assigned arbitrarily as long as $P \leq n_{\text{sim}} + n_{\text{BC}} + n_V$. Although (x, t) is a deterministic vector, for the PC² regression here it is assumed to be a random vector with uniform distribution \mathcal{U} . This is important primarily because it dictates that we use Legendre polynomial basis functions for both space and time according to the Wiener-Askey scheme. Consequently, all deterministic input variables (x, t) must be transformed to standardized space, i.e. $\boldsymbol{\xi} \sim \mathcal{U}[-1, 1]$, via an operator \mathcal{T} . That is, we define the boundary points and virtual points by $\boldsymbol{\xi}_{\text{BC}} = \mathcal{T}(\boldsymbol{x}_{\text{BC}}, t_{\text{BC}})$ and $\boldsymbol{\xi}_V = \mathcal{T}(\boldsymbol{x}_V, t_V)$, respectively. Without loss of generality,

uniform distributions are assumed in this paper for the sake of convenience since Legendre polynomials can be efficiently constructed, and their derivatives are easily computed. However, the uniform distribution assumption does not preclude the use of non-uniform statistical sampling methods or the use of sparse data obtained from measurements. On the contrary, it has been shown that importance sampling can lead to improved PCE accuracy [39]. One may also employ active learning approaches to identify an optimal sampling scheme for the given problem [21]. On the other hand, employing the uniform distribution and Legendre polynomials may lead to slow convergence or affect the sparsity/stability of the constructed PCE in some applications. In such cases, one could assume a more appropriate distribution and associated basis. Generally speaking, one can assume arbitrary distributions as long as it is possible to differentiate the associated basis functions to obtain derivatives up to n th order.

3.1. Physical Constraints & Lagrange Multipliers

The constrained optimization problem defined by Eq. (10) can be solved efficiently using the method of Lagrange multipliers. The Lagrangian function takes the following form:

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{1}{2} \mathcal{M}(\boldsymbol{\beta}) + \sum_{b=1}^{n_{\text{BC}}} \lambda_b (a_b^T \boldsymbol{\beta} - c_b) + \sum_{v=1}^{n_v} \lambda_v (a_v^T \boldsymbol{\beta} - c_v). \quad (11)$$

Boundary conditions are prescribed by the boundary operator \mathcal{B} , evaluated for the PCE at n_{BC} points $\boldsymbol{\xi}_{\text{BC}}$, and the corresponding vector \mathbf{c}_{BC} consisting of n_{BC} rows determined by $c_b = g(\boldsymbol{\xi}_{\text{BC}}^{(b)})$. Similarly, ODE/PDE constraints are given by the differential operator \mathcal{L} , evaluated for the PCE at n_v virtual points $\boldsymbol{\xi}_v$, and the corresponding vector \mathbf{c}_v consisting of n_v rows given by $c_v = u(\boldsymbol{\xi}_v^{(v)})$. We then assemble these constraints into a matrix \mathbf{A} where $a_b = \{a_b^j = \mathcal{B}(\Psi_j(\boldsymbol{\xi}_{\text{BC}}^{(b)})), j = 0, \dots, P-1\}$ form the first n_{BC} rows and the vectors $a_v = \{a_v^j = \mathcal{L}(\Psi_j(\boldsymbol{\xi}_v^{(v)})), j = 0, \dots, P-1\}$ form the remaining n_v rows of the matrix. Therefore, instead of solving the coefficients $\boldsymbol{\beta}$ by Eq. (7), we construct the following system of linear equations reflecting the OLS solution with physical constraints obtained from the Karush–Kuhn–Tucker (KKT) conditions:

$$\underbrace{\begin{bmatrix} \Psi^T \Psi & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix}}_{\text{KKT matrix}} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \Psi^T \mathbf{Y} \\ \mathbf{c} \end{bmatrix}. \quad (12)$$

Constructing the \mathbf{A} matrix requires derivatives of the PCE model. We notice, however, that the PCE and its derivatives have the same coefficients $\boldsymbol{\beta}$. The constraints can therefore be imposed by efficiently computing the term-wise derivatives of the basis functions as follows:

$$\frac{\partial^n f}{\partial x_i^n} = \frac{\partial^n [\sum_{\alpha \in \mathcal{A}} \beta_\alpha \Psi_\alpha(\boldsymbol{\xi})]}{\partial \xi_i^n} \Delta_\Gamma^n = \sum_{\alpha \in \mathcal{A}} \beta_\alpha \frac{\partial^n \Psi_\alpha(\boldsymbol{\xi})}{\partial \xi_i^n} \Delta_\Gamma^n \quad (13)$$

where Δ_Γ reflects the scaling of the time-space variable x_i to the standardized variable ξ_i , i.e. $\Delta_\Gamma = 2/(x_{\text{max}} - x_{\text{min}})$ for Legendre polynomials defined on $\xi_i \in [-1, 1]$ orthonormal to $X_i \sim \mathcal{U}[x_{\text{min}}, x_{\text{max}}]$. Derivatives of the Legendre polynomials can be obtained efficiently using various numerical implementations, e.g. SciPy [40] employed in this paper, or one can find closed-form expressions for the derivatives of normalized Hermite and Legendre polynomials [41]. Note that the number of terms in the PCE constraints (columns of \mathbf{A}) is identical to the number of terms in the original PCE and it is also possible to perform basic arithmetic operations on the basis directly to satisfy the prescribed operator \mathcal{L} using the PCE alone (i.e. we do not need to solve the PDEs to satisfy the constraints). The total computational cost of solving the KKT system is $\mathcal{O}((n_v + n_{\text{BC}} + P)^3)$. We further require that $n_v + n_{\text{BC}} + n_{\text{sim}} \geq P$ to avoid solving an under-determined system. Subject to this constraint, it is possible to generate an arbitrarily large number of virtual points (similar to PINNs) leading to a better conditioned KKT matrix [42]. However, this may lead to a very large data matrix that results in a highly over-determined system, which increases the computational cost of the solution. Therefore, we use the minimum number of virtual points obtained as $n_v = P - n_{\text{BC}} - n_{\text{sim}}$, which assures a well-determined KKT system of equations.

Naturally, the imposed physical constraints can significantly reduce n_{sim} , which is crucial for costly models. On the other hand, imposing constraints (particularly Dirichlet boundary conditions) may necessitate higher-order polynomial approximations and, for sparse PCE solutions, may require a larger basis set. As a result, for certain problems this may lead to Runge’s phenomenon. Additional studies are necessary to determine when such phenomena arise. Moreover, the number and positions of the virtual and boundary samples (in addition to the training points) could significantly affect the convergence. This problem of optimal experimental design has been extensively studied for standard PCE [39, 43, 44, 45] and has further led to the development of various techniques for sequential sampling [21, 23, 46]. Similar studies could inform best practices in determining the training points, virtual points, and boundary points for PC². Such studies are beyond the scope of this work.

3.2. Approximation Error Estimation

Throughout the training process, it is important to estimate the PC² accuracy. This is especially crucial for the adaptive sparse PC² construction employed in this paper (and developed in Section 5), where it is necessary to iteratively compare several surrogate models differing in maximum polynomial orders and sparsity in order to choose the best approximation. Ideally, the ED should be divided into validation and training sets, but this might be extremely computationally demanding for engineering applications with complex numerical models. Therefore, in the field of UQ, it is often preferred to estimate the approximation error directly from the training set without additional sampling. A common choice of error measure is the mean squared error (MSE), which is well-known from machine learning and statistics. However, driving MSE down on the training data set alone may lead to over-fitting. Alternatively, one of the most widely-used methods in UQ is the leave-one-out cross-validation (LOO-CV) error, which can be obtained analytically from a single conventional PCE [47]. However, since the PC² is designed for a very small (or even missing, i.e. $n_{\text{sim}} = 0$) ED, the standard MSE or LOO-CV may not be appropriate. Moreover, the LOO-CV cannot be computed analytically for PC² and therefore it can only be applied through a brute force implementation at high computational cost. Instead, we use an extended MSE reflecting error in the given PDE and boundary conditions, similar to the error measure applied for PINNs [1], to assess error during the training process. The extended MSE ϵ consists of three terms – error in the approximated function ϵ_u (“data error”), error associated with failure to obey the PDE constraints $\epsilon_{\mathcal{L}}$ (“PDE error”), and error in the boundary conditions $\epsilon_{\mathcal{B}}$ (“BC error”), expressed as:

$$\epsilon = \epsilon_u + \epsilon_{\mathcal{L}} + \epsilon_{\mathcal{B}}. \quad (14)$$

The tendency to over-fitting is mitigated by the second and third components of the error measure ϵ – the mean squared error in the given PDE and boundary conditions. Note that ϵ is evaluated from the training set in each iteration of the adaptive algorithm for construction of PC². Upon completion of the training, the PC² model accuracy is assessed by computing ϵ from a large validation set, as described further in the numerical examples.

4. PC² for Uncertainty Quantification

In this section, we extend the deterministic formulation of PC² to perform UQ – which, along with improvements in computational efficiency discussed above, is an essential motivation for using PCE as opposed to other physically constrained ML methods such as PINNs. Consider the general stochastic partial differential equation given by

$$\begin{aligned} \mathcal{L}(x, t, \mathcal{X}(\omega); u(x, t, \mathcal{X}(\omega))) &= f(x, t, \mathcal{X}(\omega)), \quad \forall x \in \mathcal{D}, t \in \mathcal{T}, \omega \in \Omega \\ \mathcal{B}(x, t, \mathcal{X}(\omega); u(x, t, \mathcal{X}(\omega))) &= g(x, t, \mathcal{X}(\omega)), \quad \forall x \in \partial\mathcal{D}, t \in \mathcal{T}, \omega \in \Omega \end{aligned} \quad (15)$$

where the meaning of the symbols is identical to Eq. (9) and $\mathcal{X}(\omega) \in \mathbb{R}^d$ is a d -dimensional random vector having sample space Ω . We now aim to solve the PCE regression problem described above as constrained by Eq. (15). That is, we define the objective function as

$$\begin{aligned} \mathcal{M}(\boldsymbol{\beta}) &= \min_{\boldsymbol{\beta}} \sum_{j=1}^{n_{\text{sim}}} [Y^j - u^{\text{PCE}}(x^j, t^j, \mathbf{X}^j)]^2 = \min_{\boldsymbol{\beta}} \|Y - \boldsymbol{\Psi}\boldsymbol{\beta}\|^2 \\ \text{s.t. } \mathcal{L}(x_V, t_V, \mathcal{X}(\omega); u(x_V, t_V, \mathcal{X}(\omega))) &= f(x_V, t_V, \mathcal{X}(\omega)), \\ \mathcal{B}(x_{\text{BC}}, t_{\text{BC}}, \mathcal{X}(\omega); u(x_{\text{BC}}, t_{\text{BC}}, \mathcal{X}(\omega))) &= g(x_{\text{BC}}, t_{\text{BC}}, \mathcal{X}(\omega)) \end{aligned} \quad (16)$$

where we define two discrete sets of n_{BC} boundary points $(x_{\text{BC}}, t_{\text{BC}}, \mathbf{X})$ and n_{V} virtual points $(x_{\text{V}}, t_{\text{V}}, \mathbf{X})$. As in the deterministic formulation, all input variables are transformed to standardized space according to the Wiener-Askey scheme, i.e. $\xi_{\text{BC}} = \mathcal{T}(x_{\text{BC}}, t_{\text{BC}}, \mathbf{X})$ and $\xi_{\text{V}} = \mathcal{T}(x_{\text{V}}, t_{\text{V}}, \mathbf{X})$, but more general arbitrary PCE basis functions may be selected. Note that the boundary and virtual points retain the random vector \mathcal{X} , and do not (necessarily) contain specific boundary or virtual points for \mathcal{X} . This is because physical constraints are typically expressed in terms of the physical variables (x, t) and are not (in general) expressed in terms of the random variables contained in \mathcal{X} . That is, \mathcal{X} does not necessarily affect the constraints associated with either \mathcal{L} and \mathcal{B} (although it can). As a result, the random vector \mathcal{X} usually serves only to increase the dimension of the input random vector and thus the solution of the Eq. (16) follows the same form as the deterministic formulation given in Eq. (12).

If, on the other hand, the physical constraints associated with the PDE or BCs are expressed in terms of the random variables in \mathcal{X} , virtual and/or boundary points can be used to enforce these constraints in the same manner as performed above. If these are expressed as equality constraints, the KKT equations to solve the constrained optimization do not change. On the other hand, if these constraints are expressed as inequalities (e.g. non-negative coefficients) then more advanced optimizers are required. This will be the topic of future work.

The resulting constrained PCE is expressed as a function of the complete set of physical and random variables $\xi = \mathcal{T}(x, t, \mathcal{X})$. The form of the PCE as a linear summation over orthonormal polynomials allows for powerful and efficient post-processing. In particular, once a PCE approximation is created, it is possible to directly estimate statistical moments of the output from the expansion. The first statistical moment (the mean value) is simply the first deterministic coefficient of the expansion $\mu_Y = \langle \mathcal{Y}^1 \rangle = \beta_0$. The second raw statistical moment, $\langle \mathcal{Y}^2 \rangle$, can be estimated by

$$\begin{aligned} \langle \mathcal{Y}^2 \rangle &= \int \left[\sum_{\alpha \in \mathcal{A}} \beta_{\alpha} \Psi_{\alpha}(\xi) \right]^2 p_{\xi}(\xi) \, d\xi = \sum_{\alpha_1 \in \mathcal{A}} \sum_{\alpha_2 \in \mathcal{A}} \beta_{\alpha_1} \beta_{\alpha_2} \int \Psi_{\alpha_1}(\xi) \Psi_{\alpha_2}(\xi) p_{\xi}(\xi) \, d\xi \\ &= \sum_{\alpha \in \mathcal{A}} \beta_{\alpha}^2 \int \Psi_{\alpha}(\xi)^2 p_{\xi}(\xi) \, d\xi = \sum_{\alpha \in \mathcal{A}} \beta_{\alpha}^2 \langle \Psi_{\alpha}, \Psi_{\alpha} \rangle = \sum_{\alpha \in \mathcal{A}} \beta_{\alpha}^2. \end{aligned} \quad (17)$$

Considering the orthonormality of the polynomials, it is possible to obtain the variance $\sigma_Y^2 = \langle \mathcal{Y}^2 \rangle - \mu_Y^2$ as the sum of all squared deterministic coefficients except the intercept (which represents the mean value). Note that the estimation of higher statistical central moments, specifically skewness and kurtosis, are more complicated since they require triple and quad products. These can be obtained analytically only for certain polynomial families, e.g. formulas for Hermite and Legendre polynomials (and their combination) can be found in [48]. Moreover, it can be shown that PCE takes the form of the Hoeffding-Sobol decomposition and thus Sobol indices can be obtained analytically by applying Eq. (17) to selected subsets of the PCE terms [16].

As can be seen, a main advantage of PCE over other surrogate models lies in the convenience of performing uncertainty quantification, i.e. the above analytical moment and sensitivity estimates. The PC² framework inherits this convenience when appropriate conditions are established. As noted, the PC² is expressed in terms of both the random variables \mathcal{X} and the deterministic space-time coordinates (x, t) . Therefore, to properly estimate statistical moments or sensitivity indices, we must condition the PCE on (x, t) , yielding local space-time statistics. Numerical estimation of the local mean value $\mathbb{E}[u|x, t]$ and variance $\sigma_{[u|x, t]}^2$ are based on a simple rationale: deterministic values can be specified directly as constants in the PCE. Following the concept of reduced PCE [48], the local variance $\sigma_{[u|x, t]}^2$ depends only on PCE terms containing the random variables \mathcal{X} , while all other terms are constants whose sum is equal to the local expected value $\mathbb{E}[u|x, t]$.

Formally stated, the random variables form a subset of all M input variables $\{x, t, \mathcal{X}\}$, i.e., $\mathcal{X} \subset \{x, t, \mathcal{X}\}$. We define the index subset $\mathbf{u} \subseteq \mathbf{I} = \{1, \dots, M\}$ and its complement \mathbf{u}^c as

$$\mathbf{u} = \{i \in \mathbf{I} : \xi_i \in \mathcal{T}(\mathcal{X})\}, \quad \mathbf{u}^c = \{i \in \mathbf{I} : \xi_i \in \mathcal{T}(\{x, t\})\}. \quad (18)$$

For the sake of clarity, the set of multivariate basis polynomials dependent on \mathcal{X} and its complement are

$$\mathcal{A}_{\mathcal{X}} = \{\alpha \in \mathcal{A} : \alpha_k \neq 0 \leftrightarrow k \in \mathbf{u}\}, \quad \mathcal{A}_{\sim \mathcal{X}} = \{\alpha \in \mathcal{A} : \alpha_k \neq 0 \leftrightarrow k \in \mathbf{u}^c\}. \quad (19)$$

The local variance is further obtained by a simple post-processing of the terms containing the random variables. First, it is necessary to find a unique basis, defined by the unique set of polynomials from $\mathcal{A}_{\mathcal{X}}$. Although

the original basis set \mathcal{A} contains only unique terms, after the reduction there could be identical terms, e.g. two terms differing only in polynomial degrees associated with variables in \mathbf{u}^c . The coefficients of terms with duplicate multi-indices are then summed and assigned to the unique term in the reduced PCE. The constants arising in mixed PCE terms containing deterministic and random variables must be added to the coefficients associated with the unique basis. Finally, the local mean and variance are obtained directly from the coefficients of the reduced PCE, i.e. the expected value is obtained from the coefficients associated with terms containing only deterministic variables in complement set $\mathcal{A}_{\sim\mathcal{X}}$ and the variance is obtained from the coefficients associated with terms in $\mathcal{A}_{\mathcal{X}}$ as:

$$\mathbb{E}[u|\mathbf{x}, t] = \beta_0 + \sum_{\mathbf{a} \in \mathcal{A}_{\sim\mathcal{X}}} \beta_{\mathbf{a}} \quad (20)$$

$$\sigma_{[u|\mathbf{x}, t]}^2 = \sum_{\mathbf{a} \in \mathcal{A}_{\mathcal{X}}} \beta_{\mathbf{a}}^2 \quad (21)$$

For the sake of clarity, the reduced PCE concept is illustrated in the following simple example for a PCE approximation with $p = 2$ containing a deterministic space variable x and a single random variable \mathcal{X} :

$$u(x, \mathcal{X}) = \underbrace{\beta_0 + \beta_{1,0}\psi_1^x(\mathcal{T}(x)) + \beta_{2,0}\psi_2^x(\mathcal{T}(x))}_{\mathbb{E}[u|x]} + \beta_{0,1}\psi_1^{\mathcal{X}}(\mathcal{X}) + \beta_{0,2}\psi_2^{\mathcal{X}}(\mathcal{X}) + \beta_{1,1}\psi_1^x(\mathcal{T}(x))\psi_1^{\mathcal{X}}(\mathcal{X})$$

$$u(\mathcal{X}|x) = \mathbb{E}[u|x] + \underbrace{(\beta_{0,1} + \beta_{1,1}\psi_1^x(\mathcal{T}(x)))}_{\beta_1} \psi_1^{\mathcal{X}}(\mathcal{X}) + \underbrace{\beta_{0,2}}_{\beta_2} \psi_2^{\mathcal{X}}(\mathcal{X}) \rightarrow \sigma_{[u|x]}^2 = \beta_1^2 + \beta_2^2$$

Here, the conditional mean and variance are specifically shown as the components with dependence only on the spatial variable x and the sum of the squares of the coefficients whose polynomials depend on the random variable \mathcal{X} , respectively.

5. Numerical Algorithms

In this section, we present computationally efficient algorithms to solve for the PC^2 coefficients. We then extend the approach to use a reduced basis set defined through LAR. Finally, an iterative scheme is proposed to fit the PC^2 constrained by nonlinear PDEs. Implementation of the proposed approach is simple and straightforward: it requires simply setting up and solving the KKT system of linear equations, which is compatible with existing non-intrusive regression methods. Therefore, it is possible to use various adaptive sparse non-intrusive algorithms to solve for the PC^2 expansion. The main three algorithms are presented in the following paragraphs.

The basic algorithm for assembling the KKT equations begins by building sub-matrix \mathbf{A} containing the given physical constraints in the form of boundary/initial conditions evaluated at ξ_{BC} and the PDE evaluated at ξ_{V} . The sub-matrix \mathbf{A} prescribes all constraints and initial conditions on $\boldsymbol{\beta}$, leading to efficient and accurate estimates even for very low-size (or even missing) ED. Then, the unknown coefficient vector $\boldsymbol{\beta}$ is obtained as the solution of the linear system in Eq. (12) where the KKT matrix is composed of \mathbf{A} and the information sub-matrix $\boldsymbol{\Psi}^T \boldsymbol{\Psi}$. The whole process is detailed in Algorithm 1.

Since KKT-based estimation of the PCE coefficients $\boldsymbol{\beta}$ is computationally efficient, Algorithm 1 can be easily combined with adaptive sparse solvers such as LAR. The complete algorithm integrating the KKT solver into LAR is provided in Algorithm 2. The LAR-KKT algorithm combines information from the ED and physical constraints to identify the optimal basis for the PC^2 approximation as schematically depicted in Fig. 1. Information obtained from the ED is specifically utilized to identify the most suitable basis set by standard LAR. Physical constraints are further imposed on the $\boldsymbol{\beta}$ estimates and a suitable error measure ϵ (Eq. (14)) is then used to select the most accurate PC^2 . Note that the KKT solution can be generally combined with any existing adaptive sparse regression algorithm by simply replacing standard OLS with the solution of the KKT system.

The solution is more complicated for non-linear PDEs because the derivatives of the PCE depend on the PCE coefficients themselves. Therefore, it is necessary to know the complete form of the PCE, and thus $\boldsymbol{\beta}$. It is not sufficient to simply take derivatives of the basis functions. Naturally, the PCE and its derivatives evaluated at the virtual samples are not known at the start of the algorithm. Therefore, it is necessary to initially construct the PCE

Algorithm 1 KKT solver with virtual samples

Input: ED ($\{x, t, \mathcal{X}\}$ and \mathcal{Y}), Ψ , boundary conditions ($\xi_{BC}, \mathbf{c}_{BC}$), $\mathcal{L}, \mathcal{B}, f$

- 1: **for** $b \leftarrow 1$ to n_{BC} **do**
- 2: **for** $j \leftarrow 1$ to P **do**
- 3: evaluate $a_b^j = \mathcal{B}(\Psi_j(\xi_{BC}^{(b)}))$
- 4: append a_b^j to \mathbf{A}
- 5: **end for**
- 6: **end for**
- 7: sample $n_V = P - n_{BC} - n_{sim}$ virtual points ξ_V (MC, LHS etc.)
- 8: **for** $v \leftarrow 1$ to n_V **do**
- 9: **for** $j \leftarrow 1$ to P **do**
- 10: evaluate $a_v^j = \mathcal{L}(\Psi_j(\xi_V^{(v)}))$
- 11: append a_v^j to \mathbf{A}
- 12: **end for**
- 13: evaluate $c_v = f(\xi_V^{(v)})$
- 14: append c_v to \mathbf{c}_V
- 15: **end for**
- 16: assemble vector $\mathbf{c} = [\mathbf{c}_{BC}, \mathbf{c}_V]$
- 17: construct KKT normal equations according to Eq. (12)
- 18: solve the system by OLS

Output: PCE coefficients β

Algorithm 2 LAR-KKT algorithm

Input: ED ($\{x, t, \mathcal{X}\}$ and \mathcal{Y}), Ψ , boundary conditions ($\xi_{BC}, \mathbf{c}_{BC}$), $\mathcal{L}, \mathcal{B}, f, p_{max}$

- 1: **for** $p \leftarrow 1$ to p_{max} **do**
- 2: generate set of basis functions \mathcal{A}
- 3: identify the sequence of the most important basis functions by LAR
- 4: **for** $i \leftarrow 1$ to P **do**
- 5: construct Ψ
- 6: estimate β by KKT (Algorithm 1)
- 7: get approximation error $\epsilon_{p,i}$
- 8: **end for**
- 9: **end for**

Output: $\{\Psi, \beta\}$ associated to the lowest $\epsilon_{p,i}$

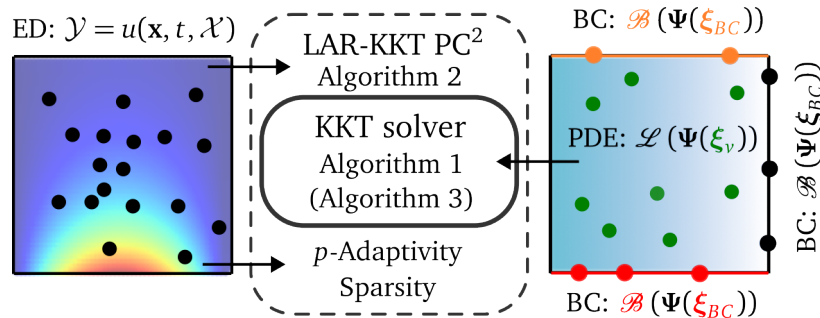


Figure 1: Graphical representation of LAR-KKT algorithm combining p -adaptivity, the LAR algorithm and the KKT solver with virtual/boundary samples.

Algorithm 3 Iterative KKT solver for non-linear PDEs

Input: ED ($\{x, t, \mathcal{X}\}$ and \mathcal{Y}), Ψ , boundary conditions ($\xi_{\text{BC}}, \mathbf{c}_{\text{BC}}$), \mathcal{L}, \mathcal{B}

- 1: get \mathbf{c}_{BC} and β_0 from Algorithm 1 without \mathcal{L}
- 2: sample $n_V = P - n_{\text{BC}} - n_{\text{sim}}$ virtual points ξ_V (MC, LHS etc.)
- 3: **for** $v \leftarrow 1$ to n_V **do**
- 4: **for** $j \leftarrow 1$ to P **do**
- 5: evaluate $a_v^j = \mathcal{L}_{\beta_0}(\Psi_j(\xi_V^{(v)}))$
- 6: append a_v^j to \mathbf{A}
- 7: **end for**
- 8: evaluate $c_v = f(\xi_V^{(v)})$
- 9: append c_v to \mathbf{c}_V
- 10: **end for**
- 11: assemble vector $\mathbf{c} = [\mathbf{c}_{\text{BC}}, \mathbf{c}_V]$
- 12: **for** $i \leftarrow 1$ to n_{iter} **do**
- 13: construct KKT normal equations
- 14: get β_i by least squares
- 15: **for** $v \leftarrow 1$ to n_V **do**
- 16: **for** $j \leftarrow 1$ to P **do**
- 17: evaluate $a_v^j = \mathcal{L}_{\beta_i}(\Psi_j(\xi_V^{(v)}))$
- 18: append a_v^j to \mathbf{A}
- 19: **end for**
- 20: evaluate $c_v = u(\xi_V^{(v)})$
- 21: append c_v to \mathbf{c}_V
- 22: **end for**
- 23: **end for**

Output: PCE coefficients $\beta = \beta_{n_{\text{iter}}}$

(estimate the coefficients β) without virtual samples and iteratively improve the estimate by evaluating the PDE at the virtual samples. An iterative KKT-based algorithm for simple non-linear PDEs is presented in Algorithm 3.

6. Numerical Experiments

To present the capabilities of the proposed PC^2 approach, several examples are presented for various types of PDEs. First, PC^2 is applied to three 1D examples of different nature, including an inhomogeneous ODE, an ODE with BCs of arbitrary order, and a non-linear ODE. PC^2 is then used to approximate the solution to the 2D wave equation and the heat equation with an uncertain input variable. The full PC^2 is constructed using Algorithm 1 (KKT) and the sparse adaptive PC^2 is constructed using Algorithm 2 (LAR-KKT). In both cases, the PC^2 is trained to minimize the extended MSE in Eq. (14) across a set of training points, virtual points, and boundary points. The results are compared to the standard iterative unconstrained LAR method [28] trained using the LOO-CV implemented in UQPy [49, 50] considering identical input and output quantities as the trained PC^2 . All compared methods are extended by p adaptivity ($p \in [5, 25]$) governed by ϵ for PC^2 and ϵ_u for the unconstrained LAR-PCE to obtain the highest possible accuracy.

The ED, containing n_{sim} realizations of the original mathematical model (ground-truth solution of the ODE/PDE), is generated by Latin Hypercube Sampling (LHS). Note that, for deterministic problems n_{sim} defines the number of points in space and time at which the PDE is solved, while for stochastic problems n_{sim} corresponds to the number of realizations of the complete input random vector (containing random variables and space/time variables) for which the PDE is solved. The sets of n_V virtual samples and n_{BC} boundary samples are generated by Crude Monte Carlo (MC) sampling unless stated otherwise. Note that Dirichlet BCs are included in the ED for the standard LAR-PCE method for a fair comparison, though additional information from the PDE and higher-order BCs can only be incorporated in PC^2 . Convergence is measured by two global quantities characterizing error across

the entire input domain computed from a large validation set after completion of model training. The first is the mean squared error where we report convergence of the extended MSE from Eq. (14) as well as its individual constituents ϵ_u , $\epsilon_{\mathcal{L}}$, and $\epsilon_{\mathcal{B}}$. Again, this is computed on the validation set, and is therefore denoted ϵ_{mean} to distinguish it from the extended MSE computed on the training data. The second is the maximum squared error given by:

$$\epsilon_{\text{max}} := \max_{\mathbf{x}, t} \left[\left(u(\mathbf{x}, t, |\mathbf{X}) - \tilde{u}(\mathbf{x}, t, |\mathbf{X}) \right)^2 \right] \quad (22)$$

where again, we show the total ϵ_{max} as well as the individual contributions from ϵ_u , $\epsilon_{\mathcal{L}}$, and $\epsilon_{\mathcal{B}}$.

All numerical results were replicated for 100 independent trials and, for each example, statistical results are presented with convergence plots showing the mean $\pm \sigma$ intervals for ϵ_{mean} and ϵ_{max} and their constituents. The validation sets contain 1000 MC samples for 1D examples and 100^M MC samples for higher dimensional examples, together with 100 samples on the boundaries for the estimation of $\epsilon_{\mathcal{B}}$. Convergence is plotted with respect to increasing training set size, n_{sim} , since evaluations of the original model typically dominate the computational cost of surrogate modeling. Although the number of virtual samples, determined as $n_v = P - n_{\text{BC}} - n_{\text{sim}}$, varies in each iteration of the adaptive LAR-KKT algorithm, their sampling and evaluations come at little computational cost since we need only apply the differential operator to the basis functions evaluated at ξ_v and apply the corresponding values of the PDE source terms.

Finally, detailed insight is also illustrated by a local error measure represented by the absolute difference between the target quantity of interest q and its approximation \tilde{q} as:

$$|\epsilon(\mathbf{x}, t)| := |q(\mathbf{x}, t) - \tilde{q}(\mathbf{x}, t)| \quad (23)$$

For this quantity of interest, we typically consider the approximated function itself, the mean or standard deviation of the solution.

6.1. 1D Poisson Equation: Inhomogeneous ODE

The first example shows the application of the PC² to a simple 1D inhomogeneous ODE. Homogeneous and inhomogeneous ODEs/PDEs solved by PC² differ only in the vectors \mathbf{c}_v – containing specific values for inhomogeneous PDEs and zeros for homogenous PDEs. Naturally, \mathbf{c}_v could contain constants or results of functions of $(\mathbf{x}_v, t_v, \mathcal{X}(\omega))$. The Poisson equation and the assumed BC are as follows:

$$\begin{aligned} \Delta u(x) &= 2, & x \in [-1, 1], \\ \frac{\partial u(x)}{\partial x} \Big|_{x=1} &= 4, & u(-1) = 0. \end{aligned} \quad (24)$$

where Δ is the Laplace operator and $n_{\text{BC}} = 2$ corresponding to constraints prescribed at the 2 boundary points.

The numerical results are summarized in Fig. 2. Note that although the figure contains $\pm \sigma$ intervals, these are not visible for the KKT method due to extremely low variance in the obtained accuracy. The accuracy is measured by three the quantities $\epsilon_u, \epsilon_{\mathcal{L}}, \epsilon_{\mathcal{B}}$ and their sum representing total error (ϵ) from Eq. (14) evaluated on the validation set. Note that while ϵ_u is typically a sufficient measure in standard surrogate modeling, for physics-informed surrogate models it is also important to measure $\epsilon_{\mathcal{L}}$ and $\epsilon_{\mathcal{B}}$ to ensure the physical constraints are satisfied. Standard LAR shows convergence to the exact solution measured by ϵ_u after $n_{\text{sim}} = 8$ samples. Meanwhile, the KKT method is exact even for $n_{\text{sim}} = 2$ realizations of the ground-truth solution. The KKT-LAR method converges in $n_{\text{sim}} = 4$ samples while having a large variance for $n_{\text{sim}} = 3$ samples. This behavior is also seen in the linear ODE example presented next, where the effect of sparsity is insignificant. It is therefore generally more efficient to use the standard KKT method for simple linear ODEs. Although all compared methods converge to the exact solution for $n_{\text{sim}} = 10$, the standard LAR method leads to significant errors for lower n_{sim} , where although it has almost perfect accuracy measured by ϵ_u for $n_{\text{sim}} = 4$, it still has significant errors in the remaining physically constrained criteria. Meanwhile, KKT and LAR-KKT converge consistently in total ϵ and each of its components.

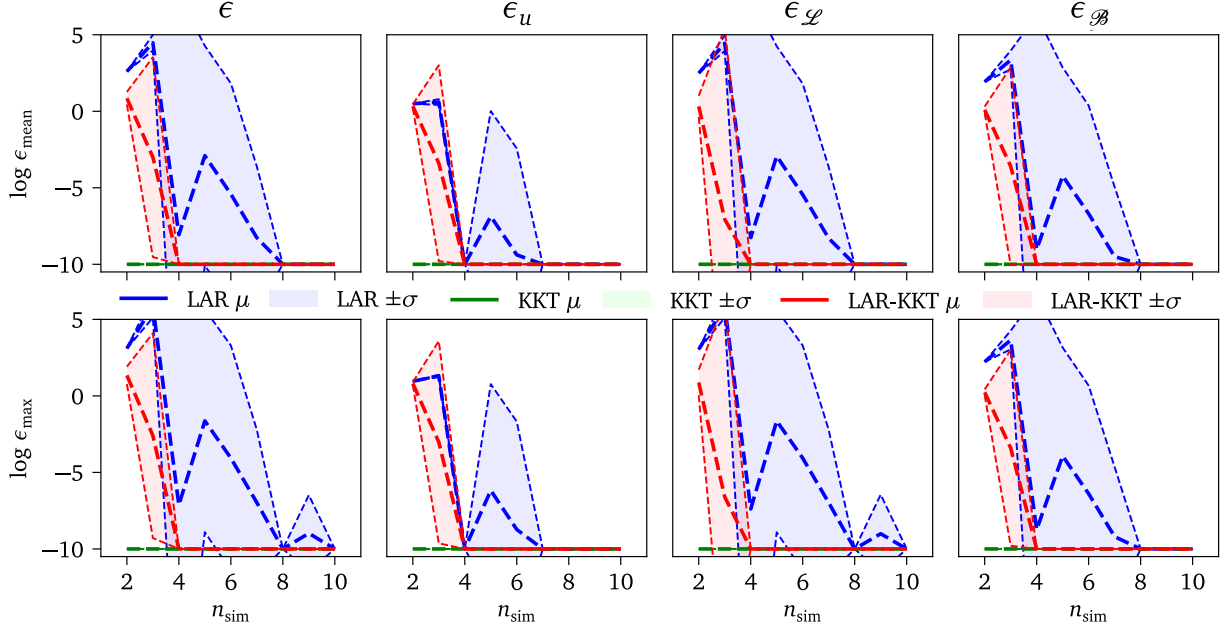


Figure 2: Numerical results for the 1D Poisson equation for increasing number of samples in the domain. The top row shows the mean squared errors and the bottom row shows the maximum squared errors obtained by the simple KKT (Algorithm 1), LAR-KKT (Algorithm 2) and standard LAR methods. Each column corresponds to a specific component of the total error measure ϵ .

6.2. 1D Euler Equation: Arbitrary BC

Using PC^2 it is possible to solve PDEs containing derivatives of arbitrary order in \mathcal{L} and \mathcal{B} if the n th derivative of the PCE basis functions exists. This example presents the 1D Euler beam equation with given BCs as follows:

$$\begin{aligned} \frac{\partial^4 u(x)}{\partial x^4} + 1 &= 0, & x \in [0, 1], \\ u(0) &= 0, & u'(0) = 0 & u''(1) = 0, & u'''(1) = 0 \end{aligned} \quad (25)$$

where $n_{BC} = 4$ corresponding to the constraints prescribed at the two boundary points. This ODE with the prescribed BCs represents the deflection of a cantilever beam with a uniformly distributed load having intensity $q = 1$. The obtained results are shown in Fig. 3 in the identical form as in the previous example. The convergence trends of the compared methods are similar to the previous example, though the differences between standard LAR and PC^2 are more significant. PC^2 convergence is consistent with respect to all criteria as can be clearly seen in convergence plots of the LAR-KKT method, while the standard LAR approach leads to similar overfitting up to $n_{sim} = 6$ similar to the previous example.

6.3. 1D Logistic Equation: Non-linear ODE

The last example presenting the capabilities of the proposed PC^2 for ODEs is a non-linear logistic equation commonly representing population growth. Non-linear ODEs/PDEs solved by PC^2 require iterative construction of the KKT system using Algorithm 3, though it is not applicable to strongly non-linear systems. For more complicated equations, it is necessary to use more advanced and more computationally expensive optimizers. The 1D logistic equation with a single BC ($n_{BC} = 1$) is given in the following form:

$$\begin{aligned} \frac{\partial u(x)}{\partial x} &= u(x)(1 - u(x)), & x \in [-5, 5], \\ u(0) &= 0.5 \end{aligned} \quad (26)$$

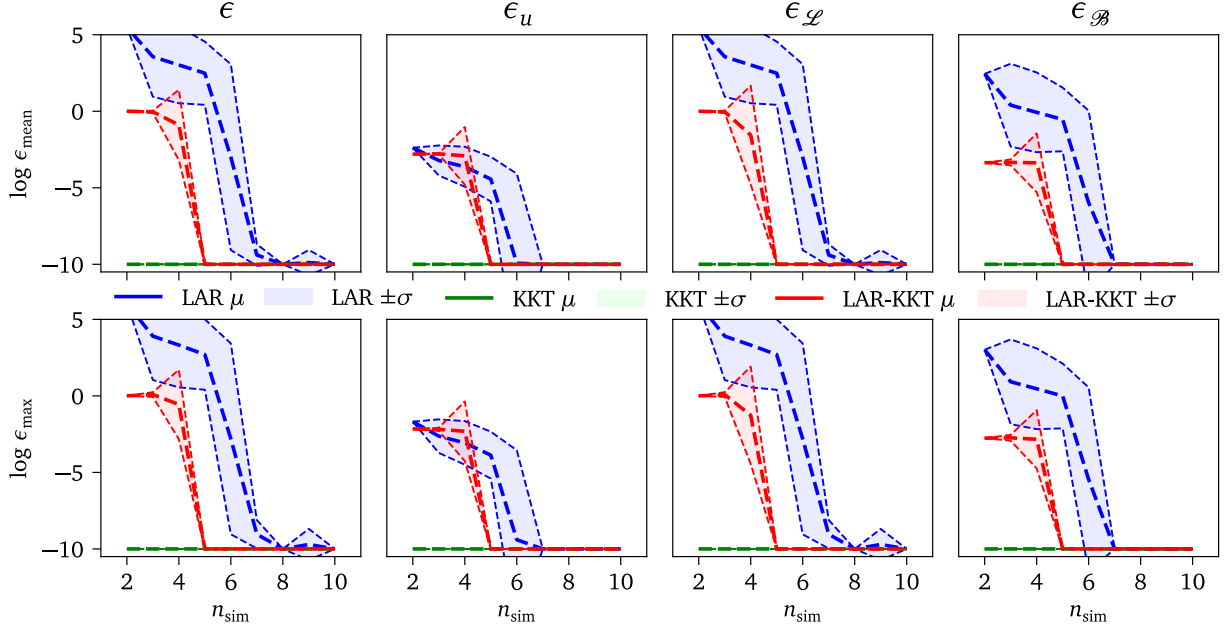


Figure 3: Numerical results for the 1D Euler equation for increasing number of samples in the domain. The top row shows the mean squared errors and the bottom row shows maximum squared errors obtained by the simple KKT (Algorithm 1), LAR-KKT (Algorithm 2) and standard LAR methods. Each column corresponds to a specific component of the total error measure ϵ .

The numerical results are summarized in Fig. 4. The convergence trend is similar for all methods, but PC^2 achieves significantly higher accuracy. Note that the simple KKT solver is the most accurate for a very low number of samples, though for increasing number of samples the benefit of the sparse solution obtained by the LAR-KKT solver leads to superior accuracy. Interestingly, the convergence trend of the standard LAR is very similar to the LAR-KKT, although there is a significant difference in the absolute accuracy. Regardless of n_{sim} , the accuracy of LAR at the given BCs is nearly perfect, possibly indicating that it is overfitting. Meanwhile, both PC^2 methods consistently converge in all presented criteria. Relatively slow convergence in $\epsilon_{\mathcal{L}}$ and $\epsilon_{\mathcal{B}}$ obtained by the KKT solution clearly show the benefit of sparse solvers in this example, since some basis functions with high p are most suitable for an accurate PCE approximation. However, since the cardinality of the full basis set used in the KKT method is high, it requires a large n_{sim} for accurate coefficient estimates.

6.4. Wave equation: Time-dependent PDE

The previous 1D examples clearly show the advantages of PC^2 for various types of ODEs. Naturally, the PC^2 can be used for problems with a larger number of input variables. This will be presented through the time-dependent wave equation given by:

$$\begin{aligned} \frac{\partial^2 u(x, t)}{\partial t^2} &= 4 \frac{\partial^2 u(x, t)}{\partial x^2}, & x \in [0, 1], t \in [0, 2] \\ \frac{\partial u(x, 0)}{\partial t} &= 0, & u(0, t) = u(1, t) = 0, & u(x, 0) = \sin(\pi x). \end{aligned} \quad (27)$$

Although the extension of PC^2 for n -dimensional problems is straightforward since it is based on standard PCE, there is one significant difference in comparison to the 1D examples: the boundary and initial conditions are now always prescribed point-wise and thus the accuracy of PC^2 is also dependent on the number and position of the n_{BC} boundary samples. In this paper, we use the DeepXDE python package [51] for deterministic sampling on the boundaries and thus their position does not influence the variance of the convergence plots. Convergence

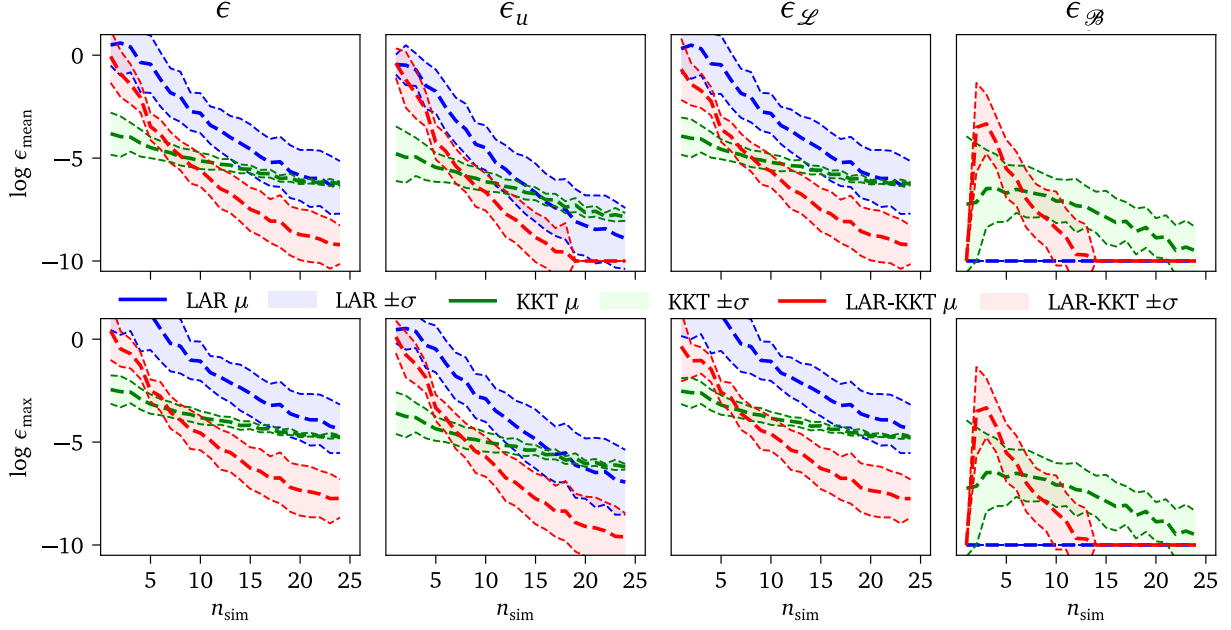


Figure 4: Numerical results for the 1D logistic equation for increasing number of samples in the domain. The top row shows the mean squared errors and the bottom row shows the maximum squared errors obtained by the simple KKT (Algorithm 1), LAR-KKT (Algorithm 2) and standard LAR methods. Each column corresponds to a specific component of the total error measure ϵ .

plots for $n_{\text{BC}} = 10$ samples on the boundaries are shown in Fig. 5. PC^2 converges rapidly to a very accurate approximation, although the ϵ_u for very small n_{sim} is identical to LAR. However, the PDE error $\epsilon_{\mathcal{L}}$ is negligible for PC^2 , but LAR does not respect the PDE, and as a result, it does not converge to the original model. Note that, although not shown, the difference between the PC^2 and LAR is significantly higher for increasing n_{BC} since the Neumann BC can only be imposed in PC^2 .

A detailed comparison of the PC^2 and standard LAR solutions for a selected realization are shown in Fig. 6. The top row shows the solution of the wave equation for a selected t where the dashed line represents the analytical reference solution and the colored line shows the solution by LAR (left) and PC^2 (right). The approximations are discretized at 200 points and their colors correspond to the logarithm of the local squared errors of the PDE, $\log \epsilon_{\mathcal{L}}$. The bottom row shows approximations of the wave equation over the entire input space together with given the ED, \mathbf{x}_{BC} and \mathbf{x}_{V} . It is clear that the standard LAR does not produce a sufficiently accurate approximation of $u(x, t)$, which is also reflected by the high $\epsilon_{\mathcal{L}}$. Furthermore, even when LAR produces similar errors ϵ_u , its PDE errors $\epsilon_{\mathcal{L}}$ are much larger as illustrated by a second realization in Fig. 7. In this case, LAR leads to $\epsilon_u < 10^{-3}$, though it is clear from the colors of $\log \epsilon_{\mathcal{L}}$ that the prescribed PDE is not respected. Note that PC^2 leads to near perfect approximations in both selected realizations.

6.5. Heat Equation: Uncertainty Quantification

The last numerical example presents a main advantage of PC^2 – analytical UQ of the original model containing deterministic and random variables as proposed in Section 4. This example is thus divided to two parts: (i) a convergence study assuming all input variables to be deterministic, which shows the efficiency of PC^2 as a surrogate model; and (ii) UQ of the given model with a random variable input.

We specifically study the heat equation with uniformly distributed coefficient of thermal diffusivity \mathcal{D} taking

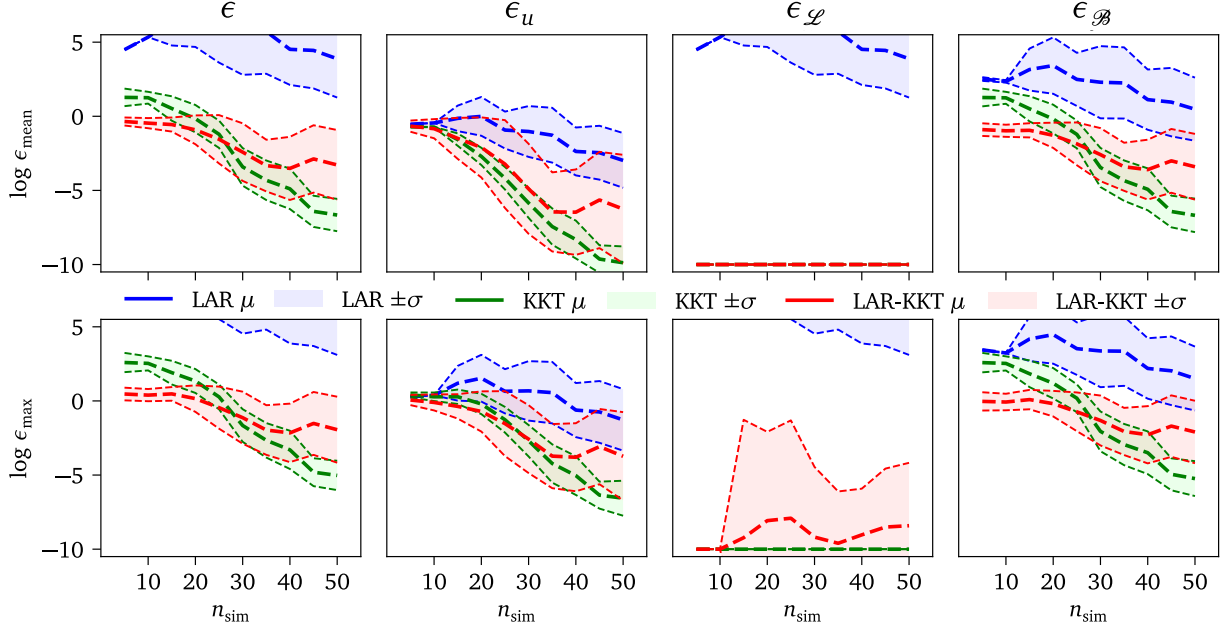


Figure 5: Numerical results for the wave equation for increasing number of samples in the domain. The top row shows the mean squared errors and the bottom row shows the maximum squared errors obtained by the simple KKT (Algorithm 1), LAR-KKT (Algorithm 2) and standard LAR methods. Each column corresponds to a specific component of the total error measure ϵ .

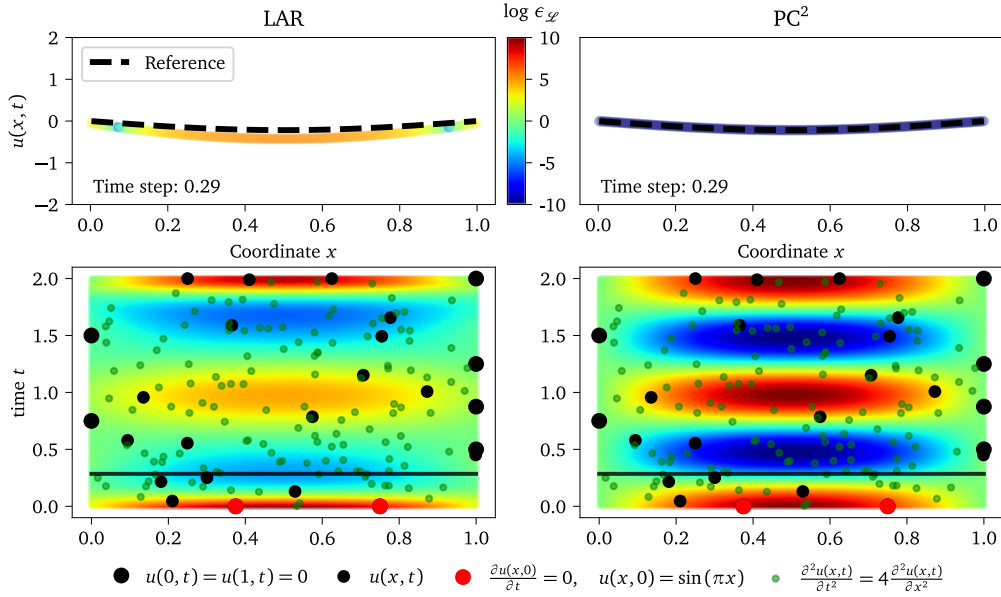


Figure 6: Selected realization of LAR (left) and PC^2 (right) for the 2D wave equation. The left pane shows significant error ϵ_u and $\epsilon_{\mathcal{L}}$ for the standard LAR. The right pane shows a very accurate solution from PC^2 using an identical ED, x_{BC} and x_V .

the following form:

$$\begin{aligned} \frac{\partial u(x, t)}{\partial t} &= \mathcal{D} \frac{\partial^2 u(x, t)}{\partial x^2}, & x \in [0, 1], \quad t \in [0, 1], \quad \mathcal{D} \sim \mathcal{U}[0.2, 0.8] \\ u(0, t) &= u(1, t) = 0, & u(x, 0) &= \sin(\pi x) \end{aligned} \quad (28)$$

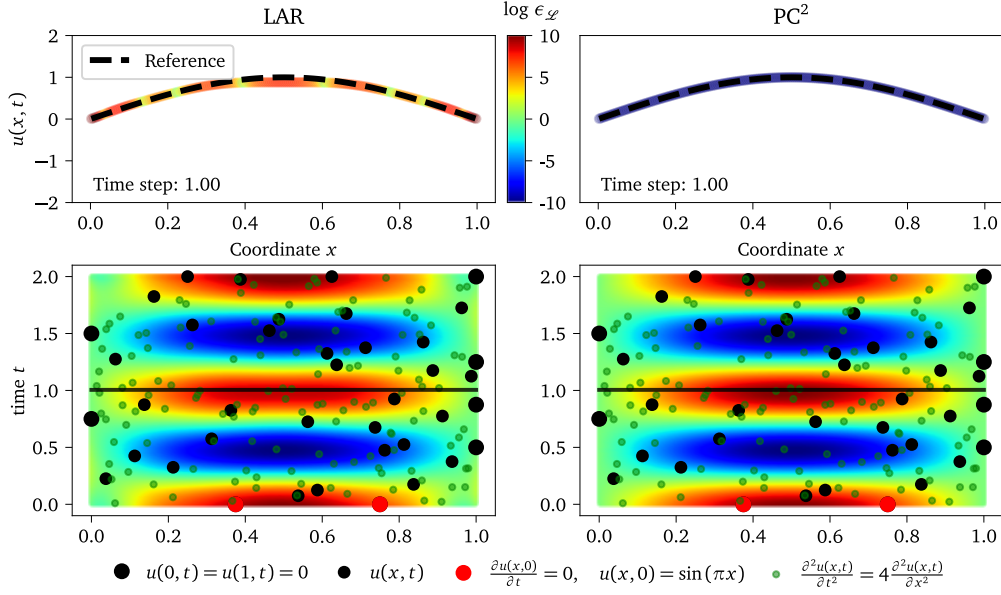


Figure 7: Selected realization of LAR (left) and PC^2 (right) for the 2D wave equation. The left and right panes show similar errors ϵ_u (bottom) but PC^2 shows much lower error in the PDE error $\epsilon_{\mathcal{L}}$ (top) using an identical ED, x_{BC} and x_V .

Deterministic value of coefficient of thermal diffusivity

In the first part of this example, we assume a deterministic value for the coefficient of thermal diffusivity $\mathcal{D} = 0.4$. The results in Fig. 8 shows the convergence for $n_{BC} = 10$ and increasing number of samples n_{sim} in the space-time domain. The benefit of imposing the physical constraints can be clearly seen by comparing the convergence plots. The convergence rate for the standard LAR is significantly lower than PC^2 , while both PC^2 algorithms achieve near-perfect accuracy after $n_{sim} = 15$. Note that LAR-KKT has a higher variance since the selected sparse set of basis functions is dependent on the given ED, as expected.

Uncertain coefficient of thermal diffusivity

In the second part, there are 2 deterministic variables (space coordinate x and time t) and one random variable $\mathcal{D} \sim \mathcal{U}[0.2, 0.8]$. To efficiently quantify solution uncertainty, PC^2 is based only on $n_{BC} = 90$ boundary samples and a set of virtual points. We do not run the deterministic simulation using samples of \mathcal{D} . In other words, $n_{sim} = 0$. This scenario is of significant practical interest, because the traditional solution of a PDE with uncertain inputs requires repeated solution of the PDE with different realizations of the random variables, which could be costly. The number of samples on the boundaries is selected to be large to specifically highlight the advantages of PC^2 for UQ, since the general convergence for PC^2 was shown in the previous deterministic case. An empty ED ($n_{sim} = 0$) leads to very efficient UQ, since it requires no model evaluations. Moreover, we can easily modify the stochastic model of the input variables – increasing the number of random variables, changing their distributions etc. – without re-evaluating the model, as would be necessary for standard surrogate models.

Once the PC^2 model is trained, we can use it to easily estimate the local means, variances, and quantiles (e.g. $\pm\sigma$) as depicted in Fig. 9, which compares the mean solution and mean $\pm\sigma$ for PC^2 with Monte Carlo simulation using LHS with $n_{sim} = 10^5$ samples. Note that there are negligible errors in the local mean values. The estimated local variance is also compared to the LHS solution in Fig. 10. Although the PC^2 is based only on BC and PDE samples, it leads to a very accurate approximation of local statistics. Moreover, the PC^2 model also reflects the uncertainty in \mathcal{D} very well in predictions. Fig. 11 shows the PC^2 approximations in space-time and errors for selected values of \mathcal{D} .

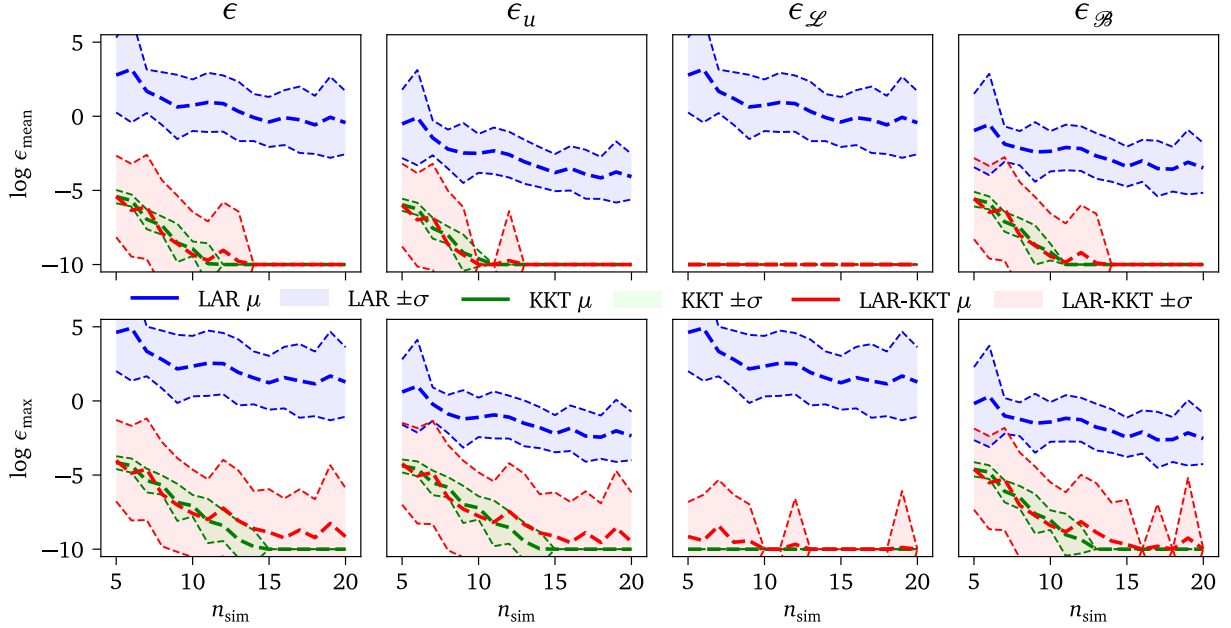


Figure 8: Numerical results for the heat equation for $n_{BC}=10$ and increasing number of samples in the time-space domain. The top row shows the mean squared errors and the bottom row shows the maximum squared errors obtained by the simple KKT (Algorithm 1), LAR-KKT (Algorithm 2) and standard LAR methods. Each column corresponds to a specific component of the total error measure ϵ .

7. Conclusions & Further Work

A novel methodology for constructing physics-informed non-intrusive regression-based PCE, referred to as PC^2 was proposed in this paper. Physical constraints in the form of ordinary and partial differential equations and their boundary conditions are imposed in a constrained least squares solution to assemble a KKT system of linear equations using the method of Lagrange multipliers. The proposed approach does not significantly increase the computational cost to estimate the PCE coefficients, but improves accuracy considerably. The proposed solution can be further employed with existing adaptive algorithms in lieu of conventional ordinary least squares. An algorithm based on least angle regression is then proposed to achieve a constrained sparse solution with p -adaptivity. The presented approach was developed specifically for UQ of costly mathematical models of physical systems using a small ED. PC^2 enables analytical UQ similar to standard PCE by simple post-processing of the coefficients, though the PC^2 must first be reduced to exclude the influence of deterministic variables. From the obtained numerical results, it is clear that PC^2 achieves significant improvement in approximation accuracy in terms of surrogate model prediction error, but also with respect to error in the PDE and boundary conditions for low-to-mid size ED with little additional computational cost. We compared two algorithms for PC^2 construction – a sparse LAR-KKT solver and a KKT-based PCE with a full set of basis functions, and it can be concluded that while LAR-KKT does not always lead to superior results, its benefits are more significant with increasing dimension of input random vector. Note that both algorithms can be applied in tandem in practical applications, since the PC^2 using the full set of basis functions increases the computational cost of LAR-KKT just by one additional iteration. Finally, it was shown in the final numerical example that it is possible to create a PC^2 surrogate using only a set of virtual samples and boundary conditions, which leads to extremely efficient UQ in comparison to other existing techniques.

This work opens the door to several open questions that should be investigated in further research. First, the proposed approach is limited to problems with linear and weakly nonlinear differential constraints. More complicated non-linear PDEs will require advanced optimization techniques and thus PC^2 must be extended also for broader classes of applications and suitable optimization techniques should be identified. Moreover, the

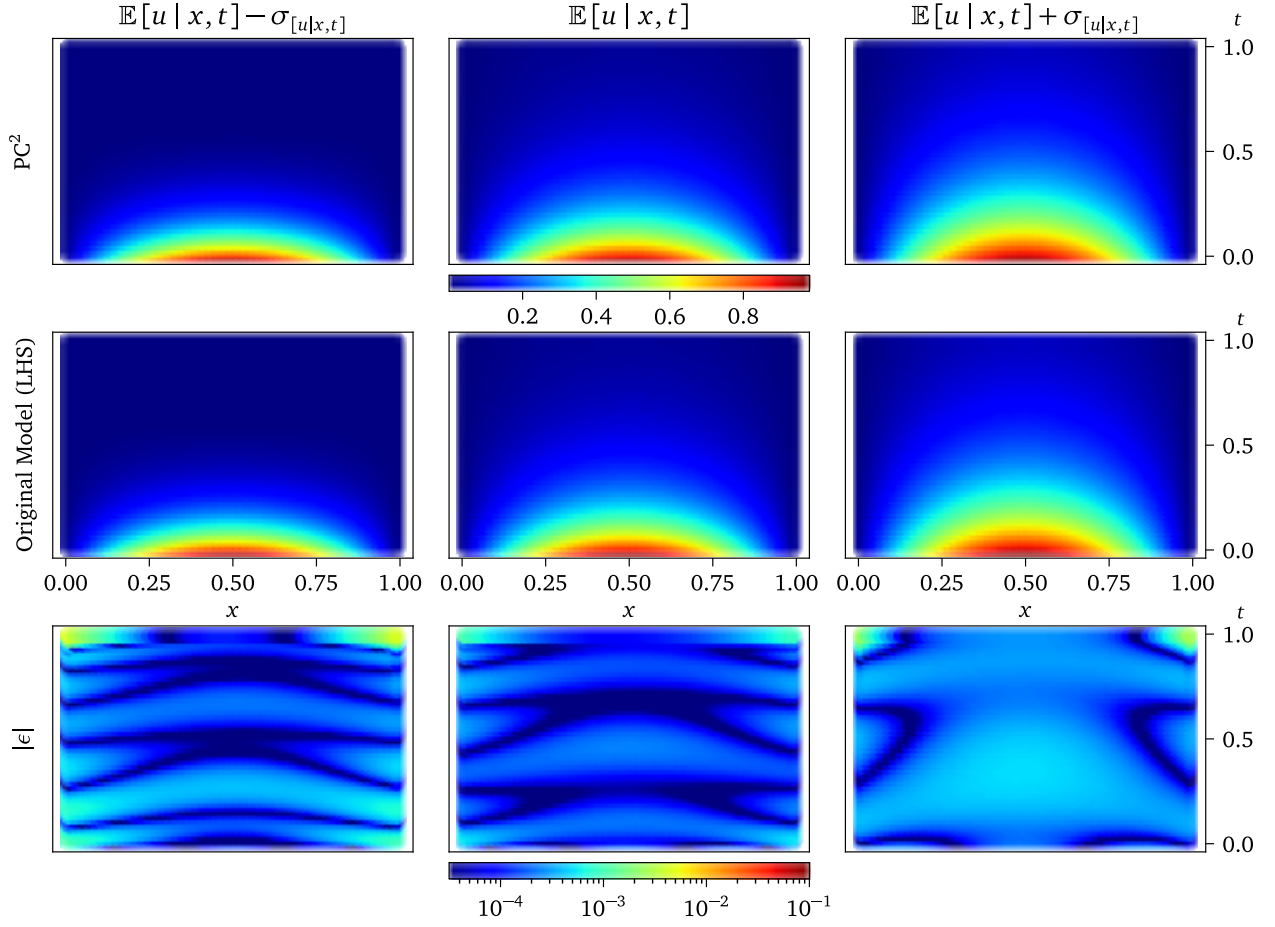


Figure 9: Estimates of the local mean values at given coordinates $\mathbb{E}[u | x, t]$ (middle), and local $\mathbb{E}[u | x, t] \pm \sigma_{[u|x,t]}$ quantiles (left and right). Comparison of the analytical solution by PC² (top), Monte Carlo simulation using the original model (middle), and their absolute difference (bottom).

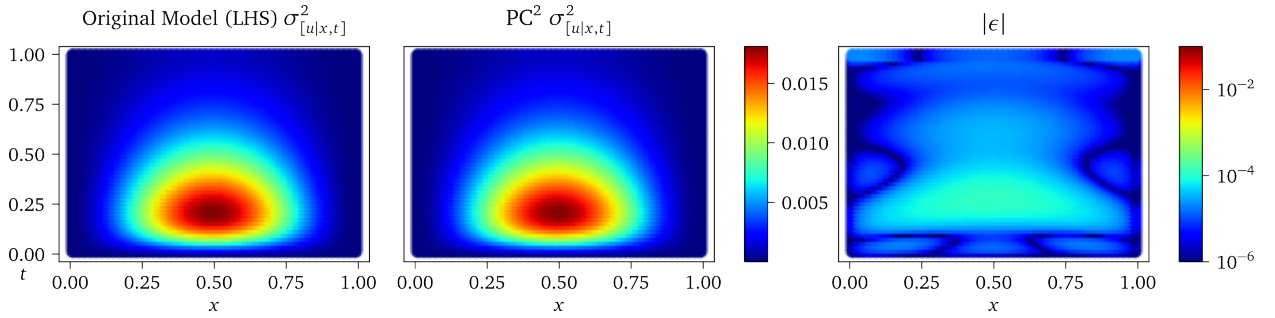


Figure 10: Estimates of the local variance at given coordinates $\sigma^2_{[u|x,t]}$. Comparison of the numerical estimation by LHS (left), the analytical solution by PC² (middle), and their absolute difference (right).

use of polynomial basis functions restricts the application of PC² to sufficiently smooth operations, which could be alleviated by coupling the proposed method with domain-decomposition techniques and/or multi-element techniques [25, 32, 24]. Further, PC² can be combined with optimal sampling strategies for both the ED and the virtual/boundary points and active learning techniques for sample selection such as the Θ criterion[21] recently proposed by authors of this paper. Finally, the presented examples are low-dimensional. Moving forward, it will

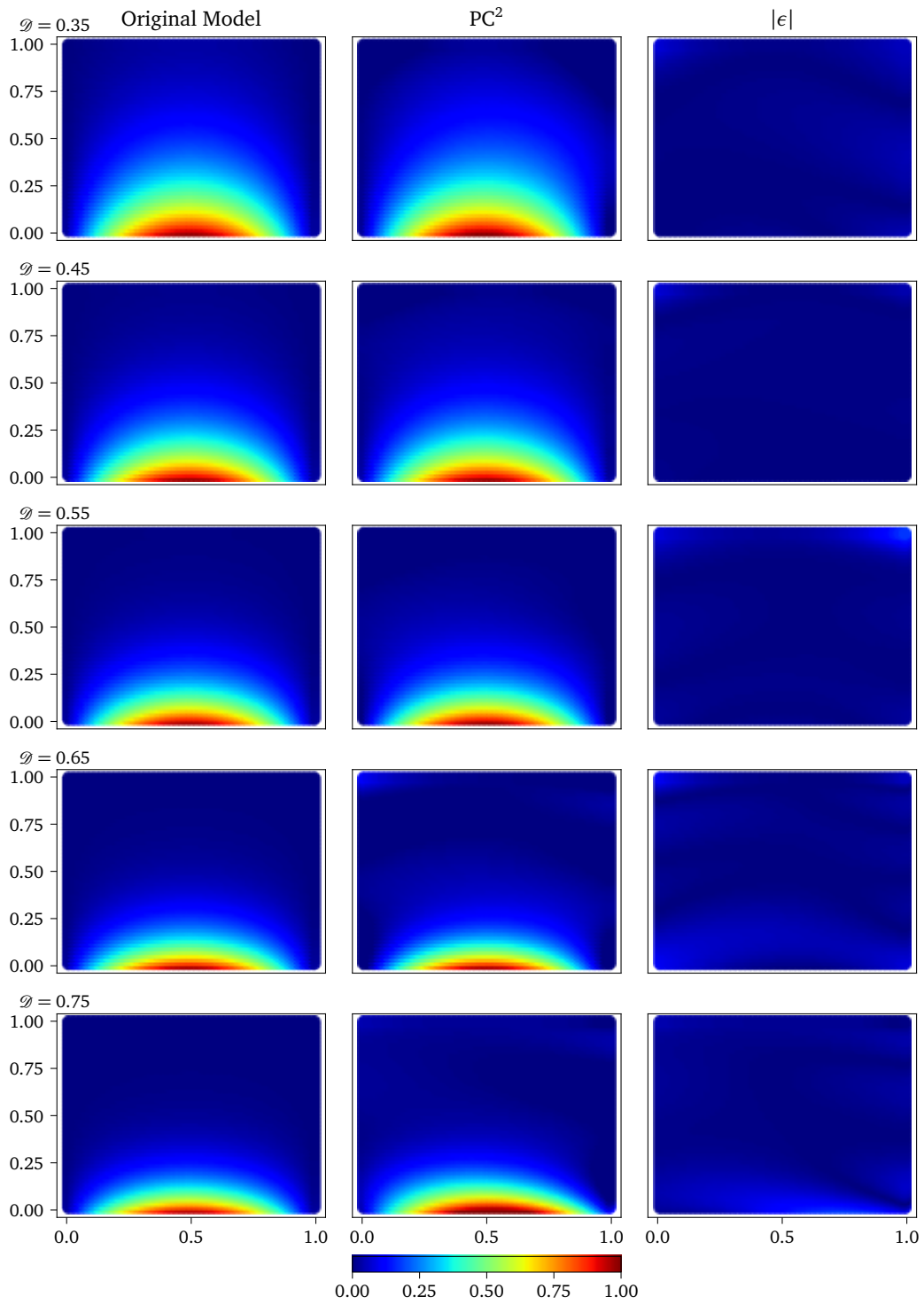


Figure 11: Prediction of the heat equation solution by PC² based on $n_{\text{BC}} = 90$ samples for different values of ϑ . Each row shows the original function, the PC² approximation, and the error corresponding to a selected realization of ϑ .

be necessary to investigate the scalability of the proposed approach with increasing input dimension – particularly with respect to the dimension of the random variable inputs. However, since PC^2 is based on a well-established non-intrusive regression-based approach, it is generally possible to apply existing techniques for high-dimensional PCEs based on the sparsity-of-effects principle [14]. Moreover, the curse of dimensionality could also be mitigated by using techniques such as stochastic domain decomposition [52]. Future work will focus on investigating the scalability of PC^2 and ultimately extending the method for higher-dimensional applications.

Acknowledgments

The first author acknowledges the financial support provided by the Czech Science Foundation under project number 23-04712S. Additionally, a part of this research was conducted during the research stay of the first author at Johns Hopkins University supported by the stipend and partnership program of Brno University of Technology for Excellence 2023. MDS and HS have been supported by the Defense Threat Reduction Agency, Award HDTRA1202000.

References

- [1] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, *Nature Reviews Physics* 3 (6) (2021) 422–440.
- [2] H. Lee, I. S. Kang, Neural algorithm for solving differential equations, *Journal of Computational Physics* 91 (1) (1990) 110–131. doi:[https://doi.org/10.1016/0021-9991\(90\)90007-N](https://doi.org/10.1016/0021-9991(90)90007-N).
- [3] A. Meade, A. Fernandez, The numerical solution of linear ordinary differential equations by feedforward neural networks, *Mathematical and Computer Modelling* 19 (12) (1994) 1–25. doi:[https://doi.org/10.1016/0895-7177\(94\)90095-7](https://doi.org/10.1016/0895-7177(94)90095-7).
- [4] I. E. Lagaris, A. Likas, D. I. Fotiadis, Artificial neural networks for solving ordinary and partial differential equations, *IEEE transactions on neural networks* 9 (5) (1998) 987–1000.
- [5] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational physics* 378 (2019) 686–707.
- [6] S. Goswami, A. Bora, Y. Yu, G. E. Karniadakis, Physics-informed deep neural operators, *arXiv preprint arXiv:2207.05748* (2022).
- [7] N. B. Erichson, M. Muehlebach, M. W. Mahoney, Physics-informed autoencoders for Lyapunov-stable fluid flow prediction, *arXiv preprint arXiv:1905.10866* (2019).
- [8] E. Kharazmi, Z. Zhang, G. E. Karniadakis, hp-vpinns: Variational physics-informed neural networks with domain decomposition, *Computer Methods in Applied Mechanics and Engineering* 374 (2021) 113547. doi:<https://doi.org/10.1016/j.cma.2020.113547>.
- [9] J. Sirignano, K. Spiliopoulos, Dgm: A deep learning algorithm for solving partial differential equations, *Journal of computational physics* 375 (2018) 1339–1364.
- [10] L. P. Swiler, M. Gulian, A. L. Frankel, C. Safta, J. D. Jakeman, A survey of constrained Gaussian process regression: Approaches and implementation challenges, *Journal of Machine Learning for Modeling and Computing* 1 (2) (2020).
- [11] G. Pang, G. E. Karniadakis, Physics-informed learning machines for partial differential equations: Gaussian processes versus neural networks, *Emerging Frontiers in Nonlinear Science* (2020) 323–343.
- [12] H. Sharma, J. A. Gaffney, D. Tsapetis, M. D. Shields, Learning thermodynamically constrained equations of state with uncertainty, *APL Machine Learning* 2 (1) (2024) 016102. doi:[10.1063/5.0165298](https://doi.org/10.1063/5.0165298).
- [13] N. Wiener, The homogeneous chaos, *American Journal of Mathematics* 60 (4) (1938) 897–936. doi:[10.2307/2371268](https://doi.org/10.2307/2371268).
- [14] N. Lüthen, S. Marelli, B. Sudret, Sparse polynomial chaos expansions: Literature survey and benchmark, *SIAM/ASA Journal on Uncertainty Quantification* 9 (2) (2021) 593–649. doi:[10.1137/20M1315774](https://doi.org/10.1137/20M1315774).
- [15] E. Torre, S. Marelli, P. Embrechts, B. Sudret, Data-driven polynomial chaos expansion for machine learning regression, *Journal of Computational Physics* 388 (2019) 601–623.
- [16] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, *Reliability Engineering & System Safety* 93 (7) (2008) 964–979. doi:[10.1016/j.ress.2007.04.002](https://doi.org/10.1016/j.ress.2007.04.002).
- [17] T. Crestaux, O. L. Maitre, J.-M. Martinez, Polynomial chaos expansion for sensitivity analysis, *Reliability Engineering & System Safety* 94 (7) (2009) 1161–1172. doi:[10.1016/j.ress.2008.10.008](https://doi.org/10.1016/j.ress.2008.10.008).
- [18] N.-Z. Chen, C. Guedes Soares, Spectral stochastic finite element analysis for laminated composite plates, *Computer methods in Applied Mechanics and Engineering* 197 (51) (2008) 4830–4839. doi:[10.1016/j.cma.2008.07.003](https://doi.org/10.1016/j.cma.2008.07.003).
- [19] N. Lüthen, S. Marelli, B. Sudret, Automatic selection of basis-adaptive sparse polynomial chaos expansions for engineering applications, *International Journal for Uncertainty Quantification* 12 (3) (2022) 49–74.
- [20] N. Fajraoui, S. Marelli, B. Sudret, Sequential design of experiment for sparse polynomial chaos expansions, *SIAM/ASA Journal on Uncertainty Quantification* 5 (1) (2017) 1061–1085. doi:[10.1137/16m1103488](https://doi.org/10.1137/16m1103488).
- [21] L. Novák, M. Vořechovský, V. Sadílek, M. D. Shields, Variance-based adaptive sequential sampling for polynomial chaos expansion, *Computer Methods in Applied Mechanics and Engineering* 386 (2021) 114105. doi:<https://doi.org/10.1016/j.cma.2021.114105>.
- [22] M. Thapa, S. B. Mulani, R. W. Walters, Adaptive weighted least-squares polynomial chaos expansion with basis adaptivity and sequential adaptive sampling, *Computer methods in Applied Mechanics and Engineering* 360 (2020) 112759. doi:[10.1016/j.cma.2019.112759](https://doi.org/10.1016/j.cma.2019.112759).

- [23] D. Loukrezis, A. Galetzka, H. De Gersem, Robust adaptive least squares polynomial chaos expansions in high-frequency applications, *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields* 33 (6) (2020) e2725. doi:<https://doi.org/10.1002/jnm.2725>.
- [24] X. Wan, G. E. Karniadakis, An adaptive multi-element generalized polynomial chaos method for stochastic differential equations, *Journal of Computational Physics* 209 (2) (2005) 617–642. doi:[10.1016/j.jcp.2005.03.023](https://doi.org/10.1016/j.jcp.2005.03.023).
- [25] L. Novák, M. D. Shields, V. Sadílek, M. Vořechovský, Active learning-based domain adaptive localized polynomial chaos expansion, *Mechanical Systems and Signal Processing* 204 (2023) 110728. doi:<https://doi.org/10.1016/j.ymsp.2023.110728>.
- [26] S. Marelli, P.-R. Wagner, C. Lataniotis, B. Sudret, Stochastic spectral embedding, *International Journal for Uncertainty Quantification* 11 (2) (2021) 25–47. doi:[10.1615/int.j.uncertaintyquantification.2020034395](https://doi.org/10.1615/int.j.uncertaintyquantification.2020034395).
- [27] S.-K. Choi, R. V. Grandhi, R. A. Canfield, C. L. Pettit, Polynomial chaos expansion with latin hypercube sampling for estimating response variability, *AIAA Journal* 42 (6) (2004) 1191–1198. doi:[10.2514/1.2220](https://doi.org/10.2514/1.2220).
- [28] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, *Journal of Computational Physics* 230 (6) (2011) 2345–2367. doi:[10.1016/j.jcp.2010.12.021](https://doi.org/10.1016/j.jcp.2010.12.021).
- [29] D. Xiu, G. E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, *SIAM Journal on Scientific Computing* 24 (2) (2002) 619–644. doi:[10.1137/S1064827501387826](https://doi.org/10.1137/S1064827501387826).
- [30] C. Soize, R. Ghanem, Physical systems with random uncertainties: Chaos representations with arbitrary probability measure, *SIAM Journal on Scientific Computing* 26 (2) (2004) 395–410. doi:[10.1137/S1064827503424505](https://doi.org/10.1137/S1064827503424505).
- [31] X. Wan, G. E. Karniadakis, Multi-element generalized polynomial chaos for arbitrary probability measures, *SIAM Journal on Scientific Computing* 28 (3) (2006) 901–928. doi:[10.1137/050627630](https://doi.org/10.1137/050627630).
- [32] A. Galetzka, D. Loukrezis, N. Georg, H. De Gersem, U. Römer, An hp-adaptive multi-element stochastic collocation method for surrogate modeling with information re-use, *International Journal for Numerical Methods in Engineering* 124 (12) (2023) 2902–2930. doi:<https://doi.org/10.1002/nme.7234>.
- [33] S. Oladyshkin, W. Nowak, Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion, *Reliability Engineering & System Safety* 106 (2012) 179–190. doi:<https://doi.org/10.1016/j.res.2012.05.002>.
- [34] R. Ahlfeld, B. Belkouchi, F. Montomoli, Samba: Sparse approximation of moment-based arbitrary polynomial chaos, *Journal of Computational Physics* 320 (2016) 1–16. doi:<https://doi.org/10.1016/j.jcp.2016.05.014>.
- [35] R. G. Ghanem, P. D. Spanos, *Stochastic finite elements: a spectral approach*, Springer New York, 1991. doi:[10.1007/978-1-4612-3094-6](https://doi.org/10.1007/978-1-4612-3094-6).
- [36] A. Cohen, G. Migliorati, Optimal weighted least-squares methods, *The SMAI journal of computational mathematics* 3 (2017) 181–203. doi:[10.5802/smai-jcm.24](https://doi.org/10.5802/smai-jcm.24).
- [37] A. C. Narayan, J. Jakeman, T. Zhou, A Christoffel function weighted least squares algorithm for collocation approximations, *Math. Comput.* 86 (2017) 1913–1947. doi:[10.1090/mcom/3192](https://doi.org/10.1090/mcom/3192).
- [38] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, *The Annals of Statistics* 32 (2) (2004) 407–451. doi:[10.2307/3448465](https://doi.org/10.2307/3448465).
- [39] J. Hampton, A. Doostan, Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies, *Journal of Computational Physics* 280 (2015) 363–386. doi:[10.1016/j.jcp.2014.09.019](https://doi.org/10.1016/j.jcp.2014.09.019).
- [40] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: fundamental algorithms for scientific computing in Python, *Nature Methods* 17 (2020) 261–272. doi:[10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- [41] B. Sudret, C. Mai, Computing derivative-based global sensitivity measures using polynomial chaos expansions, *Reliability Engineering & System Safety* 134 (2015) 241–250. doi:<https://doi.org/10.1016/j.res.2014.07.009>.
- [42] G. Migliorati, F. Nobile, E. von Schwerin, R. Tempone, Approximation of quantities of interest in stochastic pdes by the random discrete l^2 projection on polynomial spaces, *SIAM Journal on Scientific Computing* 35 (3) (2013) A1440–A1460. doi:[10.1137/120897109](https://doi.org/10.1137/120897109).
- [43] A. Cohen, G. Migliorati, Optimal weighted least-squares methods, *The SMAI journal of computational mathematics* 3 (2017) 181–203. doi:[10.5802/smai-jcm.24](https://doi.org/10.5802/smai-jcm.24).
- [44] J. D. Jakeman, A. Narayan, T. Zhou, A generalized sampling and preconditioning scheme for sparse approximation of polynomial chaos expansions, *SIAM Journal on Scientific Computing* 39 (3) (2017) A1114–A1144. doi:[10.1137/16M1063885](https://doi.org/10.1137/16M1063885).
- [45] M. Hadigol, A. Doostan, Least squares polynomial chaos expansion: A review of sampling strategies, *Computer methods in Applied Mechanics and Engineering* 332 (2018) 382–407. doi:[10.1016/j.cma.2017.12.019](https://doi.org/10.1016/j.cma.2017.12.019).
- [46] J. Hampton, A. Doostan, Basis adaptive sample efficient polynomial chaos (BASE-PC), *Journal of Computational Physics* 371 (2018) 20–49. doi:[10.1016/j.jcp.2018.03.035](https://doi.org/10.1016/j.jcp.2018.03.035).
- [47] G. Blatman, B. Sudret, An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis, *Probabilistic Engineering Mechanics* 25 (2) (2010) 183–197. doi:[10.1016/j.probenmech.2009.10.003](https://doi.org/10.1016/j.probenmech.2009.10.003).
- [48] L. Novák, On distribution-based global sensitivity analysis by polynomial chaos expansion, *Computers & Structures* 267 (2022) 106808. doi:[10.1016/j.compstruc.2022.106808](https://doi.org/10.1016/j.compstruc.2022.106808).
- [49] A. Olivier, D. Giovanis, B. Aakash, M. Chauhan, L. Vandanapu, M. D. Shields, UQpy: A general purpose python package and development environment for uncertainty quantification, *Journal of Computational Science* 47 (2020) 101204.
- [50] D. Tsapetis, M. D. Shields, D. G. Giovanis, A. Olivier, L. Novak, P. Chakroborty, H. Sharma, M. Chauhan, K. Kontolati, L. Vandanapu, D. Loukrezis, M. Gardner, Uqpy v4.1: Uncertainty quantification with Python, *SoftwareX* 24 (2023) 101561. doi:<https://doi.org/10.1016/j.softx.2023.101561>.
- [51] L. Lu, X. Meng, Z. Mao, G. E. Karniadakis, DeepXDE: A deep learning library for solving differential equations, *SIAM Review* 63 (1) (2021) 208–228. doi:[10.1137/19M1274067](https://doi.org/10.1137/19M1274067).
- [52] Y. Chen, J. Jakeman, C. Gittelson, D. Xiu, Local polynomial chaos expansion for linear differential equations with high dimensional

random inputs, SIAM Journal on Scientific Computing 37 (1) (2015) A79–A102. [doi:10.1137/140970100](https://doi.org/10.1137/140970100).

3.5 UQpy v4.1: Uncertainty Quantification with Python

TSAPETIS, D.; SHIELDS, M.; GIOVANIS, D.; OLIVIER, A.; NOVÁK, L.; CHAKROBORTY, P.; SHARMA, H.; CHAUHAN, M.; KONTOLATI, K.; VANDANAPU, L.; LOUKREZIS, D.; GARDNER, M. UQpy v4.1: Uncertainty quantification with Python. *SoftwareX*, 2023, vol. 24. ISSN: 2352-7110. (WoS-AIS: Q2)
DOI: 10.1016/j.softx.2023.101561

Description

The paper presents the new version of UQPy package for Python containing various methods for UQ of physical systems including statistical sampling, reliability analysis, transformations, surrogate models, sensitivity analysis, etc. The new version of UQPy is now developed by professional software developers, and thus it offers a user-friendly structure and stable solutions of given problems. The software package contains all of the algorithms presented in this thesis.

A part of the paper is also a publicly available documentation of the software package: <https://uqpyproject.readthedocs.io/en/latest/>
and publicly available code: <https://github.com/SURGroup/UQpy>

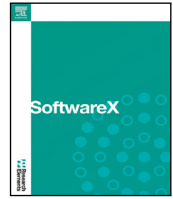
Role of the author

Percentage of contribution: 30%

Lukáš Novák is a member of an international team of developers and authors of the paper. His main contribution is in the surrogate modeling part, and he is the main developer of the PCE module of the presented UQPy package.

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original Software Publication



UQpy v4.1: Uncertainty quantification with Python

Dimitrios Tsapetis^a, Michael D. Shields^{a,*}, Dimitris G. Giovanis^a, Audrey Olivier^b,
Lukas Novak^c, Promit Chakroborty^a, Himanshu Sharma^a, Mohit Chauhan^a, Katiana Kontolati^a,
Lohit Vandanapu^a, Dimitrios Loukrezis^d, Michael Gardner^e

^a Department of Civil and Systems Engineering, Johns Hopkins University, Baltimore, MD, 21218, USA^b Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, Los Angeles, CA, 90089, USA^c Faculty of Civil Engineering, Brno University of Technology, Veveri 331/95, Brno 60200, Czech Republic^d Department of Electrical Engineering and Information Technology, Technical University of Darmstadt, Darmstadt, Germany^e Department of Civil and Environmental Engineering, University of California, Davis, CA, 95616, USA

ARTICLE INFO

Keywords:

Uncertainty quantification
Continuous integration

ABSTRACT

This paper presents the latest improvements introduced in Version 4 of the UQpy, Uncertainty Quantification with Python, library. In the latest version, the code was restructured to conform with the latest Python coding conventions, refactored to simplify previous tightly coupled features, and improve its extensibility and modularity. To improve the robustness of UQpy, software engineering best practices were adopted. A new software development workflow significantly improved collaboration between team members, and continuous integration and automated testing ensured the robustness and reliability of software performance. Continuous deployment of UQpy allowed its automated packaging and distribution in system agnostic format via multiple channels, while a Docker image enables the use of the toolbox regardless of operating system limitations.

Code metadata

Current code version	v4.1.1
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-23-00337
Code Ocean compute capsule	https://codeocean.com/capsule/9924279/tree
Legal Code License	MIT Licence
Code versioning system used	git
Software code languages, tools, and services used	Python, MPI
Compilation requirements, operating environments & dependencies	see <code>setup.py</code>
If available Link to developer documentation/manual	https://uqpyproject.readthedocs.io
Support forum for questions	https://github.com/SURGroup/UQpy/discussions

Software metadata

Current software version	4.1.1
Permanent link to executables of this version	pypi.org/project/UQpy github.com/SURGroup/UQpy/releases anaconda.org/conda-forge/uqpy
Legal Software License	MIT License
Computing platforms/Operating Systems	Linux, OS X, Microsoft Windows
Installation requirements & dependencies	see <code>setup.py</code>
If available, link to user manual - if formally published include a reference to the publication in the reference list	https://uqpyproject.readthedocs.io

* Corresponding author.

E-mail address: michael.shields@jhu.edu (Michael D. Shields).2352-7110/© 2023 Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).
<https://doi.org/10.1016/j.softx.2023.101561>

Received 22 May 2023; Received in revised form 20 August 2023; Accepted 12 October 2023

1. Motivation and significance

Uncertainty Quantification (UQ) is the science of characterizing, quantifying, managing, and reducing uncertainties in mathematical, computational and physical systems. Depending on the sources of uncertainty, UQ provides a multitude of methodologies to quantify their effects. For instance, given the probability distribution for the inputs to a computational model, forward uncertainty propagation methods aim to estimate the distributions or statistics of resulting quantities of interest. Inverse UQ, on the other hand, aims to infer uncertainties in input quantities given limitations and uncertainties in the observed system response, e.g. for model calibration from experimental data. Numerous related tasks fall under the broad classification of UQ including sensitivity analysis, which aims to quantify the influence of multiple inputs to a system, and reliability analysis which aims to estimate (and sometimes minimize) the probability of failure of the system.

A major challenge in UQ is to reduce the high computational expense associated with many repeated model evaluations. This can be achieved through advances in sampling, development of computationally inexpensive surrogate models (or metamodels), and by leveraging high performance computing. To address these challenges, multiple software packages and libraries have been developed. Some of the most comprehensive libraries for UQ include OpenTurns [1], Koralı [2], MUQ [3], UQTK [4], Dakota [5], OpenCossan [6] and UQLab [7]. These software are developed in either C, C++ programming languages or Matlab and although many provide bindings to Python (to differing extents), they are not generally suitable for direct extension in Python, which is one of the most widely used languages in the scientific community.

Apart from these general purpose UQ libraries, several packages that target specific applications or with more limited scope are available. In R, the DiceDesign [8] package aids experimental design, while DiceKriging and DiceOptim [9] use Kriging for metamodeling and surrogate-based optimization, respectively. The Matlab code FERUM [10], developed at UC Berkeley and IFMA, Clermont, serves as a general purpose finite element structural reliability code, while SUMOToolbox [11] is a framework for global surrogate modeling and adaptive sampling. The Engineering Risk Analysis group at TU Munich also provides a collection of Matlab and Python routines [12] related to the group's research. Specifically in Python, several focused libraries have been developed. UncertaintyPy [13] was developed for UQ in computational neuroscience. PyROM framework [14] provides a user-friendly way to implement model reduction techniques. The ChaosPy package provides UQ functionality centered around polynomial chaos expansions. Bayesian calibration algorithms are implemented in SPUX [15] and ABCpy [16] and sensitivity analyses by SALib [17]. PyMC [18] provides a simple Python interface that allows its user to create Bayesian models and fit them using Markov Chain Monte Carlo methods. PyGPC [19] library is based on generalized polynomial chaos theory and provides capabilities for uncertainty and sensitivity analysis of computational models. Three of the latest additions are PyApprox [20], which provides wide-ranging functionality, NeuralUQ [21] focused on UQ in neural network models, and Fortuna that provides uncertainty estimates, classification and prediction for production systems.

UQpy aims to provide a comprehensive UQ library with wide-ranging capabilities spanning the areas discussed above, as well as a development environment for creating new UQ methodologies. The UQpy package was originally introduced in [22], where the overall structure of v3 was described. Since then, the authors have reworked the UQpy architecture with the goal to simplify its structure, enhance its extensibility, and make it more robust. The updated architecture of the library rendered it not backwards compatible, as the strategy for construction of classes has changed. Yet porting older solutions to the new structure can be performed in a straightforward manner. This restructuring resulted in the current version we present here, v4.1.

The first task carried out towards v4.1 was to restructure the file system. The previous structure, which maintained a single Python file per module, had reached size limitations and made it cumbersome for the team members to add new functionalities or update existing ones. In the reorganization, a directory was created for each module, which contains, in a hierarchical structure, subdirectories for specific functionalities, with one file dedicated to each class. Slight modifications were also made to the existing code to ensure compliance with PEP8 by renaming modules, classes, and function signatures. Instead of monolithic classes per functionality, each component was split into a separate class with a dedicated abstract baseclass, where applicable. This choice reduced code complexity, provided a standardized way of extending components, and enabled the construction of the final functionality, using object composition and inheritance.

The second step to improve internal and external collaboration was to deprecate the “branch-per-developer” strategy and move to a feature-based branch structure using the Github Flow. This removed unnecessary redundancies and complications when multiple people are working on related functionalities. At the same time, the workflow is now directly combined with testing automation and Continuous Integration/Continuous Delivery (CI/CD) workflows. Unit tests were implemented throughout the software, achieving code coverage greater than 80%. The CI pipeline includes linting, code quality checks, and automated semantic versioning, while the CD pipeline packages and distributes the code via multiple channels, such as PyPI, conda-forge, and Docker images. This CI/CD pipelines are explained in more detail in Section 3.

The documentation was revamped to reflect the new hierarchical structure of the code, with embedded examples serving as tutorials to quickly familiarize users with the code functionality. Specifically, for each class, a gallery of examples is created using the sphinx-gallery extension [23]. The users can now download the examples in both Jupyter notebook and Python format or directly interact with the example in a dedicated Binder environment. Finally, several new functionalities were introduced either by the development team or external collaborations, thus boosting UQpy's capabilities.

2. Software description

2.1. Software architecture

UQpy is a Python-based toolbox that provides a series of computational methodologies and algorithms for wide-ranging UQ problems. The core of UQpy is based on state-of-the-art Python libraries, specifically NumPy [24], which is the most fundamental package supporting array and linear algebra operations, SciPy [25], that provides algorithms for optimization, integration and basic statistics, and scikit-learn [26], which includes various tools for supervised and unsupervised learning. UQpy is split into eleven modules, nine of which address specific tasks in UQ and which will be discussed in detail in the following section. A module that enables necessary simulations in all other modules, called `run_model`, aids in the batch execution of both Python and third-party computational models and includes functionality for parallelization via MPI for high performance computing. Finally, a `utilities` module contains various functions that are common to multiple modules.

2.2. Software modules

In this section, all existing modules of UQpy will be briefly introduced, with emphasis on software updates compared to v3. The basic module structure and capabilities are illustrated in Fig. 1. The detailed UML diagrams for all modules are included in the UQpy documentation allowing architecture visualization.

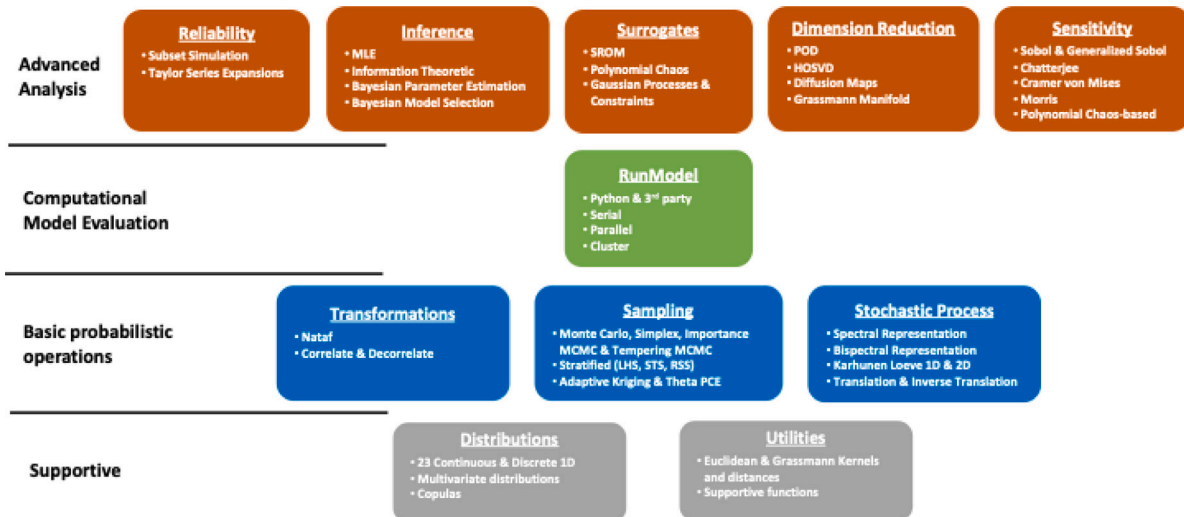


Fig. 1. Structure and capabilities of the UQpy modules.

2.2.1. distributions module

The distributions module serves as the basis for most probabilistic operations in UQpy. It is fully compatible with `scipy` distributions and enables users to create probability distribution objects. Compared to the previous version, the baseclass hierarchy was simplified. An abstract baseclass `Distribution` serves as the interface for creating all subsequent distributions. Depending on the dimensionality of the distribution this baseclass is further refined into `Distribution1D` and `DistributionND` for univariate and multivariate distributions respectively, while the `Distribution1D` is further subclassed into `DistributionsContinuous1D` and `DistributionsDiscrete1D` for continuous and discrete random variables. Within this structure, 23 distinct distributions are implemented. A Copula baseclass with two implementations enables users to add dependence between 1D distribution objects. All baseclasses can be easily extended by users to implement any distribution of their choice by simply creating a new child class for the distribution and implementing the requisite methods.

2.2.2. sampling module

This module provides a wide range of methods to draw samples of random variables. The following classes enable Monte Carlo simulation and variance reduction methods: `MonteCarloSampling`, `SimplexSampling`, `ImportanceSampling`, and `StratifiedSampling`. The `StratifiedSampling` class has been refactored as a parent class for all stratified sampling approaches with `LatinHypercubeSampling`, `TrueStratifiedSampling`, and `RefinedStratifiedSampling` as child classes, all of which utilize a common `Strata` class for geometric decomposition of the domain. Markov Chain Monte Carlo (MCMC) methods are included, with the MCMC abstract baseclass serving as the common interface and 7 different methodologies implemented as subclasses. The latest version includes two new implementations of parallel and sequential tempering MCMC algorithms. Additional MCMC methods can be implemented by the user by simply creating a new subclass with the requisite methods. The module also includes the `AdaptiveKriging` for adaptive sample generation for Gaussian process surrogate modeling (see Section 2.2.9) using specified (and custom) learning functions. Compared to v3, all learning functions have been extracted as separate classes, with a common `LearningFunction` baseclass, allowing users to easily create custom implementations.

2.2.3. transformations module

This module contains isoprobabilistic transformations of random variables. Except for updates in naming conventions, this module retained the previous functionality with the `Nataf`, `Correlate`, and `Decorrelate` transformations being available.

2.2.4. stochastic_process module

This module supports the simulation of univariate, multivariate, and multidimensional Gaussian and non-Gaussian stochastic processes, with the latest addition since v3 being the two-dimensional Karhunen-Loève Expansion in the `KarhunenLoeveExpansion2D` class. All pre-existing classes of `SpectralRepresentation`, `BispectralRepresentation` and `KarhunenLoeveExpansion` have been updated to conform with PEP8 Python coding standards.

2.2.5. run_model module

This module is not directly related to any specific UQ operations, yet it is an integral part of the UQpy software. It lies at its core and supports the execution of either Python or third-party computational models at specified sampling points.

UQpy interfaces Python models directly, by importing and executing the code. On the other hand, UQpy interfaces with third-party software models through ASCII text files to introduce uncertainties in their inputs and uses a standardized scripting format for model execution. In both cases, UQpy supports serial and parallel execution. Parallel execution allows the execution of different samples simultaneously, with options for local and cluster execution. Local parallel execution uses MPI and the `mpi4py` library to distribute the random samples among tasks that are processed independently. In this case, the model evaluation cannot invoke MPI internally. In cluster enabled parallelization, with the aid of a bash script, a tiling of the jobs can be performed to include both shared and distributed memory parallelism, while enabling the user to work with different HPC schedulers.

2.2.6. dimension_reduction module

In the update from v3 to v4.1, the `dimension_reduction` module was rewritten from scratch. The existing `DirectPOD` and `SnapshotPOD` methods were reworked to comply with the latest Python coding conventions and the `HigherOrderSVD` class was added. To support Grassmann manifold projections and operations, a series of classes were added. The `GrassmannProjection` class serves as the parent for classes that project data arrays onto the manifold, with the `SVDprojection` subclass currently available. After the data have been projected, operations such as computing the Karcher mean or

Fréchet variance are available with the aid of the `GrassmannOperations` class. Interpolation can be performed on the manifold with the `GrassmannInterpolation` class. Special attention was given to the `DiffusionMaps` class, where the kernel computation was extracted and delegated to a hierarchy of kernel classes in the `utilities` modules for broader use in future development of other kernel-based methods. More detail can be found in Section 2.2.11.

2.2.7. inference module

The functionality of the `inference` module was retained from v3 to v4.1 but restructured. The previous `InferenceModel` class, which defines the model on which inference is performed, has now been split into three separate classes depending on the specific model type, namely `DistributionModel`, `LogLikelihoodModel` and `ComputationalModel`, all under the revised `InferenceModel` baseclass. For information theoretic-based model selection using the `InformationModelSelection` class, the information criteria have been extracted as separate classes, `AIC`, `BIC`, `AICc`, under a new common `InformationCriterion` baseclass. The remaining functionality of `MLE`, `BayesParameterEstimation`, and `BayesModelSelection` was updated according to the newly adopted coding conventions and, for Bayesian evidence computation, the `EvidenceMethod` baseclass has been established with the `HarmonicMean` subclass defined and allowing straightforward implementation of new Bayesian evidence methods as distinct subclasses.

2.2.8. reliability module

Modifications to the `reliability` module were made to ensure compliance with the latest Python coding guidelines. The first and second-order reliability methods, `FORM` and `SORM`, were restructured as subclasses under a common `TaylorSeries` baseclass to remove code redundancies. The existing `SubsetSimulation` class was retained and revised to match best practices.

2.2.9. surrogates module

One of the most heavily refactored modules in the latest version is `surrogates`. Generally, surrogate models are now developed under the abstract `Surrogate` baseclass. The previously existing `Kriging` class was removed entirely and is now replaced with the more general `GaussianProcessRegression`, which includes the functionality to perform regression or interpolation (Kriging). Kernels are extracted as separate classes, with the abstract baseclass `Kernel` (from the `utilities` module), serving as an interface. The `RBF` and `Matern` kernels have been implemented. For use with `GaussianProcessRegression`, multiple regression methods are implemented as subclasses under the `Regression` baseclass. The newest addition to `GaussianProcessRegression` is the ability to add constraints using the virtual point method. These constraints are implemented under the `Constraints` baseclass, which makes adding new constraints straightforward by implementing a new subclass with the requisite methods.

The `PolynomialChaosExpansion` class was rewritten from scratch (now as a subclass of `Surrogate`) to resolve performance issues. Two new baseclasses were introduced. The `Polynomials` baseclass defines sets of orthogonal polynomials as subclasses, including the `Hermite` and `Legendre` polynomial classes. The `PolynomialBasis` baseclass establishes a set of subclasses to define the polynomial basis, e.g. using a classical tensor product basis `TensorProductBasis`, or introducing new ways to reduce the basis computation such as the `TotalDegreeBasis` and `HyperbolicBasis` classes. This makes the code easily extensible to include new means of basis construction. All regression methods were united as subclasses under the `Regression` baseclass, again making it more easily extended for new methods, and the computationally efficient `LeastAngleRegression` was added.

Lastly, the `SRM` method was retained and updated to conform with the latest Python software development practices.

2.2.10. sensitivity module

In v3, the `sensitivity` module only contained the `MorrisSensitivity` method. This module significantly benefited from the extensibility introduced in UQpy with v4.1. The `Sensitivity` abstract baseclass now contains the first major contribution from external collaborators introduced in a set of subclasses that include `SobolSensitivity`, `GeneralizedSobolSensitivity`, `ChatterjeeSensitivity`, and `CramerVonMisesSensitivity`. Additionally, the updated polynomial chaos expansion code in the `surrogates` module (see Section 2.2.9), allows the computation of first and total order sensitivity indices with reduced computational cost through the `PceSensitivity` class, which takes advantage of a fitted `PolynomialChaosExpansion` object.

2.2.11. utilities module

The new `utilities` module contains code that may be used in multiple modules. This currently contains two abstract baseclasses, the `Kernel` baseclass and the `Distance` baseclass for computing kernels and measures of distance, respectively. Within each baseclass, there are two additional baseclasses for Euclidean and Grassmannian kernels/distances. Several kernels and distances have been added as subclasses and new ones can be easily developed by writing a new subclass with the requisite methods.

3. Continuous integration

UQpy v3 [22] was developed using the flexible standards of an academic software, which challenged the ability of the team to collaborate and develop new features using a streamlined workflow. To this end, the latest version was fully restructured to enhance its extensibility, while modern software development practices were introduced to support collaboration and ensure code robustness and quality. The standard of Github Flow was adopted as development strategy. The master branch of the Github repository always contains the latest stable version. A `Development` branch is now used for merging all newly developed functionality and bug fixes. New versions of the software are released when a pull request is merged from the `Development` branch to `master`. For developing new features, a `feature-{functionality}` branch is created from the latest `Development` state and merged back once complete. The case is similar for bug fixes, with branches following the `bugfix-{bug}` naming convention. The aforementioned workflow enables a consistent way of treating new functionality or addressing errors arising during development.

To ensure the code quality of all previously implemented features, the development team enforced unit testing practices. Since the functionality implemented in UQpy is inherently stochastic, and its randomness stems from random number generators, a process of setting the seed to ensure test reproducibility is adopted. All previous functionalities are tested against benchmark problems to achieve a minimum of 80% line coverage. To ensure that the code coverage directive is enforced, Azure Pipelines were used to automatically run all tests and compute coverage when a commit is pushed to the Github repository. The static code analyzer `Pylint` is also used to enforce coding standards and ensure that no syntax errors are allowed. In addition to these checks, a code quality tool named `SonarCloud` is used to eliminate code vulnerabilities. This tool is triggered when creating a pull request and automatically detects any code smells, bugs or code duplications introduced, and fails when exceeding a predefined threshold. For a pull request to be acceptable, all test, linting, and code quality must satisfy minimum acceptance criteria and must pass a detailed code review from the code owners. Only then will the additions be merged to `Development` and subsequently to `master` branch.

Apart from the Continuous Integration process mentioned above, that ensure the robustness of UQpy, a set of Continuous Deployment

(CD) actions are triggered. The first action is to evoke the `GitVersion` tool, which traverses the Git history of the code and determines the version of the code automatically, as a sequence of numbers `v{Major}.{Minor}.{Patch}`. Using the computed version, the code is packaged and automatically distributed to Python Packaging Index (PyPI), Github release inside the repository, as well as a Docker image that contains the latest UQpy version.

Finally, a structured logging framework was established – in lieu of `print` commands triggered by `if` statements that were previously used to indicate errors or faults – that allows users to select the required level of severity tracked during code execution. Six different levels of severity are available in Python, namely `NOTSET`, `DEBUG`, `INFO`, `WARN`, `ERROR`, and `CRITICAL`, with the `ERROR` being the default case in UQpy. The users can choose a more verbose setting by opting for the `INFO` severity level. Logging output is then directed to their sinks of choice e.g. Terminal, Logfile, Http Streams, etc.

4. Impact

The latest version of UQpy modernizes the software to meet best practices in scientific software development, while also updating and improving functionality. This makes the package easier to use and more robust, broadens the classes of problems that it can solve, and greatly enhances the development experience. These points are critical to the widespread adoption of UQ in scientific applications. This robust yet friendly Python library is both user- and developer-friendly and provides core functionality to casual users, state-of-the-art methods for advanced users, and a carefully designed environment for developers of UQ methods. With the advent of version 4, we have seen the user-base increase as the library has been adopted by external UQ teams, and have now successfully integrated updates from third-party developers — both of which serve to advance the field of UQ.

To summarize, the entire package has been restructured from a single-file per area to a module hierarchy. Wherever possible, suboptions inside algorithms were extracted using the Strategy design pattern to enhance encapsulation and allow users to select their functionality in a more clear and straightforward manner. Baseclasses are now used throughout the code, which provides interfaces for the implementation of new algorithmic alternatives. To enhance the team collaboration efforts, the already existing version control and Github repository were supported with a CI/CD pipeline that automates software testing and code quality checks to ensure the best scientific output, while each new merge to the master is followed by package releases to PyPI, `conda-forge`, and Dockerhub image repository.

Compared to the other existing UQ packages, many of which have been listed above, the aim of UQpy is twofold. First of all, we aim to provide an extensive *fully Python-based* UQ library that addresses the wide-ranging needs of the scientific community. At the same time, we want to provide a toolbox that allows its straightforward extension with new functionalities and its use in real-world UQ applications. The developments outlined here represent significant advancements toward these two objectives.

5. Conclusions

In this work, the open-source library for uncertainty quantification UQpy and specifically the latest v4.1 was introduced. All changes and updates to the modules of the library were explained in detail, with one of the most significant being the new software development and continuous integration workflow. The latest version enables users and external collaborators to expedite the development of new features using UQpy as a platform. This is proven by the new functionalities introduced from both the development team, as well as external collaborators.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Michael Shields reports financial support was provided by US Department of Energy. Michael Shields reports financial support was provided by Defense Threat Reduction agency. Michael Shields reports financial support was provided by National Science Foundation.

Data availability

No data was used for the research described in the article.

Acknowledgments

The authors gratefully acknowledge the following individuals who contributed to code development in UQpy: Ulrich Römer, TU Braunschweig; Julius Schultz, TU Braunschweig; Prateek Bhustali, TU Braunschweig. The authors of this work have been supported by the following sources: U.S. Dept. of Energy Award Number DE-SC0020428 (DT, MDS, DG, KK) and Defense Threat Reduction Agency, United States Award Number DTRA1-20-2-0001 (DT, MDS, DG, HS); INL Laboratory Directed Research & Development (LDRD) Program under DOE Idaho Operations Office Contract DE-AC07-05ID14517 (PC); National Science Foundation, United States under award numbers 1652044 (LV) and 1930389 (MC). The content of the information does not necessarily reflect the position or the policy of the federal government, and no official endorsement should be inferred.

Appendix. UQpy future extensions

UQpy is continuously updated with new capabilities developed as a product of the research conducted in the group and through external contributions. When possible, existing capabilities are also made more robust and computationally efficient to address the needs of computationally intensive research objectives in UQ. The following describes tentative plans for future development in UQpy.

Existing modules are planned to be expanded with the following capabilities:

- **Dimension Reduction:** We are looking to enrich the module with advanced manifold projection methods including kernel PCA [27], principal geodesic analysis (PGA) [28], and Isomap [29] among others, while also adding new Euclidean/Grassmannian kernels [30–32].
- **Sampling:** We aim to expand the `StratifiedSampling` class with existing generalized of stratified sampling methods developed in the group such as: Partially Stratified Sampling (PSS) [33], Latinized Stratified Sampling (LSS) [33], and Hierarchical Latin Hypercube Sampling (HLHS) [34].
- **Surrogates:** We plan to enrich the Gaussian Processes with additional constraints such as monotonicity, convexity and differential constraints [35].
- **Surrogates:** We plan to incorporate the novel physics constrained Polynomial Chaos Expansions developed in the group [36,37].
- **Surrogates:** We plan to add Geometric Harmonics [38] for out-of-sample extensions on a manifold.
- **Sensitivity:** We are currently building the capability for Gaussian Process-based Sobol indices estimates [39].

In addition to these extension we envision the development of new modules that will further enhance the capabilities of UQpy as a general purpose uncertainty quantification library.

- **Multi-fidelity Modeling:** This module will include multi-fidelity modeling algorithms that include those based on control variates [40] as well as those using Gaussian process corrections and model selection – including methods developed in the group [41–43].
- **Neural Networks:** Algorithms for Bayesian Neural Networks [44], physics-informed Neural Networks [45] and Deep Operator Networks [46] are under development within the group and are planned to be incorporated in the future versions of UQpy – likely by coupling with other open-source Python libraries that contain computationally efficient Neural Network implementations [26, 47,48].
- **Multi-scale Modeling:** A module for stochastic hierarchical multi-scale modeling is currently under development that advances deterministic capabilities developed in this field [49].

These future enhancements will further solidify UQpy as a leading software in the field and will be made possible by the streamlined architecture, development, and integration processes described herein. Please note that the list above is subject to change in future releases.

References

- [1] Baudin M, Dutfoy A, Iooss B, Popelin A-L. OpenURNS: An industrial software for uncertainty quantification in simulation. In: Ghanem R, Higdon D, Owhadi H, editors. Handbook of uncertainty quantification. Cham: Springer International Publishing; 2017, p. 2001–38. http://dx.doi.org/10.1007/978-3-319-12385-1_64.
- [2] Martin SM, Wälchli D, Arampatzis G, Economides AE, Karnakov P, Koumoutsakos P, Korali: Efficient and scalable software framework for Bayesian uncertainty quantification and stochastic optimization. *Comput Methods Appl Mech Engrg* 2022;389:114264. <http://dx.doi.org/10.1016/j.cma.2021.114264>.
- [3] Parno M, Davis A, Seelinger L. MUQ: The MIT uncertainty quantification library. *J Open Source Softw* 2021;6(68):3076. <http://dx.doi.org/10.21105/joss.03076>.
- [4] Debusschere BJ, Najm HN, Pébay PP, Knio OM, Ghanem RG, Le Maître OP. Numerical challenges in the use of polynomial chaos representations for stochastic processes. *SIAM J Sci Comput* 2004;26(2):698–719. <http://dx.doi.org/10.1137/S1064827503427741>.
- [5] Dalbey K, Eldred MS, Geraci G, Jakeman JD, Maupin KA, Monschke JA, Seidl DT, Swiler LP, Tran A, Menhorn F, Zeng X. Dakota a multilevel parallel object-oriented framework for design optimization parameter estimation uncertainty quantification and sensitivity analysis: Version 6.12 theory manual. 2020. <http://dx.doi.org/10.2172/1630693>.
- [6] Patelli E, Broggi M, de Angelis M, Beer M. OpenCossan: An efficient open tool for dealing with epistemic and aleatory uncertainties. 2014, <http://dx.doi.org/10.1061/9780784413609.258>.
- [7] Marelli S, Sudret B. UQLab: A Framework for Uncertainty Quantification in Matlab, pp. 2554–2563 [arXiv:https://ascelibrary.org/doi/pdf/10.1061/9780784413609.257](https://ascelibrary.org/doi/pdf/10.1061/9780784413609.257) <http://dx.doi.org/10.1061/9780784413609.257>.
- [8] Dupuy D, Helbert C, Franco J. DiceDesign and DiceEval: Two r packages for design and analysis of computer experiments. *J Stat Softw* 2015;65(11):1–38. <http://dx.doi.org/10.18637/jss.v065.i11>.
- [9] Roustant O, Ginsbourger D, Deville Y. DiceKriging, DiceOptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *J Stat Softw* 2012;51(1):1–55. <http://dx.doi.org/10.18637/jss.v051.i01>.
- [10] Bourinet J-M, Matrand C, Dubourg V. A review of recent features and improvements added to FERUM software. In: Furuta H, Frangopol DM, Shinozuka M, editors. Proc. 10th International Conference on Structural Safety and Reliability (ICOSSAR 2009), Osaka, Japan, September 13–17, 2009. CRC Press; 2009.
- [11] Gorissen D, Couckuyt I, Demeester P, Dhaene T, Crombecq K. A surrogate modeling and adaptive sampling toolbox for computer based design. *J Mach Learn Res* 2010;11(68):2051–5. URL <http://jmlr.org/papers/v11/gorissen10a.html>.
- [12] Technical university of munich, TUM school of engineering and design, engineering risk analysis group. 2023, [link]. URL <https://www.cee.ed.tum.de/era/software/>.
- [13] Tennoe S, Haines G, Einevoll GT. Uncertainpy: A python toolbox for uncertainty quantification and sensitivity analysis in computational neuroscience. *Front Neuroinform* 2018;12. <http://dx.doi.org/10.3389/fninf.2018.00049>.
- [14] Puzryev V, Ghommem M, Meka S. PyROM: A computational framework for reduced order modeling. *J Comput Sci* 2019;30:157–73. <http://dx.doi.org/10.1016/j.jocs.2018.12.004>.
- [15] Šukys J, Kattwinkel M. SPUX: Scalable particle Markov chain Monte Carlo for uncertainty quantification in stochastic ecological models. 2017, [arXiv:1711.01410](https://arxiv.org/abs/1711.01410).
- [16] Dutta R, Schoengens M, Pacchiardi L, Ummadisingu A, Widmer N, Künzli P, Onnela J-P, Mira A. ABCpy: A high-performance computing perspective to approximate Bayesian computation. *J Stat Softw* 2021;100(7):1–38. <http://dx.doi.org/10.18637/jss.v100.i07>.
- [17] Herman J, Usher W. SALib: An open-source Python library for sensitivity analysis. *J Open Source Softw* 2017;2(9):97. <http://dx.doi.org/10.21105/joss.00097>.
- [18] Salvatier J, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3. *PeerJ Comput Sci* 2016;2:e55. <http://dx.doi.org/10.7717/peerj-cs.55>.
- [19] Weise K, Poßner L, Müller E, Gast R, Knösche TR. Pygpc: A sensitivity and uncertainty analysis toolbox for Python. *SoftwareX* 2020;11:100450. <http://dx.doi.org/10.1016/j.softx.2020.100450>.
- [20] Jakeman JD. PyApprox: Enabling efficient model analysis. 2022, <http://dx.doi.org/10.2172/1879614>, URL <https://www.osti.gov/biblio/1879614>.
- [21] Zou Z, Meng X, Psaros AF, Karniadakis GE. NeuralUQ: A comprehensive library for uncertainty quantification in neural differential equations and operators. 2022, [arXiv:2208.11866](https://arxiv.org/abs/2208.11866).
- [22] Olivier A, Giovanis DG, Aakash B, Chauhan M, Vandanapu L, Shields MD. UQpy: A general purpose python package and development environment for uncertainty quantification. *J Comput Sci* 2020;47:101204. <http://dx.doi.org/10.1016/j.jocs.2020.101204>.
- [23] Nájera Ó, Larson E, Estève L, Varoquaux G, Liu L, Grobler J, de Andrade ES, Huldgraf C, Gramfort A, Jas M, Hoffmann T, Grisel O, Varoquaux N, Goullart E, Luessi M, Lee A, Vanderplas J, Hoffmann T, Caswell TA, Sullivan B, Batula A, jaeilepp, Robitaille T, Appelhoff S, Kunzmann P, Geier M, Lars, Sunden K, Stańczak D, Shih AY. Sphinx-gallery/sphinx-gallery: Release v0.7.0. 2020, <http://dx.doi.org/10.5281/zenodo.3838216>.
- [24] Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. Array programming with NumPy. *Nature* 2020;585(7825):357–62. <http://dx.doi.org/10.1038/s41586-020-2649-2>.
- [25] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, Moore VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, Vijaykumar A, Bardelli AP, Rothberg A, Hilboll A, Kloeckner A, Scopatz A, Lee A, Rokem A, Woods CN, Fulton C, Masson C, Häggström C, Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik DV, Olivetti E, Martin E, Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold G-L, Allen GE, Lee GR, Audren H, Probst I, Dietrich JP, Silterra J, Webber JT, Slavič J, Nothman J, Buchner J, Kulick J, Schönberger JL, de Miranda Cardoso JV, Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Tritz K, Thoma M, Newville M, Kümmerer M, Bolingbroke M, Tartre M, Pak M, Smith NJ, Nowaczyk N, Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT, Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T, Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T, Krauss T, Upadhyay U, Halchenko YO, Vázquez-Baeza Y, 1.0 Contributors S. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 2020;17(3):261–72. <http://dx.doi.org/10.1038/s41592-019-0686-2>.
- [26] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011;12(85):2825–30, URL <http://jmlr.org/papers/v12/pedregosa11a.html>.
- [27] Schölkopf B, Smola A, Müller K-R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput* 1998;10(5):1299–319. <http://dx.doi.org/10.1162/089976698300017467>, [arXiv:https://direct.mit.edu/neco/article-pdf/10/5/1299/813905/089976698300017467.pdf](https://direct.mit.edu/neco/article-pdf/10/5/1299/813905/089976698300017467.pdf).
- [28] Fletcher P, Lu C, Pizer S, Joshi S. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans Med Imaging* 2004;23(8):995–1005. <http://dx.doi.org/10.1109/TMI.2004.831793>.
- [29] Tenenbaum JB, Silva Vd, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science* 2000;290(5500):2319–23. <http://dx.doi.org/10.1126/science.290.5500.2319>.
- [30] Harandi MT, Salzmann M, Jayasumana S, Hartley R, Li H. Expanding the family of grassmannian kernels: An embedding perspective. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. Computer vision – ECCV 2014. Cham: Springer International Publishing; 2014, p. 408–23.
- [31] Hamm J, Lee D. Extended Grassmann Kernels for subspace-based learning. In: Koller D, Schuurmans D, Bengio Y, Bottou L, editors. Advances in neural information processing systems, Vol. 21. Curran Associates, Inc.; 2008, URL https://proceedings.neurips.cc/paper_files/paper/2008/file/e7f8a7fb0b77bcb3b283af5be021448f-Paper.pdf.
- [32] Hong J, Chen H, Lin F. Disturbance Grassmann Kernels for subspace-based learning. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. KDD '18, New York, NY, USA: Association for Computing Machinery; 2018, p. 1521–30. <http://dx.doi.org/10.1145/3219819.3219959>.

- [33] Shields MD, Zhang J. The generalization of Latin hypercube sampling. *Reliab Eng Syst Saf* 2016;148:96–108. <http://dx.doi.org/10.1016/j.ress.2015.12.002>.
- [34] Vořechovský M. Hierarchical refinement of latin hypercube samples. *Comput-Aided Civ Infrastruct Eng* 2015;30(5):394–411.
- [35] Swiler LP, Gulian M, Frankel AL, Safta C, Jakeman JD. A survey of constrained Gaussian process regression: approaches and implementation challenges. *J Mach Learn Model Comput* 2020;1(2):119–56.
- [36] Novak L, Sharma H, Shields MD. On physically-constrained non-intrusive polynomial chaos expansion. In: 5th ECCOMAS thematic conference on uncertainty quantification in computational sciences and engineering. 2023.
- [37] Sharma H, Shields MD, Novak L. Constrained non-intrusive polynomial chaos expansion for physics-informed machine learning regression. In: 14th international conference on applications of statistics and probability in civil engineering, ICASP14. 2023.
- [38] Coifman RR, Lafon S. Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions. *Appl Comput Harmon Anal* 2006;21(1):31–52. <http://dx.doi.org/10.1016/j.acha.2005.07.005>, Special Issue: Diffusion Maps and Wavelets.
- [39] Marrel A, Iooss B, Laurent B, Roustant O. Calculations of Sobol indices for the Gaussian process metamodel. *Reliab Eng Syst Saf* 2009;94(3):742–51. <http://dx.doi.org/10.1016/j.ress.2008.07.008>.
- [40] Gorodetsky AA, Geraci G, Eldred MS, Jakeman JD. A generalized approximate control variate framework for multifidelity uncertainty quantification. *J Comput Phys* 2020;408:109257. <http://dx.doi.org/10.1016/j.jcp.2020.109257>.
- [41] Chakroborty P, Dhulipala SLN, Che Y, Jiang W, Spencer BW, Hales JD, Shields MD. General multi-fidelity surrogate models: Framework and active learning strategies for efficient rare event simulation. 2022, [arXiv:2212.03375](https://arxiv.org/abs/2212.03375).
- [42] Dhulipala SL, Shields MD, Chakroborty P, Jiang W, Spencer BW, Hales JD, Labouré VM, Prince ZM, Bolisetti C, Che Y. Reliability estimation of an advanced nuclear fuel using coupled active learning, multifidelity modeling, and subset simulation. *Reliab Eng Syst Saf* 2022;226:108693. <http://dx.doi.org/10.1016/j.ress.2022.108693>.
- [43] Dhulipala SL, Shields MD, Spencer BW, Bolisetti C, Slaughter AE, Labouré VM, Chakroborty P. Active learning with multifidelity modeling for efficient rare event simulation. *J Comput Phys* 2022;468:111506. <http://dx.doi.org/10.1016/j.jcp.2022.111506>.
- [44] Olivier A, Shields MD, Graham-Brady L. Bayesian neural networks for uncertainty quantification in data-driven materials modeling. *Comput Methods Appl Mech Engrg* 2021;386:114079. <http://dx.doi.org/10.1016/j.cma.2021.114079>.
- [45] Raissi M, Perdikaris P, Karniadakis GE. Physics informed deep learning (Part I): Data-driven solutions of nonlinear partial differential equations. 2017, [arXiv:1711.10561](https://arxiv.org/abs/1711.10561).
- [46] Lu L, Jin P, Pang G, Zhang Z, Karniadakis GE. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat Mach Intell* 2021;3(3):218–29. <http://dx.doi.org/10.1038/s42256-021-00302-5>.
- [47] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mane D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viegas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X. TensorFlow: Large-scale machine learning on heterogeneous distributed systems. 2016, [arXiv:1603.04467](https://arxiv.org/abs/1603.04467).
- [48] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S. PyTorch: An imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R, editors. *Advances in neural information processing systems*, Vol. 32. Curran Associates, Inc.; 2019.
- [49] Knap J, Spear C, Leiter K, Becker R, Powell D. A computational framework for scale-bridging in multi-scale simulations. *Internat J Numer Methods Engrg* 2016;108(13):1649–66. <http://dx.doi.org/10.1002/nme.5270>, [arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.5270](https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.5270).

4 Discussion & Concluding Remarks

The increasing economical and safety requirements placed on new structures together with the increasing number of existing structures, which should be adapted to new conditions, make uncertainty quantification in civil engineering more important than ever before. Unfortunately, there still exists a gap between purely scientific methods and extremely simplified techniques for industrial applications implemented in design codes, and thus further development of suitable methods and their adaptations are needed. This thesis presented several significant results obtained in the recent years from two sub-fields of UQ: simplified methods for estimation of statistical moments, and surrogate modeling using polynomial chaos expansion. Both areas contain important UQ methods allowing for industrial applications differing in complexity of the stochastic models and the associated computational costs. While simplified methods are suitable for simple structural elements represented by costly mathematical models, surrogate models provide tools for a complex stochastic analysis and UQ of physical systems.

The first chapter of the thesis summarized the theoretical background of the novel Eigen ECoV method [19] – a simplified safety format for the design and assessment of structures represented by non-linear finite element models. Eigen ECoV is based on the identical assumptions as other existing ECoV methods already implemented in design codes, though it offers a much higher adaptivity and versatility. Moreover, the whole process of Eigen ECoV derivation is simple and straightforward, and thus can be used for the derivation for various similar ECoV methods. Finally, the role of correlation among input random variables was investigated. A lack of knowledge on the joint probability distribution is typical for industrial applications, since only marginal probability distributions are usually known, although it is necessary to define a correlation structure of the input random variables in order to completely describe an input random vector [58]. The proposed correlation interval approach consequently reveals the impact of vague information about an assumed correlation structure. The theoretical derivation of Eigen ECoV is supported by the obtained numerical results presented in the second paper – a thorough comparison of existing safety formats in several numerical examples representing realistic concrete structural members failing in several modes [20]. The comparison serves also as a guideline for engineers who would like to use advanced safety formats in industrial applications, since the paper contains all the necessary values of safety factors and input quantiles. The obtained numerical results greatly correspond to the theoretical assumptions of Eigen ECoV. The correlation interval approach clearly shows the influence of an assumed correlation matrix, and quantifies the impact of vague or incomplete definitions of correlations among input random variables. Although the

proposed method works for the investigated simple structural members very well, there is still one open research question – how to derive safety formats for structures with multiple competing failure modes? This is not a trivial task, since all existing methods are typically based on the assumption of Gaussian or Lognormal distribution of the QoI. Therefore, further work in this field will be dedicated to the investigation of suitable methods for generally multi-modal probability distributions of QoIs.

A complex stochastic analysis is typically based on the Monte Carlo simulation, and so it is necessary to create an accurate and computationally efficient surrogate model of the original mathematical model. Therefore, the second chapter of this thesis deals with the efficient construction of surrogate models for various types of functions. Specifically, PCE approximations were investigated in this thesis since they achieve a high accuracy for low-size ED. The accuracy and efficiency of a PCE construction is highly affected by the optimality of ED, i.e. the positions and the number of statistical samples. The first article presented in the second chapter is thus focused on active learning for PCE [42]. The proposed method referenced as the Θ criterion combines the exploration of the design domain with the exploitation of the PCE approximator in order to identify locations of the input space associated to significant non-linearity of the approximated mathematical model. Although the presented results were superior in comparison to the state-of-the-art techniques, there is still a strong limitation for highly non-linear functions. Therefore, the second paper attached to the second chapter is focused on the input domain decomposition and construction of localized PCEs [47]. The whole methodology is driven by the Θ criterion, which leads to optimal decomposition of the input domain with respect to local contributions to a variance of QoI. The obtained numerical results show convergence for highly non-linear functions as well as for functions with discontinuities.

PCE methods can be beneficially utilized for UQ due to their possibilities for the analytical estimation of statistical moments and sensitivity indices of QoI. The third paper attached to the second chapter presents reviewed theoretical properties of orthonormal polynomials utilized in a novel distribution-based sensitivity analysis [52]. The proposed method represents a complex sensitivity measure based on the shapes of conditional probability distributions, and it can be derived analytically from PCE coefficients. Moreover, the paper presents a concept of reduced PCE, which is further utilized for UQ in the recently proposed physics-informed PCE [55]. The physically constrained PCE also referenced as PC^2 is a promising technique in the framework of physics-informed machine learning, i.e. the construction of surrogate models constrained by physical knowledge. Such techniques lead to a superior performance in comparison to standard data-driven methods using $\mathcal{M}(\mathbf{X})$ as a black box, and they also have higher possibilities for extrapolation. Moreover,

such methods represent ideal tools for realistic engineering applications based on several observations or calculations and typically limited knowledge regarding the governing physics. These are referenced as grey box modeling in contrast to the perfectly known physical models referenced as white box modeling in the framework of scientific machine learning. The general classification of surrogate modeling with respect to the amount of available data and known physics is schematically depicted in Fig. 4.1.

The PC^2 approach is the first PCE technique reflecting the given physical constraints and boundary conditions. PC^2 connects several results obtained from the author's previous research into a novel and innovative tool for scientific machine learning and UQ. The most significant benefits of PC^2 over other physics-informed machine learning techniques are its computational efficiency in low ED regime, and its natural possibilities for UQ thanks to the concept of the reduced PCE. In this pioneering work, the general theoretical background and concept of PC^2 were introduced together with practical details for the construction of constrained approximations (e.g. basis functions of space-time variables and their derivations). The proposed methodology was validated on several differential equations including non-linear ODE and PDE with random parameters.

In its current form, PC^2 can be used for linear PDEs (or slightly non-linear) and general equality constraints. However, the proposed framework can be easily extended to general constraints including non-equality constraints including monotonicity, non-negativity, and others. Therefore, further work will be focused on the generalization of PC^2 and its computational optimization. Moreover, this research opens the door for several questions that should be investigated in further research regarding suitable sampling techniques for virtual samples and boundary samples,

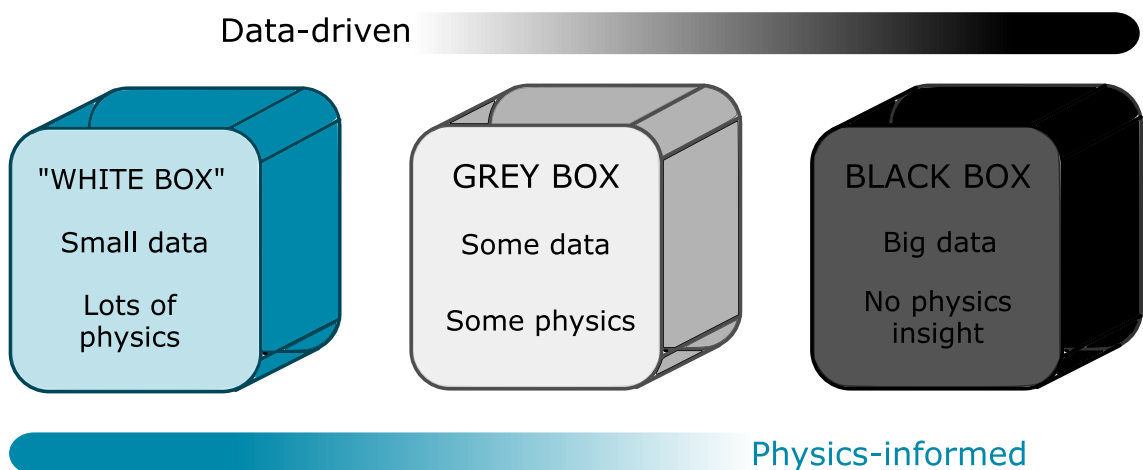


Fig. 4.1: Classifications of scientific machine learning areas.

various numerical optimization techniques, adaptivity, sparsity, general scalability, etc. ensuring further continuity of this research.

Although the developed methods represent a significant progress beyond the state-of-the-art techniques, it was necessary to provide algorithms and codes to the UQ community in order to propagate and validate the techniques. Moreover, sharing the codes together with preprints of the presented articles reflects the principles of open science, which, as the author believes, is necessary for further scientific developments. Therefore, the last paper attached to the second chapter of this thesis presents the UQPy software package containing all of the developed techniques together with various methods for UQ developed by an international team of scientists and professional software developers [57]. UQPy has gained a significant attention in the UQ community and is used by several research teams. A significant achievement can also be seen in the fact that UQPy was adopted by the Computational Modeling and Simulation Center (SimCenter), which is a part of the Natural Hazards Engineering Research Infrastructure program, USA.

This habilitation thesis summarized the recent results obtained by solving several research projects at the FCE, BUT in collaboration with international partners. The results are connected by a single task: the development of methods for *UQ in civil engineering*. It can be concluded that the presented theoretical methods based on cutting-edge techniques have the potential to become efficient tools for industrial design and the assessment of structures, as well as for further theoretical research in the fields of computational science, applied mathematics, and ultimately UQ in engineering.

Bibliography

- [1] Comité Européen de Normalisation (CEN), Brussels, Belgium. *EN 1990: Eurocode: Basis of Structural Design*, 2002.
- [2] T. J. Sullivan. *Introduction to Uncertainty Quantification*, volume 63. Springer International Publishing, 2015. doi:10.1007/978-3-319-23395-6.
- [3] Ralph C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics, USA, 2013.
- [4] Norbert Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936, 1938. doi:10.2307/2371268.
- [5] Roger G. Ghanem and Pol D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer New York, 1991. doi:10.1007/978-1-4612-3094-6.
- [6] Emilio Rosenblueth. Point estimates for probability moments. *Proceedings of the National Academy of Sciences*, 72(10):3812–3814, 1975.
- [7] John T. Christian and Gregory B. Baecher. Point-estimate method as numerical quadrature. *Journal of Geotechnical and Geoenvironmental Engineering*, 125(9):779–786, 1999.
- [8] Hendrik Schlune, Mario Plos, and Kent Gylltoft. Safety formats for nonlinear analysis tested on concrete beams subjected to shear forces and bending moments. *Engineering Structures*, 33(8):2350 – 2356, 2011.
- [9] Vladimír Červenka. Global safety format for nonlinear calculation of reinforced concrete. *Beton- und Stahlbetonbau*, 103(S1):37–42, 2008.
- [10] Lukáš Novák and Drahomír Novák. On Taylor series expansion for statistical moments of functions of correlated random variables. *Symmetry*, 12(8), 2020. doi:10.3390/sym12081379.
- [11] Michael D. McKay. Latin hypercube sampling as a tool in uncertainty analysis of computer models. In *Proceedings of the 24th Conference on Winter Simulation, WSC '92*, pages 557–564, New York, NY, USA, 1992.
- [12] Diego Lorenzo Allaix, Vincenzo Ilario Carbone, and Giuseppe Mancini. Global safety format for non-linear analysis of reinforced concrete structures. *Structural Concrete*, 14(1):29–42, 2013.

- [13] Ondřej Slowik, Drahomír Novák, Lukáš Novák, and Alfred Strauss. Stochastic modelling and assessment of long-span precast prestressed concrete elements failing in shear. *Engineering Structures*, 228:111500, 2021. doi:10.1016/j.engstruct.2020.111500.
- [14] Lukáš Novák, Drahomír Novák, and Radomír Pukl. Probabilistic and semi-probabilistic design of large concrete beams failing in shear. In *Advances in Engineering Materials, Structures and Systems: Innovations, Mechanics and Applications*. Taylor and Francis Group CRC Press, 2019.
- [15] Drahomír Novák, Lukáš Novák, Ondřej Slowik, and Alfred Strauss. Prestressed concrete roof girders: Part III – semi-probabilistic design. In *Proceedings of the Sixth International Symposium on Life-Cycle Civil Engineering (IALCCE 2018)*, pages 510–517. CRC press, Taylor and Francis group, 2018.
- [16] M. Šomodíková, L. Novák, M. Lipowczan, M. Vyhlídal, J. Doležel, D. Lehký, D. Novák, and R. Pukl. Probabilistic analysis and safety formats approaches applied for czech bridge structures under the ATCZ190 SAFEBRIDGE project. In H. Yokota and D.M. Frangopol, editors, *Proceedings of the Tenth International Conference on Bridge Maintenance, Safety and Management (IABMAS 2020)*. CRC Press, June 2020.
- [17] fib federation internationale du beton. *fib Model Code for Concrete Structures 2010*. John Wiley & Sons, Berlin, Heidelberg, 2013.
- [18] Lukas Novák and Drahomir Novák. On taylor series expansion for statistical moments of functions of correlated random variables. *Symmetry*, 12:1379, 2020.
- [19] Lukáš Novák and Drahomír Novák. Estimation of coefficient of variation for structural analysis: The correlation interval approach. *Structural Safety*, 92:102101, 2021. doi:10.1016/j.strusafe.2021.102101.
- [20] Lukáš Novák, Jan Červenka, Vladimír Červenka, Drahomír Novák, and Miroslav Sýkora. Comparison of advanced semi-probabilistic methods for design and assessment of concrete structures. *Structural Concrete*, 24(1):771–787, 2023. doi:10.1002/suco.202200179.
- [21] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math and Comput in Simulation* 55, 55(1–3):271–280, 2001.
- [22] Michael D Shields, Kirubel Teferra, Adam Hapij, and Raymond P Daddazio. Refined stratified sampling for efficient Monte Carlo based uncertainty quantification. *Reliability Engineering & System Safety*, 142:310–325, 2015.

- [23] Lixia Pan, Lukas Novák, David Lehký, Drahomir Novák, and Maosen Cao. Neural network ensemble-based sensitivity analysis in structural engineering: Comparison of selected methods and the influence of statistical correlation. *Computers & Structures*, 242:106376, 2021.
- [24] Nora Lüthen, Stefano Marelli, and Bruno Sudret. Sparse polynomial chaos expansions: Literature survey and benchmark. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):593–649, 2021. doi:10.1137/20M1315774.
- [25] D.C. Montgomery. *Design and Analysis of Experiments*. John Wiley & Sons, 2008.
- [26] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):407–451, 2004. doi:10.2307/3448465.
- [27] Géraud Blatman and Bruno Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230(6):2345–2367, 2011. doi:10.1016/j.jcp.2010.12.021.
- [28] Joel A. Tropp and Anna C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007. doi:10.1109/tit.2007.909108.
- [29] Shihao Ji, Ya Xue, and Lawrence Carin. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356, 2008. doi:10.1109/TSP.2007.914345.
- [30] W.J. Conover. On a better method for selecting input variables. unpublished Los Alamos National Laboratories manuscript, reproduced as Appendix A of “Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems” by J.C. Helton and F.J. Davis, Sandia National Laboratories report SAND2001-0417, printed November 2002., 1975. URL: <https://prod-ng.sandia.gov/techlib-noauth/access-control.cgi/2001/010417.pdf>.
- [31] Seung-Kyum Choi, Ramana V. Grandhi, Robert A. Canfield, and Chris L. Pettit. Polynomial chaos expansion with Latin hypercube sampling for estimating response variability. *AIAA Journal*, 42(6):1191–1198, 2004. doi:10.2514/1.2220.
- [32] M.E. Johnson, L.M. Moore, and D. Ylvisaker. Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 2(26):131–148, 1990. doi:10.1016/0378-3758(90)90122-B.

- [33] J.H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2(1):84–90, 1960. doi:10.1007/BF01386213.
- [34] Henri Faure and Christiane Lemieux. Generalized halton sequences in 2008: A comparative study. *ACM Transactions on Modeling and Computer Simulation*, 19(4):15:1–15:31, 2009. doi:10.1145/1596519.1596520.
- [35] I.M Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7, 1967. doi:10.1016/0041-5553(67)90144-9.
- [36] Harald Niederreiter. Point sets and sequences with small discrepancy. *Monatshefte für Mathematik*, 104(4):273–337, 1987. doi:10.1007/BF01294651.
- [37] Shu Tezuka. *Uniform Random Numbers: Theory and Practice*. The Springer International Series in Engineering and Computer Science 315. Springer, Boston, MA, 1995. doi:10.1007/978-1-4615-2317-8.
- [38] Noura Fajraoui, Stefano Marelli, and Bruno Sudret. Sequential design of experiment for sparse polynomial chaos expansions. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):1061–1085, 2017. doi:10.1137/16m1103488.
- [39] Mishal Thapa, Sameer B. Mulani, and Robert W. Walters. Adaptive weighted least-squares polynomial chaos expansion with basis adaptivity and sequential adaptive sampling. 360:112759, 2020. doi:10.1016/j.cma.2019.112759.
- [40] Paul Diaz, Alireza Doostan, and Jerrad Hampton. Sparse polynomial chaos expansions via compressed sensing and D-optimal design. 336:640–666, 2018. doi:10.1016/j.cma.2018.03.020.
- [41] Michael D. Shields. Adaptive Monte Carlo analysis for strongly nonlinear stochastic systems. *Reliability Engineering & System Safety*, 175:207–224, July 2018. doi:10.1016/j.ress.2018.03.018.
- [42] Lukáš Novák, Miroslav Vořechovský, Václav Sadílek, and Michael D. Shields. Variance-based adaptive sequential sampling for polynomial chaos expansion. *Computer Methods in Applied Mechanics and Engineering*, 386:114105, 2021. doi:10.1016/j.cma.2021.114105.
- [43] Aristeides D. Papadopoulos, Theodoros T. Zygiridis, Elias N. Glytsis, and Nikolaos V. Kantartzis. Uncertainty quantification of grating filters via a polynomial-chaos method with a variance-adaptive design domain. *Results in Optics*, 15:100627, 2024. doi:10.1016/j.rio.2024.100627.

- [44] Yangtian Li, Yangjun Luo, and Zheng Zhong. An active sparse polynomial chaos expansion approach based on sequential relevance vector machine. *Computer Methods in Applied Mechanics and Engineering*, 418:116554, 2024. doi:10.1016/j.cma.2023.116554.
- [45] Lin Chen and Hanyan Huang. Global sensitivity analysis for multivariate outputs using generalized rbf-pce metamodel enhanced by variance-based sequential sampling. *Applied Mathematical Modelling*, 126:381–404, 2024. doi:10.1016/j.apm.2023.10.047.
- [46] Jian Zhang, Weijie Gong, Xinxin Yue, Maolin Shi, and Lei Chen. Efficient reliability analysis using prediction-oriented active sparse polynomial chaos expansion. *Reliability Engineering & System Safety*, 228:108749, 2022. doi:10.1016/j.ress.2022.108749.
- [47] Lukáš Novák, Michael D. Shields, Václav Sadílek, and Miroslav Vořechovský. Active learning-based domain adaptive localized polynomial chaos expansion. *Mechanical Systems and Signal Processing*, 204:110728, 2023. doi:10.1016/j.ymsp.2023.110728.
- [48] John D. Jakeman, Akil Narayan, and Dongbin Xiu. Minimal multi-element stochastic collocation for uncertainty quantification of discontinuous functions. *Journal of Computational Physics*, 242:790–808, 2013. doi:10.1016/j.jcp.2013.02.035.
- [49] Stefano Marelli, P.-R. Wagner, Christos Lataniotis, and Bruno Sudret. Stochastic spectral embedding. *International Journal for Uncertainty Quantification*, 11(2):25–47, 2021. doi:10.1615/int.j.uncertaintyquantification.2020034395.
- [50] Gaelle Chastaing, Fabrice Gamboa, and Clémentine Prieur. Generalized Hoeffding-Sobol decomposition for dependent variables—application to sensitivity analysis. *Electronic Journal of Statistics*, 6:2420–2448, 2012. doi:10.1214/12-ejs749.
- [51] Bruno Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979, 2008. doi:10.1016/j.ress.2007.04.002.
- [52] Lukáš Novák. On distribution-based global sensitivity analysis by polynomial chaos expansion. *Computers & Structures*, 267:106808, 2022. doi:10.1016/j.compstruc.2022.106808.

- [53] E. Borgonovo. A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784, 2007. doi:10.1016/j.ress.2006.04.015.
- [54] Fabrice Gamboa, Thierry Klein, and Agnès Lagnoux. Sensitivity analysis based on Cramér–von Mises distance. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):522–548, 2018. doi:10.1137/15m1025621.
- [55] Lukáš Novák, Himanshu Sharma, and Michael D. Shields. Physics-informed polynomial chaos expansions, 2023. arXiv:2309.01697.
- [56] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.
- [57] Dimitrios Tsapetis, Michael D. Shields, Dimitris G. Giovanis, Audrey Olivier, Lukas Novak, Promit Chakroborty, Himanshu Sharma, Mohit Chauhan, Katianna Kontolati, Lohit Vandanapu, Dimitrios Loukrezis, and Michael Gardner. UQpy v4.1: Uncertainty quantification with Python. *SoftwareX*, 24:101561, 2023. doi:10.1016/j.softx.2023.101561.
- [58] Armen Der Kiureghian and Pei-Ling Liu. Structural reliability under incomplete probability information. *Journal of Engineering Mechanics*, 112:85–104, 1986. doi:10.1061/(asce)0733-9399(1986)112:1(85).