ETH*zürich*

GLOBAL SENSITIVITY ANALYSIS IN THE CONTEXT OF IMPRECISE PROBABILITIES (P-BOXES) USING SPARSE POLYNOMIAL CHAOS EXPANSIONS

R. Schöbi and B. Sudret

CHAIR OF RISK, SAFETY AND UNCERTAINTY QUANTIFICATION
STEFANO-FRANSCINI-PLATZ 5
CH-8093 ZÜRICH

Global sensitivity analysis in the context of imprecise probabilities (p-boxes) using sparse polynomial chaos expansions

R. Schöbi[1] and B. Sudret[1]

[1] *Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland*

Abstract

Global sensitivity analysis aims at determining which uncertain input parameters of a computational model primarily drives the variance of the output quantities of interest. Sobol' indices are now routinely applied in this context when the input parameters are modelled by classical probability theory using random variables. In many practical applications however, input parameters are affected by both aleatory and epistemic (so-called polymorphic) uncertainty, for which imprecise probability representations have become popular in the last decade. In this paper, we consider that the uncertain input parameters are modelled by parametric probability boxes (p-boxes). We propose interval-valued (so-called imprecise) Sobol' indices as an extension of their classical definition. An original algorithm based on the concepts of augmented space, isoprobabilistic transforms and sparse polynomial chaos expansions is devised to allow for the computation of these imprecise Sobol' indices at extremely low cost. In particular, phantoms points are introduced to build an experimental design in the augmented space (necessary for the calibration of the sparse PCE) which leads to a smart reuse of runs of the original computational model. The approach is illustrated on three analytical and engineering examples which allows one to validate the proposed algorithms against brute-force double-loop Monte Carlo simulation.

Keywords: uncertainty quantification – - global sensitivity analysis – - probability-boxes – - imprecise Sobol' indices – sparse polynomial chaos expansions

1 Introduction

Computational simulation tools, such as finite element models (FEM), are the most popular approach to model complex systems and processes in modern engineering. Such simulation tools map a set of input parameters describing the system and its environmental and operational conditions through a computational model to a so-called quantity of interest (QoI), *e.g.* performance indicators. The input parameters are often not perfectly known, due to noisy measurements, expert judgement, or intrinsic variability. Hence, each variable is modelled by an uncertainty model, as an example by a probability distribution. The uncertainty propagates through the computational model and results in an uncertain QoI.

Sensitivity analysis (SA) examines the impact of the uncertainty in the input vector of a computational model onto the uncertainty in the QoI. This is of importance in practice where the relation

1

between the input variables and the QoI is implicitly defined by a complex computer code. A large number of SA methods can be found in the literature, as reported in the reviews in Saltelli et al. (2008); Helton et al. (2006); Xu and Gertner (2008). The methods can be separated broadly into two classes, namely *local* and *global* SA methods. Local SA examines how small variations in the input variables affect the QoI, whereas global SA focuses on the entire variation of the input variables. In this paper, we look at global SA methods.

Traditionally, global SA methods are developed in the context of probability theory, *i.e.* the uncertainty in the input variables is modelled by probability distributions. However, a common situation in practice is to have only scarce or incomplete knowledge to characterize the uncertainty. This leads to the concept of epistemic uncertainty (representing the lack of data and lack of knowledge) alongside with aleatory uncertainty (natural variability) as a source of uncertainty. For such cases, *imprecise probability* frameworks have been developed as a generalization of probability theory. Among these frameworks are Bayesian hierarchical models Gelman et al. (2009), probability-boxes Ferson and Ginzburg (1996), and fuzzy distributions Möller and Beer (2004).

The number of publications for *imprecise sensitivity analysis* (ISA) remains small. Li & Mahadevan Li and Mahadevan (2016b,a), Sankararaman & Mahadevan Sankararaman and Mahadevan (2013), and Krzykacz-Hausmann Krzykacz-Hausmann (2006) describe global SA methods in the presence of Bayesian hierarchical models. Pinching and imprecise Sobol' indices are discussed in Oberguggenberger and Fellin (2005); Oberguggenberger et al. (2009). Pinching of probability-boxes is also elaborated in Ferson and Troy Tucker (2006). Helton et al. Helton et al. (2006) discuss variance-based algorithm in the context of evidence theory.

SA methods, and in particular ISA methods, usually require a large number of evaluations of the computational model for different realizations of the input vector. Hence, such analyses become intractable when the computational model is an expensive-to-evaluate function. *Meta-models*, a.k.a. *surrogate models* and *emulators*, are a popular solution to lower the computational costs by approximating the computational model by a cheap-to-evaluate analytical function. Examples include polynomial chaos expansions (PCE) (Xiu and Karniadakis, 2002; Ghanem and Spanos, 2003), Kriging (a.k.a. Gaussian process models) (Sacks et al., 1989; Santner et al., 2003), support vector machines (Gunn, 1998; Smola and Schölkopf, 2006), and artificial neural networks in the context of reliability analysis (Hurtado and Alvarez, 2001; Schueremans and Van Gemert, 2005).

A naive use of meta-models consists in the implementation of a sequential algorithm: first, calibrate a meta-model and subsequently conduct SA (or ISA) on the approximation. There are, however, more elaborate ways to use meta-models in the context of SA. As an example, based on polynomial chaos expansions (PCE) Ghanem and Spanos (2003), Sudret Sudret (2006, 2008) computes Sobol' indices and Sudret & Mai Sudret and Mai (2015) compute derivative-based global sensitivity measures (DGSM) analytically, see also Alexanderian et al Alexanderian et al. (2012) and Le Gratiet et al Le Gratiet et al. (2017). Konakli & Sudret Konakli and Sudret (2016) also obtain analytical expressions for Sobol' indices when the surrogate model is a low-rank tensor representation. However, the meta-model-based approaches are to-date limited to the probabilistic context.

In this paper, the standard Sobol' indices Sobol' (1993) are first extended to the context of probability-boxes in order to account for mixed epistemic and aleatory uncertainty in global sensitivity analysis. Moreover, making use of the PCE-based Sobol' indices Sudret (2008) as well as a recently published variation of PCE Schöbi and Sudret (2015), a novel approach is proposed to compute imprecise Sobol' indices efficiently. Applying the proposed approach, ISA are expected to become tractable for realistic engineering problems.

This paper is organized as follows. Section 2 recapitulates the basics of PCE and Sobol' indices in the standard setting of pure probabilistic modelling of input variables. Section 3 defines probability-boxes and derives the definition of imprecise Sobol' indices. In Section 4, a smart PCE model is introduced, which is later used to compute the imprecise Sobol' indices in Section 5. Three applications illustrate the workings of the proposed approach in Section 6. The paper is terminated with conclusions in Section 7.

2 Polynomial chaos expansions for global sensitivity analysis

2.1 Computational model

Let us define a computational model $\mathcal{M}$ as a mapping from the M-dimensional input space $\mathcal{D}_{\boldsymbol{X}}$ to the one-dimensional output space $\mathcal{D}_Y$:

$$\mathcal{M}: \ \boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}} \subset \mathbb{R}^M \mapsto y \in \mathcal{D}_Y \subset \mathbb{R}. \tag{1}$$

Herein, we assume that the model predicts a *scalar* quantity of interest y once it is fed with a vector of input parameters $\boldsymbol{x}$. Extensions to vector-valued models are straightforward by applying the proposed methodology to each component separately.

Due to uncertainties in the input vector, the latter is represented by a random vector $\boldsymbol{X}$ with joint cumulative distribution function (CDF) $F_{\boldsymbol{X}}$. The components of $\boldsymbol{X} = (X_1, \ldots, X_M)^{\mathsf{T}}$ are assumed independent for the sake of simplicity throughout this paper. The model response becomes a random variable Y obtained by propagating the uncertainties in the input random vector $\boldsymbol{X}$ through the computational model $\mathcal{M}$.

Uncertainty propagation is typically conducted by sampling-based algorithms such as Monte Carlo simulation (Robert and Casella, 2004). When the computational model is an expensive-to-evaluate function, the estimation of Y is costly due to the repeated evaluations of the computational model. Hence, a number of techniques have been proposed in the literature for surrogating $\mathcal{M}$. In the following section, Polynomial Chaos Expansions (Xiu and Karniadakis, 2002; Ghanem and Spanos, 2003; Soize and Ghanem, 2004) are introduced.

2.2 Sparse Polynomial Chaos Expansions

2.2.1 Polynomial approximation

Polynomial Chaos Expansions (PCE) approximate the computational model $\mathcal{M}$ by a series of multivariate polynomials that are orthogonal to the distributions of the input variables (Ghanem and Spanos, 2003; Soize and Ghanem, 2004):

$$Y \approx \mathcal{M}^{(\mathrm{P})}\left(\boldsymbol{X}\right) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}\left(\boldsymbol{X}\right), \tag{2}$$

where $\psi_{\boldsymbol{\alpha}}\left(\boldsymbol{X}\right)$ are multivariate orthonormal polynomials indexed by multi-indices $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)$, $\mathcal{A} \subset \mathbb{N}^M$ is a finite set of such indices (so-called truncation scheme), and $a_{\boldsymbol{\alpha}} \in \mathbb{R}$ are expansion coefficients. To derive the orthonormal polynomial basis, one proceeds as follows. Since the components of $\boldsymbol{X}$ are assumed independent, its joint probability density function (PDF) is the product of its marginals. Then, a functional inner product for each marginal PDF f_{X_i} is defined by:

$$\langle \phi_1, \phi_2 \rangle_i = \int_{\mathcal{D}_{X_i}} \phi_1(x)\, \phi_2(x)\, f_{X_i}(x)\, \mathrm{d}x, \tag{3}$$

3

for any two functions $\{\phi_1, \phi_2\}$ such that the integral exists. For each variable X_i, an orthonormal polynomial basis can be constructed that satisfies (Xiu and Karniadakis, 2002):

$$\langle P_j^{(i)}, P_k^{(i)} \rangle = \int_{\mathcal{D}_{X_i}} P_j^{(i)}(x) P_k^{(i)}(x) f_{X_i}(x) \, \mathrm{d}x = \delta_{jk}, \tag{4}$$

where $\left\{ P_j^{(i)}, P_k^{(i)} \right\}$ are two univariate polynomials in the ith variable and δ_{jk} is the Kronecker symbol which is equal to 1 for $j = k$ and equal to 0 otherwise. Xiu & Karniadakis (Xiu and Karniadakis, 2002) summarize orthogonal bases for classical probability distributions.

The multivariate polynomials are then obtained by tensor product of the univariate ones:

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \prod_{i=1}^{M} P_{\alpha_i}^{(i)}(x_i). \tag{5}$$

Through the tensor product construction, the orthogonality property is kept, meaning that:

$$\mathbb{E}\left[\psi_{\boldsymbol{\alpha}}(\boldsymbol{X}) \psi_{\boldsymbol{\beta}}(\boldsymbol{X})\right] \stackrel{\text{def}}{=} \int_{\mathcal{D}_{\boldsymbol{x}}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \psi_{\boldsymbol{\beta}}(\boldsymbol{X}) f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \delta_{\boldsymbol{\alpha\beta}}, \tag{6}$$

where $\delta_{\boldsymbol{\alpha\beta}}$ is the Kronecker delta with $\delta_{\boldsymbol{\alpha\beta}} = 1$ for $\boldsymbol{\alpha} = \boldsymbol{\beta}$ and $\delta_{\boldsymbol{\alpha\beta}} = 0$ otherwise.

2.2.2 Truncation schemes

The efficiency of PCE models greatly depends on the choice of the set of multi-indices $\mathcal{A}$ in Eq. (2). A popular strategy to choose $\mathcal{A}$ consists in upper-bounding the total degree of polynomials $\sum_{i=1}^{M} \alpha_i$ to a maximum value p or in using hyperbolic index sets (Blatman and Sudret, 2010):

$$\mathcal{A}_q^{M,p} = \left\{ \boldsymbol{\alpha} \in \mathbb{N}^M : \ ||\boldsymbol{\alpha}||_q \leq p \right\}, \qquad ||\boldsymbol{\alpha}||_q = \left(\sum_{i=1}^{M} \alpha_i^q \right)^{1/q}, \tag{7}$$

where $0 < q \leq 1$ is a parameter and p is the maximal total degree of the univariate polynomials retained.

2.2.3 Computation of coefficients $a_{\boldsymbol{\alpha}}$

After defining the set of candidate polynomials, the coefficients $a_{\boldsymbol{\alpha}}$ in Eq. (2) are to be determined. Based on a finite set of realizations of the input vector $\boldsymbol{X}$, denoted by $\boldsymbol{\mathcal{X}} = \left\{ \boldsymbol{\chi}^{(1)}, \ldots, \boldsymbol{\chi}^{(N)} \right\}$ and called experimental design, and the corresponding response values $\mathcal{Y} = \left\{ \mathcal{Y}^{(1)}, \ldots, \mathcal{Y}^{(N)} \right\} \stackrel{\text{def}}{=} \left\{ \mathcal{M}\left(\boldsymbol{\chi}^{(1)}\right), \ldots, \mathcal{M}\left(\boldsymbol{\chi}^{(N)}\right) \right\}$, the coefficients are calculated by discretized least-squares minimization Berveiller et al. (2006); Sudret (2007):

$$\widehat{\boldsymbol{a}} = \arg \min_{\boldsymbol{a} \in \mathbb{R}^{|\mathcal{A}|}} \frac{1}{N} \sum_{i=1}^{N} \left(\mathcal{Y}^{(i)} - \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}\left(\boldsymbol{\chi}^{(i)}\right) \right)^2, \tag{8}$$

where $n_{\mathcal{A}} = |\mathcal{A}|$ is the number of polynomials retained in the truncated expansion. The optimal coefficients $\widehat{\boldsymbol{a}}$ can be obtained by solving the linear system:

$$\widehat{\boldsymbol{a}} = \left(\mathbf{F}^{\mathsf{T}} \mathbf{F} \right)^{-1} \mathbf{F}^{\mathsf{T}} \mathcal{Y}. \tag{9}$$

In this expression, the information matrix $\mathbf{F}$ contains the values of all selected polynomials at all points of the experimental design, namely:

$$F_{ij} = \psi_j\left(\boldsymbol{\chi}^{(i)}\right), \qquad i = 1, \ldots, N, \qquad j = 1, \ldots, n_{\mathcal{A}}. \tag{10}$$

4

2.2.4 Polynomial selection

Typically for smooth functions, a small number of polynomials is sufficient to accurately represent the output of the computational model. Thus, a further reduction of the set of predictors is possible. A number of methods are available in the literature, including the *least absolute shrinkage and selection operator* (LASSO) (Tibshirani, 1996), *least angle regression* (LAR) (Efron et al., 2004), *orthogonal matching pursuit* (OMP) (Pati et al., 1993; Mallat and Zhang, 1993), and *Bayesian compressive sensing* (Sargsyan et al., 2014). Based on LAR, Blatman & Sudret Blatman and Sudret (2011) proposed the LARS algorithm which allows for an efficient selection of a small number of polynomials out of the candidate set $\mathcal{A}_q^{M,p}$ (see also Eq. (7)). For a thorough discussion of the LARS algorithm, the reader is referred to Blatman and Sudret (2011); Blatman (2009).

2.3 Sobol' indices

The so-called *Sobol' decomposition* represents the computational model by a series of summands of increasing dimension (Sobol', 1993):

$$\mathcal{M}(\boldsymbol{x}) = \mathcal{M}_0 + \sum_{i=1}^{M} \mathcal{M}_i(x_i) + \sum_{1 \leq i < j \leq M} \mathcal{M}_{ij}(x_i, x_j) + \ldots$$
$$+ \sum_{1 \leq i_1 < \ldots < i_s \leq M} \mathcal{M}_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s}) + \ldots + \mathcal{M}_{1,2,\ldots,M}(\boldsymbol{x}), \quad (11)$$

where $\mathcal{M}_0$ is a constant (mean value of the function) and where it is imposed that the integral of each summand $\mathcal{M}_{i_1,\ldots,i_s}(x_{i_1}, \ldots, x_{i_s})$ over any of its arguments is zero:

$$\int_{\mathcal{D}_{X_k}} \mathcal{M}_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s}) f_{X_k}(x_k) \, \mathrm{d}x_k = 0, \qquad 1 \leq i_1 < \ldots < i_s \leq M, \qquad k \in \{i_1, \ldots, i_s\}. \quad (12)$$

The mean value of the function can be computed as:

$$\mathcal{M}_0 = \int_{\mathcal{D}_{\boldsymbol{X}}} \mathcal{M}(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \quad (13)$$

Due to the constraints imposed in Eq. (12), the summands are orthogonal to each other so that (Homma and Saltelli, 1996):

$$\int_{\mathcal{D}_{\boldsymbol{X}}} \mathcal{M}_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s}) \mathcal{M}_{j_1 \ldots j_t}(x_{j_1,\ldots,j_t}) f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = 0, \qquad \{i_1, \ldots, i_s\} \neq \{j_1, \ldots, j_t\}. \quad (14)$$

Considering now that the input parameters are modelled by independent random variables, the *total variance* of the computational model is defined as:

$$D = \mathrm{Var}\left[\mathcal{M}(\boldsymbol{X})\right] = \int_{\mathcal{D}_{\boldsymbol{X}}} \mathcal{M}^2(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} - \mathcal{M}_0^2. \quad (15)$$

By using the properties of Eq. (14) in Eq. (11), one obtains the following decomposition of the variance:

$$D = \sum_{i=1}^{M} D_i + \sum_{1 \leq i < j \leq M} D_{ij} + \ldots + D_{1,2,\ldots,M}, \quad (16)$$

where the *partial variances* are computed as:

$$D_{i_1,\ldots,i_s} = \int_{\mathcal{D}_{\boldsymbol{X}}} \mathcal{M}_{i_1 \ldots i_s}^2(x_{i_1}, \ldots, x_{i_s}) f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, \qquad 1 \leq i_1 < \ldots < i_s \leq M, \qquad s = 1, \ldots, M. \quad (17)$$

Making use of this decomposition, the Sobol' indices are defined as the relative partial variances:

$$S_{i_1 \dots i_s} = \frac{D_{i_1 \dots i_s}}{D} = \frac{D_{i_1 \dots i_s}}{\sum_{i=1}^{M} D_i + \sum_{1 \leq i < j \leq M} D_{ij} + \dots + D_{1,2,\dots,M}}, \tag{18}$$

and by virtue of Eq. (16), it holds that:

$$\sum_{i=1}^{M} S_i + \sum_{1 \leq i < j \leq M} S_{ij} + \dots + S_{1,2,\dots,M} = 1. \tag{19}$$

The Sobol' indices measure the amount of the total variance coming from the uncertainties in a set of input parameters. In practice, it is common to compute the *first order* indices S_i, which measure the influence of each parameter taken separately. Higher order indices account for the interactive contributions to the total variance. In this sense, the *total sensitivity indices* $S_i^{(T)}$ are defined as:

$$S_i^{(T)} = \sum_{\{i_1,\dots,i_s\} \supset \{i\}} \frac{D_{i_1 \dots i_s}}{D}. \tag{20}$$

In other words, $S_i^{(T)} = 1 - S_{-i}$, where S_{-i} is the sum of all $S_{i_1 \dots i_s}$ that do not include index i.

2.4 PCE-based Sobol' indices

Let us consider now that an appropriate approximation of the computational model $\mathcal{M}$ exists in the form of a truncated PCE, as seen in Eq. (2). Then, a set of multi-indices $\mathcal{I}_{i_1,\dots,i_s}$ can be defined so that only the indices $i_1, \dots, i_s$ are non-zero, *i.e.*:

$$\mathcal{I}_{i_i,\dots,i_s} = \left\{ \boldsymbol{\alpha} \in \mathcal{A} : \begin{array}{ll} \alpha_k > 0 & \text{for all} \quad k \in \{i_1,\dots,i_s\} \\ \alpha_k = 0 & \text{for all} \quad k \notin \{i_1,\dots,i_s\} \end{array} \right\}. \tag{21}$$

The elements of Eq. (2) can be reordered according to the decomposition in Eq. (11) (Sudret, 2008):

$$\mathcal{M}^{(\mathrm{P})}(\boldsymbol{x}) = a_0 + \sum_{i=1}^{N} \left(\sum_{\boldsymbol{\alpha} \in \mathcal{I}_i} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(x_i) \right) + \sum_{1 \leq i_1 < i_2 \leq M} \left(\sum_{\boldsymbol{\alpha} \in \mathcal{I}_{i_1,i_2}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(x_{i_1}, x_{i_2}) \right) + \dots$$

$$+ \sum_{1 \leq i_1 < \dots < i_s \leq M} \left(\sum_{\boldsymbol{\alpha} \in \mathcal{I}_{i_1,\dots,i_s}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(x_{i_1}, \dots, x_{i_s}) \right) + \dots + \sum_{\boldsymbol{\alpha} \in \mathcal{I}_{1,2,\dots,M}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(x_1, \dots, x_M). \tag{22}$$

Then, the summands of Eq. (22) can be identified as summands of the Sobol' decomposition:

$$\mathcal{M}_{i_1 \dots i_s}(x_{i_1}, \dots, x_{i_s}) = \sum_{\boldsymbol{\alpha} \in \mathcal{I}_{i_1,\dots,i_s}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(x_{i_1}, \dots, x_{i_s}). \tag{23}$$

Due to the property of uniqueness, it can be concluded that Eq. (22) is indeed the Sobol' decomposition of the PCE model. Furthermore, due to the orthonormality of the PC basis, the partial variance $D_{i_1 \dots i_s}$ is readily obtained from Eq. (23) as:

$$D_{i_1 \dots i_s} = \sum_{\boldsymbol{\alpha} \in \mathcal{I}_{i_1,\dots,i_s}} a_{\boldsymbol{\alpha}}^2. \tag{24}$$

The corresponding *polynomial-chaos-based Sobol' indices*, denoted by $S_{i_1 \dots i_s}^{(\mathrm{P})}$, are then defined, as originally proposed in Sudret (2006), as:

$$S_{i_1 \dots i_s}^{(\mathrm{P})} = \sum_{\boldsymbol{\alpha} \in \mathcal{I}_{i_1,\dots,i_s}} a_{\boldsymbol{\alpha}}^2 \Bigg/ \sum_{\boldsymbol{\alpha} \in \mathcal{A}, \, \boldsymbol{\alpha} \neq \mathbf{0}} a_{\boldsymbol{\alpha}}^2 . \tag{25}$$

The *total PC-based sensitivity indices* are then computed analogously to Eq. (20):

$$S_i^{(T)(\mathrm{P})} = \sum_{\{i_1,\ldots,i_s\} \supset \{i\}} S_{i_1,\ldots,i_s}^{(\mathrm{P})}. \tag{26}$$

As seen in this section, the computation of the Sobol' indices for a PCE model boils down to working with the coefficients of the PC-expansion, which is a simple operation at almost no additional computational cost. Hence, PCE-based Sobol' indices are obtained by simple post-processing of an existing PCE model.

3 Imprecise Sobol' indices

3.1 Free and parametric p-boxes

Sobol' indices can be computed by the procedure described in the above section as soon as the input parameters are modelled by well-defined probability density functions. In practice however, probability theory is often not fully appropriate to characterize the uncertainty in the system. Probability-boxes (p-boxes) pose an intuitive generalization of probability theory. Instead of representing X by a single CDF, lower and upper bounds denoted by $\underline{F}_X$ and $\overline{F}_X$ are introduced (Ferson and Ginzburg, 1996; Ferson and Hajagos, 2004). For any value of $x \in \mathcal{D}_X$, the true-but-unknown CDF lies within these bounds such that $\underline{F}_X(x) \leq F_X(x) \leq \overline{F}_X(x)$, which form a so-called p-box.

In the literature, two types of p-boxes can be found, namely *free* and *parametric* p-boxes. Free p-boxes are defined directly though the formulation of the bounds $\{\underline{F}_X, \overline{F}_X\}$. Parametric p-boxes (a.k.a distributional p-boxes) are defined as cumulative distribution functions with interval-valued hyper-parameters: $F_X(x) = F_X(x|\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}} \subset \mathbb{R}^{n_{\boldsymbol{\theta}}}$, and $\mathcal{D}_{\boldsymbol{\Theta}}$ is the interval domain of the distribution parameters of dimension $n_{\boldsymbol{\theta}} = |\boldsymbol{\theta}|$. Precisely, $\mathcal{D}_{\boldsymbol{\Theta}} = \left[\underline{\theta}_1, \overline{\theta}_1\right] \times \ldots \times \left[\underline{\theta}_{n_{\boldsymbol{\theta}}}, \overline{\theta}_{n_{\boldsymbol{\theta}}}\right]$ denotes a hyper-rectangular domain, and $\underline{\theta}_k$ and $\overline{\theta}_k$ denote the lower and upper bounds of the interval $\left[\underline{\theta}_k, \overline{\theta}_k\right]$ for the k-th parameter of the CDF of X. Boundary curves for parametric p-boxes can be defined as for free p-boxes as follows:

$$\underline{F}_X^{(p)}(x) = \min_{\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}} F_X(x|\boldsymbol{\theta}), \qquad \overline{F}_X^{(p)}(x) = \max_{\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}} F_X(x|\boldsymbol{\theta}). \tag{27}$$

Figure 1 shows the intrinsic difference between free and parametric p-boxes. In both cases, only non-decreasing realizations of the true-but-unknown CDF are feasible. However, free p-boxes are a more general description of uncertainty than parametric p-boxes. This can be seen by the stair-case realizations, which are possible in free but not in parametric p-boxes (see also Realization #3 in Figure 1(A)).

In this paper, the focus lies on the parametric p-boxes due to their clear separation of aleatory and epistemic uncertainty: the distribution family represents the aleatory uncertainty, whereas the interval-valued hyper-parameters represent the epistemic uncertainty.

3.2 Imprecise Sobol' indices

In the context of parametric p-boxes, uncertainty quantification is typically conducted by nested Monte Carlo simulations (nMCS) algorithms where an outer loop samples the parameters $\boldsymbol{\theta}$ and an inner loop samples the conditional probability distribution $F_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\theta})$ (Eldred and Swiler, 2009; Chowdhary and Dupuis, 2013). Taking this idea into the context of sensitivity analysis, the Sobol' indices can be computed for each realization $\boldsymbol{\theta}^{(j)}$ and the corresponding $F_{\boldsymbol{X}}\left(\boldsymbol{x}\middle|\boldsymbol{\theta}^{(j)}\right)$. Each realization can result in
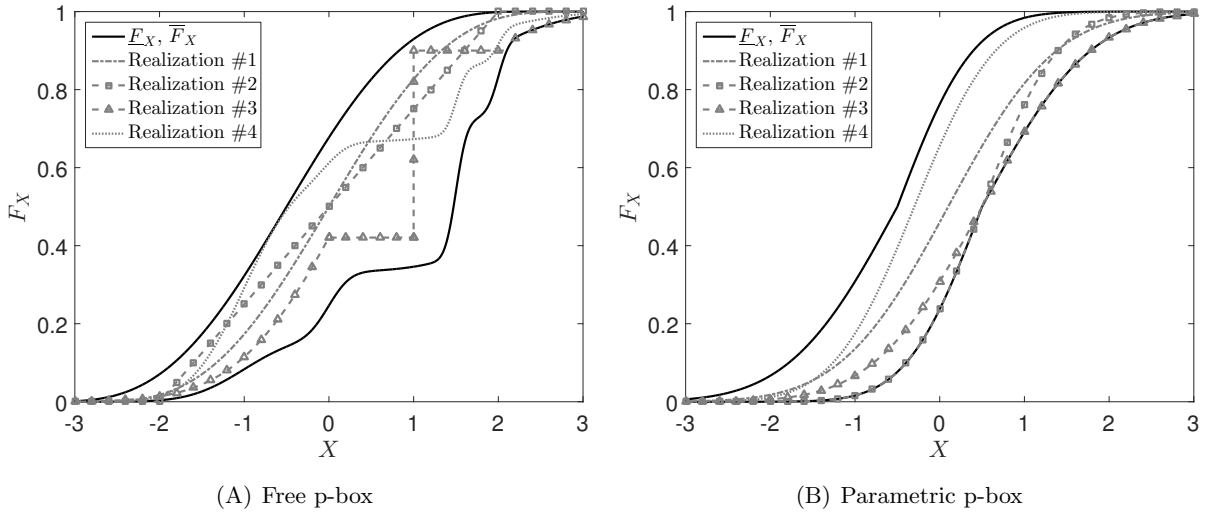
(A) Free p-box   (B) Parametric p-box

Figure 1: Examples of the two types of p-boxes

a different set of Sobol' indices. Finally, the bounds of the imprecise Sobol' indices can be determined by analysing the set of realizations.

In the sequel, we define imprecise Sobol' indices $S_{i_1 \ldots i_s}$ as a pair of bounds obtained by minimizing (resp. maximizing) the classical Sobol' indices over the range of the hyper-parameters $\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}$:

$$\underline{S}_{i_1 \ldots i_s} = \min_{\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}} S_{i_1 \ldots i_s} (\boldsymbol{\theta}), \qquad \overline{S}_{i_1 \ldots i_s} = \max_{\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}} S_{i_1 \ldots i_s} (\boldsymbol{\theta}), \tag{28}$$

where $S_{i_1 \ldots i_s} \in \left[\underline{S}_{i_1 \ldots i_s}, \overline{S}_{i_1 \ldots i_s} \right]$ are the interval-valued Sobol' indices and $S_{i_1 \ldots i_s} (\boldsymbol{\theta})$ is the corresponding Sobol' index for the conditional input distribution $F_{\boldsymbol{X}} (\boldsymbol{x} | \boldsymbol{\theta})$. Similarly to Eq. (26), the bounds of the total Sobol' indices can be obtained by:

$$\underline{S}_j^{(T)} = \min_{\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}} S_j^{(T)} (\boldsymbol{\theta}), \qquad \overline{S}_j^{(T)} = \max_{\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}} S_j^{(T)} (\boldsymbol{\theta}). \tag{29}$$

3.3 PCE-based indices

From the definitions above, it can be seen that computing the two bounds of imprecise Sobol' indices requires solving two optimization problems over the hyper-parameter space $\mathcal{D}_{\boldsymbol{\Theta}}$. In order to speed up the process, we propose to use polynomial chaos expansions as presented in Section 2.2.

The use of the PCE meta-models would reduce the computational effort of computing Sobol' indices conditioned on a particular $\boldsymbol{\theta}$ but does not remove the repeated computation of the variance decomposition for many $\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}$ to solve the optimization problems. In other words, when adopting a brute force procedure, a large number of PCE models should be calibrated, each of them with its own experimental design, own evaluations of the costly computational model $\mathcal{M}$, etc. Hence in the following, we introduce a smart PCE based on a *single experimental design* that is sufficient to estimate any $S_i (\boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}$.

4 Augmented polynomial chaos expansions for parametric distributions

4.1 Augmented space

Due to the clear separation of aleatory and epistemic uncertainty in the formulation of parametric p-boxes, these sources of uncertainty can be treated as separate entities, *i.e.* $\boldsymbol{X}$ and $\boldsymbol{\Theta}$. Hence, the QoI can be interpreted as a function of the augmented input vector $(\boldsymbol{X}, \boldsymbol{\Theta})$. The corresponding map then reads:

$$W = \mathcal{M}^{(\text{aug})}(\boldsymbol{X}, \boldsymbol{\Theta}), \tag{30}$$

where the *augmented computational model* is based on the original computational model as $\mathcal{M}^{(\text{aug})}(\boldsymbol{x}, \boldsymbol{\theta}) \equiv \mathcal{M}(\boldsymbol{x})$.

The components of the augmented input vector are dependent on each other because $\boldsymbol{X}$ depends on $\boldsymbol{\Theta}$ as a result of the hierarchical formulation of the parametric p-box. The dependencies are visualized in Figure 2(A). In this paper, however, the input variables to the computational model are assumed to be independent. In order to obtain independent variables in the augmented input vector, an isoprobabilistic transform is used that maps $(\boldsymbol{X}, \boldsymbol{\Theta})$ to $\boldsymbol{V}$, which shall be a vector with independent components. Assuming that the components X_i are independent and that the parameter intervals in $\boldsymbol{\Theta}$ are independent too (*i.e.* $\mathcal{D}_{\boldsymbol{\Theta}}$ is a hyper-rectangular domain), the parametric p-box $\boldsymbol{X}$ can be recast as:

$$X_i = F_{X_i}^{-1}(C_i|\boldsymbol{\theta}_i), \quad i = 1, \dots, M. \tag{31}$$

where $\boldsymbol{C} = (C_1, \dots, C_M)$ is a vector of independent uniform distributions and where $C_i \sim \mathcal{U}(0, 1)$ describes the CDF value of X_i. Then, the augmented vector $\boldsymbol{V} = (\boldsymbol{C}, \boldsymbol{\Theta})$ is a vector with independent components that allows to define explicitly any realization of $\boldsymbol{X}$ through Eq. (31).

(A) Original dependencies (B) Independent input variables
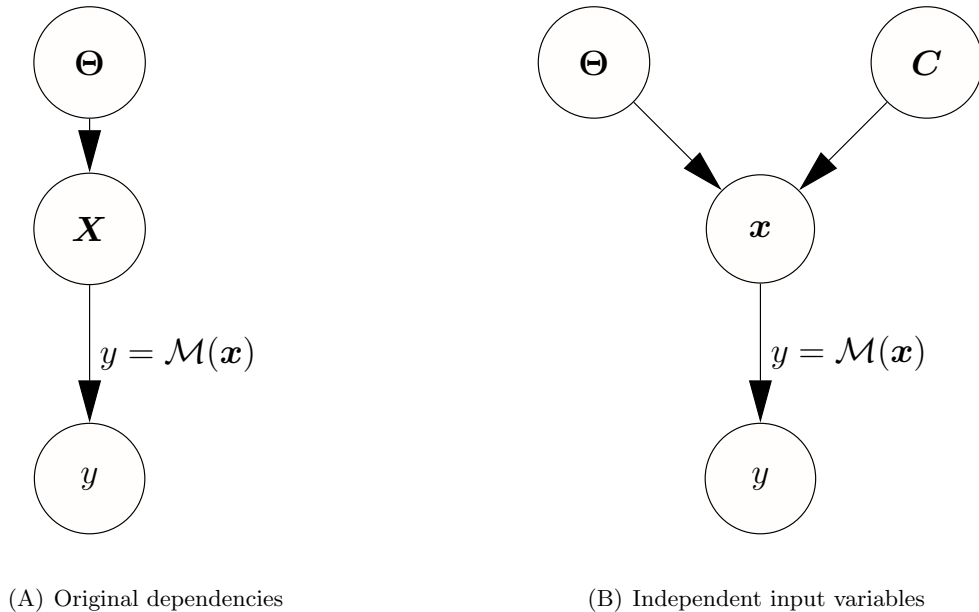
Figure 2: Parametric p-box – dependency structures before and after isoprobabilistic transform

The corresponding probabilistic graphical model transform is visualized in Figure 2(B). Note that

$\boldsymbol{x}$ is used in lower case due to the deterministic transform of $\boldsymbol{V} = (\boldsymbol{C}, \boldsymbol{\Theta}) \mapsto \boldsymbol{X}$. In Figure 2(A), the transform of $\boldsymbol{\Theta}$ to $\boldsymbol{X}$ is not deterministic, hence, the upper case $\boldsymbol{X}$.

4.2 Augmented PCE

The augmented computational model $\mathcal{M}^{(\text{aug})}$ can be formulated as a function of vector $\boldsymbol{V}$ as follows:

$$W = \mathcal{M}^{(\text{aug})}\left(T\left(\boldsymbol{V}\right)\right), \tag{32}$$

where T denotes the isoprobabilistic transform from $\boldsymbol{V}$ to $(\boldsymbol{X}, \boldsymbol{\Theta})$. When interpreting the epistemic intervals in $\boldsymbol{\Theta}$ as uniform distributions, a sparse PCE model can be calibrated to meta-model the augmented computational model (analogously to Eq. (2)):

$$W \approx \mathcal{M}^{(\text{PCE})}\left(\boldsymbol{V}\right) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \ \psi_{\boldsymbol{\alpha}}\left(\boldsymbol{V}\right), \tag{33}$$

where the dimensionality is $M = |\boldsymbol{V}| = |\boldsymbol{C}| + |\boldsymbol{\Theta}| \equiv |\boldsymbol{X}| + |\boldsymbol{\Theta}|$. The meta-model should be trained with an experimental design $\mathfrak{V} = \left\{\mathfrak{v}^{(1)}, \dots, \mathfrak{v}^{(N)}\right\}$ and the corresponding responses:

$$\mathfrak{W} = \left\{\mathfrak{W}^{(1)} = \mathcal{M}^{(\text{aug})}\left(T\left(\mathfrak{v}^{(1)}\right)\right), \dots, \mathfrak{W}^{(N)} = \mathcal{M}^{(\text{aug})}\left(T\left(\mathfrak{v}^{(N)}\right)\right)\right\},$$

by using Eq. (32). In order to efficiently calibrate a sparse PCE meta-model, the reader is referred to Section 2.2 for further details. Note that in this setup, the input vector $\boldsymbol{V}$ is sampled in the same way for the $\boldsymbol{C}$- and the $\boldsymbol{\Theta}$-components.

4.3 Phantom points

4.3.1 Isoprobabilistic transform

The number of dimensions in the augmented input space $\mathcal{D}_{\boldsymbol{V}}$ is larger than the one of the original input space $\mathcal{D}_{\boldsymbol{X}}$ due to the epistemic uncertainty. However, Eq. (31) leads to an interesting feature of the augmented PCE. For a given realization of $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}}$, there is an infinite number of realizations $\boldsymbol{v}$ that satisfy Eq. (31). In other words, the inverse operation of Eq. (31) is non-unique. This characteristics can be exploited when constructing the experimental design of the augmented PCE model.

Consider again the experimental design $\mathfrak{V}$. Each sample $\mathfrak{v}^{(j)}$ can be transformed into an equivalent $\boldsymbol{\chi}^{(j_0)}$ using Eq. (31). A number of realizations $\boldsymbol{v} = (\boldsymbol{c}, \boldsymbol{\tau})$ can be generated by sampling $\boldsymbol{\tau} \in \mathcal{D}_{\boldsymbol{\Theta}}$ and inserting them into the inverse isoprobabilistic transform to obtain realizations $\boldsymbol{c}$ that all correspond to the single point $\boldsymbol{\chi}^{(j_0)}$ in the original experimental design. In particular, for each dimension $i = 1, \dots, M$ and experimental design sample $j = 1, \dots, N$, n_{ph} samples can be generated from $\theta_i \in \mathcal{D}_{\boldsymbol{\Theta}_i}$ and collected in $\left\{\tau_i^{(j)(k)}, k = 1, \dots, n_{ph}\right\}$. Then, the inverse transform consists of evaluating simple CDFs and reads:

$$\mathfrak{c}_i^{(j)(k)} = F_{X_i}\left(\chi_i^{(j)} \Big| \tau_i^{(j)(k)}\right), \tag{34}$$

where each experimental design sample point is $\boldsymbol{\chi}^{(j)} = \left(\chi_1^{(j)}, \dots, \chi_M^{(j)}\right)$. The new components in dimension i of vector $\mathfrak{v}^{(j)}$ reads then:

$$\left\{\mathfrak{v}_i^{(j)(k)} = \left(\mathfrak{c}_i^{(j)(k)}, \tau_i^{(j)(k)}\right), \ k = 1, \dots, n_{ph}\right\}. \tag{35}$$

When pursuing this procedure for all $i = 1, \dots, M$ dimensions of the input vector, new samples of the experimental design are generated. These samples are called *phantom points* (Schöbi and Sudret,

2015) because they contribute to the experimental design $\mathfrak{V}$ without increasing the number of model evaluations $\mathcal{M}$, which is once and for all N. Indeed, the n_{ph} points in Eq. (35) all correspond to a *single* $\chi^{(j_0)}$ in the original experimental design, meaning a *single run* $\mathcal{Y}^{(j_0)} = \mathcal{M}\left(\chi^{(j_0)}\right)$. This has a large effect on the global efficiency when considering that model evaluations are dominating the total computational costs. The experimental design with phantom points and the corresponding responses then read:

$$
\left\{
\begin{array}{l}
\mathfrak{V} = \left\{ \mathfrak{v}^{(j)(k)}, \quad j = 1, \dots, N, \quad k = 1, \dots, n_{ph} \right\} \\[2mm]
\mathfrak{W} = \left\{ \mathfrak{W}^{(j)(1)} = \dots = \mathfrak{W}^{(j)(n_{ph})}, \quad j = 1, \dots, N \right\}
\end{array}
\right\},
\tag{36}
$$

which consists of $n_{ph} \times N$ samples at a cost of exactly N evaluations of the computational model $\mathcal{M}$.

For illustration purposes, consider a parametric p-box defined by a Gaussian distribution family with interval-valued mean and standard deviation. As an example, it is assumed that $\mu_X \in [-1, 1]$ and $\sigma_X \in [0.5, 1.0]$. Figure 3 shows the bounds of the parametric p-box (see Eq. (27)) as well as the experimental sample point $\chi^{(0)} = 0$. Furthermore, the dotted and dashed-dotted blue lines show two realizations of the CDF in the parametric p-box, namely $\{\mu_X = -0.5, \sigma_X = 1\}$ and $\{\mu_X = 0.5, \sigma_X = 0.5\}$, respectively. The phantom points corresponding to $\chi^{(0)}$ are then computed through the CDF of the Gaussian distribution:

$$
\mathfrak{v}^{(0)(1)} = \left\{ \mu_X = -0.5, \sigma_X = 1, c = F_{\mathcal{N}}\left(\chi^{(0)} \Big| -0.5, 1\right) \approx 0.6915 \right\},
\tag{37}
$$

$$
\mathfrak{v}^{(0)(2)} = \left\{ \mu_X = 0.5, \sigma_X = 0.5, c = F_{\mathcal{N}}\left(\chi^{(0)} \Big| 0.5, 0.5\right) \approx 0.1587 \right\},
\tag{38}
$$

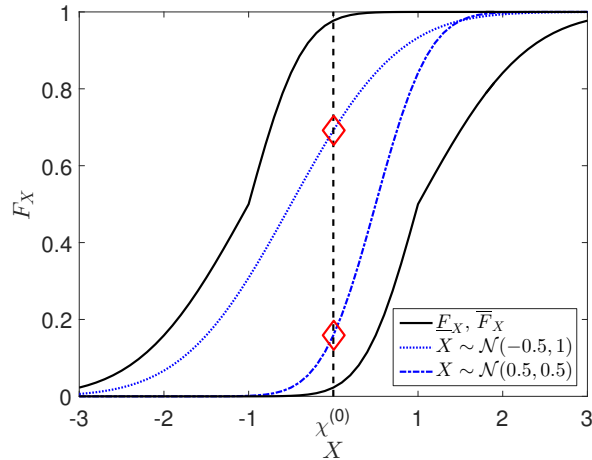which are highlighted by the red diamonds in Figure 3.

Figure 3: Illustration of phantom points on a parametric p-box $X \sim \mathcal{N}\left(\mu_X, \sigma_X\right)$, where $\mu_X \in [-1, 1]$ and $\sigma_X \in [0.5, 1.0]$.

As a summary, a parametric p-box $F_x\left(x | \boldsymbol{\theta}; \boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}\right)$ can be equivalently represented by $(n_{\boldsymbol{\theta}} + 1)$ independent random variables in the augmented space. The extended model $\mathcal{M}^{(\mathrm{aug})}$ in the augmented space is formulated as a function of a n_{aug}-dimensional input vector, where $n_{\mathrm{aug}} = M + \sum_{i=1}^{M} n_{\boldsymbol{\theta}}^{(i)}$, $n_{\boldsymbol{\theta}}^{(i)}$ being the number of parameters $\boldsymbol{\theta}$ characterizing the ith p-box. However, using the "phantom points" trick, an experimental design of arbitrary size $n_{ph} \times N$ can be computed in the augmented input space at the cost of only N evaluations of the original model $\mathcal{M}$. This leads to a remarkably efficient computational procedure. Further computational details are given below.

4.3.2 Case of unbounded distributions

In the case of unbounded distributions, the phantom points can be generated as described above. However, the CDF used in Eq. (34) is a highly non-linear function due to its mapping of an unbounded to a bounded domain, *i.e.* $X \in [-\infty, \infty] \mapsto C \in [0, 1]$. In order to avoid the non-linearity in this mapping and hence expecting more accurate meta-models when using low-degree polynomials, an auxiliary variable ξ is introduced replacing the uniform distribution C. Then, an auxiliary isoprobabilistic transform is formulated for the mapping $(\xi, \boldsymbol{\Theta}) \mapsto X$, which shall be less non-linear. In the following, such an auxiliary transform is derived for the case of Gaussian, lognormal, Gumbel, and Weibull parametric p-boxes.

Gaussian p-box. Consider a Gaussian parametric p-box where the mean value μ_X and standard deviation σ_X are given in intervals. Instead of using the uniform distribution C, a standard Gaussian variable $\xi \sim \mathcal{N}(0, 1)$ can be used in the independent vector $\boldsymbol{V} = (\xi, \mu_X, \sigma_X)$. Then, the isoprobabilistic transform reads:

$$X = \mu_X + \sigma_X \cdot \xi. \tag{39}$$

This transform is simpler than the inverse CDF formulation in Eq. (31) because the unbounded variable X is modelled by means of the unbounded variable ξ. In fact, the transform features a linear mapping from X to ξ.

Lognormal p-box. Considering a lognormal parametric p-box with interval-valued parameters (λ_X, ζ_X), such that $\log X \sim \mathcal{N}(\lambda_X, \zeta_X)$, then:

$$X = e^{\lambda_X + \zeta_X \cdot \xi}, \tag{40}$$

where ξ follows a standard Gaussian distribution, and (λ_X, ζ_X) are obtained by the usual equations from the mean value μ_X and standard deviation σ_X. Note that a lognormal p-box defined by lower and upper bounds on mean μ_X and standard deviation σ_X leads to a non-rectangular domain in the (λ_X, ζ_X)-plane. This is, however, not an issue since the experimental designs will be sampled in the (μ_X, σ_X)-space.

Gumbel p-box. The CDF of a Gumbel distribution is defined as:

$$F_{\mathcal{GU}}(x|\alpha, \beta) = \exp\left[-\exp\left(-(x - \alpha)/\beta\right)\right], \tag{41}$$

where $\alpha = \mu_X - \beta\gamma_e$ and $\beta = \sigma_X \sqrt{6}/\pi$ are the distribution hyper-parameters (μ_X and σ_X denote the mean and standard deviation) and $\gamma_e = 0.5772\dots$ is the Euler-Mascheroni constant. Then, a standard Gumbel variable is denoted by $\varpi \sim \mathcal{GU}(\alpha = 0, \beta = 1)$. It follows that a distribution with arbitrary hyper-parameter values (α, β) can be formulated in terms of a standard Gumbel distribution as follows:

$$F_{\mathcal{GU}}(x|\alpha, \beta) = F_{\mathcal{GU}}(\varpi|0, 1), \tag{42}$$

which reduces to:

$$(x - \alpha)/\beta = \varpi. \tag{43}$$

Hence, a standard Gumbel distribution can be transformed into an arbitrary Gumbel distribution by:

$$X = \alpha + \beta \cdot \varpi, \tag{44}$$

which resembles the construction for Gaussian variables in Eq. (39) with different transformation parameters. Accordingly, a Gumbel p-box can be represented by three random variables, including two uniform and one standard Gumbel.

Weibull p-box. The CDF of a Weibull distribution is defined as:

$$F_{\mathcal{W}}(x|\alpha, \beta) = 1 - \exp\left[-(x/\alpha)^\beta\right], \tag{45}$$

where $x \in [0, \infty]$ and (α, β) are its hyper-parameters. Then, a standard Weibull distribution is denoted by $\varpi \sim \mathcal{W}(\alpha = 1, \beta = 1)$, which also corresponds to a standardized exponential distribution. It follows that a distribution with arbitrary hyper-parameter values (α, β) can be formulated in terms of an exponential distribution as follows:

$$X = \alpha \cdot \varpi^{1/\beta}. \tag{46}$$

Accordingly, a Weibull p-box can be represented by three random variables including two uniform and one exponential.

4.3.3 Bounded distributions

When dealing with distributions that are bounded on two sides, *i.e.* $X \in [\underline{x}, \overline{x}]$, the phantom points cannot be generated by the approach detailed in the previous section. The problem is illustrated in Figure 4, which shows a parametric p-box consisting of a uniform distribution with interval-valued distribution bounds, *i.e.* $X \sim \mathcal{U}(a, b)$, where $a \in [1, 2]$ and $b \in [3, 4]$. Consider the experimental design point $\chi^{(0)} = 3.5$ and the two realizations with $\{a = 1.2, b = 3.8\}$ and $\{a = 1, b = 3.25\}$. For the first realization, the phantom point for $\chi^{(0)}$ reads:

$$\mathfrak{v}^{(0)(1)} = \left\{a = 1.2, \, b = 3.8, \, c = F_{\mathcal{U}}\left(\chi^{(0)} \Big| 1.2, 3.8\right) \approx 0.885\right\}.$$

For the second realization, however, the phantom point lies outside of the support interval $X \in [1, 3.25]$. Hence, the construction of a phantom point is not meaningful in this case. For $\chi^{(0)} = 3.5$ in particular, a phantom point is only feasible when $a \in [1, 2]$ and $b \geq 3.5$ at the same time.
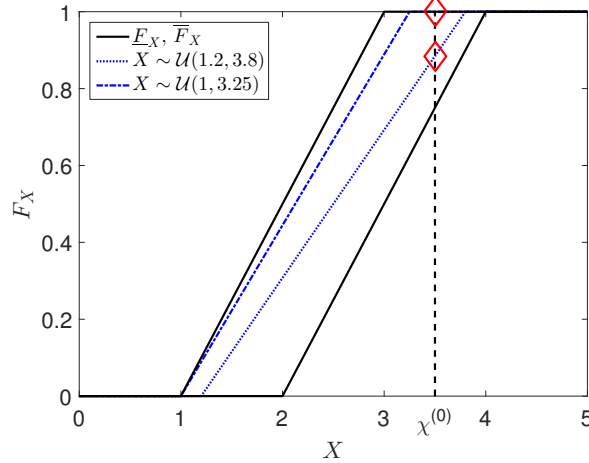

Figure 4: Illustration of phantom points for bounded distributions. $X \sim \mathcal{U}(a, b)$, where $a \in [1, 2]$ and $b \in [3, 4]$.

In other words, the generation of phantom points is submitted to constraints in case of bounded parametric distributions. In order to generate phantom points efficiently, these constraints must be taken into account. Therefore, Eq. (34) is transformed into:

$$\mathfrak{c}_i^{(j)(k)} = F_{X_i}\left(\chi_i^{(j)} \Big| \boldsymbol{\tau}_i^{(j)(k)}\right) \qquad \text{s.t.} \qquad \chi_i^{(j)} \in \left[\underline{x}_{\boldsymbol{\tau}_i^{(j)(k)}}, \, \overline{x}_{\boldsymbol{\tau}_i^{(j)(k)}}\right], \tag{47}$$

13

where $X_{\boldsymbol{\tau}_i^{(j)(k)}}$ denotes a conditional distribution defined within the bounds $X_{\boldsymbol{\tau}_i^{(j)(k)}} \in \left[\underline{x}_{\boldsymbol{\tau}_i^{(j)(k)}}, \ \overline{x}_{\boldsymbol{\tau}_i^{(j)(k)}} \right]$. When $\chi_i^{(j)}$ is outside the interval bounds, no phantom point is generated for the realization $\boldsymbol{\tau}_i^{(j)(k)}$.

5 Imprecise Sobol' indices by PCE

5.1 Reordering

When an augmented PCE model is available as in Eq. (33), the imprecise Sobol' indices can be obtained by post-processing this augmented PCE model. The imprecise Sobol' indices shall be depending on $\boldsymbol{C}$ and conditioned on $\boldsymbol{\Theta}$. In order to do so, the augmented PCE model is recast as follows:

$$W \approx \mathcal{M}^{(\mathrm{P})}\left(\boldsymbol{C}, \boldsymbol{\Theta}\right) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \, \psi_{\boldsymbol{\alpha}_{\boldsymbol{C}}}\left(\boldsymbol{C}\right) \, \psi_{\boldsymbol{\alpha}_{\boldsymbol{\Theta}}}\left(\boldsymbol{\Theta}\right), \tag{48}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_{\boldsymbol{C}}, \boldsymbol{\alpha}_{\boldsymbol{\Theta}})$ forms the same index set as in Eq. (33) and $\psi_{\boldsymbol{\alpha}}\left(\boldsymbol{V}\right) = \psi_{\boldsymbol{\alpha}_{\boldsymbol{C}}}\left(\boldsymbol{C}\right) \cdot \psi_{\boldsymbol{\alpha}_{\boldsymbol{\Theta}}}\left(\boldsymbol{\Theta}\right)$ are the corresponding multivariate orthonormal polynomials. With this notation, we emphasize that any polynomial of $(\boldsymbol{C}, \boldsymbol{\Theta})$ can be cast as a product of two polynomials of $\boldsymbol{C}$ and $\boldsymbol{\Theta}$, respectively, due to the initial tensor product construction. Then, the computational model can be further rearranged, for a given value $\boldsymbol{\theta}$ as a function of $\boldsymbol{C}$ only:

$$W\left(\boldsymbol{\theta}\right) \approx \mathcal{M}^{(\mathrm{P})}\left(\boldsymbol{C}, \boldsymbol{\theta}\right) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \left(a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}_{\boldsymbol{\Theta}}}\left(\boldsymbol{\theta}\right)\right) \, \psi_{\boldsymbol{\alpha}_{\boldsymbol{C}}}\left(\boldsymbol{C}\right) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}, \boldsymbol{\theta}}(\boldsymbol{\theta}) \, \psi_{\boldsymbol{\alpha}_{\boldsymbol{C}}}\left(\boldsymbol{C}\right), \tag{49}$$

where $a_{\boldsymbol{\alpha}, \boldsymbol{\theta}}(\boldsymbol{\theta}) = a_{\boldsymbol{\alpha}} \, \psi_{\boldsymbol{\alpha}_{\boldsymbol{\Theta}}}\left(\boldsymbol{\theta}\right)$ is a new coefficient dependent on the value of $\boldsymbol{\theta}$. Note that several $\boldsymbol{\alpha}_{\boldsymbol{\Theta}}$ may correspond to the same $\boldsymbol{\alpha}_{\boldsymbol{C}}$ in $\boldsymbol{\alpha}$. Hence, Eq. (49) can be further rewritten to a model with proper variance decomposition as in Eq. (22):

$$W\left(\boldsymbol{\theta}\right) \approx \mathcal{M}^{(\mathrm{P})}\left(\boldsymbol{C}, \boldsymbol{\theta}\right) = \sum_{\boldsymbol{\alpha}_{\boldsymbol{C}}^* \in \mathcal{A}_{\boldsymbol{C}}^*} \left(\sum_{i=1,\ldots,n_{\mathcal{A}}} \mathbb{I}_{\left\{\boldsymbol{\alpha}_{\boldsymbol{C}}^{(i)} = \boldsymbol{\alpha}_{\boldsymbol{C}}^*\right\}} \left(\boldsymbol{\alpha}_{\boldsymbol{C}}^{(i)}\right) \, a_{\boldsymbol{\alpha}^{(i)}} \, \psi_{\boldsymbol{\alpha}_{\boldsymbol{\Theta}}^{(i)}}\left(\boldsymbol{\theta}\right) \right) \psi_{\boldsymbol{\alpha}_{\boldsymbol{C}}^*}\left(\boldsymbol{C}\right), \tag{50}$$

where i marks the $1, \ldots, n_{\mathcal{A}}$ multi-indices in $\mathcal{A}$, $\mathbb{I}$ is the indicator function with $\mathbb{I} = 1$ for $\boldsymbol{\alpha}_{\boldsymbol{C}}^{(i)} = \boldsymbol{\alpha}_{\boldsymbol{C}}^*$ and $\mathbb{I} = 0$ otherwise, $\mathcal{A}_{\boldsymbol{C}}^*$ is the set of unique multi-indices $\boldsymbol{\alpha}_{\boldsymbol{C}}$ in $\mathcal{A}$, and $\boldsymbol{\alpha}^{(i)} = \left(\boldsymbol{\alpha}_{\boldsymbol{C}}^{(i)}, \boldsymbol{\alpha}_{\boldsymbol{\Theta}}^{(i)}\right)$. Rearranging the terms to obtain the same structure as in Eq. (22) leads to the following coefficients depending on $\boldsymbol{\theta}$:

$$a_{\boldsymbol{\alpha}_{\boldsymbol{C}}^*}\left(\boldsymbol{\theta}\right) = \sum_{i=1,\ldots,n_{\mathcal{A}}} \mathbb{I}_{\left\{\boldsymbol{\alpha}_{\boldsymbol{C}}^{(i)} = \boldsymbol{\alpha}_{\boldsymbol{C}}^*\right\}} \left(\boldsymbol{\alpha}_{\boldsymbol{C}}^{(i)}\right) \, a_{\boldsymbol{\alpha}^{(i)}} \, \psi_{\boldsymbol{\alpha}_{\boldsymbol{\Theta}}^{(i)}}\left(\boldsymbol{\theta}\right). \tag{51}$$

The simplified decomposition then reads:

$$W_{\boldsymbol{\theta}} \approx \mathcal{M}^{(\mathrm{P})}\left(\boldsymbol{C}, \boldsymbol{\Theta} = \boldsymbol{\theta}\right) = \sum_{\boldsymbol{\alpha}_{\boldsymbol{C}}^* \in \mathcal{A}_{\boldsymbol{C}}^*} a_{\boldsymbol{\alpha}_{\boldsymbol{C}}^*}(\boldsymbol{\Theta} = \boldsymbol{\theta}) \, \psi_{\boldsymbol{\alpha}_{\boldsymbol{C}}^*}\left(\boldsymbol{C}\right). \tag{52}$$

5.2 Optimization

Finally, in order to obtain the extreme values of the Sobol' indices, the optimizations in Eq. (28) use the reformulated model in Eq. (52). Then, the bounds of imprecise Sobol' indices can be computed by solving two optimization problems:

$$\underline{S}_{i_1 \ldots i_s}^{(\mathrm{P})} = \min_{\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}} \frac{D_{i_1 \ldots i_s}(\boldsymbol{\theta})}{D(\boldsymbol{\theta})} = \min_{\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}} \left[\sum_{\boldsymbol{\alpha}_{\boldsymbol{C}}^* \in \mathcal{I}_{i_1 \ldots i_s}} a_{\boldsymbol{\alpha}_{\boldsymbol{C}}^*}^2\left(\boldsymbol{\theta}\right) \Bigg/ \sum_{\boldsymbol{\alpha}_{\boldsymbol{C}}^* \in \mathcal{A}_{\boldsymbol{C}}^*, \ \boldsymbol{\alpha}_{\boldsymbol{C}}^* \neq \boldsymbol{0}} a_{\boldsymbol{\alpha}_{\boldsymbol{C}}^*}^2\left(\boldsymbol{\theta}\right) \right], \tag{53}$$

$$\overline{S}_{i_1\ldots i_s}^{(\mathrm{P})} = \max_{\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}} \frac{D_{i_1\ldots i_s}(\boldsymbol{\theta})}{D(\boldsymbol{\theta})} = \max_{\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\Theta}}} \left[\sum_{\boldsymbol{\alpha}_C^* \in \mathcal{I}_{i_1\ldots i_s}} a_{\boldsymbol{\alpha}_C^*}^2(\boldsymbol{\theta}) \bigg/ \sum_{\boldsymbol{\alpha}_C^* \in \mathcal{A}_C^*, \; \boldsymbol{\alpha}_C^* \neq \mathbf{0}} a_{\boldsymbol{\alpha}_C^*}^2(\boldsymbol{\theta}) \right], \tag{54}$$

In a nutshell, once the augmented PCE has been calibrated, the conditional Sobol' indices $S_{i_1\ldots i_s}(\boldsymbol{\theta})$ are simply *ratios* of multivariate orthogonal polynomials in $\boldsymbol{\theta}$. The two optimization problems in Eqs. (53) and (54) can then be solved at reasonable cost through global optimization algorithms, such as genetic or simulated annealing algorithms.

5.3 Connection to Bayesian hierarchical models

The imprecise Sobol' indices computed here for parametric p-boxes can be extended to Bayesian hierarchical models with little modifications. When interpreting $\boldsymbol{\theta}$ as distributions rather than interval-valued quantities, Eqs. (53) and (54) converge to a single equation describing a distribution of Sobol' indices:

$$S_{i_1\ldots i_s}^{(\mathrm{P})}(\boldsymbol{\Theta}) = \frac{D_{i_1\ldots i_s}(\boldsymbol{\Theta})}{D(\boldsymbol{\Theta})} = \sum_{\boldsymbol{\alpha}_C^* \in \mathcal{I}_{i_1\ldots i_s}} a_{\boldsymbol{\alpha}_C^*}^2(\boldsymbol{\Theta}) \bigg/ \sum_{\boldsymbol{\alpha}_C^* \in \mathcal{A}_C^*, \; \boldsymbol{\alpha}_C^* \neq \mathbf{0}} a_{\boldsymbol{\alpha}_C^*}^2(\boldsymbol{\Theta}). \tag{55}$$

As a direct consequence, the *distribution* of any Sobol' index $S_{i_1\ldots i_s}^{(\mathrm{P})}$ due to the epistemic uncertainty in $\boldsymbol{\theta}$ could be obtained at low cost by Monte Carlo sampling of Eq. (55).

6 Applications

6.1 Step-by-step illustration of algorithm

6.1.1 Problem statement

A simple two-dimensional function is considered here to visualize the algorithm for computing the imprecise Sobol' indices. The analytical function is defined as:

$$f_1(\boldsymbol{x}) = x_1 \cdot x_2, \tag{56}$$

where x_1 and x_2 are modelled by parametric p-boxes. The input variables are modelled by Gaussian distributions with interval-valued mean and standard deviation, namely $\mu_i \in [-1, 1]$ and $\sigma_i \in [0.5, 1.0]$. As seen in the sequel, this function allows us to derive analytically the conditional Sobol' indices as well as their bounds.

6.1.2 Imprecise Sobol' indices

Each input parameter may be cast as $X_i = \mu_i + \sigma_i \cdot \xi_i$ where ξ_i is a standard normal variable. Then, the augmented computational model can be written as:

$$f_1^{(\mathrm{aug})}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}) = (\mu_1 + \sigma_1 \cdot \xi_1) \cdot (\mu_2 + \sigma_2 \cdot \xi_2)$$
$$= (\mu_1 \cdot \mu_2) + (\mu_2 \cdot \sigma_1) \cdot \xi_1 + (\mu_1 \cdot \sigma_2) \cdot \xi_2 + (\sigma_1 \cdot \sigma_2) \cdot \xi_1 \cdot \xi_2. \tag{57}$$

The variance of the response variable conditioned on the values of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ then reads:

$$D(\boldsymbol{\mu}, \boldsymbol{\sigma}) \equiv \mathrm{Var}\left[f_1 | \boldsymbol{\mu}, \boldsymbol{\sigma}\right] = (\mu_2 \sigma_1)^2 + (\mu_1 \sigma_2)^2 + (\sigma_1 \sigma_2)^2. \tag{58}$$

Hence, the first order Sobol' indices can be computed as a function of $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma})$:

$$S_1(\boldsymbol{\theta}) = \frac{(\mu_2 \sigma_1)^2}{(\mu_2 \sigma_1)^2 + (\mu_1 \sigma_2)^2 + (\sigma_1 \sigma_2)^2}, \qquad S_2(\boldsymbol{\theta}) = \frac{(\mu_1 \sigma_2)^2}{(\mu_2 \sigma_1)^2 + (\mu_1 \sigma_2)^2 + (\sigma_1 \sigma_2)^2}, \tag{59}$$

15

The analytical derivation of the extreme Sobol' indices results in the following bounds for the first order Sobol' indices:

$$\underline{S}_1 = \underline{S}_2 = 0.0, \qquad \overline{S}_1 = \overline{S}_2 = 0.8. \tag{60}$$

As an example, the minimal first order Sobol' index for X_1 is obtained by setting $\mu_2 = 0$. Similarly to the first order indices, the total order Sobol' indices can be written as:

$$S_1^{(T)}(\boldsymbol{\theta}) = \frac{(\mu_2\sigma_1)^2 + (\sigma_1\sigma_2)^2}{(\mu_2\sigma_1)^2 + (\mu_1\sigma_2)^2 + (\sigma_1\sigma_2)^2}, \qquad S_2^{(T)}(\boldsymbol{\theta}) = \frac{(\mu_1\sigma_2)^2 + (\sigma_1\sigma_2)^2}{(\mu_2\sigma_1)^2 + (\mu_1\sigma_2)^2 + (\sigma_1\sigma_2)^2}, \tag{61}$$

The bounds of the total order Sobol' indices read:

$$\underline{S}_1^{(T)} = \underline{S}_2^{(T)} = 0.2, \qquad \overline{S}_1^{(T)} = \overline{S}_2^{(T)} = 1.0. \tag{62}$$

As an example, the minimal total order Sobol' index for X_1 is obtained by setting $\mu_1 = \pm 1$, $\sigma_1 = 0.5$, $\mu_2 = 0$, and $\sigma_2 = 1$. Due to the symmetry of the problem definition, *i.e.* computational model $f_1^{(\mathrm{aug})}$ and input p-boxes X_i, the Sobol' indices are symmetrical in (μ_i, σ_i). Note that in this example, the first and total order Sobol' indices are not identical due to the interactive term $(\sigma_1 \cdot \sigma_2) \cdot \xi_1 \cdot \xi_2$ in Eq. (57).

6.1.3 Augmented PCE-based Sobol' indices

Using the MATLAB-based uncertainty quantification framework UQLAB (Marelli and Sudret, 2014, 2015), the augmented PCE of $f_1^{(\mathrm{aug})}$ and the parametric p-boxes results in the following set of multi-indices $\mathcal{A}$:

$$\mathcal{A} = \begin{pmatrix} \boldsymbol{\alpha}^{(1)} \\ \vdots \\ \boldsymbol{\alpha}^{(10)} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}, \tag{63}$$

which contains ten vectors $\boldsymbol{\alpha}$ and where the indices in each row correspond to the augmented input vector $(\mu_1, \sigma_1, \xi_1, \mu_2, \sigma_2, \xi_2)$. Splitting the multi-index set $\mathcal{A}$ to the contributions of the epistemic $(\mu_1, \sigma_1, \mu_2, \sigma_2)$ and aleatory (ξ_1, ξ_2) variables leads to:

$$\mathcal{A}_{\boldsymbol{\Theta}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \qquad \mathcal{A}_{\boldsymbol{C}} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}. \tag{64}$$

Note that according to Eq. (57), only four elements should be part of $\mathcal{A}$. However, due to the normalization to a standard space (uniform distributions to $[-1, 1]$ and Gaussian to standard normal space), an additional six elements are present.

$\mathcal{A}_C^*$ summarizes the set of unique row vectors $\boldsymbol{\alpha}_C$ in $\mathcal{A}_C$ and contains the following four elements:

$$
\mathcal{A}_C^* = \begin{pmatrix} \boldsymbol{\alpha}_C^{*(1)} \\ \boldsymbol{\alpha}_C^{*(2)} \\ \boldsymbol{\alpha}_C^{*(3)} \\ \boldsymbol{\alpha}_C^{*(4)} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{65}
$$

The final coefficients are composed as follows (see also Eq. (51)):

$$
\begin{align}
a_{\boldsymbol{\alpha}_C^{*(1)}}(\boldsymbol{\theta}) &= a_{\boldsymbol{\alpha}^{(1)}}\psi_{\boldsymbol{\alpha}_\Theta^{(1)}}(\boldsymbol{\theta}) + a_{\boldsymbol{\alpha}^{(5)}}\psi_{\boldsymbol{\alpha}_\Theta^{(5)}}(\boldsymbol{\theta}), \tag{66} \\
a_{\boldsymbol{\alpha}_C^{*(2)}}(\boldsymbol{\theta}) &= a_{\boldsymbol{\alpha}^{(2)}}\psi_{\boldsymbol{\alpha}_\Theta^{(2)}}(\boldsymbol{\theta}) + a_{\boldsymbol{\alpha}^{(6)}}\psi_{\boldsymbol{\alpha}_\Theta^{(6)}}(\boldsymbol{\theta}) + a_{\boldsymbol{\alpha}^{(7)}}\psi_{\boldsymbol{\alpha}_\Theta^{(7)}}(\boldsymbol{\theta}) + a_{\boldsymbol{\alpha}^{(10)}}\psi_{\boldsymbol{\alpha}_\Theta^{(10)}}(\boldsymbol{\theta}), \tag{67} \\
a_{\boldsymbol{\alpha}_C^{*(3)}}(\boldsymbol{\theta}) &= a_{\boldsymbol{\alpha}^{(3)}}\psi_{\boldsymbol{\alpha}_\Theta^{(3)}}(\boldsymbol{\theta}) + a_{\boldsymbol{\alpha}^{(8)}}\psi_{\boldsymbol{\alpha}_\Theta^{(8)}}(\boldsymbol{\theta}), \tag{68} \\
a_{\boldsymbol{\alpha}_C^{*(4)}}(\boldsymbol{\theta}) &= a_{\boldsymbol{\alpha}^{(4)}}\psi_{\boldsymbol{\alpha}_\Theta^{(4)}}(\boldsymbol{\theta}) + a_{\boldsymbol{\alpha}^{(9)}}\psi_{\boldsymbol{\alpha}_\Theta^{(9)}}(\boldsymbol{\theta}), \tag{69}
\end{align}
$$

where Eqs. (66) and (67) can be simplified to:

$$
\begin{align}
a_{\boldsymbol{\alpha}_C^{*(1)}}(\boldsymbol{\theta}) &= a_{\boldsymbol{\alpha}^{(1)}} + a_{\boldsymbol{\alpha}^{(5)}}\psi_{\boldsymbol{\alpha}_\Theta^{(5)}}(\boldsymbol{\theta}), \tag{70} \\
a_{\boldsymbol{\alpha}_C^{*(2)}}(\boldsymbol{\theta}) &= a_{\boldsymbol{\alpha}^{(2)}} + a_{\boldsymbol{\alpha}^{(6)}}\psi_{\boldsymbol{\alpha}_\Theta^{(6)}}(\boldsymbol{\theta}) + a_{\boldsymbol{\alpha}^{(7)}}\psi_{\boldsymbol{\alpha}_\Theta^{(7)}}(\boldsymbol{\theta}) + a_{\boldsymbol{\alpha}^{(10)}}\psi_{\boldsymbol{\alpha}_\Theta^{(10)}}(\boldsymbol{\theta}). \tag{71}
\end{align}
$$

The variance decomposition of the PCE model then reads (see also Eq. (52)):

$$
\mathcal{M}^{(\mathrm{P})}(\boldsymbol{C}, \boldsymbol{\Theta}) = \sum_{i=1,\ldots,4} a_{\boldsymbol{\alpha}_C^{*(i)}}(\boldsymbol{\Theta})\ \psi_{\boldsymbol{\alpha}_C^{(i)}}(\boldsymbol{C}), \tag{72}
$$

where $\boldsymbol{C} = (\xi_1, \xi_2)$. By optimization on $\boldsymbol{\theta}$, the PC-based imprecise Sobol' indices can be computed. In this particular case, we obtain the analytical results of Eqs. (60) and (62) exactly. Indeed, the initial model f_1 is polynomial, as well as the isoprobabilistic transforms. In the end, a 10-term PCE in the augmented space is exact.

6.2 Single-degree-of-freedom oscillator

6.2.1 Problem statement

Consider the non-linear undamped single-degree-of-freedom (SDOF) oscillator sketched in Figure 5 (Schueremans and Van Gemert, 2005; Echard et al., 2011, 2013). The corresponding computational model reads:

$$
f_2(r, F_1, t_1, c_1, c_2, m) = 3r - \left| \frac{2F_1}{m\omega_0^2} \sin\left(\frac{\omega_0 T_1}{2}\right) \right|, \tag{73}
$$

where m is the mass, $\{c_1, c_2\}$ are the spring constants of the primary and secondary springs, r is the displacement at which the secondary spring yields, t_1 is the duration of the loading, F_1 is the amplitude of the force and $\omega_0 = \sqrt{\frac{c_1 + c_2}{m}}$ is the natural frequency of the oscillator.

The input vector is modelled by a mix of probabilistic variables and parametric p-boxes accounting for the different levels of knowledge. The description of the input variables is provided in Table 1. It is assumed that the spring stiffnesses and the mass are well-known. Hence $\{c_1, c_2, m\}$ are modelled by regular ("precise") CDFs. On the other side, knowledge on $\{r, F_1, t_1\}$ is supposed to be scarce. Hence, these variables are modelled by parametric p-boxes. As seen in Table 1, the parametric p-box is characterized by a distribution function with interval-valued mean value but constant standard deviation.
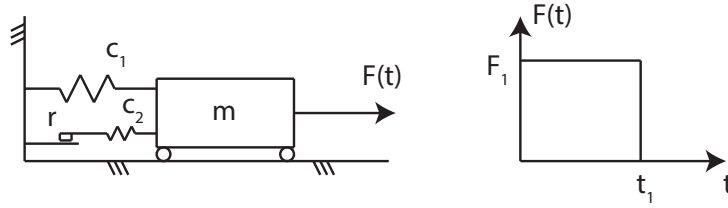
17

Figure 5: SDOF oscillator – geometry sketch and definition of the variables

Table 1: SDOF oscillator – definition of the parametric p-boxes in the input vector.

X_i	Distribution	Mean value	Standard deviation
r	Gaussian	$[0.49, 0.51]$	0.05
F_1	Gaussian	$[0.8, 1.2]$	0.2
t_1	Gaussian	$[0.95, 1.05]$	0.2
c_1	Gaussian	1	0.1
c_2	Gaussian	0.1	0.01
m	Gaussian	1	0.05

6.2.2 Analysis

In order to estimate the imprecise Sobol' indices, an augmented PCE model is trained for an augmented input space of dimension $M_{\mathrm{aug}} = 6+3 = 9$. The number of model evaluations N is varied to investigate the influence of the experimental design size. Note that the number of phantom points is set to $n_{ph} = 10$ for any N in order to take full advantage of the phantom points. The sparse PCE is built with hyperbolic index sets ($q = 0.75$) and degree-adaptive LARS with a maximal total polynomial degree $p = 10$. The bounds of the imprecise Sobol' indices are then obtained through MATLAB's genetic optimization algorithm.

A reference solution is obtained by using augmented PCE with a large sample size, *i.e.* $N = 1000$ and $n_{ph} = 10$. This results in an accurate meta-model and hence an accurate estimate of the imprecise Sobol' indices.

6.2.3 Results

The estimate for the first-order Sobol' indices S_i are summarized in Table 2. The imprecise Sobol' indices are shown for the cases of $N = \{30, 50, 100, 200\}$. The larger the experimental design, the closer the estimates of bounds of the Sobol' indices to the reference values. In fact, an experimental design of $N = 50$ is sufficient to obtain results with sufficient accuracy in this example.

In order to estimate the accuracy of the augmented PCE models, the relative generalization error is derived as the mean squared error between prediction and exact response value relative to the variance of the response variable. Based on a large validation set of $n = 10^6$ samples generated from the augmented input vector $\boldsymbol{V}$, the error estimate reads:

$$\widehat{err}_{gen} = \frac{\sum_{i=1}^{n} \left(w_i - w_i^{(\mathrm{P})} \right)^2}{\sum_{i=1}^{n} (w_i - \mu_w)^2}, \tag{74}$$

where $\mathbb{W} = \{w_1, \ldots, w_n\}$ are the exact response values corresponding to the validation set $\mathbb{V} =$

Table 2: SDOF oscillator – imprecise first-order Sobol' indices estimated by augmented PCE ($n_{ph} = 10$).

	Reference	Augmented PCE			
		$N = 30$	$N = 50$	$N = 100$	$N = 200$
S_r	[0.220, 0.307]	[0.312, 0.313]	[0.223, 0.305]	[0.220, 0.307]	[0.220, 0.307]
S_{F_1}	[0.308, 0.459]	[0.373, 0.375]	[0.306, 0.450]	[0.308, 0.459]	[0.308, 0.460]
S_{t_1}	[0.215, 0.413]	[0.214, 0.215]	[0.224, 0.415]	[0.216, 0.414]	[0.215, 0.413]
S_{c_1}	[0.017, 0.034]	[0.030, 0.030]	[0.017, 0.032]	[0.017, 0.034]	[0.017, 0.034]
S_{c_2}	[0.000, 0.000]	[0.001, 0.001]	[0.000, 0.001]	[0.000, 0.000]	[0.000, 0.000]
S_m	[0.003, 0.006]	[0.008, 0.009]	[0.003, 0.006]	[0.003, 0.006]	[0.003, 0.006]
$\widehat{err}_{gen}$		$1.05 \cdot 10^{-1}$	$2.39 \cdot 10^{-4}$	$1.63 \cdot 10^{-5}$	$1.65 \cdot 10^{-6}$

$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$, $w_i^{(\mathrm{P})}$ are the corresponding values predicted with the augmented PCE model, and $\mu_w = \frac{1}{n} \sum_{i=1}^{n} \mathbb{w}_i$ is the mean value of the samples in $\mathbb{W}$.

The values of the relative generalization error are listed in Table 2. As indicated for the bounds of the imprecise Sobol' indices, the augmented PCE model is accurate for large experimental designs. Values smaller than $\widehat{err}_{gen} \approx 10^{-4}$ indicate a high prediction accuracy.

Figure 6(A) illustrates the imprecise Sobol' indices obtained by $N = 200$ and $n_{ph} = 10$ in a bar plot. The solid (black) part of the bars represents the lower bound of the first order Sobol' indices whereas the hollow bar represents the upper bound. For comparison to a purely probabilistic approach (*i.e.* no epistemic uncertainty), the parametric p-boxes are "pinched" into classical probability distributions. The CDFs use the central value of the intervals as hyper-parameter: $r \sim \mathcal{N}(0.50, 0.05)$, $F_1 \sim \mathcal{N}(1.0, 0.2)$, and $t_1 \sim \mathcal{N}(1.00, 0.2)$. The resulting Sobol' indices are shown by blue markers in the bars. As expected, the imprecise Sobol' indices envelope the probabilistic estimates due to the fact that the "probabilistic case" is one possible realization of the parametric p-boxes.

(A) Bar plot

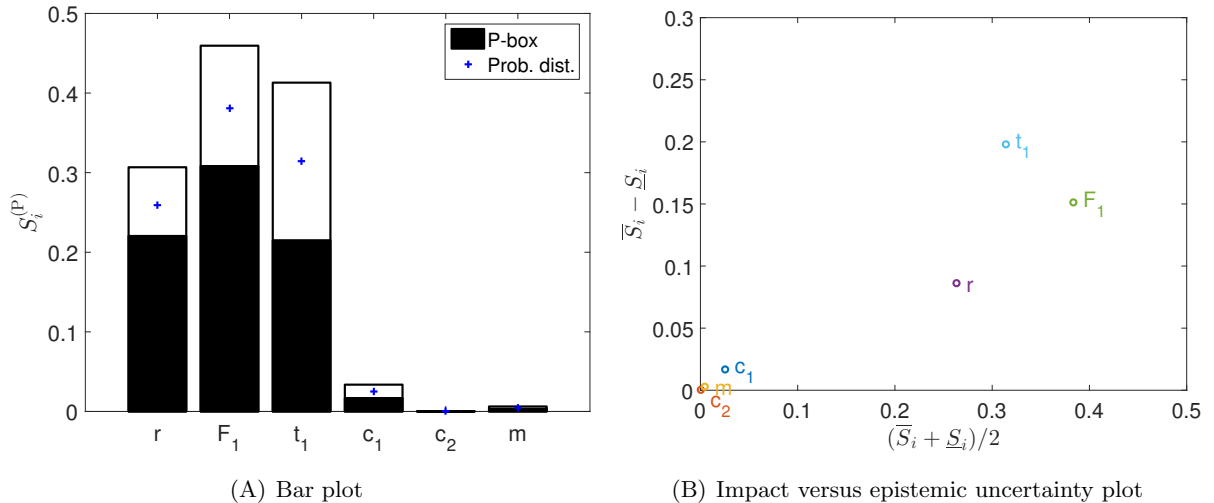(B) Impact versus epistemic uncertainty plot

Figure 6: SDOF oscillator – imprecise Sobol' indices (first order) based on an augmented PCE model with $N = 200$ and $n_{ph} = 10$.

A different visualization of the imprecise Sobol' indices is shown in Figure 6(B). On the horizontal

axis, the central value of the Sobol' indices is displayed, which is computed by $\left(\overline{S}_i + \underline{S}_i\right)/2$. Hence, a variable placed on the right hand side of the plot has a large influence on the uncertainty in the response variable Y. On the vertical axis, the epistemic uncertainty of the Sobol' indices is shown defined as $\overline{S}_i - \underline{S}_i$. A variable placed at the upper end of the plot has a large epistemic uncertainty attached to the Sobol' indices. In this example, F_1 has the largest overall influence, whereas t_1 has the largest epistemic uncertainty in the estimate. Variables whose representative point lies close to the lower left corner are unimportant (c_1, c_2, and m in the current example).

6.3   Simply supported truss

6.3.1   Problem statement

Hurtado Hurtado (2013) introduced the two-dimensional truss structure shown in Figure 7 for the purpose of reliability analysis. The truss is subjected to seven loads P_i which are modelled with parametric p-boxes. The loads are defined by lognormal distributions with mean value $\mu_{P_i} \in [95, 105]$ kN and standard deviation $\sigma_{P_i} \in [13, 17]$ kN. The geometry of the structure and the material properties are assumed constant. The modulus of elasticity is $E = 200 \cdot 10^9$ Pa, whereas the cross section area varies along the different bars: $A = 0.00535$ m^2 for bars marked by ●, $A = 0.0068$ m^2 for bars marked by ○, and $A = 0.004$ m^2 for the remaining bars. The quantity of interest is the deflection at mid-span, denoted by u_4.
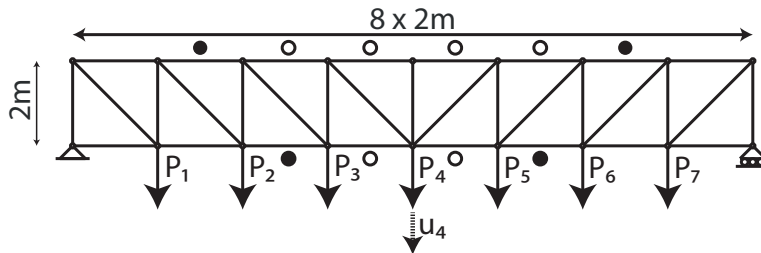

Figure 7: Truss structure – sketch of the geometry and definition of the input variables.

6.3.2   Analysis

The deflection of the truss is computed by a finite element model (FEM), which is implemented in a MATLAB-based in-house code. The FEM interprets each bar as a bar element, whereas the loads are modelled as point loads at the intersections of the corresponding bars as indicated in Figure 7.

In order to estimate the imprecise Sobol' indices, an augmented PCE model is computed based on an experimental design of $N = 100$ samples generated by Latin-hypercube sampling in the unit-cube and subsequent isoprobabilistic transform to domain $\mathcal{D}_{\boldsymbol{V}}$. The original computational model requires a seven-dimensional input vector describing the loading forces. Due to the definition of the parametric p-boxes, the augmented PCE model requires a 21-dimensional input vector ($M_{\text{aug}} = 21$). The number of phantom points n_{ph} is varied to examine the influence of these points onto the accuracy of the computed Sobol' indices. The bounds of the Sobol' indices are obtained by using MATLAB's built-in genetic algorithm with default settings as an optimization tool.

The maximum deflection computed by the FE model corresponds to a monotone function of the load parameters. Hence, the extreme cases of the first order Sobol' indices can be obtained by setting

one variable to $\{\mu_{P_i} = 105 \text{ kN}, \sigma_{P_i} = 17 \text{ kN}\}$ and all others to $\{\mu_{P_i} = 95 \text{ kN}, \sigma_{P_i} = 13 \text{ kN}\}$ or vice versa. The reference solution of the Sobol' indices are then obtained by PCE and a large experimental design.

6.3.3 Results

The imprecise first-order Sobol' indices are visualized in Figure 8 as a function of n_{ph}. The filled and hollow bars represent $\underline{S}_i$ and $\overline{S}_i$, respectively. The estimated values (plotted with orange bars) are compared to the reference solution (black bars). As n_{ph} increases, the extreme values of the Sobol' indices converge to the true values without adding additional FEM runs. In this example, $n_{ph} = 5$ is sufficient to obtain reliable results. This impression is confirmed by the relative generalization error which is provided in the caption of the figures. In the case of $n_{ph} = 3$ and the corresponding relative generalization error of $\widehat{err}_{gen} = 8.4 \cdot 10^{-4}$, the imprecise PCE-based analysis provides usable estimates of the imprecise Sobol' indices. In total, the full set of Sobol' indices together with their ranges of epistemic uncertainty is obtained for a total cost of 100 finite element analyses of the truss structure, meaning several orders of magnitude less than any Monte Carlo-based double loop approach.

Similarly as for the SDOF oscillator example, Figure 9 shows the impact of aleatory vs. epistemic uncertainty for the truss structure. This plot illustrates nicely the symmetry of the imprecise Sobol' indices with respect to the central load P_4, i.e. the points representing P_3 and P_5, P_2 and P_6, P_1 and P_7 are identical. Moreover, the amount of imprecision is proportional to the absolute impact for all seven loads in this example.

7 Conclusions

Sensitivity analysis is a popular tool to analyse the input-output relationship of computational models. In this paper, imprecise Sobol' indices are discussed as a measure of global sensitivity in the context of parametric probability-boxes for modelling the model input parameters. The clear separation of aleatory and epistemic uncertainty in probability-boxes allows for an intuitive interpretation of the resulting interval-valued Sobol' indices. Therein, the absolute value represents the importance of each parameter with respect to the aleatory uncertainty whereas the range of the interval represents the epistemic uncertainty.

In order to speed up the computations of imprecise Sobol' indices, an augmented polynomial chaos expansion (PCE) model is introduced. The imprecise Sobol' indices are obtained by simply post-processing the augmented PCE model. In particular, the extreme values of the Sobol' indices are computed by optimizing on the coefficients of the augmented PCE model without rerunning the computational model. Hence, the imprecise Sobol' indices can be evaluated efficiently. Furthermore, the introduction of phantom points counteracts the increased complexity of the augmented PCE model due to the augmented input vector.

The proposed algorithm allows for analysing realistic engineering problems as shown with the single-degree-of-freedom oscillator and the truss structure examples. A small experimental design is sufficient to estimate the imprecise PCE-based Sobol' indices accurately. Moreover, the "phantom points trick" allows for a further increase in meta-model accuracy and at the same time a more efficient estimation of the Sobol' indices.
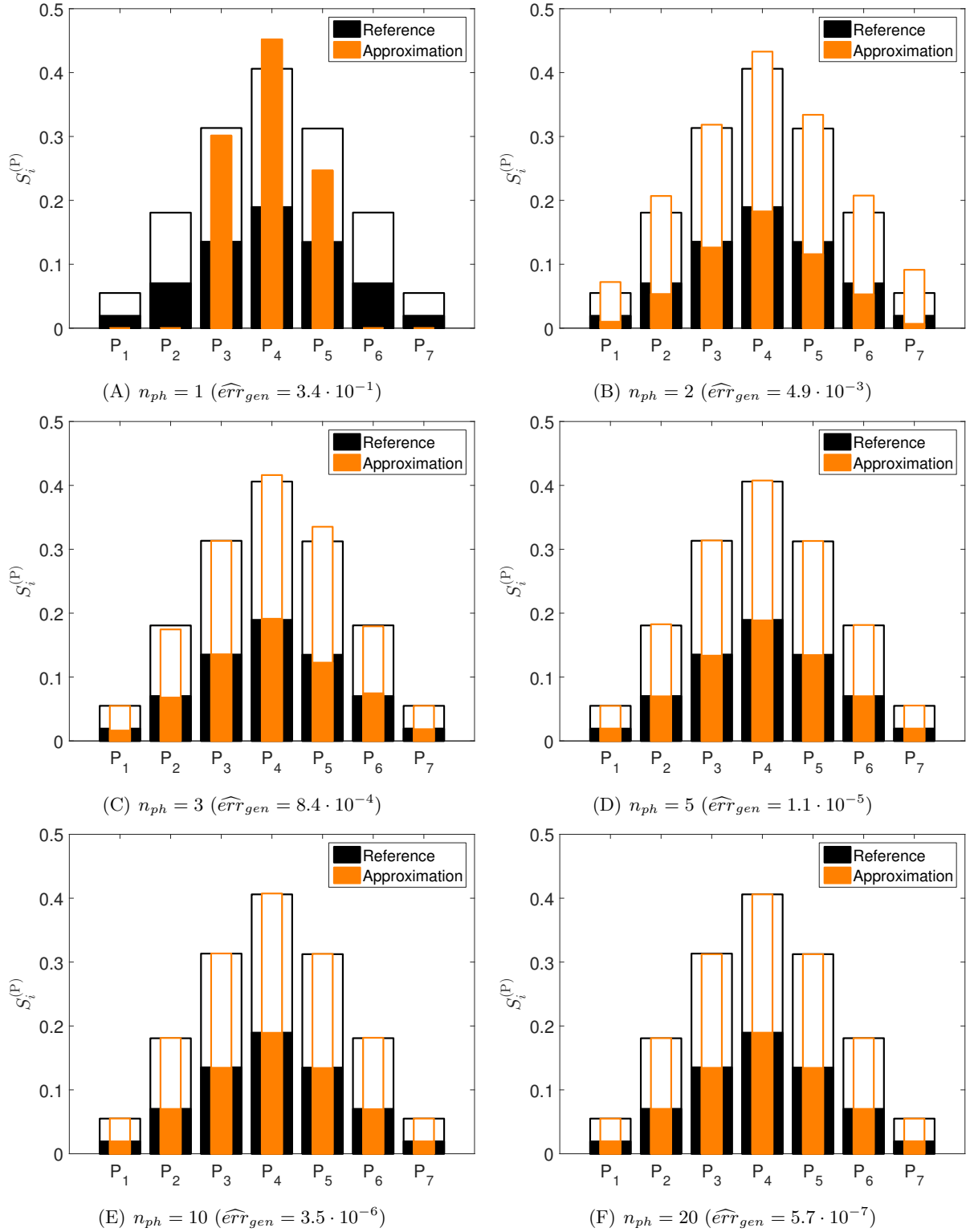
Figure 8: Simply supported truss structure – imprecise Sobol' indices (first order) – augmented-PCE estimates (orange bars) versus the reference values (black bars).
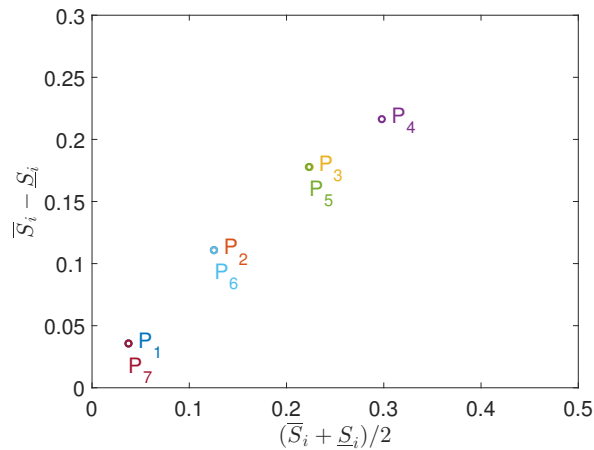
Figure 9: Simply supported truss structure – impact of aleatory vs. epistemic uncertainty plot

References

References

Alexanderian, A., J. Winokur, I. Sraj, A. Srinivasan, M. Iskandarani, W. C. Thacker, and O. M. Knio (2012, jun). Global sensitivity analysis in an ocean general circulation model: a sparse spectral projection approach. *Comput. Geosci. 16*(3), 757–778.

Berveiller, M., B. Sudret, and M. Lemaire (2006). Stochastic finite elements: a non intrusive approach by regression. *Eur. J. Comput. Mech. 15*(1-3), 81–92.

Blatman, G. (2009). *Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis*. Ph. D. thesis, Université Blaise Pascal, Clermont-Ferrand, France.

Blatman, G. and B. Sudret (2010). An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Prob. Eng. Mech. 25*(2), 183–197.

Blatman, G. and B. Sudret (2011). Adaptive sparse polynomial chaos expansion based on Least Angle Regression. *J. Comput. Phys 230*(6), 2345–2367.

Chowdhary, K. and P. Dupuis (2013). Distinguishing and integrating aleatoric and epistemic variation in uncertainty quantification. *ESAIM Math. Model. Numer. Anal. 47*(3), 635–662.

Echard, B., N. Gayton, and M. Lemaire (2011). AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation. *Structural Safety 33*(2), 145–154.

Echard, B., N. Gayton, M. Lemaire, and N. Relun (2013). A combined importance sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models. *Reliab. Eng. Sys. Safety 111*, 232–240.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. *Ann. Stat. 32*(2), 407–499.

Eldred, M. S. and L. P. Swiler (2009). Efficient algorithms for mixed aleatory-epistemic uncertainty quantification with application to radiation-hardened electronics part I: algorithms and benchmark results. Technical Report SAND2009-5805, Sandia National Laboratories.

Ferson, S. and L. R. Ginzburg (1996). Different methods are needed to propagate ignorance and variability. *Reliab. Eng. Sys. Safety 54* (2-3), 133–144.

Ferson, S. and J. G. Hajagos (2004). Arithmetic with uncertain numbers: rigorous and (often) best possible answers. *Reliab. Eng. Sys. Safety 85* (1-3), 135–152.

Ferson, S. and W. Troy Tucker (2006). Sensitivity analysis using probability bounding. *Reliab. Eng. Sys. Safety 91* (10-11), 1435–1442.

Gelman, A., J. Carlin, H. Stern, and D. Rubin (2009). *Bayesian Data Analysis* (second ed.). Chapman & Hall/CRC.

Ghanem, R. and P. Spanos (2003). *Stochastic Finite Elements: A Spectral Approach* (2nd ed.). Courier Dover Publications, Mineola.

Gunn, S. (1998). Support vector machines for classification and regression. Technical Report ISIS-1-98, Dpt. of Electronics and Computer Science, University of Southampton, UK.

Helton, J., J. Johnson, W. Oberkampf, and C. Sallaberry (2006). Sensitivity analysis in conjunction with evidence theory representations of epistemic uncertainty. *Reliab. Eng. Sys. Safety 91* (10-11), 1414–1434.

Helton, J., J. Johnson, C. Sallaberry, and C. Storlie (2006). Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliab. Eng. Sys. Safety 91* (10-11), 1175–1209.

Homma, T. and A. Saltelli (1996). Importance measures in global sensitivity analysis of non linear models. *Reliab. Eng. Sys. Safety 52*, 1–17.

Hurtado, J. E. (2013). Assessment of reliability intervals under input distributions with uncertain parameters. *Prob. Eng. Mech. 32*, 80–92.

Hurtado, J. E. and D. A. Alvarez (2001). Neural-network-based reliability analysis: a comparative study. *Comput. Method. Appl. M. 191* (1-2), 113–132.

Konakli, K. and B. Sudret (2016). Global sensitivity analysis using low-rank tensor approximations. *Reliab. Eng. Sys. Safety 156*, 64–83.

Krzykacz-Hausmann, B. (2006). An approximate sensitivity analysis of results from complex computer models in the presence of epistemic and aleatory uncertainties. *Reliab. Eng. Sys. Safety 91* (10-11), 1210–1218.

Le Gratiet, L., S. Marelli, and B. Sudret (2017). *Metamodel-based sensitivity analysis: polynomial chaos expansions and Gaussian processes*, Chapter 8, Handbook on Uncertainty Quantification (Ghanem, R. and Higdon, D. and Owhadi, H. (Eds.). Springer.

Li, C. and S. Mahadevan (2016a). Relative contributions of aleatory and epistemic uncertainty sources in time series prediction. *Int. J. Fatigue 82*, 474–486.

Li, C. and S. Mahadevan (2016b). Role of calibration, validation, and relevance in multi-level uncertainty integration. *Reliab. Eng. Sys. Safety 148*, 32–43.

Mallat, S. and Z. Zhang (1993). Matching pursuits with time-frequency dictionaries. *IEEE T. Signal. Proces. 41*(12), 3397–3415.

Marelli, S. and B. Sudret (2014). UQLab: a framework for uncertainty quantification in MATLAB. In *Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014), Liverpool, United Kingdom*.

Marelli, S. and B. Sudret (2015). UQLab user manual – Polynomial chaos expansions. Technical report, Chair of Risk, Safety & Uncertainty Quantification, ETH Zurich. Report # UQLab-V0.9-104.

Möller, B. and M. Beer (2004). *Fuzzy Randomness*. Springer.

Oberguggenberger, M. and W. Fellin (2005). Assessing the sensitivity of failure probabilities: a random set approach. In G. Augusti, G. Schuëller, and M. Ciampoli (Eds.), *Proc. 9th Int. Conf. Struct. Safety and Reliability (ICOSSAR'2005), Rome, Italy*, Millpress, Rotterdam, pp. 1755–1760.

Oberguggenberger, M., J. King, and B. Schmelzer (2009). Classical and imprecise probability methods for sensitivity analysis in engineering: a case study. *Int. J. Approx. Reason. 50*(4), 680–693.

Pati, Y., R. Rezaiifar, and P. Krishnaprasad (1993). Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. *Proc. 27th Asilomar Conference on Signals, Systems and Computers*, 40–44.

Robert, C. and G. Casella (2004). *Monte Carlo statistical methods (2nd Ed.)*. Springer Series in Statistics. Springer Verlag.

Sacks, J., W. Welch, T. Mitchell, and H. Wynn (1989). Design and analysis of computer experiments. *Stat. Sci. 4*, 409–435.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008). *Global Sensitivity Analysis – The Primer*. Wiley.

Sankararaman, S. and S. Mahadevan (2013). Separating the contributions of variability and parameter uncertainty in probability distributions. *Reliab. Eng. Sys. Safety 112*, 187–199.

Santner, T., B. Williams, and W. Notz (2003). *The Design and Analysis of Computer Experiments*. Springer, New York.

Sargsyan, K., C. Safta, H. Najm, B. Debusschere, D. Ricciuto, and P. Thornton (2014). Dimensionality reduction for complex models via Bayesian compressive sensing. *Int. J. Uncer. Quant. 4*(1), 63–93.

Schöbi, R. and B. Sudret (2015). Propagation of uncertainties modelled by parametric p-boxes using sparse polynomial chaos expansions. In *Proc. 12th Int. Conf. App. Stat. Prob. Civil Eng. (ICASP12), Vancouver, Canada*.

Schueremans, L. and D. Van Gemert (2005). Benefit of splines and neural networks in simulation based structural reliability analysis. *Structural Safety 27*(3), 246–261.

Smola, A. and B. Schölkopf (2006). A tutorial on support vector regression. *Stat. Comput. 14*, 199–222.

Sobol', I. (1993). Sensitivity estimates for nonlinear mathematical models. *Math. Modeling & Comp. Exp. 1*, 407–414.

Soize, C. and R. Ghanem (2004). Physical systems with random uncertainties: chaos representations with arbitrary probability measure. *SIAM J. Sci. Comput. 26*(2), 395–410.

Sudret, B. (2006). Global sensitivity analysis using polynomial chaos expansions. In P. Spanos and G. Deodatis (Eds.), *Proc. 5th Int. Conf. on Comp. Stoch. Mech (CSM5), Rhodos, Greece*.

Sudret, B. (2007). *Uncertainty propagation and sensitivity analysis in mechanical models – Contributions to structural reliability and stochastic spectral methods*. Université Blaise Pascal, Clermont-Ferrand, France. Habilitation à diriger des recherches, 173 pages.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Sys. Safety 93*, 964–979.

Sudret, B. and C. Mai (2015). Computing derivative-based global sensitivity measures using polynomial chaos expansions. *Reliab. Eng. Sys. Safety 134*, 241–250.

Tibshirani, R. (1996). Regression shrinkage and selection via LASSO. *J. Roy. Stat. Soc. B 58*, 267–288.

Xiu, D. and G. Karniadakis (2002). The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput. 24*(2), 619–644.

Xu, C. and G. Gertner (2008). Uncertainty and sensitivity analysis for models with correlated parameters. *Reliab. Eng. Sys. Safety 93*, 1563–1573.