POLYNOMIAL-CHAOS-BASED KRIGING

Roland Schöbi,[1,*] *Bruno Sudret,*[1] *& Joe Wiart*[2]

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

[2] *WHIST Lab, Institut Mines Telecom, 46 rue Barrault 75634 Paris Cedex 13, France*

Computer simulation has become the standard tool in many engineering fields for designing and optimizing systems, as well as for assessing their reliability. Optimization and uncertainty quantification problems typically require a large number of runs of the computational model at hand, which may not be feasible with high-fidelity models directly. Thus surrogate models (a.k.a meta-models) have been increasingly investigated in the last decade. Polynomial chaos expansions (PCE) and Kriging are two popular nonintrusive meta-modeling techniques. PCE surrogates the computational model with a series of orthonormal polynomials in the input variables where polynomials are chosen in coherency with the probability distributions of those input variables. A least-square minimization technique may be used to determine the coefficients of the PCE. Kriging assumes that the computer model behaves as a realization of a Gaussian random process whose parameters are estimated from the available computer runs, i.e., input vectors and response values. These two techniques have been developed more or less in parallel so far with little interaction between the researchers in the two fields. In this paper, PC-Kriging is derived as a new nonintrusive meta-modeling approach combining PCE and Kriging. A sparse set of orthonormal polynomials (PCE) approximates the global behavior of the computational model whereas Kriging manages the local variability of the model output. An adaptive algorithm similar to the least angle regression algorithm determines the optimal sparse set of polynomials. PC-Kriging is validated on various benchmark analytical functions which are easy to sample for reference results. From the numerical investigations it is concluded that PC-Kriging performs better than or at least as good as the two distinct meta-modeling techniques. A larger gain in accuracy is obtained when the experimental design has a limited size, which is an asset when dealing with demanding computational models.

KEY WORDS: *emulator, Gaussian process modeling, Kriging, meta-modeling, polynomial chaos expansions, PC-Kriging, benchmark functions*

1. INTRODUCTION

Modern engineering makes a large use of computer simulation in order to design systems of ever-increasing complexity and assess their performance. As an example, let us consider a structural engineer planning a new structure. An essential part of his work is to predict the behavior of the not-yet-built structure based on some assumptions, and available information about, e.g., acceptable dimensions, loads to be applied to the structure, and material properties. These ingredients are basically the input parameters of computational models that predict the performance of the system under various conditions. The models help the engineer analyze/understand the behavior of the structure and eventually optimize the design in order to comply with safety and serviceability constraints.

Similar conditions can also be found in many other scientific and engineering disciplines. The common point is the simulation of the behavior of a physical process or system by dedicated algorithms which provide a numerical solution to the governing equations. These simulation algorithms aim at reproducing the physical process with the highest possible fidelity. As an example, finite element models have become a standard tool in modern civil and

---

*Correspond to Roland Schöbi, E-mail: schoebi@ibk.baug.ethz.ch, URL: http://www.ibk.ethz.ch/su/index_EN

mechanical engineering. Due to high fidelity, such models typically exploit the available computer power, meaning that a single run of the model may take hours to days of computing, even when using a large number of CPUs.

An additional layer of complexity comes from the fact that most input parameters of such computational models are not perfectly known in practice. Some parameters (e.g., material properties, applied loads, etc.) may exhibit natural variability so that the exact value to be used for simulating the behavior of a particular system is not known in advance (this is referred to as *aleatory uncertainty*). Some others may have a unique value which is however not directly measurable and prone to lack of knowledge (*epistemic uncertainty*). In a probabilistic setup these parameters are modeled by random variables with prescribed joint probability density function, or more generally, by random fields. The goal of *uncertainty propagation* is to assess the effect of the input uncertainty onto the model output, and consequently onto the performance of the system under consideration [1, 2].

Propagating uncertainties usually requires a large number of repeated calls to the model for different values of the input parameters, for instance through a Monte Carlo simulation procedure. Such an approach usually require thousands to millions of runs which is not affordable even with modern high-performance computing architectures. To circumvent this problem, *surrogate models* may be used, which replace the *original* computational model by an easy-to-evaluate function [3–5]. These surrogate models, also known as *response surfaces* or *meta-models*, are capable of quickly predicting responses to new input realizations. This allows for conducting analyses which require a large number of model evaluations, such as structural reliability and optimization, in a reasonable time.

Among the various options for constructing meta-models, this paper focuses on *nonintrusive approaches*, meaning that the computational model is considered as a "black-box model": once an input vector (i.e., a realization of the random input in the process of uncertainty propagation) is selected, the model is run and provides an output vector of *quantities of interest*. No additional knowledge on the inner structure of the computer code is assumed. Popular types of meta-models that may be built from a limited set of runs of the original model (called the *experimental design* of computations) include polynomial chaos expansions (PCE) [6], Gaussian process modeling (also called Kriging) [7–10], and support vector machines [11–14], which have been extensively investigated in the last decade. Polynomial chaos expansions and Kriging are specifically of interest in this paper.

Polynomial chaos expansions (PCE), also know as spectral expansions, approximate the computational model by a series of multivariate polynomials which are orthogonal with respect to the distributions of the input random variables. Traditionally, spectral expansions have been used to solve partial differential equations in an *intrusive* manner [6]. In this setup truncated expansions are inserted into the governing equations and the expansion coefficients are obtained using a Galerkin scheme. This pioneering approach called *spectral stochastic finite element method* (SSFEM), was later developed by [15–19], among others. These intrusive methods require specific, problem-dependent algorithmic developments though. Because it is not always possible and/or feasible to treat a computational model intrusively, especially when legacy codes are at hand in an industrial context, nonintrusive polynomial chaos expansions were developed. So-called *projections methods* were developed by [20–23], see a review in [24]. *Least-square minimization* techniques have been introduced by [25–27]. Further developments which combine spectral expansions and compressive sensing ideas have lead to so-called *sparse polynomial chaos expansions* [28–34]. This is the approach followed in this paper. Further recent applications of PCE to structural reliability analysis and design optimization can be found in [35–37].

The second meta-modeling technique of interest in this paper is Kriging, which originates from interpolating geographical data in mining [38] and is today also known as Gaussian process modeling [8, 9]. The Kriging meta-model is interpreted as the realization of a Gaussian process. Practical applications can be found in many fields, such as structural reliability analysis [39–44] and design optimization [45–47]. The implementation of the Kriging meta-modeling technique can be found in, e.g., the Matlab toolbox *DACE* [48] and the more recent R toolbox *DiceKriging* [49, 50].

So far the two distinct meta-modeling approaches have been applied in various fields rather independently. To our knowledge, there has not been any attempt to combine PCE and Kriging in a systematic way yet. To bridge the gap between the two communities, this paper aims at combining the two distinct approaches into a new and more powerful meta-modeling technique called *Polynomial-Chaos-Kriging* (PC-Kriging). As seen in the sequel the combination of the characteristics and the advantages of both approaches leads to a more accurate and flexible algorithm which will be detailed in this paper.

The paper is organized as follows. PCE and Kriging are first summarized in Section 2 and Section 3, respectively. Section 4 introduces the new meta-modeling approach PC-Kriging as the combination of the two distinct approaches. The accuracy of the new and traditional meta-modeling approaches is compared in Section 5 on a set of benchmark analytical functions.

2. POLYNOMIAL CHAOS EXPANSIONS

2.1 Problem Definition

Consider the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where Ω denotes the event space equipped with σ-algebra $\mathcal{F}$ and the probability measure $\mathcal{P}$. Random variables are denoted by capital letters $X(\omega) : \Omega \mapsto \mathcal{D}_X \subset \mathbb{R}$ and their realizations denoted by the corresponding lower case letters, e.g., x. Random vectors (e.g., $\boldsymbol{X} = \{X_1, \ldots, X_M\}^\mathsf{T}$) and their realizations (e.g., $\boldsymbol{x} = \{x_1, \ldots, x_M\}^\mathsf{T}$) are denoted by bold faced capital and lowercase letters, respectively.

In this context, consider a system whose behavior is represented by a computational model $\mathcal{M}$ which maps the M-dimensional input parameter space to the one-dimensional output space, i.e., $\mathcal{M} : \boldsymbol{x} \in \mathcal{D}_X \subset \mathbb{R}^M \mapsto y \in \mathbb{R}$ where $\boldsymbol{x} = \{x_1, \ldots, x_M\}^\mathsf{T}$. As the input vector $\boldsymbol{x}$ is assumed to be affected by uncertainty, a probabilistic framework is introduced. Due to uncertainties in the input vector, it is represented by a random vector $\boldsymbol{X}$ with given joint probability density function (PDF) $f_{\boldsymbol{X}}$. For the sake of simplicity the components are assumed independent throughout the paper, so that the joint PDF may be written as the product of the marginal PDFs denoted by f_{X_i}, $i = 1, \ldots, M$. Note that the case of dependent input variables can easily be addressed by using an isoprobabilistic transform first, such as the Nataf or Rosenblatt transform [29]. The output of the model is a random variable Y obtained by propagating the input uncertainty in $\boldsymbol{X}$ through the computational model $\mathcal{M}$:

$$Y = \mathcal{M}(\boldsymbol{X}). \tag{1}$$

In this paper we consider that the computational model is a *deterministic* mapping from the input to the output space, i.e., repeated evaluations with the same input value $\boldsymbol{x}_0 \in \mathcal{D}_X$ lead to the same output value $y_0 = \mathcal{M}(\boldsymbol{x}_0)$.

Provided that the output random variable Y is a second-order variable (i.e., $\mathbb{E}\left[Y^2\right] < +\infty$), it can be cast as the following polynomial chaos expansion (PCE) [6, 51]:

$$Y \equiv \mathcal{M}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} \boldsymbol{a}_{\boldsymbol{\alpha}} \, \psi_{\boldsymbol{\alpha}}(\boldsymbol{X}), \tag{2}$$

where $\{\boldsymbol{a}_{\boldsymbol{\alpha}}, \, \boldsymbol{\alpha} \in \mathbb{N}^M\}$ are coefficients of the multivariate orthonormal polynomials $\psi_{\boldsymbol{\alpha}}(\boldsymbol{X})$ in coherency with the distribution of the input random vector $\boldsymbol{X}$, $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_M\}$ is the multi-index and M is the number of input variables (dimensions). Since the components of $\boldsymbol{X}$ are independent, the joint probability density function $f_{\boldsymbol{X}}$ is the product of the margins f_{X_i}. Then a *functional inner product* for each marginal PDF f_{X_i} is defined by

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

for any two functions $\{\phi_i, \phi_2\}$ such that the integral exists. For each variable $i = 1, \ldots, M$ an orthonormal polynomial basis can be constructed which satisfies [15]:

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

where $P_j^{(i)}$, $P_k^{(i)}$ are two candidate univariate polynomials in the ith variable, $\mathcal{D}_i$ is the support of the random variable X_i and δ_{jk} is the Kronecker delta which is equal to 1 for $j = k$ and equal to 0 otherwise. [15] summarize various orthonormal bases for some classical PDFs, some of which are summarized in Table 1.

TABLE 1: Classical orthogonal/orthonormal polynomials (as presented in [52])

Distribution	PDF	Orthogonal polynomials	Orthonormal basis
Uniform	$\mathbf{1}_{]-1,1[}(x)/2$	Legendre $P_k(x)$	$P_k(x)/\sqrt{\dfrac{1}{2k+1}}$
Gaussian	$\dfrac{1}{\sqrt{2\pi}}e^{-x^2/2}$	Hermite $H_{e_k}(x)$	$H_{e_k}(x)/\sqrt{k!}$
Gamma	$x^a e^{-x}\mathbf{1}_{\mathbb{R}^+}(x)$	Laguerre $L_k^a(x)$	$L_k^a(x)/\sqrt{\dfrac{\Gamma(k+a+1)}{k!}}$
Beta	$\mathbf{1}_{]-1,1[}(x)\dfrac{(1-x)^a(1+x)^b}{B(a)B(b)}$	Jacobi $J_k^{a,b}(x)$	$J_k^{a,b}(x)/\mathcal{J}_{a,b,k}$
			$\mathcal{J}_{a,b,k}^2 = \dfrac{2^{a+b+1}}{2k+a+b+1}\dfrac{\Gamma(k+a+1)\Gamma(k+b+1)}{\Gamma(k+a+b+1)\Gamma(k+1)}$

The multivariate polynomials in Eq. (2) are then composed of univariate polynomials by tensor product, i.e., by multiplying the various polynomials in each input variable:

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{X}) = \prod_{i=1}^{M} \psi_{\alpha_i}^{(i)}(X_i), \tag{5}$$

where $\psi_{\alpha_i}^{(i)}$ is the polynomial of degree α_i in the ith variable.

The main idea of PCE is then to surrogate the computational model by an infinite series of polynomials as shown in Eq. (2). In practice, it is not feasible to handle infinite series, thus the need of a truncation scheme. Such a truncation scheme corresponds to a set of multi-indexes $\boldsymbol{\alpha} \in \mathcal{A} \subset \mathbb{N}^M$ such that the system response is accurately approximated with respect to some error measure [29, 32]:

$$Y \approx Y^{(\text{PCE})} \stackrel{\text{def}}{=} \mathcal{M}^{(\text{PCE})}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \boldsymbol{a}_{\boldsymbol{\alpha}}\,\psi_{\boldsymbol{\alpha}}(\boldsymbol{X}). \tag{6}$$

There are several ways to select *a priori* a truncation set $\mathcal{A}$. A simple and commonly applied scheme consists of upper-bounding the total degree of polynomials to a maximal value p. The total degree of polynomials is defined by

$$|\boldsymbol{\alpha}| = \sum_{i=1}^{M} \alpha_i. \tag{7}$$

In this case the set of multi-indices is denoted by $\mathcal{A}^{M,p} = \{\boldsymbol{\alpha} \in \mathbb{N}^M : |\boldsymbol{\alpha}| \leq p\}$ where p is the maximal total polynomial degree. The cardinality of the set $\mathcal{A}$ reads

$$\left|\mathcal{A}^{M,p}\right| = \frac{(M+p)!}{M!\,p!}. \tag{8}$$

This cardinality grows polynomially with both M and p. Such a truncation scheme thus leads to nontractable problems if the response is highly nonlinear in its input parameters (need for a large p) and/or if the size of the input vector $\boldsymbol{X}$ is large (say, $M > 10$). This problem is referred to as the *curse of dimensionality*.

In [29, 53] a more restrictive truncation scheme called *hyperbolic truncation set* was proposed. The authors observed that many systems tend to have only low-degree interaction polynomials and thus it is not necessary to compute all interaction terms of higher polynomial degree. The hyperbolic index set is based on the following q-norm:

$$\mathcal{A}_q^{M,p} \equiv \{\boldsymbol{\alpha} \in \mathbb{N}^M : \|\boldsymbol{\alpha}\|_q \leq p\}, \tag{9}$$

where

$$\|\boldsymbol{\alpha}\|_q \equiv \left(\sum_{i=1}^{M} \alpha_i^q\right)^{1/q}, \tag{10}$$

$0 < q \le 1$ is a tuning parameter and p is the maximal total degree of the polynomials. A decreasing q leads to a smaller number of interactive polynomials, i.e., a smaller set of polynomials. When $q \to 0$, only univariate polynomials are left in the set of polynomials which is called an additive model [52]. For the sake of illustration, the retained polynomial indices $\boldsymbol{\alpha} \in \mathcal{A}_q^{M,p}$ of a two-dimensional input space ($M = 2$) and varying p and q are illustrated in Fig. 1. The indices denoted by $\bullet$ are part of $\mathcal{A}_q^{M,p}$ and the solid black line represents $\|\boldsymbol{\alpha}\|_q = p$. Note that for $q = 1$, the hyperbolic index sets are equivalent to the total degree index set [see Eq. (7)].

2.2 Computation of the Coefficients

After defining the set of candidate polynomials, the next step is to determine the expansion coefficients $\boldsymbol{a_\alpha}$ of each multivariate polynomial $\psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$. In this paper we consider only *nonintrusive* methods which are based on repeatedly evaluating the model $\mathcal{M}$ over a set of input realizations $\boldsymbol{\mathcal{X}} = \{\boldsymbol{\chi}^{(1)}, \ldots, \boldsymbol{\chi}^{(N)}\}$, the so-called *experimental design*. Different nonintrusive methods have been proposed in the last decade to calibrate PC meta-models, namely projection [20–22], stochastic collocation [23, 24], and least-square minimization methods [26, 29, 32, 54, 55]. In this paper we adopt the least-square minimization method. The expansion coefficients $\boldsymbol{a} = \{\boldsymbol{a_\alpha}, \boldsymbol{\alpha} \in \mathcal{A} \subset \mathbb{N}^M\}$ are calculated by minimizing the expectation of the least-squares residual:

$$\boldsymbol{a} = \arg\min_{\boldsymbol{a} \in \mathbb{R}^{|\mathcal{A}|}} \mathbb{E}\left[\left(Y - \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \boldsymbol{a_\alpha}\,\psi_{\boldsymbol{\alpha}}(\boldsymbol{X})\right)^2\right]. \tag{11}$$
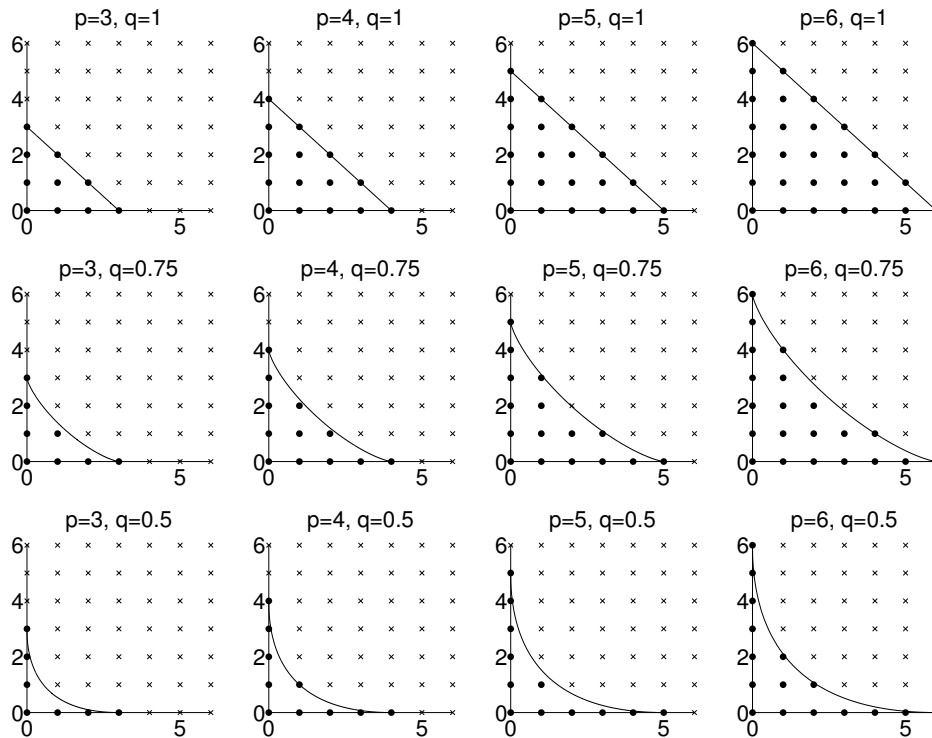

FIG. 1: Representation of a hyperbolic index set $\boldsymbol{\alpha} \in \mathcal{A}_q^{M,p}$ for various p and q ($M = 2$).

In practice the expectation in Eq. (11) is evaluated by an empirical sample-based estimator. Denoting by $\mathcal{Y} = \{\mathcal{M}(\boldsymbol{\chi}^{(1)}), \ldots, \mathcal{M}(\boldsymbol{\chi}^{(N)})\} \equiv \{\mathcal{Y}^{(1)}, \ldots, \mathcal{Y}^{(N)}\}$ the set of outputs of the exact model $\mathcal{M}$ for each point in the experimental design $\boldsymbol{\mathcal{X}}$, the discretized least-squares error minimization problem derived from Eq. (11) reads

$$\hat{\boldsymbol{a}} = \underset{\boldsymbol{a} \in \mathbb{R}^{|\mathcal{A}|}}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} \left(\mathcal{Y}^{(i)} - \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \boldsymbol{a}_{\boldsymbol{\alpha}} \, \psi_{\boldsymbol{\alpha}}(\boldsymbol{\chi}^{(i)}) \right)^2 . \tag{12}$$

The optimal expansion coefficients $\hat{\boldsymbol{a}}$ may be computed by solving the linear system

$$\hat{\boldsymbol{a}} = (\mathbf{F}^{\mathsf{T}} \mathbf{F})^{-1} \mathbf{F}^{\mathsf{T}} \mathcal{Y}, \tag{13}$$

where $\mathbf{F}$ is the information matrix of size $N \times |\mathcal{A}|$ whose generic term reads:

$$F_{ij} = \psi_j(\boldsymbol{\chi}^{(i)}), \quad i = 1, \ldots, N, \;\; j = 1, \ldots, |\mathcal{A}|. \tag{14}$$

Typically for smooth functions, a small number of polynomials is able to represent accurately the output of the computational model. Thus a further reduction of the set of predictors in high-dimensional space is possible. Various types of generalized regression algorithms have been proposed in the literature, namely the *least absolute shrinkage operator* (LASSO) [56], the *Least Angle Regression* (LAR) [57], *low-rank approximations* [33, 58], and compressive sensing [59]. In the applications of this paper, the LAR algorithm is used in combination with hyperbolic index sets.

Given a PCE meta-model, i.e., a set of polynomials $\psi_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathcal{A}$, and the corresponding parameters $\boldsymbol{a}_{\boldsymbol{\alpha}}$, the response of a new sample $\boldsymbol{x} \in \mathcal{D}_X$ may be eventually predicted by Eq. (6):

$$y^{(\mathrm{PCE})} = \mathcal{M}^{(\mathrm{PCE})}(\boldsymbol{x}). \tag{15}$$

2.3 Error Estimation

As seen in Eq. (6), PCEs are approximations of the exact computational model and thus the prediction at new input samples leads to some residual error. That is why error measures are developed to quantify the deviation between the exact output $\mathcal{Y} = \{\mathcal{Y}^{(i)}, i = 1, \ldots, N\}$ and the meta-model output $\mathcal{Y}^{(\mathrm{PCE})} = \mathcal{M}^{(\mathrm{PCE})}(\boldsymbol{\mathcal{X}})$. The *generalization error* (also called L^2-error) is the expectation of the squared output residuals [60], i.e.,

$$Err_{gen} = \mathbb{E}\left[\left(Y - Y^{(\mathrm{PCE})} \right)^2 \right], \tag{16}$$

where $Y^{(\mathrm{PCE})}$ corresponds to the truncated series [see Eq. (6)] and the expectation is defined with respect to the PDF of the input variables $\boldsymbol{X}$. If the computational model $\mathcal{M}$ is inexpensive to evaluate, the generalization error can be estimated accurately using an auxiliary validation set $\mathbb{X} = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}\}$, which is sampled from the input distribution $f_{\boldsymbol{X}}$. The estimate of the generalization error then reads

$$\widehat{Err}_{gen} = \frac{1}{n} \sum_{i=1}^{n} \left(\mathcal{M}(\boldsymbol{x}^{(i)}) - \mathcal{M}^{(\mathrm{PCE})}(\boldsymbol{x}^{(i)}) \right)^2 . \tag{17}$$

However, this rarely happens in real applications since the very purpose of building a meta-model is to avoid evaluating $\mathcal{M}$ on a large sample set $\mathbb{X}$. Note that in Section 5 though, the various meta-modeling techniques are compared on analytical benchmark functions, making it possible to use such an estimate of the generalization error.

When the use of a large validation set is not affordable, the *empirical error* based on the available experimental design $\boldsymbol{\mathcal{X}}$ may be defined:

$$Err_{emp} \equiv \frac{1}{N} \sum_{i=1}^{N} \left(\mathcal{Y}^{(i)} - \mathcal{M}^{(\mathrm{PCE})}(\boldsymbol{\chi}^{(i)}) \right)^2 . \tag{18}$$

Normalizing the empirical error by the variance of the output values leads to the *relative empirical error* which is defined as

$$\epsilon_{emp} \equiv \frac{\sum_{i=1}^{N} \left(\mathcal{Y}^{(i)} - \mathcal{M}^{(\text{PCE})}(\chi^{(i)}) \right)^2}{\sum_{i=1}^{N} \left(\mathcal{Y}^{(i)} - \mu_{\mathcal{Y}} \right)^2}, \tag{19}$$

where $\mu_{\mathcal{Y}}$ is the mean value of the output values $\mathcal{Y}$. The empirical error which is based on the experimental design generally underestimates the generalization error. In particular, if the number of polynomials $|\mathcal{A}|$ is close to the number of samples N in the experimental design, the empirical error tends to zero whereas the true (generalization) error does not. This phenomenon is called *overfitting* (in the extreme case where $N = |\mathcal{A}|$ the predictors may interpolate the experimental design points and thus the empirical error vanishes). Hence the *leave-one-out* (LOO) error has been proposed as an estimate of the generalization error [61, 62]. The general formulation of the leave-one-out error is

$$Err_{LOO}^{(\text{PCE})} \equiv \frac{1}{N} \sum_{i=1}^{N} \left(\mathcal{Y}^{(i)} - \mathcal{M}_{(-i)}^{(\text{PCE})}(\chi^{(i)}) \right)^2, \tag{20}$$

where $\mathcal{M}_{(-i)}^{(\text{PCE})}(\cdot)$ is a PCE model built from the sample set $\mathcal{X}^{(-i)} = \mathcal{X} \backslash \chi^{(i)} \equiv \{\chi^{(j)}, j = 1, \ldots, i-1, i+1, \ldots, N\}$ and $\mathcal{Y} = \{\mathcal{Y}^{(i)}, i = 1, \ldots, N\}$ are the response values of the exact computational model. The LOO error is a special case of the *leave-k-out cross-validation* error [63], which discards k samples from the initial experimental design to build up a model and predict the error at the k discarded samples.

In theory the computational cost of the LOO error is proportional to the number of samples N since it would require the determination of N PCE meta-models corresponding to each experimental design $\mathcal{X}^{(-i)}$. In the special case of linearly parameterized regression, which is the case for PCE, it is possible to calculate the LOO error analytically *without* building N separate models. The LOO error reads (see, e.g., [53, 64] for the proof)

$$Err_{LOO}^{(\text{PCE})} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\mathcal{Y}^{(i)} - \mathcal{M}^{(\text{PCE})}(\chi^{(i)})}{1 - h_i} \right)^2, \tag{21}$$

where h_i is the ith diagonal term of the matrix $\mathbf{F} \left(\mathbf{F}^{\mathsf{T}} \mathbf{F} \right)^{-1} \mathbf{F}^{\mathsf{T}}$ and the information matrix $\mathbf{F}$ is defined in Eq. (14). Note that the PCE used in Eq. (21) is built only once from the full experimental design $\mathcal{X}$.

3. KRIGING

3.1 Problem Definition

The second meta-modeling technique in this paper is Kriging, also known as *Gaussian process modeling*, which assumes that the response of a computational model is a realization of a Gaussian random process [9], i.e.,

$$\mathcal{M}(\boldsymbol{x}) \approx \mathcal{M}^{(\text{K})}(\boldsymbol{x}) = \boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{f}(\boldsymbol{x}) + \sigma^2 Z(\boldsymbol{x}), \tag{22}$$

where $\boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{f}(\boldsymbol{x}) = \sum_{j=1}^{P} \beta_j f_j(\boldsymbol{x})$ is the mean value of the Gaussian process, also called trend, with coefficients $\boldsymbol{\beta}$, σ^2 is the Gaussian process variance and $Z(\boldsymbol{x})$ is a zero-mean, unit-variance stationary Gaussian process. The zero-mean Gaussian process $Z(\boldsymbol{x})$ is fully determined by the auto-correlation function between two input sample points $R(\boldsymbol{x}, \boldsymbol{x}') = R(|\boldsymbol{x} - \boldsymbol{x}'|; \theta)$ due to stationarity, where θ are hyper-parameters to be computed.

Various correlation functions can be found in the literature [8, 9], some of which are the *linear*, *exponential*, *Gaussian* (also called *squared exponential*), and *Matérn* autocorrelation function. In this paper the Matérn autocorrelation function is mainly used as it is a generalization of the exponential and the Gaussian autocorrelation functions. The general Matérn kernel of degree ν is defined as [65]

$$R(|\boldsymbol{x} - \boldsymbol{x}'|; \boldsymbol{l}, \nu) = \prod_{i=1}^{M} \frac{1}{2^{\nu-1} \Gamma(\nu)} \left(\sqrt{2\nu} \frac{|x_i - x_i'|}{l_i} \right)^{\nu} \kappa_{\nu} \left(\sqrt{2\nu} \frac{|x_i - x_i'|}{l_i} \right), \tag{23}$$

where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are two sample points in the input space $\mathcal{D}_X$, $\boldsymbol{l} = \{l_i > 0, \ i = 1, \ldots, M\}$ are the scale parameters (also called *correlation lengths*), $\nu \geq 1/2$ is the shape parameter, $\Gamma(\cdot)$ is the Euler Gamma function, and $\kappa_\nu(\cdot)$ is the modified Bessel function of the second kind (also known as Bessel function of the third kind). In many publications the shape parameter is set to either $\nu = 3/2$ or $\nu = 5/2$ which simplifies Eq. (23) to [49]:

$$R(|\boldsymbol{x} - \boldsymbol{x}'|; \boldsymbol{l}, \nu = 3/2) = \prod_{i=1}^{M} \left(1 + \frac{\sqrt{3}\,|x_i - x_i'|}{l_i}\right) \exp\left(-\frac{\sqrt{3}\,|x_i - x_i'|}{l_i}\right), \tag{24}$$

$$R(|\boldsymbol{x} - \boldsymbol{x}'|; \boldsymbol{l}, \nu = 5/2) = \prod_{i=1}^{M} \left(1 + \frac{\sqrt{5}\,|x_i - x_i'|}{l_i} + \frac{5(x_i - x_i')^2}{3\,l_i^2}\right) \exp\left(-\frac{\sqrt{5}\,|x_i - x_i'|}{l_i}\right). \tag{25}$$

Apart from the correlation part in Eq. (22) there is also a trend part $\boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{f}(\boldsymbol{x})$. Three different flavors of Kriging are defined in the literature [8–10], namely simple, ordinary, and universal Kriging according to the choice of the trend. *Simple Kriging* assumes that the trend has a known constant value, i.e., $\boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{f}(\boldsymbol{x}) = \beta_0$. In *ordinary Kriging* the trend has a constant but unknown value, i.e., $P = 1$, $f_1(\boldsymbol{x}) = 1$, and β_1 is unknown. The most general and flexible formulation is *universal Kriging*, which assumes that the trend is composed of a sum of P pre-selected functions $f_k(\boldsymbol{x})$, i.e.,

$$\boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{f}(\boldsymbol{x}) = \sum_{k=1}^{P} \beta_k f_k(\boldsymbol{x}), \tag{26}$$

where β_k is the trend coefficient of each function. Note that simple and ordinary Kriging are special cases of universal Kriging. As discussed later in this paper, one approach to set up a trend is to use a sparse set of polynomials, which defines a new variant of universal Kriging.

3.2 Calibration of the Kriging Model

Given a value for the auto-correlation hyper-parameters $\hat{\boldsymbol{\theta}}$, the calibration of the Kriging model parameters $\{\boldsymbol{\beta}(\hat{\boldsymbol{\theta}}), \sigma_y^2(\hat{\boldsymbol{\theta}})\}$ may be computed using an *empirical best linear unbiased estimator* (BLUE). The optimization yields an analytical expression as a function of $\hat{\boldsymbol{\theta}}$:

$$\boldsymbol{\beta}(\hat{\boldsymbol{\theta}}) = \left(\mathbf{F}^{\mathsf{T}} \mathbf{R}^{-1} \mathbf{F}\right)^{-1} \mathbf{F}\, \mathbf{R}^{-1} \mathcal{Y}, \tag{27}$$

$$\sigma_y^2(\hat{\boldsymbol{\theta}}) = \frac{1}{N} \left(\mathcal{Y} - \mathbf{F}\, \boldsymbol{\beta}\right)^{\mathsf{T}} \mathbf{R}^{-1} \left(\mathcal{Y} - \mathbf{F}\, \boldsymbol{\beta}\right), \tag{28}$$

where $\mathcal{Y} = \{\mathcal{Y}^{(i)}, \ i = 1, \ldots, N\}$ are model responses of the exact computational model on the experimental design $\boldsymbol{\mathcal{X}} = \{\boldsymbol{\chi}^{(i)}, \ i = 1, \ldots, N\}$, $\mathbf{R}_{ij} = R(|\boldsymbol{\chi}^{(i)} - \boldsymbol{\chi}^{(j)}|; \hat{\boldsymbol{\theta}})$ is the correlation matrix and $\mathbf{F}_{ij} = f_j(\boldsymbol{\chi}^{(i)})$ is the information matrix.

In recent developments, the optimal correlation parameters $\hat{\boldsymbol{\theta}}$ may be determined by either a maximum-likelihood-estimate (denoted by *ML*) [47, 66] or by leave-one-out cross-validation (*CV*) [67]. The optimal parameters are determined through a minimization which reads

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\min_{\boldsymbol{\theta}} \left[\frac{1}{N} \left(\mathcal{Y} - \mathbf{F}\, \boldsymbol{\beta}\right)^{\mathsf{T}} \mathbf{R}^{-1} \left(\mathcal{Y} - \mathbf{F}\, \boldsymbol{\beta}\right) (\det \mathbf{R})^{1/N}\right], \tag{29}$$

$$\hat{\boldsymbol{\theta}}_{CV} = \arg\min_{\boldsymbol{\theta}} \left[\mathcal{Y}^{\mathsf{T}} \mathbf{R}(\boldsymbol{\theta})^{-1} \mathrm{diag}\left(\mathbf{R}(\boldsymbol{\theta})^{-1}\right)^{-2} \mathbf{R}(\boldsymbol{\theta})^{-1} \mathcal{Y}\right]. \tag{30}$$

The comparison of both approaches shows that ML is preferable to CV in well-specified cases, i.e., when the meta-model autocorrelation function family is identical to the autocorrelation function of the computational model. For

practical problems, i.e., assuming a black-box model, the autocorrelation function family is not known with certainty. In this case CV shall lead to more robust results than ML, as discussed in [67].

Determining the optimal correlation parameters in Eqs. (29) and (30) is a complex multi-dimensional minimization problem. Optimization algorithms can be cast into two distinct categories: local and global optimization algorithms. Local methods are usually gradient-based algorithms such as the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [68–70] and its modifications [71]. Global methods are algorithms such as genetic algorithms [72] and differential evolution algorithms [73, 74]. The best optimization algorithm is problem dependent and in many cases not known *a priori*.

The optimal correlation parameters are then used for predicting the model response at new samples of the input space. By assumption, the prediction of a Kriging model of a new point $\boldsymbol{x}$ is a Gaussian random variable with mean $\mu_{\hat{y}}(\boldsymbol{x})$ and variance $\sigma^2_{\hat{y}}(\boldsymbol{x})$:

$$\mu_{\hat{y}}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta} + \boldsymbol{r}(\boldsymbol{x})^\mathsf{T}\mathbf{R}^{-1}\left(\mathcal{Y} - \mathbf{F}\boldsymbol{\beta}\right), \tag{31}$$

$$\sigma^2_{\hat{y}}(\boldsymbol{x}) = \sigma^2_y \left(1 - \langle \boldsymbol{f}(\boldsymbol{x})^\mathsf{T} \boldsymbol{r}(\boldsymbol{x})^\mathsf{T}\rangle \left[\begin{array}{cc} \mathbf{0} & \mathbf{F}^\mathsf{T} \\ \mathbf{F} & \mathbf{R} \end{array}\right]^{-1} \left[\begin{array}{c} \boldsymbol{f}(\boldsymbol{x}) \\ \boldsymbol{r}(\boldsymbol{x}) \end{array}\right]\right), \tag{32}$$

where $r_i(\boldsymbol{x}) = R(|\boldsymbol{x} - \boldsymbol{\chi}^{(i)}|; \boldsymbol{\theta})$ is the correlation between the new sample $\boldsymbol{x}$ and the sample $\boldsymbol{\chi}^{(i)}$ of the experimental design. The prediction mean is used as the surrogate to the original model $\mathcal{M}$, whereas the variance gives a *local* error indicator about the precision. Note that the Kriging model interpolates the data, i.e.,

$$\mu_{\hat{y}}(\boldsymbol{\chi}^{(i)}) = \mathcal{M}(\boldsymbol{\chi}^{(i)}), \quad \sigma^2_{\hat{y}}(\boldsymbol{\chi}^{(i)}) = 0, \quad \forall \boldsymbol{\chi}^{(i)} \in \boldsymbol{\mathcal{X}}.$$

Apart from this procedure to calibrate the meta-model and predict model responses, the Kriging meta-modeling technique has been developed further in recent years. The latest developments in Kriging are contributed in the aspects of optimal estimation of the hyper-parameters [67, 75, 76], the use of adaptive kernels [77, 78], and the use of additive auto-correlation kernels [79–81].

3.3 Error Estimation

A local error measure for any sample $\boldsymbol{x}$ is given by the prediction variance $\sigma^2_{\hat{y}}(\boldsymbol{x})$ in Eq. (32). This information is useful to detect regions where the prediction accuracy is low. Adding new samples to the experimental design $\boldsymbol{\mathcal{X}}$ in the regions with high prediction variance may lead to an overall increase in the accuracy of the meta-model in that region. This characteristics is exploited when devising adaptive experimental designs in structural reliability analysis, see [42, 46, 82, 83].

A simple global error measure of the accuracy of the meta-model [such as Eq. (19) for PC expansions] is not available for Kriging due to its interpolating properties, which make the empirical error vanish (considering no nugget effect in its auto-correlation function). Thus one approach to a global error measure is the *leave-one-out* (LOO) error

$$Err_{LOO}^{(\mathrm{K})} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathcal{Y}^{(i)} - \mu_{\hat{y},(-i)}(\boldsymbol{\chi}^{(i)})\right)^2, \tag{33}$$

where $\mu_{\hat{y},(-i)}(\boldsymbol{\chi}^{(i)})$ is the prediction mean $\mu_{\hat{y}}$ of sample $\boldsymbol{\chi}^{(i)}$ by a Kriging meta-model based on the experimental design $\boldsymbol{\mathcal{X}}^{(-i)} = \boldsymbol{\mathcal{X}}\backslash\boldsymbol{\chi}^{(i)}$ and $\mathcal{Y} = \{\mathcal{Y}^{(i)}, i = 1,\ldots,N\}$ is the exact model response. [84] derived an analytical solution for the LOO error for universal Kriging without computing the N meta-models explicitly in the same spirit as Eq. (21) for PC expansions. The prediction mean and variance are given by

$$\mu_{\hat{y},(-i)} = -\sum_{j=1,j\neq i}^{N}\frac{\mathbf{B}_{ij}}{\mathbf{B}_{ii}}\mathcal{Y}^{(j)} = -\sum_{j=1}^{N}\frac{\mathbf{B}_{ij}}{\mathbf{B}_{ii}}\mathcal{Y}^{(j)} + \mathcal{Y}^{(i)}, \tag{34}$$

$$\sigma^2_{\hat{y},(-i)} = \frac{1}{\mathbf{B}_{ii}}, \tag{35}$$

where $\mathbf{B}$ is a square matrix of size $(N + P)$ with N and P denoting the number of samples in the experimental design and the number of polynomials in the trend part, respectively:

$$\mathbf{B} = \left[\begin{array}{cc} \sigma^2 \mathbf{R} & \mathbf{F} \\ \mathbf{F}^\mathsf{T} & \mathbf{0} \end{array} \right]^{-1}, \tag{36}$$

where σ^2 is the Kriging variance for the full experimental design $\mathcal{X}$ estimated by Eq. (28). A generalized version of this algorithm called *v-fold cross-correlation error* can be found in [84].

3.4 PCE as a Particular Case of Universal Kriging

PC expansions can be interpreted as Kriging models where the samples of the experimental design are uncorrelated, i.e., where the auto-correlation function consists of the Dirac function:

$$R(|\boldsymbol{x} - \boldsymbol{x}'|) = \delta(\boldsymbol{x} - \boldsymbol{x}'). \tag{37}$$

The correlation matrix is then the identity matrix of size N, i.e., $\mathbf{R} = \mathbf{I}_N$. This reduces Eq. (27) to Eq. (13) and Eq. (31) to Eq. (2).

Further, it can be shown that the LOO error in Eqs. (33)-(35) reduces to Eq. (21) by the following derivations. Consider the symmetric partitioned matrix $\mathbf{C}$ and the corresponding inverse $\mathbf{D}$, i.e., $\mathbf{D} = \mathbf{C}^{-1}$ which are defined as

$$\mathbf{C} = \left[\begin{array}{cc} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}^\mathsf{T} & \mathbf{C}_{22} \end{array} \right], \qquad \mathbf{D} = \left[\begin{array}{cc} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}_{12}^\mathsf{T} & \mathbf{D}_{22} \end{array} \right],$$

where $\mathbf{C}_{11}$, $\mathbf{D}_{11}$ (respectively, $\mathbf{C}_{22}$, $\mathbf{D}_{22}$) are square matrices with dimension N (respectively, P). Using block matrix inversion one can derive:

$$\mathbf{D}_{11} = \mathbf{C}_{11}^{-1} + \mathbf{C}_{11}^{-1} \mathbf{C}_{12} \left(\mathbf{C}_{22} - \mathbf{C}_{12}^\mathsf{T} \mathbf{C}_{11}^{-1} \mathbf{C}_{12} \right)^{-1} \mathbf{C}_{12}^\mathsf{T} \mathbf{C}_{11}^{-1}. \tag{38}$$

In the context of the LOO error, taking $\mathbf{C} \equiv \mathbf{B}$ in Eq. (36), $\mathbf{C}_{11} = \sigma^2 \mathbf{I}_N$, $\mathbf{C}_{12} = \mathbf{F}$, $\mathbf{C}_{22} = \mathbf{0}_P$:

$$\begin{aligned} \mathbf{D}_{11} &= \frac{1}{\sigma^2} \mathbf{I}_N + \frac{1}{\sigma^2} \mathbf{I}_N \mathbf{F} \left(\mathbf{0}_P - \mathbf{F}^\mathsf{T} \frac{1}{\sigma^2} \mathbf{I}_N \mathbf{F} \right)^{-1} \mathbf{F}^\mathsf{T} \frac{1}{\sigma^2} \mathbf{I}_N \\ &= \frac{1}{\sigma^2} \mathbf{I} + \frac{1}{\sigma^2} \mathbf{F} \left(-\mathbf{F}^\mathsf{T} \mathbf{F} \right)^{-1} \mathbf{F}^\mathsf{T} = \frac{1}{\sigma^2} \left(\mathbf{I} - \mathbf{F} \left(\mathbf{F}^\mathsf{T} \mathbf{F} \right)^{-1} \mathbf{F}^\mathsf{T} \right). \end{aligned} \tag{39}$$

Then, the LOO error in Eq. (33) combined with Eq. (34) and the above inverse formulation of the $\mathbf{B}$ matrix reads:

$$\begin{aligned} Err_{LOO}^{\mathsf{K}} &= \frac{1}{N} \sum_{i=1}^{N} \left(\mathcal{Y}^{(i)} + \sum_{j=1}^{N} \frac{B_{ij}}{B_{ii}} \mathcal{Y}^{(j)} - \mathcal{Y}^{(i)} \right)^2 = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{B_{ii}} \sum_{j=1}^{N} B_{ij} \mathcal{Y}^{(j)} \right)^2 \\ &= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{B_{ii} \sigma^2} \sum_{j=1}^{N} \left(\mathbf{I} - \mathbf{F} \left(\mathbf{F}^\mathsf{T} \mathbf{F} \right)^{-1} \mathbf{F}^\mathsf{T} \right)_{ij} \mathcal{Y}^{(j)} \right)^2 \\ &= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{1 - \left(\mathrm{diag} \left[\mathbf{F} \left(\mathbf{F}^\mathsf{T} \mathbf{F} \right)^{-1} \mathbf{F}^\mathsf{T} \right] \right)_i} \cdot \left[\mathcal{Y}^{(i)} - \sum_{j=1}^{N} \left[\mathbf{F} \left(\mathbf{F}^\mathsf{T} \mathbf{F} \right)^{-1} \mathbf{F}^\mathsf{T} \right]_{ij} \mathcal{Y}^{(j)} \right] \right)^2 \\ &= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\mathcal{Y}^{(i)} - \left[\left(\mathbf{F}^\mathsf{T} \mathbf{F} \right)^{-1} \mathbf{F}^\mathsf{T} \mathcal{Y} \right]^\mathsf{T} f(\boldsymbol{\chi}^{(i)})}{1 - \left(\mathrm{diag} \left[\mathbf{F} \left(\mathbf{F}^\mathsf{T} \mathbf{F} \right)^{-1} \mathbf{F}^\mathsf{T} \right] \right)_i} \right)^2, \end{aligned} \tag{40}$$

which is equivalent to the formulation of the LOO error in Eq. (21) for the case of PCE and $f(\boldsymbol{\chi}^{(i)}) \equiv \psi(\boldsymbol{\chi}^{(i)})$. Thus, the LOO error in PCE can be seen as a special case of the LOO error in the Kriging framework.

4. PC-KRIGING

4.1 Principle

The characteristic of Kriging is to interpolate local variations of the output of the computational model as a function of the neighboring experimental design points. In contrast, polynomial chaos expansions (PCE) are used for approximating the *global* behavior of $\mathcal{M}$ using a set of orthogonal polynomials. By combining the two techniques we aim at capturing the global behavior of the computational model with the set of orthogonal polynomials in the trend of a universal Kriging model and the local variability with the Gaussian process. The new approach called *polynomial-chaos-Kriging* (PC-Kriging) combines these two distinct meta-modeling techniques and their characteristics.

Using now the standard notation for truncated polynomial chaos expansions [see Eq. (6)], we cast the PC-Kriging meta-model as follows [85]:

$$\mathcal{M}(\boldsymbol{x}) \approx \mathcal{M}^{(\text{PCK})}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \boldsymbol{a}_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) + \sigma^2 Z(\boldsymbol{x}), \tag{41}$$

where $\sum_{\boldsymbol{\alpha} \in \mathcal{A}} \boldsymbol{a}_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ is a weighted sum of orthonormal polynomials describing the mean value of the Gaussian process and $\mathcal{A}$ is the index set of the polynomials. $Z(\boldsymbol{x})$ is a zero-mean, unit-variance stationary Gaussian process defined by an autocorrelation function $R(|\boldsymbol{x} - \boldsymbol{x}'|; \boldsymbol{\theta})$ and is parametrized by a set of hyper-parameters $\boldsymbol{\theta}$.

Building a PC-Kriging meta-model consists of two parts: the determination of the optimal set of polynomials contained in the regression part (i.e., the truncation set $\mathcal{A}$) and the calibration of the correlation hyper-parameters $\boldsymbol{\theta}$ as well as the Kriging parameters $\{\sigma^2, \boldsymbol{a}_{\boldsymbol{\alpha}}\}$. The set of polynomials is determined using the least-angle-regression (LAR) algorithm as in [32] together with hyperbolic index sets to obtain sparse sets of polynomials. After the set of polynomials is fixed, the trend and correlation parameters are evaluated using the universal Kriging equations [Eqs. (27)–(30)].

4.2 Algorithm

The two distinct frameworks for PCE and Kriging can be combined in various ways. In this paper two approaches will be explained in detail, i.e., the sequential PC-Kriging (SPC-Kriging) and the optimal PC-Kriging (OPC-Kriging). Both approaches are based on the same input information, namely the experimental design $\boldsymbol{\mathcal{X}}$, the corresponding response values $\mathcal{Y}$ obtained from the computational model $\mathcal{M}(\boldsymbol{\mathcal{X}})$, the description of the stochastic input variables (joint PDF $f_{\boldsymbol{X}}$), and the parametric expression of an auto-correlation function $R(|\boldsymbol{x} - \boldsymbol{x}'|; \boldsymbol{\theta})$. The two approaches are defined as follows:

- *Sequential PC-Kriging* (SPC-Kriging): in this approach, the set of polynomials and the Kriging meta-model are determined *sequentially*. The assumption behind this procedure is that the optimal set of polynomials found by the LAR algorithm in the context of pure PCE can be used directly as an optimal trend for the universal Kriging. In a first step the optimal set of polynomials is determined using the PCE framework: $\mathcal{A}$ is found by applying the LAR procedure as in [32]. The set of multivariate orthonormal polynomials $\mathcal{A}$ is then embedded into a universal Kriging model as the trend. The universal Kriging meta-model is calibrated using Eqs. (27)–(30).

  At the end of the algorithm the accuracy of the meta-model can be measured by the LOO error given in Eq. (33) or, when using a validation set, by the sample-based generalization error in Eq. (17). The SPC-Kriging algorithm is illustrated in Fig. 2, in which the white boxes represent the required input information and the blue boxes represent the computational tasks. Given a calibrated SPC-Kriging model, the response of new input realizations (i.e., the prediction) is computed by Eq. (31) and Eq. (32).

- *Optimal PC-Kriging* (OPC-Kriging): in this approach, the PC-Kriging meta-model is obtained iteratively. The set of orthonormal multivariate polynomials is determined by LAR algorithm in the same way as in SPC-Kriging. Yet the LAR algorithm results in a list of ranked polynomials which are chosen depending on their correlation to the current residual at each iteration in decreasing order. OPC-Kriging consists of an iterative algorithm where each polynomial is added one-by-one to the trend part. In each iteration, the coefficients of the
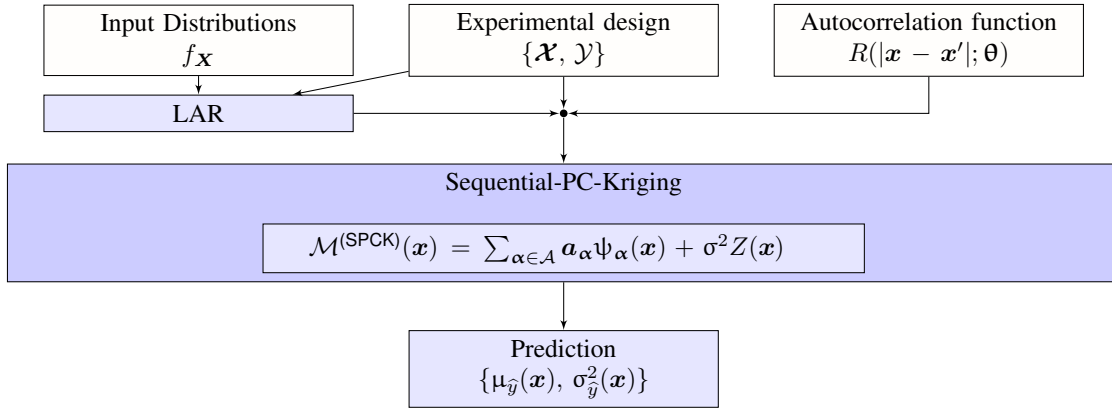
FIG. 2: Flowchart for sequential-PC-Kriging (SPC-Kriging).

trend and the parameters of the auto-correlation function are calibrated. In the end, a number $|\mathcal{A}|$ of different PC-Kriging models are available. The $|\mathcal{A}|$ meta-models are then compared in terms of the LOO error [Eq. (33)]. The optimal PC-Kriging meta-model is then chosen as the one with minimal LOO error.

Figure 3 illustrates the OPC-Kriging algorithm in a flowchart. The notation $\mathcal{M}^{(K)(Q)}(\boldsymbol{x})$ means a universal Kriging meta-model where the trend is modeled by the Q first polynomials selected in $\mathcal{A}$ according to the ranking obtained by LAR. Note that the red box represents the universal Kriging model which eventually minimizes the LOO error ($Q = P^*$) and which is thus finally chosen. The final box marks the prediction of new model responses which is computed by Eq. (31) and Eq. (32).
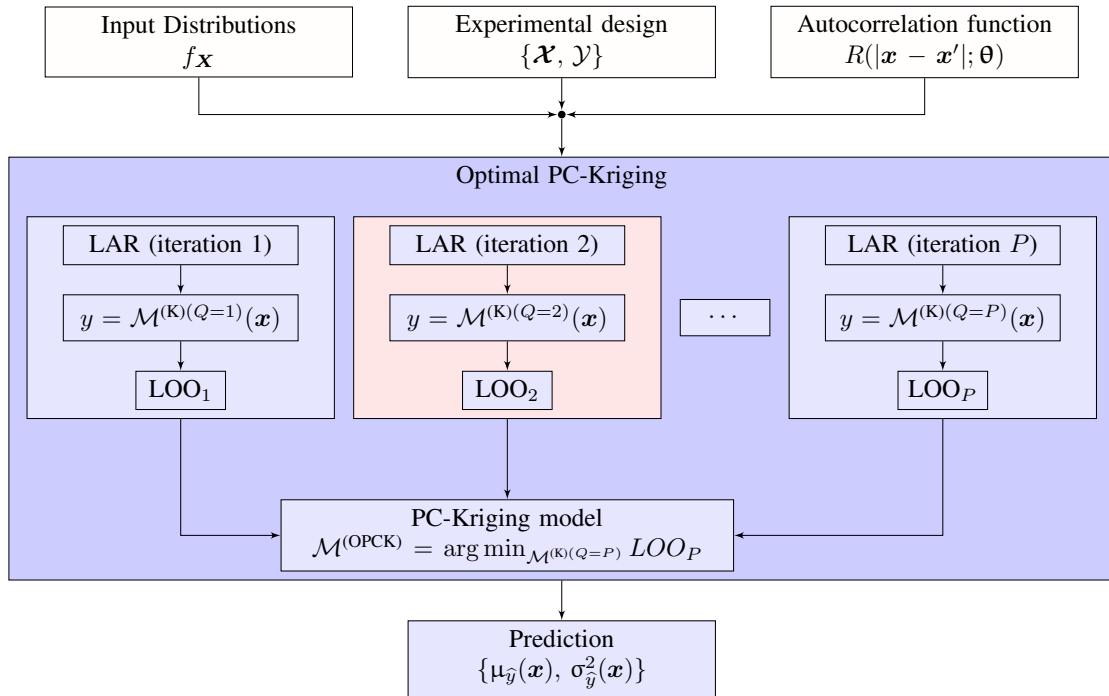
FIG. 3: Flowchart for optimal PC-Kriging (OPC-Kriging).

As a summary, PC-Kriging can be viewed as a universal Kriging meta-model with a nonstandard trend part. Thus the error estimates from Kriging in Section 3.3 are valid here with no modification. In particular, the LOO error in Eq. (33) is computed to compare different PC-Kriging models and to compute the optimal meta-model in OPC-Kriging. The performance of both approaches and the comparison to the traditional PCE and Kriging approaches is now illustrated for a variety of benchmark analytical functions.

5. ANALYTICAL BENCHMARK FUNCTIONS

5.1 Setup

Easy-to-evaluate functions are tested to verify and validate the new PC-Kriging approach by comparing the meta-model to the exact model response. In this paper, six analytical functions of different dimensionality are illustrated, namely four with uniformly distributed input variables (i.e., Ishigami, Sobol', Rosenbrock, and Morris functions) and two with Gaussian input variables (i.e., Rastrigin and O'Hagan function). References to the original use of the benchmark functions in the literature are given below.

The first four analytical functions use uniformly distributed random variables in the input space. The *Ishigami* function is a smooth function with three independent input parameters commonly used for benchmarking methods in global sensitivity analysis.

$$f_1(\boldsymbol{x}) = \sin x_1 + 7 \, \sin^2 x_2 + 0.1 \, x_3^4 \, \sin x_1, \tag{42}$$

where $X_i \sim \mathcal{U}(-\pi, \pi)$, $i = 1, 2, 3$. The *Sobol'* function is also well-known sensitivity analysis because the Sobol' indices are easy to derive analytically [86]:

$$f_2(\boldsymbol{x}) = \prod_{i=i}^{8} \frac{|4 \, x_i - 2| + c_i}{1 + c_i}, \tag{43}$$

where $X_i \sim \mathcal{U}(0, 1)$, $i = 1, \ldots, 8$, and $c = (1, 2, 5, 10, 20, 50, 100, 500)^\mathsf{T}$ as in [27]. Due to the absolute value operator in the enumerator the function behaves nonsmoothly at the point $x_i = 0.5$. The *Rosenbrock* function is a polynomial function with a 2-dimensional input space [87]:

$$f_3(\boldsymbol{x}) = 100 \, \left(x_2 - x_1^2\right)^2 + \left(1 - x_1\right)^2, \tag{44}$$

where $X_i \sim \mathcal{U}(-2, 2)$, $i = 1, 2$. The last function considered is the *Morris* function, which is defined by [88]

$$f_4(\boldsymbol{x}) = \sum_{i=1}^{20} \beta_i \, w_i + \sum_{i<j}^{20} \beta_{ij} \, w_i w_j + \sum_{i<j<l}^{20} \beta_{ijl} \, w_i w_j w_l + 5 \, w_1 w_2 w_3 w_4, \tag{45}$$

where $X_i \sim \mathcal{U}(0, 1)$, $i = 1, \ldots, 20$, and $w_i = 2 \, (x_i - 1/2)$ for all i except for $i = 3, 5, 7$ where $w_i = 2 \, (1.1 \, x_i / (x_i + 0.1) - 1/2)$. The coefficients are defined as: $\beta_i = 20$, $i = 1, \ldots, 10$; $\beta_{ij} = -15$, $i, j = 1, \ldots, 6$; $\beta_{ijl} = -10$, $i, j, l = 1, \ldots, 5$. The remaining coefficients are set equal to $\beta_i = (-1)^i$ and $\beta_{ij} = (-1)^{i+j}$ as in [53].

Two other benchmark functions of independent Gaussian variables are also studied. The *Rastrigin* function has a two-dimensional input space and is defined by [89]

$$f_5(\boldsymbol{x}) = 10 - \sum_{i=1}^{2} (x_i^2 - 5 \, \cos(2\pi \, x_i)), \tag{46}$$

where $X_i \sim \mathcal{N}(0, 1)$, $i = 1, 2$. The last function is the *O'Hagan* function, which is defined by [90]

$$f_6(\boldsymbol{x}) = \boldsymbol{a_1^\mathsf{T}} \boldsymbol{x} + \boldsymbol{a_2^\mathsf{T}} \sin(\boldsymbol{x}) + \boldsymbol{a_3^\mathsf{T}} \cos(\boldsymbol{x}) + \boldsymbol{x}^\mathsf{T} \boldsymbol{Q} \boldsymbol{x}, \tag{47}$$

where $X_i \sim \mathcal{N}(0, 1)$, $i = 1, \ldots, 15$. The vectors $\{\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{a_3}\}$ and matrix $\boldsymbol{Q}$ are defined in [90].

Note that the functions $f_1 - f_4$ have uniform input random variables. Accordingly the PC trend in PC-Kriging is built up from multivariate Legendre polynomials. In contrast f_5, f_6 have Gaussian input random variables. Thus the PC trend is modeled by multivariate Hermite polynomials (see Table 1).

5.2 Analysis

At the beginning of each algorithm the experimental design is generated with the Latin-hypercube sampling technique [91]. Then the meta-modeling is processed applying the four previously discussed meta-modeling techniques, i.e., ordinary Kriging, PCE, SPC-Kriging and OPC-Kriging. Note that for the Kriging meta-models, the maximum likelihood formulation [Eq. (29)] in combination with the gradient-based BFGS optimization algorithm is used in order to compute the optimal correlation parameters.

Their performance is compared by means of the *relative generalization error*, which is defined as the ratio between the generalization error [Eq. (16)] and the output variance:

$$\epsilon_{gen} = \frac{\mathbb{E}\left[\left(Y - Y^{(\mathsf{PCE})}\right)^2\right]}{\mathrm{Var}\left[Y\right]}. \tag{48}$$

The error is estimated here using a large validation set $\mathbb{X} = \{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(n)}\}$ of size $n = 10^5$, which results in

$$\widehat{\epsilon}_{gen} \approx \frac{\sum_{i=1}^{n}\left(\mathcal{M}(\boldsymbol{x}^{(i)}) - \widehat{\mathcal{M}}(\boldsymbol{x}^{(i)})\right)^2}{\sum_{i=1}^{n}\left(\mathcal{M}(\boldsymbol{x}^{(i)}) - \mu_y\right)^2}, \tag{49}$$

where μ_y is the mean value of the set of exact model responses over the validation set $\mathbb{Y} = \{y^{(1)}, \dots, y^{(n)}\} \equiv \{\mathcal{M}(\boldsymbol{x}^{(1)}), \dots, \mathcal{M}(\boldsymbol{x}^{(n)})\}$. For all Kriging-based approaches, the meta-model $\widehat{\mathcal{M}}(\boldsymbol{x})$ is the prediction mean $\mu_{\hat{y}}(\boldsymbol{x})$. Note that the samples in $\mathbb{X}$ follow the distribution of the input variables $\boldsymbol{X}$ in order to obtain a reliable error estimate.

For each experimental setup, the analysis is replicated to account for the statistical uncertainties in the experimental design. Fifty independent runs of the full analysis are carried out and the results are represented using boxplots. In a box plot the central mark represents the median value of the 50 runs, the edges are the 25th and 75th percentile denoted by q_{25} and q_{75}. The whiskers describe the boundary to the outliers. Outliers are defined as the values smaller than $q_{25} - 1.5\,(q_{75} - q_{25})$ or larger than $q_{75} + 1.5\,(q_{75} - q_{25})$.

5.3 Results

5.3.1 Visualization of PC-Kriging's behavior

The different types of meta-models are illustrated in Fig. 4, which shows a two-dimensional contour plot of the output of the Rastrigin function ($N = 128$ samples) [Fig. 4(a)] and its approximations by PCE, ordinary Kriging, and PC-Kriging [Fig. 4(b)–4(d)]. The Rastrigin function has a highly oscillatory behavior on the entire input space as seen in Fig. 4(a). This behavior is difficult to meta-model with a small number of samples because many local minima/maxima are missed out.

The analytical formulation of the Rastrigin function is a combination of a quadratic component and a high-frequency trigonometric component. The PCE model in Fig. 4(c) captures the global characteristic of the function,
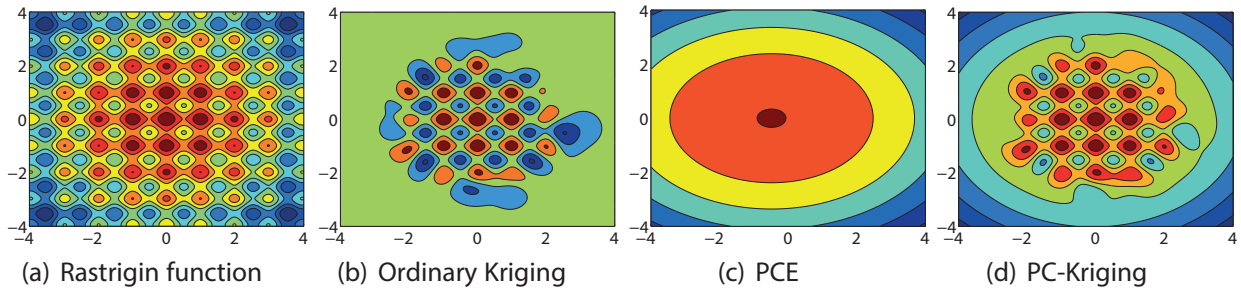

(a) Rastrigin function   (b) Ordinary Kriging   (c) PCE   (d) PC-Kriging

FIG. 4: Rastrigin function—Visual composition of PC-Kriging.

i.e., the quadratic component, whereas the ordinary Kriging model in Fig. 4(b) approximates the local characteristics, i.e., the high-frequency trigonometric component. Finally, the combination of PCE and Kriging leads to a better meta-model as shown in Fig. 4(d).

Note that the meta-models in Fig. 4 have a high accuracy around the origin of the coordinate system due to the definition of the input vector PDF as standard normal distributions ($X_i \sim \mathcal{N}(0,1)$).

5.3.2 Small Experimental Design

The four meta-modeling techniques are compared for the six analytical functions using experimental designs of increasing size. The number of samples is chosen so that it yields a large range of relative generalization errors on the second axis. The results are illustrated in Fig. 5–10. In each figure (a) shows the ordinary Kriging model and (b) shows the PCE model. The new approaches SPC-Kriging and OPC-Kriging are shown in (c) and (d), respectively.

Figure 5 shows the relative generalization error for the various meta-modeling techniques for the Ishigami function. For a small sample size of $N = 20$ samples, ordinary Kriging performs best with respect to the median value of the box plots. SPC-Kriging and OPC-Kriging perform worse due to the effect of overfitting (see also Section 2.3). The number of parameters to be estimated in the case of PC-Kriging is larger than in the case of ordinary Kriging due to the number of polynomials in the trend of the Kriging model. Thus, OPC-Kriging (and also SPC-Kriging) are more
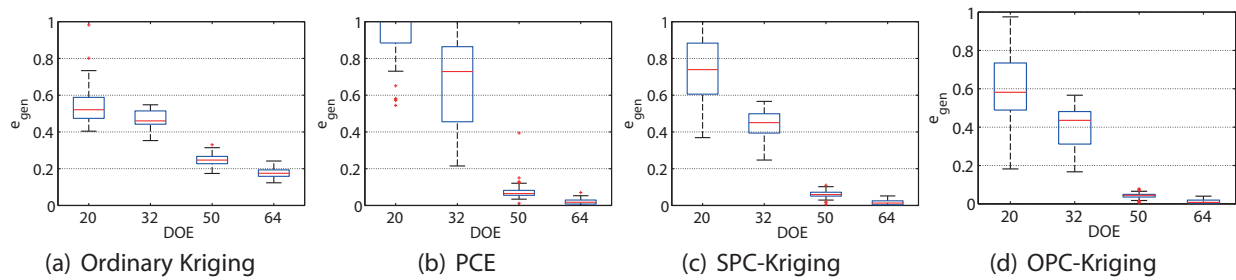
(a) Ordinary Kriging (b) PCE (c) SPC-Kriging (d) OPC-Kriging

FIG. 5: Ishigami function—Relative generalization error [Eq. (49)] for the various meta-modeling approaches.

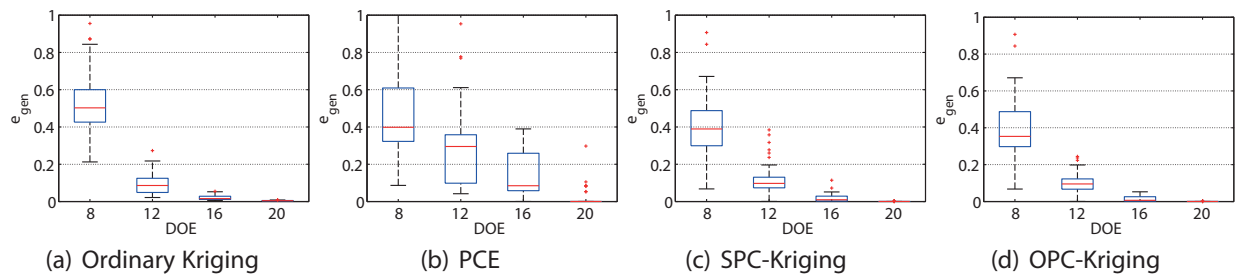(a) Ordinary Kriging (b) PCE (c) SPC-Kriging (d) OPC-Kriging

FIG. 6: Rosenbrock function—Relative generalization error [Eq. (49)] for the various meta-modeling approaches.

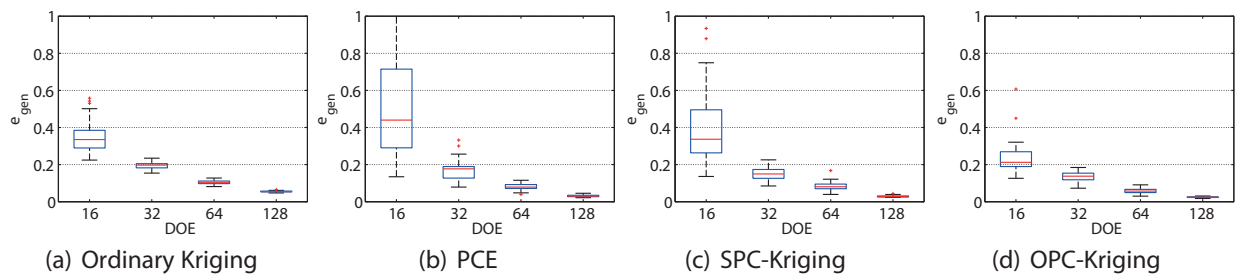(a) Ordinary Kriging (b) PCE (c) SPC-Kriging (d) OPC-Kriging

FIG. 7: Sobol' function—Relative generalization error [Eq. (49)] for the various meta-modeling approaches.
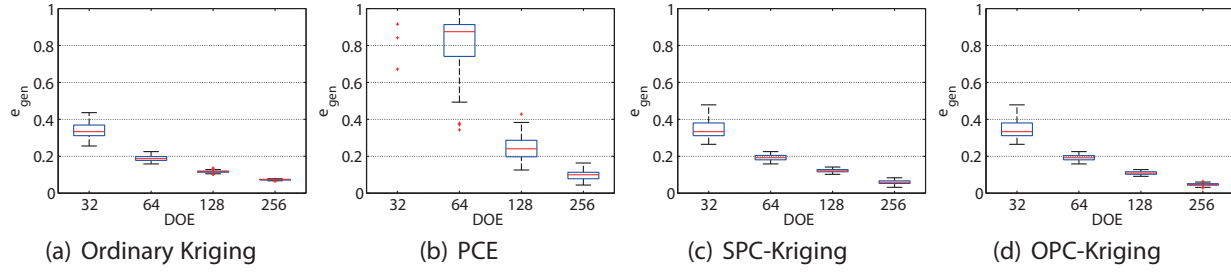
FIG. 8: Morris function—Relative generalization error [Eq. (49)] for the various meta-modeling approaches.
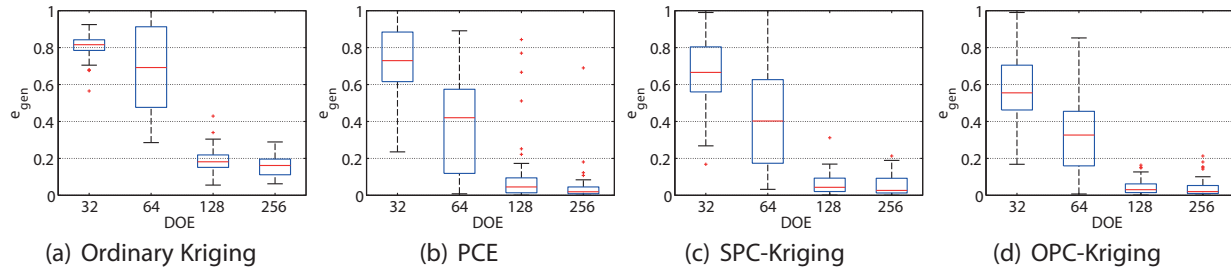

FIG. 9: Rastrigin function—Relative generalization error [Eq. (49)] for the various meta-modeling approaches.
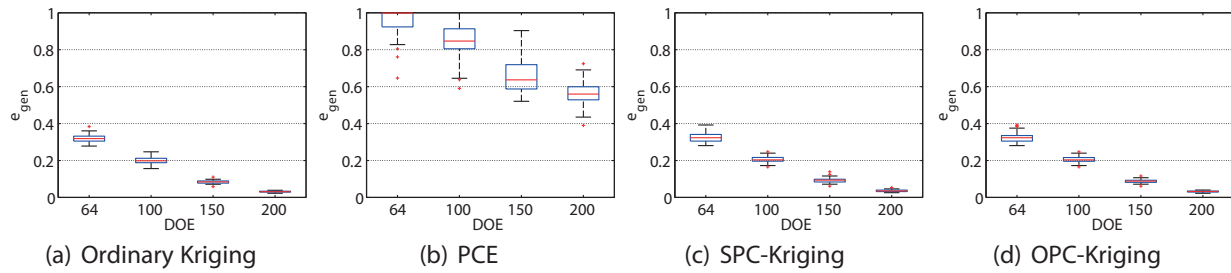

FIG. 10: O'Hagan function—Relative generalization error [Eq. (49)] for the various meta-modeling approaches.

prone to overfitting than ordinary Kriging for small experimental designs. When the number of samples is increased, however, the two PC-Kriging approaches perform better than the traditional approaches because their median value and their variation of the error are lower. For the large sample sizes ($N \geq 50$), PC-Kriging performs similarly to PCE, though slightly better. OPC-Kriging is slightly more accurate than SPC-Kriging over the whole range of sample sizes.

Figure 6 presents the results of the Rosenbrock function, which is a purely polynomial function and can be modeled accordingly with a small number of polynomials based on a small number of points in the experimental design (e.g., in the case of PCE). This is the reason why the number of points lies within $N = 8, \ldots, 20$. Highly accurate surrogate models are obtained with only 20 samples. For small sample sizes OPC-Kriging performs best among the four techniques in terms of the relative generalization error.

The Sobol' function is more complex than the two previous functions because of the dimensionality ($M = 8$) and the nonsmooth behavior at $x_i = 0.5$. Thus more samples are needed to obtain a similar range of relative generalization errors compared to the previous functions, as seen in Fig. 7. Behaviors for the larger sample sizes ($N = 64, 128$) are very similar among the meta-modeling approaches, although PC-Kriging performs slightly better than the traditional PCE and ordinary Kriging approaches. For very small sample sizes ($N = 16, 32$), OPC-Kriging performs significantly better than the others.

Figure 8 shows the results for the Morris function. A large experimental design is required to properly surrogate the computational model because of the high dimensionality of the input vector X and the amount of interactive terms

of different input variables X_i in the analytical formulation [see Eq. (45)]. The relative generalization error of the two PC-Kriging approaches resembles more the one of ordinary Kriging than the one of PCE in this case. PCE is not capable of modeling this analytical function with a small number of samples.

The results associated with the Rastrigin function are shown in Fig. 9. Despite the low dimensionality of the input ($M = 2$), many samples are needed to obtain small error estimates. This is because the function output is highly oscillatory over the entire input space as previously illustrated in Fig. 4. In comparison to Section 5.3.1, which describes the qualitative performance of PC-Kriging on the Rastrigin function, the quantitative benefit of combining PCE and Kriging becomes visible in Fig. 9: PC-Kriging performs better than the traditional approaches. Ordinary Kriging performs the worst followed by PCE. OPC-Kriging has statistically the lowest relative generalization errors over the whole range of experimental design sizes.

Figure 10 displays the results associated with the O'Hagan function. Similarly to the Morris function, the performance of PC-Kriging in the case of the O'Hagan function resembles that of ordinary Kriging whereas PCE performs worse than the other three approaches. Over the entire displayed range of experimental designs in Fig. 10, the performance of OPC-Kriging is slightly better than the performance of SPC-Kriging and ordinary Kriging. Note that meta-modeling the O'Hagan function requires less samples in the experimental design to obtain the same accuracy as in the case of the Rastrigin function despite the fact that the O'Hagan function has a 15-dimensional input space and that both functions are very smooth.

Summarizing the results of all six analytical functions in Figs. 5–10, the proposed PC-Kriging approaches perform better than or at least as good as the traditional PCE and Kriging approaches. Note that for functions like O'Hagan (Fig. 10) and Morris (Fig. 8) the performance of PC-Kriging is more similar to Kriging than PCE, whereas for the other functions the performance of PC-Kriging resembles more that of PCE. As one could expect, there is no general rule so as to decide whether PCE or Kriging provide the most accurate meta-models for a given experimental design. The advantage of PC-Kriging is to perform as least as well as the best of the two.

The combination of PCE and Kriging and its increased accuracy comes with a higher computational cost. The traditional ordinary Kriging and PCE approaches have the lowest computational cost, SPC-Kriging has an intermediate and OPC-Kriging has the highest cost. The high cost of OPC-Kriging originates from the iterative character of the algorithm and the accompanying repetitive calibration of Kriging models. OPC-Kriging computes LAR and calibrates P Kriging models with increasing complex trend, whereas ordinary Kriging consists of a single calibration of a Kriging model with constant trend. For a single calibration of a surrogate of the Ishigami function (experimental design of size $N = 128$ samples) the ratio of computational times when comparing PCE to ordinary Kriging, SPC-Kriging and OPC-Kriging is approximately $1 : 5$, $1 : 20$, and $1 : 200$, respectively.

Note that it is intended to apply these techniques to realistic problems where the evaluation of the exact computational model response lasts much longer than the computation of a meta-model. The apparent computational overload of OPC-Kriging will not be anymore an issue in many practical applications.

5.3.3 Large Experimental Designs

When the resources for experiments are limited, the focus lies on doing as few computational model runs as possible as discussed in the previous section. In order to describe the entire behavior of the meta-modeling approaches, the Ishigami function is also studied now for larger experimental designs in order to assess the convergence of the various schemes. Results are shown in Fig. 11. This figure illustrates the evolution of the relative generalization error from small to large sample sizes on the *logarithmic* (base 10) scale. The error measure decreases fast when enlarging the sample set because the Ishigami function is composed of sine and cosine functions which can be approximated well with series of polynomials, e.g., as in a Taylor expansion. Thus PCE is the dominating effect for PC-Kriging.

For large sample sizes, ordinary Kriging is outperformed by the other three approaches. Kriging in general works well with small sample sizes. If too many samples are used, the interpolation algorithm becomes unstable due to singularities and bad conditioning in the auto-correlation matrix. If a large sample size is available, regional Kriging models on a subset of samples, e.g., the neighboring samples, are more suitable [84].

The performance of PC-Kriging and also PCE for a large number of samples (here 128 and 256 samples in Fig. 11) is on the order of magnitude of the machine computation precision. Errors around $\epsilon_{gen} \approx 10^{-12}$ originate
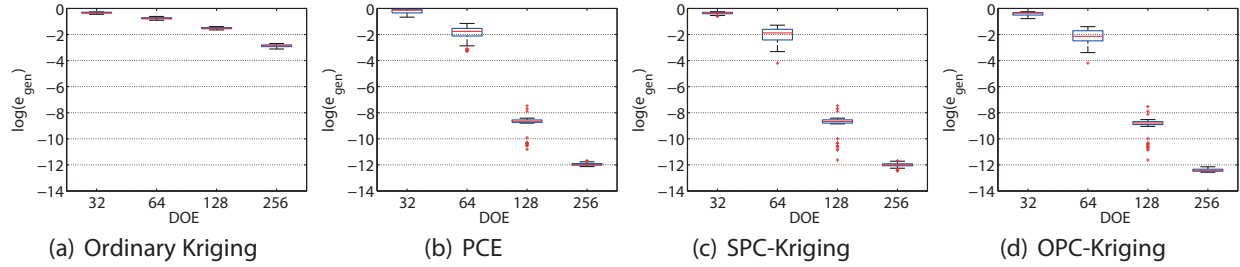
(a) Ordinary Kriging   (b) PCE   (c) SPC-Kriging   (d) OPC-Kriging

FIG. 11: The Ishigami function—Relative generalization error [Eq. (49)] associated with large experimental designs for the various meta-modeling approaches.

from numerical round-off errors which are not reducible by adding more samples. It is questionable though, whether in reality such a high accuracy in the meta-model prediction is needed.

5.3.4 Evolution of the error measures

The OPC-Kriging algorithm includes the tracking of the LOO error to optimally choose the sparse set of orthonormal polynomials. The evolution of the LOO error for the Ishigami function and a sample size of $N = 128$ samples is presented in Fig. 12. The experimental-design-based LOO error (dashed, red line) is compared to the relative generalization error which is shown as the solid black line.

The first point to notice is that the LOO error slightly underpredicts the true value of the relative generalization error for all sizes of polynomial sets. This is because the LOO error is based solely on the information contained in the experimental design samples whereas the relative generalization error is based on a large validation set ($n = 10^5$). Although there is an inherent underprediction, the overall behavior of the two error measures is similar. Thus the choice of the optimal set of polynomials for the OPC-Kriging can be based on the LOO error, which is obtained as a by-product of the procedure used to fit the parameters of the PC-Kriging model. In the example case of Fig. 12, choosing only half of the polynomials, i.e., $P = 27$ leads to a meta-model which is almost as accurate as using all 56 polynomials. The optimal set of polynomials can be chosen at the point where the decrease in LOO error becomes insignificant. This reduces the number of polynomials needed and thus also reduces the complexity of the OPC-Kriging meta-model.
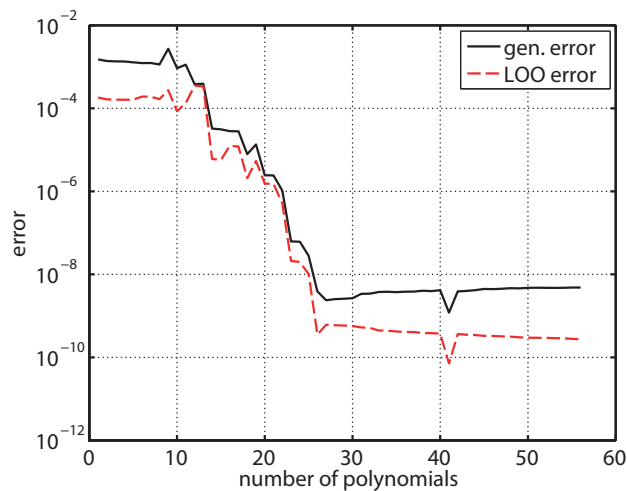

FIG. 12: Evolution of the leave-one-out error (LOO) and relative generalization error inside the OPC-Kriging algorithm as a function of the number of polynomials in the regression part.

6. CONCLUSION

In the context of increasing computational power, computational models in engineering sciences have become more and more complex. Many analyses such as reliability assessment or design optimization require repeated runs of such computational models which may be infeasible due to resource limitations. To overcome these limitations computational models are approximated by easy-to-evaluate functions called meta-models.

This paper summarized the principles of two popular nonintrusive meta-modeling techniques, namely polynomial chaos expansions (PCE) and Kriging (also called Gaussian process modeling). Then the combination of the two approaches to a new meta-modeling approach called polynomial-chaos-Kriging (PC-Kriging) is proposed. Two formulations of PC-Kriging are introduced in this paper, namely optimal-PC-Kriging (OPC-Kriging) and sequential-PC-Kriging (SPC-Kriging). SPC-Kriging employs first a least-angle-regression (LAR) algorithm to determine the optimal sparse set of orthonormal polynomials in the input space. Then, the set of polynomials is used as the trend of a universal Kriging meta-model. OPC-Kriging employs the same least-angle-regression algorithm as SPC-Kriging, yet iteratively adds polynomials to the trend part of the universal Kriging model one-by-one, and fit the hyper-parameters of the auto-correlation function in each iteration. In this case polynomials are added to the trend in the order they are selected by the LAR algorithm. Based on the LOO error the best meta-model is found to be the OPC-Kriging meta-model.

The performance of the four approaches (ordinary Kriging, PCE, SPC-Kriging, OPC-Kriging) is compared in terms of the relative generalization error on benchmark analytical functions. The results show that PC-Kriging is better than, or at least as good as the distinct approaches for small experimental designs. Specifically, OPC-Kriging is preferable to SPC-Kriging as it reduces the number of polynomials in the regression part and thus reduces the complexity of the meta-model, at a computational calibration cost which is however higher than that of SPC-Kriging.

The analysis of the performance of PC-Kriging is limited to some benchmark analytical functions in this paper. The ongoing research applies PC-Kriging to realistic engineering problems such as reliability analysis or design optimization. The idea of adaptive experimental designs (also called design enrichment) is introduced in order to increase the accuracy of the surrogate in some specific regions of the input space (e.g., close to the zero-level of the limit state function in reliability analysis) instead of everywhere. The initialization of the iterative algorithm is a small initial experimental design to which points are added in regions of interest. These added points will then increase the quality of the meta-model specifically in those regions. Preliminary ideas developed in [85, 92, 93] are currently investigated in detail.

ACKNOWLEDGMENTS

REFERENCES

1. De Rocquigny, E., Devictor, N., and Tarantola, S. (Eds.), *Uncertainty in Industrial Practice—A Guide to Quantitative Uncertainty Management*, John Wiley & Sons, New York, 2008.

2. Sudret, B., Uncertainty propagation and sensitivity analysis in mechanical models—contributions to structural reliability and stochastic spectral methods, Tech. Rep., Habilitation à diriger des recherches, Université Blaise Pascal, Clermont-Ferrand, France, 2007.

3. Storlie, C. B., Swiler, L. P., Helton, J. C., and Sallaberry, C. J., Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models, *Reliab. Eng. Sys. Safety*, 94:1735–1763, 2009.

4. Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, New York, 2001.

5. Forrester, A., Sóbester, A., and Keane, A., *Engineering Design via Surrogate Modelling*, Wiley, New York, 2008.

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

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

8. Rasmussen, C. E. and Williams, C. K. I., *Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning,* MIT Press, Cambridge, MA, 2006.

9. Santner, T. J., Williams, B. J., and Notz, W. I., *The Design and Analysis of Computer Experiments*, Springer Series in Statistics, Springer, New York, 2003.

10. Stein, M. L., *Interpolation of Spatial Data: Some Theory for Kriging*, Springer-Verlag, New York, 1999.

11. Gunn, S. R., Support vector machines for classification and regression, Tech. Rep. ISIS-1-98, Dpt. of Electronics and Computer Science, University of Southampton, 1998.

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

13. Vazquez, E. and Walter, E., Multi-output support vector regression, In *Proc. 13th IFAC Symp. Syst. Identification*, Rotterdam, pp. 1820–1825, 2003.

14. Clarke, S. M., Griebsch, J. H., and Simpson, T. W., Analysis of support vector regression for approximation of complex engineering analyses, In *Proc. DETC'03, Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, Chicago, Illinois, 2003.

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

16. Sudret, B., Berveiller, M., and Lemaire, M., Eléments finis stochastiques en élasticité linéaire, *Comptes Rendus Mécanique*, 332(7):531–537, 2004.

17. Wan, X. and Karniadakis, G. E., An adaptive multi-element generalized polynomial chaos method for stochastic differential equations, *J. Comput. Phys.*, 209:617–642, 2005.

18. Wan, X. and Karniadakis, G., Beyond Wiener-Askey expansions: Handling arbitrary PDFs, *J. Sci. Comput.*, 27:455–464, 2006.

19. Berveiller, M., Sudret, B., and Lemaire, M., A stochastic finite element procedure for moment and reliability analysis, *Eur. J. Comput. Mech.*, 15(7-8):825–866, 2006.

20. Ghiocel, D. and Ghanem, R., Stochastic finite element analysis of seismic soil-structure interaction, *J. Eng. Mech.*, 128:66–77, 2002.

21. Le Maître, O. P., Reagan, M., Najm, H. N., Ghanem, R. G., and Knio, O. M., A stochastic projection method for fluid flow—II. Random process, *J. Comput. Phys.*, 181:9–44, 2002.

22. Keese, A. and Matthies, H. G., Hierarchical parallelisation for the solution of stochastic finite element equations, *Comp. Struct.*, 83:1033–1047, 2005.

23. Xiu, D. and Hesthaven, J. S., High-order collocation methods for differential equations with random inputs, *SIAM J. Sci. Comput.*, 27(3):1118–1139, 2005.

24. Xiu, D., Fast numerical methods for stochastic computations: A review, *Comm. Comput. Phys.*, 5(2-4):242–272, 2009.

25. Choi, S. K., Grandhi, R. V., Canfield, R. A., and Pettit, C. L., Polynomial chaos expansion with Latin Hypercube sampling for estimating response variability, *AIAA J.*, 45:1191–1198, 2004.

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

27. Sudret, B., Global sensitivity analysis using polynomial chaos expansions, *Reliab. Eng. Syst. Safety*, 93(7):964–979, 2008.

28. Blatman, G. and Sudret, B., Sparse polynomial chaos expansions and adaptive stochastic finite elements using regression approach, *Comptes Rendus Mécanique*, 336:518–523, 2008.

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

30. Blatman, G. and Sudret, B., Efficient computation of global sensitivity indices using sparse polynomial chaos expansions, *Reliab. Eng. Sys. Safety*, 95:1216–1229, 2010.

31. Doostan, A. and Owhadi, H., A non-adapted sparse approximation of pdes with stochastic inputs, *J. Comput. Phys.*, 230(8):3015–3034, 2011.

32. Blatman, G. and Sudret, B., Adaptive sparse polynomial chaos expansion based on least angle regression, *J. Comput. Phys*, 230:2345–2367, 2011.

33. Doostan, A., Validi, A., and Iaccarino, G., Non-intrusive low-rank separated approximation of high-dimensional stochastic

models, *Comput. Methods Appl. Mech. Eng.*, 263:42–55, 2013.

34. Jakeman, J. D., Eldred, M. S., and Sargsyan, K., Enhancing ℓ_1-minimization estimates of polynomial chaos expansions using basis selection, arXiv:1407.8093, 2014.

35. Eldred, M. S., Recent advances in non-intrusive polynomial chaos and stochastic collocation methods for uncertainty analysis and design, In *50th AIAA/ASME/ASCE/AHS/ASC Struct. Struct. Dyn. Mater. Conf.*, Palm Springs, California, 2009.

36. Eldred, M. S., Webster, C. G., and Constantine, P. G., Design under uncertainty employing stochastic expansion methods, In *Proc. 12th AIAA/ISSMO Multidiscip. Anal. Optim. Conf.* Victoria, British Columbia, Canada, 2008.

37. Sarangi, S., Bodla, K. K., Garimella, S. V., and Murthy, J. Y., Manifold microchannel heat sink design using optimization under uncertainty, *Int. J. Heat Mass Tran.*, 69:92–105, 2014.

38. Krige, D. G., A statistical approach to some basic mine valuation problems on the Witwatersrand, *J. Chem. Metall. Mining Soc. South Africa*, 52(6):119–139, 1951.

39. Kaymaz, I., Application of Kriging method to structural reliability problems, *Struct. Saf.*, 27(2):133–151, 2005.

40. Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E., Sequential design of computer experiments for the estimation of a probability of failure, *Stat. Comput.*, 22:773–793, 2012.

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

42. Bichon, B. J., Eldred, M. S., Swiler, L., Mahadevan, S., and McFarland, J., Efficient global reliability analysis for nonlinear implicit performance functions, *AIAA J.*, 46(10):2459–2468, 2008.

43. Dubourg, V., Sudret, B., and Deheeger, F., Metamodel-based importance sampling for structural reliability analysis, *Prob. Eng. Mech.*, 33:47–57, 2013.

44. Dubourg, V. and Sudret, B., Meta-model-based importance sampling for reliability sensitivity analysis, *Struct. Saf.*, 49:27–36, 2014.

45. Jones, D. R., Schonlau, M., and Welch, W. J., Efficient global optimization of expensive black-box functions, *J. Glob. Optim.*, 13(4):455–492, 1998.

46. Dubourg, V., Sudret, B., and Bourinet, J. M., Reliability-based design optimization using Kriging and subset simulation, *Struct. Multidisc. Optim.*, 44(5):673–690, 2011.

47. Dubourg, V., Adaptive surrogate models for reliability analysis and reliability-based design optimization, PhD thesis, Université Blaise Pascal, Clermont-Ferrand, France, 2011.

48. Lophaven, S., Nielsen, H. B., and Søndergaard, J., DACE, A Matlab Kriging Toolbox, Tech. Rep., IMM-TR-2002-12, Technical University of Denmark, Lyngby, Denmark, 2002.

49. Roustant, O., Ginsbourger, D., and Deville, Y., DiceKriging, DiceOptim: Two R package for the analysis of computer experiments by Kriging-based metamodeling and optimisation, *J. Stat. Software*, 51:1–55, 2012.

50. Roustant, O., Ginsbourger, D., and Deville, Y., *Package DiceKriging—Kriging Methods for Computer Experiments*, St. Étienne, France, 2013.

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

52. Sudret, B., Polynomials chaos expansions and stochastic finite element methods, In: Phoon, K.-K. and Ching, J. (Ed.), *Risk Reliab. Geotech. Eng.*, chapter 5. Taylor and Francis, London, 2014.

53. Blatman, G., Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis, PhD thesis, Université Blaise Pascal, Clermont-Ferrand, 2009.

54. Chkifa, A., Cohen, A., Migliorati, G., Nobile, F., and Tempone, R., Discrete least squares polynomial approximation with random evaluations—Appliations to parametric and stochastic elliptic PDEs, Tech. Rep., MATHICSE, EPFL, Lausanne, Switzerland, 2013.

55. Migliorati, G., Nobile, F., Schwerin, E., and Tempone, R., Analysis of discrete L2 projection on polynomial spaces with random evaluations, *Found. Comput. Math.*, 14(3):419–456, 2014.

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

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

58. Peng, J., Hampton, J., and Doostan, A., A weighted ℓ_1-minimization approach for sparse polynomial chaos expansions, *J.*

Comput. Phys., 267:92–111, 2014.

59. Sargsyan, K., Safta, C., Najm, H., Debusschere, B., Ricciuto, D., and Thornton, P., Dimensionality reduction for complex models via Bayesian compressive sensing, *Int. J. Unc. Quant.*, 4:63–94, 2014.

60. Vapnik, V., *The Nature of Statistical Learning Theory*, Springer, Berlin, 1995.

61. Stone, M., Cross-validatory choice and assessment of statistical predictions, *J. R. Stat. Soc., Ser. B*, 36:111–147, 1974.

62. Geisser, S., The predictive sample reuse method with applications, *J. Am. Stat. Assoc.*, 70:320–328, 1975.

63. Allen, D., The prediction sum of squares as a criterion for selecting predictor variables, Tech. Rep. 23, Dept. of Statistics, University of Kentucky, 1971.

64. Saporta, G., *Probabilités, Analyse des Données et Statistique*, Technip, 2006.

65. Matérn, B., *Spatial Variation, Lect. Notes Stat. 36*, Brillinger, D., Frienberg, S., Gani, J., Hartigan, J., and Krickeberg, K. (Eds.), Springer-Verlag, Berlin, Heidelberg, 1986.

66. Marrel, A., Iooss, B., Van Dorpe, F., and Volkova, E., An efficient methodology for modeling complex computer codes with Gaussian processes, *Comput. Stat. Data Anal.*, 52:4731–4744, 2008.

67. Bachoc, F., Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecifications, *Comput. Stat. Data Anal.*, 66:55–69, 2013.

68. Goldfarb, D., A family of variable-metric methods derived by variational means, *Math. Comput.*, 24:23–26, 1970.

69. Fletcher, R., A new approach to variable metric algorithms, *Comput. J.*, 13:317–322, 1970.

70. Shanno, D. F., Conditioning of quasi-Newton methods for function minimization, *Math. Comput.*, 24:647–656, 1970.

71. Byrd, R. H., Hribar, M. E., and Nocedal, J., An interior point algorithm for large-scale nonlinear programming, *SIAM J. Optim.*, 9(4):877–900, 1999.

72. Goldberg, G. H., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Professional, MA, 1989.

73. Storn, R. and Price, K., Differential evolution, a simple and efficient heuristic for global optimization over continuous spaces, *J. Global Optim.*, 11(4):341–359, 1997.

74. Deng, W., Yang, X., Zou, L., Wang, M., Liu, Y., and Li, Y., An improved self-adaptive differential evolution algorithm and its application, *Chemometr. Intell. Lab.*, 128:66–76, 2013.

75. Bachoc, F., Bois, G., Garnier, J., and Martinez, J.-M., Calibration and improved prediction of computer models by universal Kriging, *Nucl. Sci. Eng.*, 176:91–97, 2014.

76. Bachoc, F., Asymptotic analysis of the role of spatial sampling for hyper-parameter estimation of Gaussian processes, *J. Multivar. Anal.*, 125:1–35, 2014.

77. Duvenaud, D. K., Nickisch, H., and Rasmussen, C. E., Additive Gaussian processes, In: Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K. (Eds.), *Advances in Neural Information Processing Systems 24*, pp. 226–234, Curran Associates, Inc., Granada, Spain, 2011.

78. Ginsbourger, D., Roustant, O., and Durrande, N., Invariances of random fields paths, with applications in Gaussian process regression, arXiv:1308.1359, 2013.

79. Durrande, N., Ginsbourger, D., and Roustant, O., Additive covariance kernels for high-dimensional Gaussian process modeling, *Ann. Fac. Sci. Toulouse Tome*, 3:481–499, 2012.

80. Durrande, N., Ginsbourger, D., and Roustant, O., ANOVA kernels and RKHS of zero mean function for model-based sensitivity analysis, *J. Multivar. Anal.*, 115:57–67, 2013.

81. Ginsbourger, D., Durrande, N., and Roustant, O., Kernels and designs for modelling invariant functions: From group invariance to additivity, Tech. Rep. hal-00731657, INRIA, 2013.

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

83. Bichon, B., McFarland, J., and Mahadevan, S., Efficient surrogate models for reliability analysis of systems with multiple failure modes, *Reliab. Eng. Sys. Safety*, 96(10):1386–1395, 2011.

84. Dubrule, O., Cross validation of Kriging in a unique neighbourhood, *Math. Geol.*, 15(6):687–698, 1983.

85. Schöbi, R. and Sudret, B., PC-Kriging: A new metamodelling method combining polynomial chaos expansions and Kriging,

In *Proc. 2nd Int. Symposium on Uncertainty Quantification and Stochastic Modeling*, Rouen, France, 2014.

86. Sobol', I. M., Sensitivity estimates for nonlinear mathematical models, *Math. Model. Comp. Exp.*, 1:407–414, 1993.

87. Rosenbrock, H., An automatic method for finding the greatest or least value of a function, *Comput. J.*, 3:175–184, 1960.

88. Morris, M. D., Factorial sampling plans for preliminary computational experiments, *Technometrics*, 33(2):161–174, 1991.

89. Rastrigin, L., *Systems of Extremal Control*, Zinatne, Riga, 1974.

90. Oakley, J. and O'Hagan, A., Probabilistic sensitivity analysis of complex models: a Bayesian approach, *J. R. Stat. Soc., Ser. B*, 66:751–769, 2004.

91. McKay, M. D., Beckman, R. J., and Conover, W. J., A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, 2:239–245, 1979.

92. Schöbi, R. and Sudret, B., PC-Kriging: A new meta-modelling method and its applications to quantile estimation, In: Li, J. and Zhao, Y. (Eds.), *Proc. 17th IFIP WG7.5 Conference on Reliability and Optimization of Structural Systems, Huangshan, China*, Taylor & Francis, 2014.

93. Schöbi, R. and Sudret, B., Combining polynomial chaos expansions and Kriging for solving structural reliability problems, In: Spanos, P. and Deodatis, G. (Eds.), *Proc. 7th Int. Conf. on Comp. Stoch. Mech (CSM7)*, Santorini, Greece, 2014.