Multielement polynomial chaos Kriging-based metamodelling for Bayesian inference of non-smooth systems

J.C. García-Merino [a,*], C. Calvo-Jurado [a], E. Martínez-Pañeda [b], E. García-Macías [b,c]

[a] *Department of Mathematics, School of Technology, Cáceres, 10003 Spain*
[b] *Department of Civil and Environmental Engineering, Imperial College, London SW7 2AZ, UK*
[c] *Department of Structural Mechanics and Hydraulic Engineering, Campus Universitario de Fuentenueva (Edificio Politécnico) 18071 Granada Spain*

ARTICLE INFO

ABSTRACT

This paper presents a surrogate modelling technique based on domain partitioning for Bayesian parameter inference of highly nonlinear engineering models. In order to alleviate the computational burden typically involved in Bayesian inference applications, a multielement Polynomial Chaos Expansion based Kriging metamodel is proposed. The developed surrogate model combines in a piecewise function an array of local Polynomial Chaos based Kriging metamodels constructed on a finite set of non-overlapping subdomains of the stochastic input space. Therewith, the presence of non-smoothness in the response of the forward model (e.g. nonlinearities and sparseness) can be reproduced by the proposed metamodel with minimum computational costs owing to its local adaptation capabilities. The model parameter inference is conducted through a Markov chain Monte Carlo approach comprising adaptive exploration and delayed rejection. The efficiency and accuracy of the proposed approach are validated through two case studies, including an analytical benchmark and a numerical case study. The latter relates the partial differential equation governing the hydrogen diffusion phenomenon of metallic materials in Thermal Desorption Spectroscopy tests.

1. Introduction

The widespread use of high performance computing (HPC) technologies has enabled the increasingly frequent adoption of computationally intensive numerical models in a myriad of disciplines in academia and industry. Such high fidelity models allow conducting virtual testing of engineering systems, avoiding the technical limitations and minimizing the costs associated with traditional experimental testing. Nevertheless, formidable challenges still arise when implementing these models into computationally demanding studies such as optimization [1], sensitivity analysis [2], model identification and calibration [3], reliability analysis [4], or robust design [5]. As a solution, a variety of surrogate models or metamodels have been proposed in the literature in recent years. Nevertheless, despite the considerable advances in the field, there remain open research challenges for their extensive use such as handling highly nonlinear model responses, limited training datasets and,

---

* Corresponding author.
E-mail address: jcgarcia@unex.es (J.C. García-Merino).

in general, the online implementation of surrogate models to enable decision-making in engineering systems [6,7]. In particular, the development of real time surrogate model-based parameter estimation approaches draws high interest in frontier research fields such as digital twins development [8] and smart maintenance of engineering systems [9].

In their broadest sense, surrogate models are computationally light black box representations of resource-intensive models. Surrogate modelling methods can be generally classified into three categories [10]: (i) projection-based methods or reduced-order models (ROMs) [11]; (ii) multi-fidelity methods [12]; and (iii) data-driven or responsive surface methods (RSMs). Projection-based techniques project the governing equations of the original model onto a low-dimensional subspace. Hence, although ROMs have the advantage of retaining the physics underlying the model, these are limited to situations where access to the governing equations is granted, which is not the case in many practical applications when using commercial software. Multi-fidelity methods are built by simplifying the underlying physics or reducing the numerical resolution. These methods may also render some difficulties, being highly case-dependent and requiring specialized expertise to find a suitable trade-off between prediction accuracy and computational burden. Such difficulties have fostered rapid developments of RSMs in recent years as a non-intrusive technique with great flexibility in a wide range of applications. The key advantage of these methods relies in the fact that the forward model does not need to be modified and, therefore, it can be essentially treated as a black box [13]. Among the broad variety of RSMs available in the literature, some of the most popular ones are Kriging [14], radial basis functions (RBF) [15], support vector regression (SVR) [16], artificial neural networks (ANN) [17], Gaussian process (GP) regression [18], polynomial chaos expansions (PCE) [2] and Polynomial Chaos Expansion based Kriging (PCK) [19]. These models are trained by exploiting a set of realizations of parameters of interest of the model, called the experimental design (ED), and the corresponding model evaluations, also called quantities of interest (QoI). Therefore, the accuracy of data-driven surrogate models is highly determined by the dimensions of the design space and the number and distribution of the training samples in the ED [20]. In engineering practice, obtaining the ED constitutes the most time-consuming part since it requires the evaluation of the computationally intensive forward model at each sample point. Choosing high quality EDs is thus critical to achieve high accuracy in the surrogate model construction with the least possible number of training samples. Sampling approaches can generally be divided into static (one-shot) and sequential methods. One-shot sampling generates the training sample points in one single step, and common approaches include fractional designs and orthogonal arrays [21]. While these techniques offer easy implementation and minimal computational cost, the determination of the optimal sample size may be troublesome when the behaviour of the forward model is unknown. To minimize such difficulties, a number of sequential sampling strategies have been introduced, including adaptive and space-filling sequential methods [22]. On one hand, space-filling sequential designs such as Latin Hypercube Sampling (LHS), Sobol, Hammersley and Halton sampling [23] generate samples iteratively to attain good coverage of the parametric domain. On the other hand, adaptive sampling techniques iteratively draw new samples in regions of the parametric space with large prediction errors, enabling to account for local refinements in the ED (refer to references [22,24] for a thorough state-of-the-art review).

A second major challenge of non-intrusive surrogate models regards the difficulties involved in the fitting of non-smooth models exhibiting unsteadiness, sparseness or large perturbations. In these cases, adaptive solutions accommodating local relevant refinements of the response surface are required. A large volume of research has been conducted in the last decade to address this issue, giving origin to a number of advanced surrogate modelling techniques such as multi-resolution generalized PCE [25], domain partitioning [26], sparse grid collocation [27], Voronoi tesselations [28], local search at trust regions [29], clustering-based partitioning [30], ensembles of surrogate models [31], multi-element generalized polynomial chaos (ME-gPC) method [27] and multi-element probabilistic collocation methods (ME-PCM) [32]. In this light, different partitioning techniques can be found in the literature in the context of the pure GPs [33,34], PCE [35], and PCE-based mapping of likelihood functions in parameter inference applications [25,36]. In those works, different domain partition criteria were proposed based on dissimilarity measurements between regions [34], data density [33], or maximum residual differences [35,36]. In general, these approaches usually generate series of subdomains where the non-smooth response surface can be assumed locally smooth, thus enabling the definition of local surrogate models contributing to the global response in a piecewise fashion.

The development of cost-efficient surrogate models opens vast new opportunities for real-time parameter estimation applications. Probabilistic Bayesian approaches are particularly attractive owing to their ability to assess the effects of uncertainties on the model parameters and the derived response predictions, as well as their robustness to noise pollution and efficiency to deal with ill-conditioning and ill-posedness. Such excellent features have fostered their implementation in multiple fields such as structural identification [37], geotechnical problems [38], material characterization [39], and bioengineering [40], just to mention a few. In general, Bayesian parameter estimation approaches exploit experimental data to infer the posterior probability distribution functions (PDFs) of certain unknown model parameters through the Bayes' theorem. Nonetheless, the direct evaluation of the posterior PDFs requires solving the possibly high-dimensional integral related to the evidence of the model. Therefore, except for some trivial cases, posterior PDFs often need to be approximated numerically. Markov chain methods constitute the most widespread set of techniques to extract series of samples to estimate the posterior PDFs, allowing to sample from a large class of high-dimensional distributions. Popular procedures for Markov chain Monte Carlo (MCMC) sampling are the Metropolis-Hastings [41] and Gibbs algorithms [42]. The basic idea of these techniques is to construct a Markov chain with a stationary distribution resembling the posterior distribution, in such a way that a sample of the joint PDF of the model parameters can be obtained by collecting the states of the chain. These methods guarantee asymptotic convergence to the exact PDF, although a considerably large number of iterations are

typically required to achieve convergence, which compromises the computational efficiency of the inference [43]. To alleviate the computational burden in classical MCMC techniques, a variety of more efficient sampling algorithms have been proposed in recent years, including Sequential Monte Carlo (SMC) [44], Transitional MCMC [45], and Bayesian broad learning [46]. Notwithstanding these advances, their elevated computational cost remains a critical limitation when high-fidelity models are considered in the inference problem. Herein is where surrogate models offer an efficient solution to conduct cost-efficient Bayesian inference while retaining the accuracy of high-fidelity models. The enormous potentials of this approach are evidenced by the increasing number of research studies reported in recent years. It is worth noting the work by Schneider et al. [47] who proposed a Bayesian procedure using rational PCE metamodels of the response of dynamic systems in the frequency domain, and demonstrated its effectiveness for the identification of a cross-laminated timber plate. Xing et al. [48] developed an additive GP model for multi-fidelity surrogate modelling inserted into a non-parametric Bayesian approach with a closed-form solution for the predictive posterior PDF. The accuracy and flexibility of their approach were validated with several benchmark problems, including the identification of a solid oxide fuel cell model and an elbow-shape pipe under turbulent mixing flow conditions. Ierimonti and co-authors [49] proposed a Kriging-based conjugate Bayesian identification methodology for online damage identification of an instrumented monumental building, the Consoli Palace in Gubbio (Italy). Del Val et al. [50] implemented an MCMC approach to infer the catalytic recombination parameters of reusable thermal protection materials exploiting plasma wind tunnel measurements. Interestingly, instead of bypassing a high-fidelity model, those authors approximated the likelihood function in the Bayesian inference using a GP surrogate model to accelerate the parameters estimation.

In light of the literature review above, it is apparent that the fields of surrogate modelling and its applications for fast parameter estimation have experienced considerable advances in recent years. Nonetheless, light metamodels capable of bypassing highly nonlinear models for online parameter estimation are yet to be fully developed. In this context, the present work presents a novel multielement surrogate model extending the PCK model proposed in Schobi et al. [19] for online Bayesian parameter inference of non-smooth models exhibiting highly nonlinear spatial variability. To do so, a simple but efficient and flexible regular block partitioning approach is adopted. On this basis, the space domain is partitioned into distinct subregions with splitting directions defined after a preliminary sensitivity analysis. The algorithm prioritizes the partition of those variables providing highest sensitivities in terms of the Sobol indices, which can be readily extracted from the coefficients of a global PCE [51]. Regarding the construction of the local PCK surrogate models, the optimal order of the polynomials in the PCE is automatically identified by a model selection technique for sparse linear models, the least-angle regression (LAR) algorithm set out by Efron et al. [52]. Then, the optimal PCE is inserted into a Kriging predictor as the trend term, while the stochastic term is fitted through a genetic algorithm (GA) global optimization approach. The global model response is obtained by combining the local surrogate models in a piecewise fashion. In order to minimize the computational burden in the construction of the surrogate models, the ED is obtained with the adaptive Monte Carlo-Intersite-proj-th (MIPT) approach [22]. Finally, the surrogate model is used for Bayesian parameter estimation using a cost-efficient adaptive MCMC with delayed rejection (DRAM) algorithm developed by Haario et al. [53]. The effectiveness of the proposed approach is validated through two benchmark case studies: (i) an analytical benchmark; (ii) and a partial differential equation (PDE) for thermal desorption spectroscopy (TDS) experiments of hydrogen in metals. The latter represents a formidable example of a model exhibiting a non-smooth behaviour in the shape of nonlinearities and unsteadiness. The presented results and discussion demonstrate the robustness of the proposed approach to the presence of measurement noise and multimodality in the posterior distributions. Overall, the developed surrogate model assisted Bayesian inference approach represents a comprehensive solution for fast parameter inference of computationally intensive mathematical models, offering wide applicability across disciplines such as the identification of structural systems, geotechnical problems, Bioengineering, or Astrophysics, among others.

The remainder of this paper is organized as follows. Section 2 outlines the theoretical formulation of the proposed approach. Section 3 presents the numerical results and discussion. In particular, two case studies are investigated, namely a benchmark analytical model and a numerical model of the TDS analysis of hydrogen desorption in metals. Finally, Section 4 discusses the contributions of this work and presents the main concluding remarks.

2. Theoretical formulation

The main purpose of this section is to present the theoretical fundamentals of the proposed surrogate model-based Bayesian inference approach. The general methodology is sketched in Fig. 1 and comprises four main steps, namely (i) domain partitioning; (ii) construction of the local surrogate models; (iii) surrogate models assemblage; and (iv) validation.

2.1. Surrogate modelling: polynomial chaos expansion based Kriging

Let (Ω, Σ, μ) be a probability space, with $\Omega \subset \mathbb{R}^M$ denoting the event space equipped with a σ-algebra Σ of subsets of Ω and a probability measure μ such that $\mu(\Omega) = 1$. Let $\mathcal{M} : \Omega \subset \mathbb{R}^M \to \mathbb{R}$ be a computational model mapping a vector of input variables $\mathbf{x} = [x_1, \ldots, x_M]^T$ into the output variable or quantity of interest $y \in \mathbb{R}$. The goal of surrogate modelling is to approximate the forward model $\mathcal{M}$, which is typically computationally intensive, by a computationally inexpensive function $\hat{\mathcal{M}}$. In this paper, PCK metamodels combining PCE and Kriging are adopted. In this light, PCE is used to approximate the global behaviour of the computational model $\mathcal{M}$ while the Kriging metamodel captures its local behaviour.
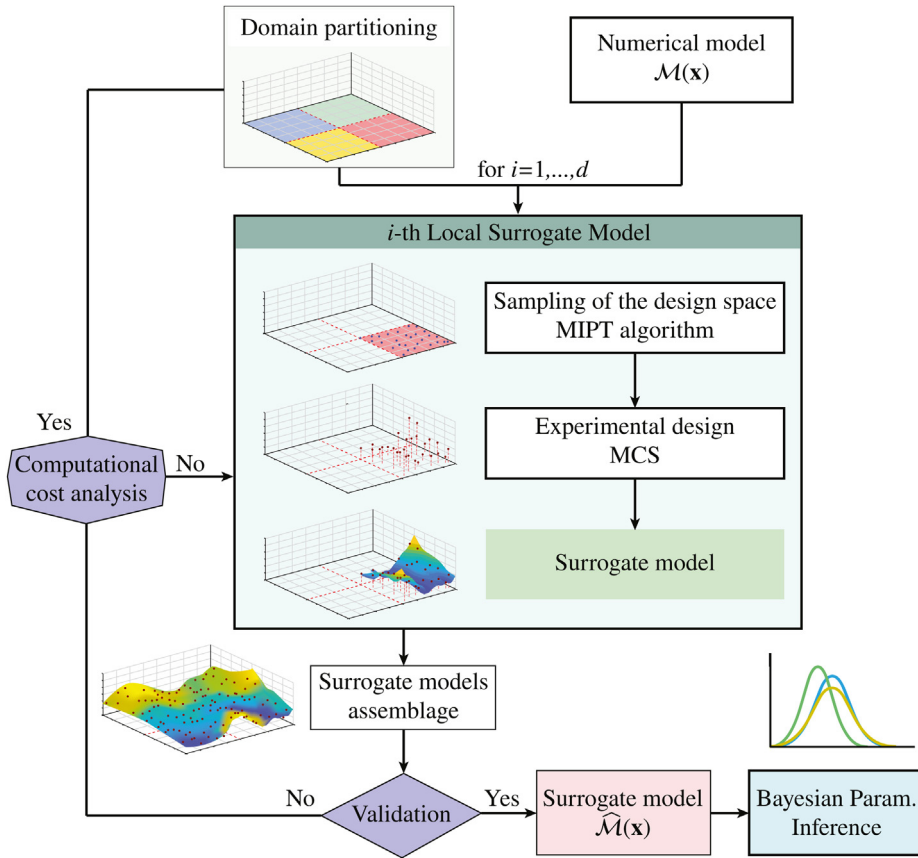
Fig. 1. Flowchart of the proposed PCK metamodel to perform Bayesian Inference.

2.1.1. Polynomial chaos expansion

Assume that the input vector $\mathbf{x} \in \Omega$ is constituted by M independent random variables components $\{x_i\}_{i=1}^M$ with probability density functions μ_{X_i}. Then, PCE represents the output response $y \in \mathbb{R}$ as the infinite expansion of $\mathcal{M}$ over an orthonormal basis of multivariate polynomials $\Psi_{\boldsymbol{\alpha}}$ as:

$$y = \mathcal{M}(\mathbf{x}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\mathbf{x}), \tag{1}$$

where $a_{\boldsymbol{\alpha}}$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)$, $\alpha_i \in \mathbb{N}$ are the coefficients of the expansion. Orthonormal basis families associated with a variety of standard distributions can be found in reference [54]. Multivariate polynomials $\Psi_{\boldsymbol{\alpha}}$ can be expressed in terms of a family of univariate polynomials $\left\{ \psi_j^{(i)}, \ j \in \mathbb{N} \right\}$ as $\Psi_{\boldsymbol{\alpha}}(\mathbf{x}) = \prod_{i=1}^M \psi_{\alpha_i}^{(i)}(x_i)$. Polynomials $\psi_j^{(i)}$ are also orthonormal with respect to the marginal distribution, that is:

$$\mathbb{E}\left[\psi_j^{(i)}(x_i), \ \psi_k^{(i)}(x_i)\right] = \int \psi_j^{(i)}(u) \, \psi_k^{(i)}(u) \, \mu_{x_i}(u) \, \mathrm{d}u = \delta_{j,k}, \tag{2}$$

with $\delta_{j,k}$ being the Dirac Delta function. Random variable $\mathbf{x}$ induces the probability measure $\mu_{\mathbf{x}}$ on $\left(\mathbb{R}^M, \mathcal{B}\left(\mathbb{R}^M\right)\right)$, with $\mathcal{B}(\mathbb{R})$ denoting the Borel σ-algebra $\mu_{\mathbf{x}}(B) = \mu\{\omega \in \Omega : \mathbf{x}(\omega) \in B\}$, $B \in \mathcal{B}$. The PDF of $\mathbf{x}$ is given by the probability measure $\mu_{\mathbf{x}}$ as $\mu_{\mathbf{x}}(\mathbf{u}) = \mu(\mathbf{x} \leq \mathbf{u})$, $\mathbf{u} \in \mathbb{R}^M$. Note that, since the components of $\mathbf{x}$ are independent, the joint probability function $\mu_{\mathbf{x}}(\mathbf{u})$ is given by the product of the marginal PDFs of x_i, i.e. $\mu_{\mathbf{x}}(\mathbf{u}) = \prod_{i=1}^M \mu_{x_i}(u_i)$, $\mathbf{u} = (u_1, \ldots, u_M) \in \mathbb{R}^M$. Therefore, Eq. (2) can be written in a more compact form as:

$$\mathbb{E}\left[\Psi_{\boldsymbol{\alpha}}(\mathbf{x}), \ \Psi_{\boldsymbol{\beta}}(\mathbf{x})\right] = \int \Psi_{\boldsymbol{\alpha}}(\mathbf{u}) \Psi_{\boldsymbol{\alpha}}(\mathbf{u}) \, \mu_{\mathbf{x}}(\mathbf{u}) \ \mathrm{d}\mathbf{u} = \delta_{\boldsymbol{\alpha},\boldsymbol{\beta}}, \quad \boldsymbol{\alpha}, \ \boldsymbol{\beta} \in \mathbb{N}^M. \tag{3}$$

Although the expression in Eq. (1) is exact for an infinite number of terms, in practice only a finite number can be computed and a certain truncation scheme needs to be adopted. One of the simplest approach consists in selecting all the

polynomials whose total degree $|\boldsymbol{\alpha}| = \sum\limits_{i=1}^{M} \alpha_i$ belongs to the set:

$$\mathcal{A}^{M,p} = \left\{\boldsymbol{\alpha} \in \mathbb{N}^M : 0 \le |\boldsymbol{\alpha}| \le p\right\}, \quad \text{with} \quad \text{card } \mathcal{A}^{M,p} = \binom{M+p}{p} = \frac{(M+p)!}{M!\,p!}. \tag{4}$$

For high-dimensional and non-linear problems, this truncation procedure usually leads to large numbers of polynomial coefficients and considerable computational burdens. Nevertheless, it is often observed in many practical applications that coefficients corresponding to high interaction terms between the input variables are close to zero, a phenomenon that is also known as the *sparsity-of-effect principle* [51]. To alleviate this, an hyperbolic truncation scheme can be adopted. This approach selects all multi-indices with q-norm $\|\boldsymbol{\alpha}\|_q = \left(\sum\limits_{i=1}^{M} \alpha_i^q\right)^{\frac{1}{q}}$ less than or equal to a certain model order p,

i.e. $\mathcal{A}^{M,p,q} = \left\{\boldsymbol{\alpha} \in \mathbb{N}^M : \|\boldsymbol{\alpha}\|_q \le p\right\}$. Note that the expansion tends to only maintain univariate polynomials as the q-norm decreases, thus achieving a reduction in the computational cost of the metamodel. Nonetheless, the number of terms may remain elevated since the potential sparseness in the coefficients is not being actually assessed. Following the work by Blatman and Sudret [51], the LAR algorithm [52] is adopted to further reduce the number of polynomial coefficients. In the context of PCE, LAR constructs a set of expansions incorporating an increasing number of basis polynomials $\Psi_{\boldsymbol{\alpha}}$, from 1 to $\mathcal{P} = \text{card}\left(\mathcal{A}^{M,p,q}\right)$. The resulting sequence of index sets is used to construct a family of expansions with decreasing sparseness. Finally, a cross-validation procedure can be implemented to select the best metamodel among the obtained family of expansions. In particular, the Bayesian Information Criterion (BIC) is adopted in this work. Once the optimal index set is selected, the expansion coefficients $\boldsymbol{a} = \{a_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathcal{A}^{M,p} \subset \mathbb{N}^M\}$ are obtained by minimizing the expectation of the least squared error:

$$\boldsymbol{a} = \arg\min_{\boldsymbol{a} \in \mathbb{R}^P} \mathbb{E}\left[\left(\mathcal{M}(\mathbf{x}) - \sum_{\boldsymbol{\alpha} \in \mathcal{A}^{M,p,q}} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\mathbf{x})\right)^2\right]. \tag{5}$$

In practice, Eq. (5) is calibrated on a set of N realizations $\Xi = \left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\right\}$ of the input variable $\mathbf{x}$ forming the ED. In order to obtain a representative ED, the adaptive MIPT sampling method [22] is adopted in this work. Then, the expectation operator in Eq. (5) is replaced by its discretized version using the ED realizations as follows:

$$\boldsymbol{a} = \arg\min_{\boldsymbol{a} \in \mathbb{R}^P} \frac{1}{N} \sum_{i=1}^{N} \left(\mathcal{M}\left(\mathbf{x}^{(i)}\right) - \sum_{\boldsymbol{\alpha} \in \mathcal{A}^{M,p,q}} \boldsymbol{a}_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}\left(\mathbf{x}^{(i)}\right)\right)^2. \tag{6}$$

Denoting the realizations of the output variable y by $\mathbf{y} = \{y^{(1)} = \mathcal{M}(\mathbf{x}^{(1)}), \ldots, y^{(N)} = \mathcal{M}(\mathbf{x}^{(N)})\}^{\mathrm{T}}$, the solution of the optimization problem in Eq. (6) reads:

$$\hat{\boldsymbol{a}} = \left(\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta}\right)^{-1}\boldsymbol{\Theta}^{\mathrm{T}}\mathbf{y}, \qquad \boldsymbol{\Theta} = (\boldsymbol{\Theta}_{ij}) = \left[\psi_j(\boldsymbol{x}^{(i)})\right]_{i=1,\ldots,N}^{j=1,\ldots,P}, \tag{7}$$

where $\boldsymbol{\Theta}$ denotes the information matrix calculated from the evaluation of the basis polynomials on Ξ. For the least-square minimization problem in Eq. (6) to be well posed, the size of the ED is usually selected according to the heuristic rule $N \approx 2 \cdot P$ or $3 \cdot P$ [51]. Once Eq. (6) has been solved, the predictions of the PCE surrogate model can be obtained as:

$$\hat{y} = \hat{\mathcal{M}}_{PCE}(\mathbf{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \hat{\boldsymbol{a}}_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\mathbf{x}). \tag{8}$$

2.1.2. Polynomial chaos expansion based Kriging (PCK)

The Kriging method assumes that the response of a computational model $\mathcal{M}(\mathbf{x})$ is modelled by the sum of a stochastic random process $\mathcal{Z}(\mathbf{x})$ and a regression model $\mathcal{T}(\mathbf{x})$, also called trend, in the form Fuhg et al. [22]:

$$\hat{\mathcal{M}}(\mathbf{x}) = \mathcal{T}(\mathbf{x}) + \mathcal{Z}(\mathbf{x}). \tag{9}$$

The stochastic component in Eq. (9) is fully determined by the covariance function [55]:

$$\text{Cov}\left(\mathcal{Z}(\mathbf{x}), \mathcal{Z}(\mathbf{x}')\right) = \mathbb{E}\left[\mathcal{Z}(\mathbf{x})\mathcal{Z}(\mathbf{x}')\right] = \sigma^2 R\left(|\mathbf{x} - \mathbf{x}'|; \boldsymbol{\theta}\right), \tag{10}$$

with σ^2 being the process variance, and $R\left(|\mathbf{x} - \mathbf{x}'|; \boldsymbol{\theta}\right)$ an auto-correlation function [56] between two input sample points $\mathbf{x}$ and $\mathbf{x}'$ that depends on certain hyper-parameters $\boldsymbol{\theta}$ to be computed. In this work, the Gaussian correlation function is adopted as:

$$R(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta}) = \prod_{\ell=1}^{M} \exp\left[-\theta_\ell\left(x_\ell - x'_\ell\right)^2\right]. \tag{11}$$

The trend term of the Kriging model in Eq. (9) interpolates the forward model evaluations at the ED, while the local variability is captured by the stochastic process. Depending on the form of the trend, three different versions of Kriging are typically referred to in the literature [22], including simple, ordinary and universal Kriging, which respectively correspond to polynomials of degrees 0, 1 and *N*. In this work, with the aim of combining the excellent global approximation capabilities of the PCE previously introduced in Section 2.1.1, the sparse PC expansion obtained by LAR is introduced in the shape of the trend term in Eq. (9). The resulting PCK metamodel reads:

$$\hat{\mathcal{M}}_{PCK}(\mathbf{x}) = \hat{\mathcal{M}}_{PCE}(\mathbf{x}) + \mathcal{Z}(\mathbf{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \hat{\boldsymbol{a}}_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\mathbf{x}) + \mathcal{Z}(\mathbf{x}). \tag{12}$$

The construction of the PCK metamodel in Eq. (12) consists in two steps. Firstly, the optimal set of orthonormal polynomials $\boldsymbol{\Psi}_{\boldsymbol{\alpha}}$ (for $\boldsymbol{\alpha} \in \mathcal{A}$ the truncation set) is obtained by LAR as indicated in Section 2.1.1. Secondly, the calculation of hyperparameters $\hat{\boldsymbol{\theta}}$ and the polynomial coefficients and the process variance $\{\boldsymbol{a}(\hat{\boldsymbol{\theta}}), \sigma^2(\hat{\boldsymbol{\theta}})\}$ are obtained. The optimal correlation parameters $\hat{\boldsymbol{\theta}}$ can be determined by the Maximum-Likelihood-Estimation (ML) through the following minimization problem [57]:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left[\frac{1}{N} (\mathbf{y} - \boldsymbol{\Theta}\boldsymbol{a})^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y} - \boldsymbol{\Theta}\boldsymbol{a}) (\det \mathbf{R})^{1/N} \right]. \tag{13}$$

In order to solve the optimization problem in Eq. (13), local optimization algorithms such as gradient-based methods are often used. Nonetheless, a major drawback of these techniques relates the troublesome identification of global maxima/minima, being possible to get stuck in local maxima/minima. To avoid this, a global genetic algorithm optimization procedure is used in this work. Since the correlation matrix is symmetric and positive definite, its inverse in Eq. (13) is computed by Cholesky decomposition. Then, once $\hat{\boldsymbol{\theta}}$ is computed, the polynomial coefficients and the process variance $\{\boldsymbol{a}(\hat{\boldsymbol{\theta}}), \sigma^2(\hat{\boldsymbol{\theta}})\}$ are calculated using the Empirical Best Linear Unbiased Estimator (BLUE) as [58]:

$$\boldsymbol{a}\left(\hat{\boldsymbol{\theta}}\right) = \left(\boldsymbol{\Theta}^{\mathrm{T}} \mathbf{R}^{-1} \boldsymbol{\Theta}\right)^{-1} \boldsymbol{\Theta}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{y}, \quad \sigma^2\left(\hat{\boldsymbol{\theta}}\right) = \frac{1}{N} (\mathbf{y} - \boldsymbol{\Theta}\boldsymbol{a})^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y} - \boldsymbol{\Theta}\boldsymbol{a}), \tag{14}$$

where $R_{ij} = R\left(\left|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right|; \hat{\boldsymbol{\theta}}\right)$ is the correlation matrix and $\boldsymbol{\Theta}_{ij} = \psi_j\left(\mathbf{x}^{(i)}\right)$ the information matrix evaluated at all the samples of the ED.

Effective explorative sampling

With the aim of generating representative EDs, the MIPT algorithm is adopted as a computationally efficient and easily implementable adaptive sampling technique. The main advantage of this technique compared to space-filling techniques such as LHS regards its ability to avoid local clustering of points which may consequently lead to numerical instabilities in the inverse of the Kriging correlation matrix in Eq. (14). This exploration distance-based sampling method iteratively augments the ED by adding new sampling points with maximum distance with respect to the data population in the ED among a large set of N_c random Monte-Carlo candidates. Specifically, among the candidates set $\mathcal{C} = \{\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \ldots, \boldsymbol{\xi}^{(N_c)}\}$, a new sample $\mathbf{x}^{(N+1)}$ is chosen by solving the following optimization problem:

$$\mathbf{x}^{(N+1)} = \arg\max_{\boldsymbol{\xi}^* \in \mathcal{C}} \left(\min_{\mathbf{x}^{(i)}, i=1,\ldots,N} \left\| \boldsymbol{\xi}^* - \mathbf{x}^{(i)} \right\|_2 \right), \tag{15}$$

with $\|\cdot\|_2$ the euclidean norm, i.e. $\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{M} x_i^2\right)^{1/2}$.

2.1.3. Multi-element surrogate model approach

The previously presented PCK metamodel suffers from low convergence rates when the forward model $\mathcal{M}$ exhibits non-smoothness [59]. Thus, considerably large ED sizes are often required to achieve accurate predictions. This aspect undermines the computational efficiency of the PCE-based Kriging model, which is dominated by the $\mathcal{O}(N^3)$ complexity of the Kriging predictor. In turn, this implies long construction times or even memory overflow issues when solving the optimization problem in Eq. (13). Moreover, the larger the size of the ED, the slower the evaluation of the corresponding metamodel, which reduces or vanishes the advantages of the surrogate approach. To address this issue, a multi-element PCK model inspired by the ME-gPC method by Wan and Karniadakis [60] is proposed in this work. This approach consists in the partitioning of the random input space into a finite set of non-overlapping subdomains, the construction of a local PCK surrogate model in each one following the formulation in Section 2.1.2 and, finally, assembling them into a piecewise function to obtain a global metamodel, as sketched in Fig. 1.

To systematize the partition of the domain, a simple but efficient regular block partitioning approach has been adopted. To do so, the domain partitions are conducted one variable at a time based on a preliminary sensitivity analysis. To this aim, the algorithm prioritizes the partition of those variables providing highest sensitivities in terms of the Sobol indices, which can be readily extracted from the coefficients of a global PCE [51]. In this way, priority in the partitioning is given to the direction of those parameters with highest sensitivity, i.e., those with the greatest effect on the variability of the quantity of interest *y*. Over a set a surrogate models considering an increasing number of domain partitions, the optimal number of

Table 1
Error metrics for the accuracy assessment of surrogate models over a validation set (VS) of size K (Ref. [61]).

Normalized mean-square error (NRMSE)	Normalized average absolute error (NAAE)
$\text{NRMSE} = \sum_{i=1}^{K} \left(\hat{v}^{(i)} - v^{(i)} \right)^2 \Big/ \sum_{i=1}^{K} \left(\bar{\Upsilon} - v^{(i)} \right)^2$	$\text{NAAE} = (K\sigma_{\Upsilon})^{-1} \sum_{i=1}^{K} \left| \hat{v}^{(i)} - v^{(i)} \right|$
Coefficient of Determination (R^2)	Normalized maximum absolute error (NMAE)
$R^2 = \sigma_{\Upsilon\hat{\Upsilon}}^2 \Big/ \sigma_{\Upsilon}^2 \sigma_{\hat{\Upsilon}}^2$	$\text{NMAE} = (K\sigma_{\Upsilon})^{-1} \max_{i=1}^{K} \left| \hat{v}^{(i)} - v^{(i)} \right|$

partitions is determined in terms of the quality metrics reported hereafter in Table 1. In general, for the random variable $\mathbf{x} : \Omega \to \mathcal{D}_{\mathbf{x}} \subset \mathbb{R}^M$, a decomposition is defined as:

$$\mathcal{D}_{\mathbf{x}} = \bigcup_{j \in \mathcal{J}} \mathcal{D}_j, \quad \mathcal{D}_j \cap \mathcal{D}_{j'} = \emptyset, \quad \text{if } j \neq j', \tag{16}$$

where $\chi_{\mathcal{D}_j} : \Omega \to \mathbb{R}$ denotes the indicator random variable:

$$\chi_{\mathcal{D}_j}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in \mathcal{D}_j \\ 0 & \text{otherwise} \end{cases}.$$

In this way, the global model is defined in a piecewise fashion as:

$$\hat{\mathcal{M}}_{PCK}(\mathbf{x}) = \sum_{j \in \mathcal{J}} \chi_{\mathcal{D}_j}(\mathbf{x}) \hat{\mathcal{M}}^j(\mathbf{x}). \tag{17}$$

As aforementioned, the number of partitions in this work is defined after a parametric analysis. Nevertheless, the previous formulation may be readily automated as follows. The splitting criterion of the domain is determined by a certain user-defined accuracy goal and a minimum number of samples $\mathcal{N}$ per region. Afterwards, the splitting process is performed iteratively from a PCK model built over the full parameter space Ω. In case the target accuracy has not been reached, the space is split into two regions and the ED is enriched in each of these subdomains by the MIPT algorithm until there are $\mathcal{N}$ samples in each one. Note that, given the sequential nature of MIPT, the information of the previously extracted samples is not lost. If the accuracy goal is not reached yet, a new division of the space and a new enrichment of the ED are performed.

2.1.4. Surrogate model accuracy. Complexity analysis of the algorithm

To evaluate accuracy of the developed metamodel, both local and global error metrics are considered. These metrics are computed by considering a validation set (VS) $\Lambda = \{\boldsymbol{\xi}^{(1)}, \ldots, \boldsymbol{\xi}^{(K)}\}$, $K \in \mathbb{N}$, of the parameters space (independent of the ED). Denote by $\Upsilon = \{v^{(1)} = \mathcal{M}(\boldsymbol{\xi}^{(1)}), \ldots, v^{(K)} = \mathcal{M}(\boldsymbol{\xi}^{(K)})\}$ and $\hat{\Upsilon} = \{\hat{v}^{(1)} = \hat{\mathcal{M}}_{PCK}(\boldsymbol{\xi}^{(1)}), \ldots, \hat{v}^{(K)} = \hat{\mathcal{M}}_{PCK}(\boldsymbol{\xi}^{(K)})\}$ the outputs of the VS estimated by the forward model and the metamodel, respectively. Then, the accuracy of the surrogate model can be assessed by using the error metrics like those collected in Table 1. In this table, $\bar{\Upsilon}$ and $\sigma_{\Upsilon} = \sqrt{\left(\sum_{i=1}^{K} \left(\bar{\Upsilon} - v^{(i)} \right)^2 \right) / (K-1)}$ denote the arithmetic mean and the quasi standard deviation of Υ, respectively. Term $\sigma_{\Upsilon\hat{\Upsilon}}$ represents the covariance of $(\Upsilon, \hat{\Upsilon})$, and σ_{Υ}^2 and $\sigma_{\hat{\Upsilon}}^2$ indicate the variance of Υ and $\hat{\Upsilon}$, respectively. Note that the error metric NMAE in Table 1 provides a local estimation of accuracy, while NRMSE, NAAE, and R^2 represent global accuracy measures.

In addition to the error metrics shown in Table 1, and to verify the whole rate of convergence of the proposed model to the unknown function on untried points, we are interested in bounding the *maximum PCK-predictive error* over the domain $\mathcal{D}_{\mathbf{x}} \subset \mathbb{R}^M$:

$$\sup_{\mathbf{x} \in \mathcal{D}_{\mathbf{x}}} |\mathcal{M}(\mathbf{x}) - \hat{\mathcal{M}}_{PCK}(\mathbf{x})|, \tag{18}$$

where

$$\hat{\mathcal{M}}_{PCK}(\mathbf{x}) = \sum_{j \in \mathcal{J}} \chi_{\mathcal{D}_j}(\mathbf{x}) \hat{\mathcal{M}}_{PCK}^j(\mathbf{x}) =$$

$$\sum_{j \in \mathcal{J}} \chi_{\mathcal{D}_j}(\mathbf{x}) \left[\mathbf{r}^{jT}(\mathbf{x}) \mathbf{R}^{j-1} \boldsymbol{y} - \left(\boldsymbol{\Theta}^{jT} \mathbf{R}^{j-1} \mathbf{r}^j(\mathbf{x}) - \boldsymbol{\Psi}^j(\mathbf{x}) \right)^T \left(\boldsymbol{\Theta}^{jT} \mathbf{R}^{j-1} \boldsymbol{\Theta}^j \right)^{-1} \boldsymbol{\Theta}^{jT} \mathbf{R}^{j-1} \boldsymbol{y} \right],$$

is the best linear unbiased predictor (BLUP) of the model $\mathcal{M}$ response at any untried point $\mathbf{x} \in \mathcal{D}_{\mathbf{x}}$, with $\mathbf{r}^j(\mathbf{x}) = [R^j(|\mathbf{x} - \mathbf{x}^{(1)}|), \ldots R^j(|\mathbf{x} - \mathbf{x}^{(N_j)}|)]^T$ the vector of correlations between the design sites $\Xi_j = \left\{ \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N_j)} \right\} \subset \mathcal{D}_j$ and $\mathbf{x}$, and R^j the selected correlation function particularized in the jth subregion. Note that the uniform bound in Eq. (18) covers the worst case for the prediction error of the PCK model.

It has been reported in the literature [62,63] that the prediction error of the universal Kriging converges to zero under uniform metric. Adapting Theorem 2 in Wang et al. [63] to the multielement PCE-Kriging model proposed in this work, the prediction error can be stated to satisfy:

$$
\mathbb{E}\left[\sup_{\mathbf{x} \in \mathcal{D}_{\mathbf{x}}} |\mathcal{M}(\mathbf{x}) - \hat{\mathcal{M}}_{PCK}(\mathbf{x})|\right] \leq
$$
$$
\mathcal{J} \cdot \max_{j \in \mathcal{J}} \mathbb{E}\left[\sup_{\mathbf{x} \in \mathcal{D}_j} |\mathcal{M}(\mathbf{x}) - \hat{\mathcal{M}}_{PCK}(\mathbf{x})|\right] = \mathcal{O}\left(\mathcal{J} P_{\Xi_j}\left(\mathcal{P}^j A + \log^{\frac{1}{2}} P_{\Xi_j}^{-1}\right)\right), \tag{19}
$$

where $\mathcal{J}$ is the number of subdomains, $\mathcal{P}^j = \text{card}\left(A^{M,p,q}\right)$, and A is a constant depending on the eigenvalues of $\boldsymbol{\Theta}_j$. Term $P_{\Xi_j}(\mathbf{x})$ denotes the *power function* given by $P_{\Xi_j}^2(\mathbf{x}) := 1 - \mathbf{r}^{j\mathsf{T}}(\mathbf{x})\mathbf{R}^{j-1}\mathbf{r}^j(\mathbf{x})$, and $P_\Xi := \sup_{\mathbf{x} \in \mathcal{D}_{\mathbf{x}}} P_{\Xi_j}(\mathbf{x})$ is the supremum of the pointwise predictive standard deviation. It is thus reasonable to look for EDs minimizing P_{Ξ_j}. Note that the rate of convergence in Eq. (19) is a deterministic function dependent on the experimental design Ξ_j and decreasing with P_{Ξ_j}. In fact, when $N_j = \text{card}(\Xi_j)$ increases, P_{Ξ_j} tends to zero and so does the multielement PCK prediction error under the uniform metric in Eq. (19).

On the other hand, the algorithm for finding an optimizer of Eq. (13) is an iterative process involving the calculation of the inverse and determinant of a large $N \times N$ covariance matrix $R_{ij} = R\left(|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}|; \hat{\boldsymbol{\theta}}\right)$. Thus, the computational effort to obtain the solution may become impractical for large numbers N of training data points in Ξ. Note that the PCK model requires $\mathcal{O}(N^3)$ operations and has a memory complexity of the order of N^2 [34]. In this light, the splitting technique presented in Section 2.1.3 leads to substantial reductions in the computational effort. Specifically, taking $N_j = \text{card}(\Xi_j)$, with $N_j \ll N$, $\mathcal{J} \ll N_j$, the algorithm effort and the memory storage reduces to $\mathcal{J} \cdot \mathcal{O}(N_j^3) \sim \mathcal{O}(N_j^3) \ll \mathcal{O}(N^3)$ and $\mathcal{J} \cdot \mathcal{O}(N_j^2) \sim \mathcal{O}(N_j^2) \ll \mathcal{O}(N^2)$, respectively. On the other hand, the optimal order of the polynomials in the PCE is automatically identified by the LAR algorithm. It is reported in reference [52] that the LAR algorithm with M variables requires $\mathcal{O}(M^3 + N_j M^2)$ computations in any subdomain $\mathcal{D}_j$. Therefore, in our case where $M \ll N_j$, it follows that $M^3 < N_j M^2$ and, thus, $\mathcal{O}(N_j M^2) \sim \mathcal{O}(N_j)$. Hence, the computational complexity of PCE when inserted as the trend term is marginal with respect to the overall construction of the Kriging model, thereby we can deduce that the efficiency of the proposed PCE-Kriging metamodel is $\mathcal{O}(N_j^3) \ll \mathcal{O}(N^3)$.

2.2. Bayesian parameter inference via MCMC

In the Bayesian inference framework, model parameters $\boldsymbol{\theta}$ are conceived as a random variable with a certain posterior PDF π described by Bayes' theorem:

$$
\pi\left(\boldsymbol{\theta}|\mathbf{y}\right) = \frac{p\left(\mathbf{y}|\boldsymbol{\theta}\right)p\left(\boldsymbol{\theta}\right)}{p(\mathbf{y})}, \quad p(\mathbf{y}) = \int_\Omega p\left(\mathbf{y}|\boldsymbol{\theta}\right)p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}, \tag{20}
$$

where $p(\mathbf{y}|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta})$ denotes the likelihood function, $p(\boldsymbol{\theta})$ the prior distribution of the model parameters, and $p(\mathbf{y}|\mathcal{M})$ a normalizing constant, also called evidence. In the context of this work, $\mathbf{y}$ and $\boldsymbol{\theta}$ represent a set of n experimental observations and the model parameters of the metamodel to be calibrated, respectively. Errors $\boldsymbol{\varepsilon}$ between the experiment and the predictions of the surrogate model are assumed to be normally distributed with zero mean and standard deviation $\sigma_{\boldsymbol{\varepsilon}}$, that is $\mathbf{y} = \widehat{\mathcal{M}}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma_{\boldsymbol{\varepsilon}}\boldsymbol{I})$. Then, the likelihood function $\mathcal{L}(\boldsymbol{\theta})$ can be expressed as:

$$
\mathcal{L}(\boldsymbol{\theta}) = \frac{\exp\left(-\frac{1}{2\sigma_{\boldsymbol{\varepsilon}}^2}\sum_{i=1}^n |y_i - \widehat{\mathcal{M}}(\boldsymbol{\theta})|^2\right)}{\sqrt{2\pi}\sigma_{\boldsymbol{\varepsilon}}}. \tag{21}
$$

Obtaining π from Eq. (20) in analytical closed-form is infeasible in most practical applications, being MCMC methods the most popular approach to numerically characterize the PDF of the model parameters. This approach allows one to draw samples from π without computing the model evidence, which is independent from the model parameters $\boldsymbol{\theta}$. In this work, the DRAM algorithm developed by Haario et al. [53] is implemented. This approach combines delayed rejection (DR) [64] and adaptive Metropolis (AM) [65], which enhances the sampling efficiency of the sampling and enables the identification of multi-modal PDFs. Given a set of observed data samples in vector $\mathbf{d}$, the working principle of the DRAM approach can be outlined as follows:

1. Initialize the parameter set $\boldsymbol{\theta}_c = \boldsymbol{\theta}_0$ and the number T of desired samples. Set an initial point from the parameter space and the covariance of the proposal distribution $\Sigma_p = \Sigma_0$. The proposal distribution is chosen as a multivariate Gaussian distribution with mean $\boldsymbol{\theta}_c$ and covariance matrix Σ_p. Select the initial non-adaptation period n_o and set $i = 1$.
2. Propose a new parameter value $\boldsymbol{\theta}_{p,1}$ by sampling from a proposal PDF $S_1(\boldsymbol{\theta}, \boldsymbol{\theta}_c)$. Accept $\boldsymbol{\theta}_{p,1}$ with probability:

$$
\alpha_1\left(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{p,1}\right) = \min\left(1, \frac{p\left(\mathbf{d}|\boldsymbol{\theta}_{p,1}\right)S_1(\boldsymbol{\theta}_{p,1}, \boldsymbol{\theta}_c)}{p\left(\mathbf{d}|\boldsymbol{\theta}_c\right)S_1(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{p,1})}\right), \tag{22}
$$

and go to step (4). If rejected, propose a second stage move in step (3).

3. Propose a second stage move $\boldsymbol{\theta}_{p,2}$ sampling from $S_2(\boldsymbol{\theta}, \boldsymbol{\theta}_{p,1}, \boldsymbol{\theta}_c)$. This second stage proposal depends not only on the current position of the chain but also on the candidate that has just been proposed and rejected. Accept or reject $\boldsymbol{\theta}_{p,2}$ by setting:

$$\boldsymbol{\theta}_i = \begin{cases} \boldsymbol{\theta}_{p,2}, & \text{with probability } \alpha_2(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{p,1}, \boldsymbol{\theta}_{p,2}), \\ \boldsymbol{\theta}_c, & \text{with probability } 1 - \alpha_2(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{p,1}, \boldsymbol{\theta}_{p,2}), \end{cases} \tag{23}$$

with

$$\alpha_2(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{p,1}, \boldsymbol{\theta}_{p,2}) = \min\left\{1, \frac{p(\mathbf{d}|\boldsymbol{\theta}_{p,2})S_1(\boldsymbol{\theta}_{p,2}, \boldsymbol{\theta}_{p,1})S_2(\boldsymbol{\theta}_{p,2}, \boldsymbol{\theta}_{p,1}, \boldsymbol{\theta}_c)\left[1 - \alpha_1(\boldsymbol{\theta}_{p,2}, \boldsymbol{\theta}_{p,1})\right]}{p(\mathbf{d}|\boldsymbol{\theta}_c)S_1(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{p,1})S_2(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{p,1}, \boldsymbol{\theta}_{p,1})\left[1 - \alpha_1(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{p,1})\right]}\right\}. \tag{24}$$

4. Update the covariance matrix Σ_p as:

$$\Sigma_p = \begin{cases} \Sigma_0 & i \leq n_0 \\ s_d \text{cov}(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_i) & i > n_0 \end{cases} \tag{25}$$

with s_d a scaling parameter. Following [66], $s_d = 2.4^2/d$ with d being the number of fitting parameters is recommended as a good default value in most applications.

5. Go to step 2, until the desired number of samples T are obtained.

3. Numerical results and discussion

This section presents two application case studies to demonstrate the effectiveness of the proposed surrogate model-based Bayesian parameter estimation. These include a two-dimensional benchmark function and the PDE for TDS testing of hydrogen desorption in metals. The previous formulation has been implemented in Matlab environment, and all the numerical tests have carried out in a computer Intel(R) Core(TM) i9-10900X CPU@3.70 GHz with 64 GB of RAM memory. In the remainder of this section, for simplicity in the notation, the predictions of the PCK metamodels $\hat{\mathcal{M}}_{PCK}$ are noted as $\hat{\mathcal{M}}$. A q-norm value of 0.95 and Legendre polynomials of orders ranging from 2 to 6 are selected to build the PCEs in all the analyses hereafter. For the generation of the EDs, the number of random Monte Carlo candidate samples in the MIPT algorithm introduced in Section 2.1.2 is set to 25,000.

3.1. Two-dimensional Drop-Wave function

This first case study investigates the Drop-Wave function, also known as the Salomon's function [67], given by $f : \mathcal{D}_{\mathbf{x}} = [-10, 10]^2 \in \mathbb{R}^2 \rightarrow \mathbb{R}$:

$$f(x_1, x_2) = 1 - \cos\left(2\pi\sqrt{x_1^2 + x_2^2}\right) + 0.1\sqrt{x_1^2 + x_2^2}. \tag{26}$$

This function is commonly used for benchmarking optimization algorithms. Owing to its highly non-linear character, the Drop-Wave function represents an ideal case study to validate the proposed multi-element PCK metamodel. Note that the surrogate modelling of this function using conventional techniques is extremely challenging given its fast-varying gradients and irregular response as observed in Fig. 2 (a). Following Section 2.1.3, four experimental design sets $ED_i \subset \mathcal{D}_{\mathbf{x}}$, $i = 1, \ldots, 4$ containing 360, 720, 1440 and 2880 samples have been defined. In addition, three different number of domain partition schemes $\mathcal{P}^j$, $j = 1, \ldots, 3$, have been considered. These include $\mathcal{P}^1 = [-10, 10]^2$, $\mathcal{P}^2 = ([-10, 0) \cup [0, 10])^2$, and $\mathcal{P}^3 = \left(\left[-10, \frac{-10}{3}\right] \cup \left[\frac{-10}{3}, \frac{10}{3}\right] \cup \left[\frac{10}{3}, 10\right]\right)^2$, leading to a total of two, four and nine sub-domains, respectively. The number of samples has been chosen with the aim of obtaining a wide range of errors to correctly identify the convergence of the prediction error. For instance, if one takes the R^2 error metric, note that the constructed metamodels exhibit a wide range of R^2 values from 0.011 to 0.999. Additionally, the predictions by a previously reported multielement approach, the Stochastic Spectral Embedding (SSE) model proposed by Sudret and Marelli [35], are also presented as a reference solution. The SSE model is a PCE-based technique consisting of constructing a sequence of residual spectral expansions of the target model in subdomains of the input space. The implementation included in the UQLab software [68] has been used to carry out the analyses. Four different surrogate models have been built, one for each considered ED. As parameters, a q-norm value of 0.95, polynomials ranging from degree 2 to 10, and a minimum size of points per region equal to the size(ED)/120 have been selected. To sample the ED, the sequential experimental design based on the LHS implemented in UQLab has been chosen.

Hence, a total of sixteen surrogate models have been constructed. For ease in the discussion, the PCK surrogate models are specified with sub- and super-indexes denoting the size of the ED and the number of partitions, respectively, as reported in Table 2. All the surrogate models have been validated using a reasonably large VS of 20,000 samples, and the accuracy of the models has been evaluated through the accuracy metrics reported in Table 3. The computational times involved in the construction t_c of the surrogate models, as well as their evaluation times t_e and t_e(VS) for a single point and the full VS
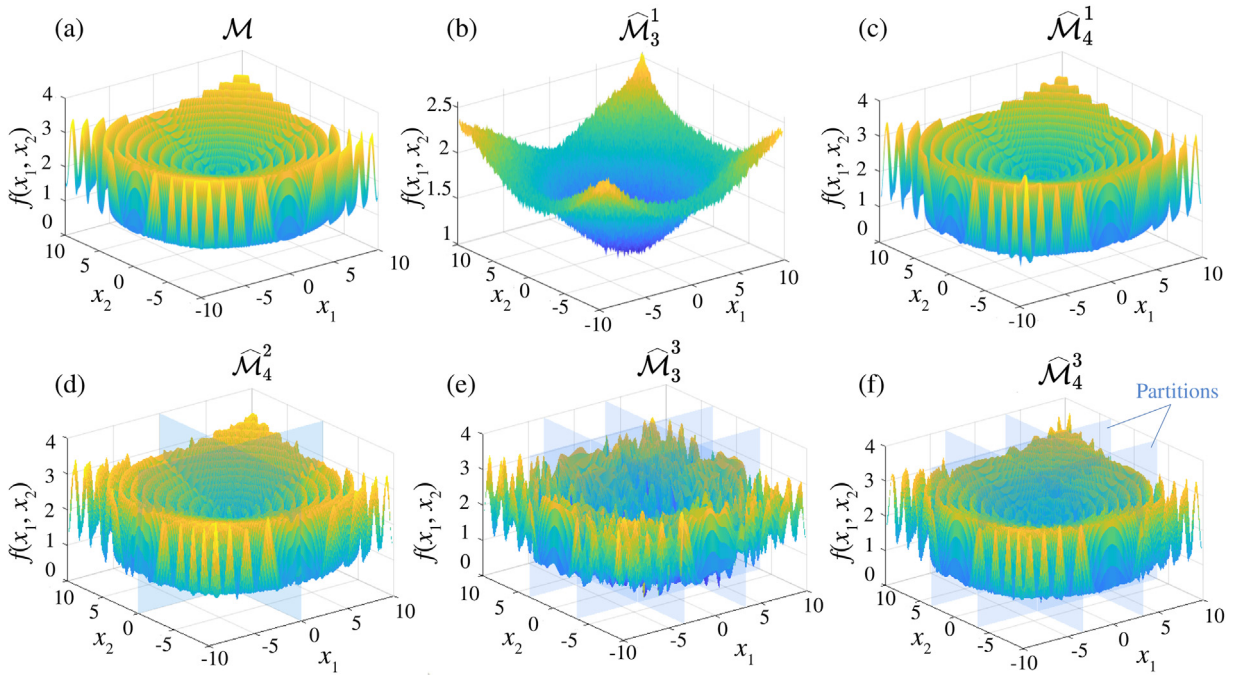
Fig. 2. Exact response surface of the Drop-Wave function (a), and predictions by surrogate models $\hat{\mathcal{M}}_1^3$ (b), $\hat{\mathcal{M}}_4^1$ (c), $\hat{\mathcal{M}}_4^2$ (d), $\hat{\mathcal{M}}_3^3$ (e) and $\hat{\mathcal{M}}_4^3$ (f).

Table 2
PCK surrogate-models constructed for the Drop-Wave function considering increasing EDs (ED_i) with varying numbers of domain partitions ($\mathcal{P}^j$) (VS of 20,000 samples).

No. of sub-domains	ED_1 (360 samples)	ED_2 (720 samples)	ED_3 (1440 samples)	ED_4 (2880 samples)
$\mathcal{P}^1$ (1 partition)	$\hat{\mathcal{M}}_1^1$	$\hat{\mathcal{M}}_2^1$	$\hat{\mathcal{M}}_3^1$	$\hat{\mathcal{M}}_4^1$
$\mathcal{P}^2$ (4 partitions)	$\hat{\mathcal{M}}_1^2$	$\hat{\mathcal{M}}_2^2$	$\hat{\mathcal{M}}_3^2$	$\hat{\mathcal{M}}_4^2$
$\mathcal{P}^3$ (9 partitions)	$\hat{\mathcal{M}}_1^3$	$\hat{\mathcal{M}}_2^3$	$\hat{\mathcal{M}}_3^3$	$\hat{\mathcal{M}}_4^3$
SSE (reference)	$\hat{\mathcal{M}}_1^{SSE}$	$\hat{\mathcal{M}}_2^{SSE}$	$\hat{\mathcal{M}}_3^{SSE}$	$\hat{\mathcal{M}}_4^{SSE}$
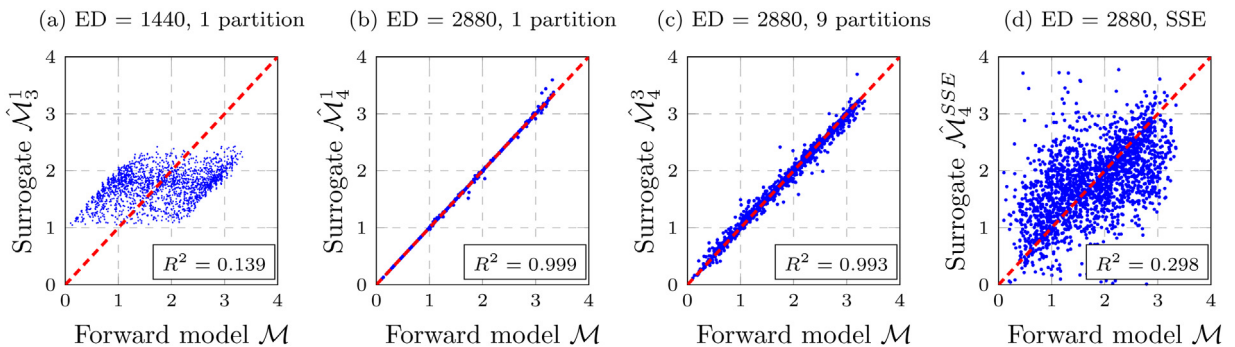
Fig. 3. Forward model evaluations versus surrogate model predictions for the Drop-Wave function. $\hat{\mathcal{M}}_3^1$ (a), $\hat{\mathcal{M}}_4^1$ (b), $\hat{\mathcal{M}}_4^3$ (c) and $\hat{\mathcal{M}}_4^{SSE}$ (d) (VS of 20,000 samples).

have been also computed and collected to compare their effectiveness. The comparison of the exact response surface of the Drop-Wave function and the predicted ones by five of the constructed surrogate models is depicted in Fig. 2.

Fig. 3 shows the scatter plots of the forward model evaluated on the VS versus four metamodels, $\hat{\mathcal{M}}_3^1$ (a), $\hat{\mathcal{M}}_4^1$ (b), $\hat{\mathcal{M}}_4^3$ (c) and $\hat{\mathcal{M}}_4^{SSE}$ (d). The first two metamodels consider the whole design space, while the last two account for partition approaches. It is noted in this figure that the best approximations are found for $\hat{\mathcal{M}}_4^1$ (no partitions) and $\hat{\mathcal{M}}_4^3$ (9 partitions with EDs of 2880 samples). The predictions by these models exhibit low scatter around the diagonal line (perfect metamodel) with coefficients of determination R^2 very close to 1. The limited efficiency of the SSE model in this case study is evidenced by the large scatter of its predictions along the diagonal in Fig. 3(d). Interestingly, note that the predictions by $\hat{\mathcal{M}}_4^1$ slightly

Table 3
Accuracy and computational efficiency analysis of proposed surrogate models applied to the Drop-Wave function. Terms t_c, t_e and t_e(VS) denote the time of construction, average point evaluation, and evaluation on the full validation set, respectively.

| Model	NAAE	NMAE	NRMSE	R^2	t_c [s]	t_e [s]	t_e(VS) [s]
$\hat{\mathcal{M}}_1^1$	1.030E + 0	1.736E−4	1.589E + 0	0.024	34.1	5.069E−5	1.013
$\hat{\mathcal{M}}_2^1$	8.340E−1	1.001E−4	8.701E−1	0.134	151.3	2.014E−4	4.028
$\hat{\mathcal{M}}_3^1$	8.321E−1	8.159E−5	8.627E−1	0.139	661.8	9.301E−4	18.601
$\hat{\mathcal{M}}_4^1$	2.876E−3	2.707E−5	2.651E−4	0.999	2905.7	5.078E−3	101.540
$\hat{\mathcal{M}}_1^2$	9.162E−1	1.399E−4	12.080E−1	0.043	20.3	6.840E−6	0.138
$\hat{\mathcal{M}}_2^2$	8.978E−1	1.412E−4	11.980E−1	0.823	44.4	1.996E−5	0.399
$\hat{\mathcal{M}}_3^2$	5.448E−1	1.355E−4	4.974E−1	0.551	131.8	5.656E−5	1.131
$\hat{\mathcal{M}}_4^2$	2.256E−2	3.976E−5	2.559E−3	0.997	594.7	2.066E−4	4.132
$\hat{\mathcal{M}}_1^3$	8.757E−1	2.044E−4	11.230E−1	0.079	28.9	3.590E−6	0.071
$\hat{\mathcal{M}}_2^3$	7.765E−1	1.231E−4	8.622E−1	0.193	41.8	6.875E−6	0.138
$\hat{\mathcal{M}}_3^3$	4.404E−1	1.297E−4	3.551E−1	0.645	86.4	1.513E−5	0.303
$\hat{\mathcal{M}}_4^3$	4.345E−2	4.667E−5	6.712E−3	0.993	244.7	4.750E−5	0.950
$\hat{\mathcal{M}}_1^{SSE}$	1.249E + 0	6.463E−4	2.684E + 0	0.011	42.9	3.529E−5	0.706
$\hat{\mathcal{M}}_2^{SSE}$	1.191E + 0	1.302E−3	2.733E + 0	0.024	33.9	3.401E−5	0.681
$\hat{\mathcal{M}}_3^{SSE}$	8.838E−1	1.714E−3	1.535E + 0	0.113	30.6	3.266E−5	0.653
$\hat{\mathcal{M}}_4^{SSE}$	6.748E−1	5.694E−4	1.041E + 0	0.298	32.5	3.101E−5	0.620
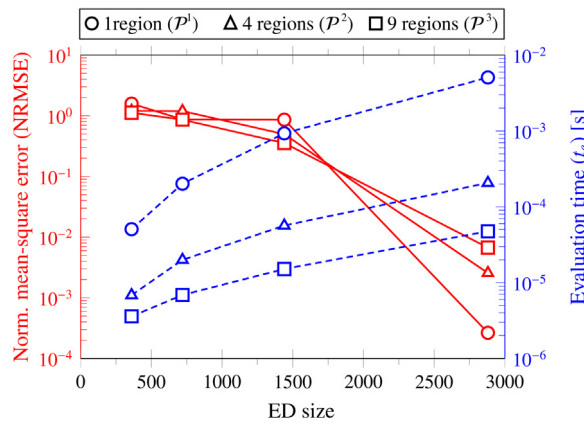
Fig. 4. Performance assessment of PCK surrogate models for the Drop-Wave function (VS of 20 000 samples).

outperform those obtained with $\hat{\mathcal{M}}_4^3$, while higher numbers of partitions do not seem not to systematically improve the prediction accuracy. Nevertheless, the computational times involved in the construction and evaluation of $\hat{\mathcal{M}}_4^1$ are, respectively, about 10 and 100 times those required by $\hat{\mathcal{M}}_4^3$ (see Table 3). It is extracted from this analysis that the selection of the optimal surrogate model must be conducted by balancing the computational burden and the fitting accuracy. In this regard, Fig. 4 investigates the computational efficiency in terms of t_e versus prediction accuracy (NRMSE) for all the considered surrogate models. In this figure, it is trivially observed that as the size of the ED increases, both computational time and accuracy of the metamodels increase. It is important to highlight that the consideration of higher number of subdomains leads to lower evaluation times for all the considered EDs. This is explained by the implementation of the LAR algorithm to extract optimal sets of polynomials in the PCE, and in particular, thanks to the reductions in the computational cost involved in the construction of the Kriging predictor (see Section 2.1.4). When inspected in a partition-wise fashion, the forward model exhibits a smoother behaviour, in such a way that the PCE requires less high-order polynomials to reproduce its behaviour. This results in more compact expansions, which also decreases the cost in the computation of the correlation matrix in the Kriging metamodel. On the other hand, note that the higher the order in the PCE, the larger the number of samples that are required in the ED to fit the expansion with accuracy. In this light, to reach a comparable accuracy to the one achieved by the Multi-Element PCK ($R^2 > 0.99$) through SSE, it is necessary to increase the degree of the polynomial expansion up to 14 and to sample a ED with more than 45,000 training points. These results strengthen the comparatively superior convergence rate of the proposed approach for this class of problems with highly non-linear spatial variability. As previously detailed in Section 2.1.4, the computational complexity of the proposed PCK model is $\mathcal{O}(N_j^3)$. This is due to the Cholesky decomposition of the correlation matrix **R** in Eq. (13). Therefore, as the ED increases, the computational cost involved in the determination of the stochastic hyper-parameters of the Kriging model and its evaluation rises dramatically. On the other hand, the dependence of the accuracy of the metamodel with the number of partitions is not so clear. It is noted that the accuracy of metamodels trained with a larger number of subdomains is higher compared to those with no
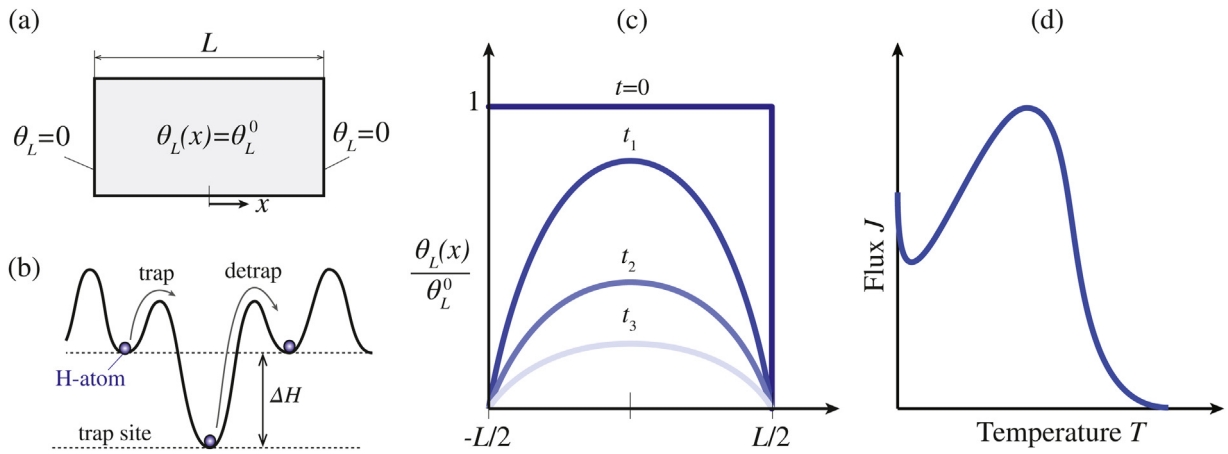
Fig. 5. (a) A schematic illustration of initial and boundary conditions in a TDS test. (b) Schematic definition of binding energy in a one-dimensional diffusion path. (c) Transient solution curves of the normalised lattice occupancy fraction θ_L/θ_L^0 at different times t along the specimen's thickness. (d) A schematic of typical hydrogen desorption flux versus temperature curves obtained in a TDS test.

partitions for limited to moderate EDs. Nevertheless, when the size of the ED goes from moderate to large, the accuracy diminishes.

3.2. Thermal desorption spectroscopy (TDS) of hydrogen in metals

This last section reports the use of the proposed surrogate model for the Bayesian identification of hydrogen desorption and trapping characteristics in metallic materials. Hydrogen embrittlement (HE) refers to the loss of ductility and toughness of metallic alloys induced by hydrogen atoms deposited at lattice sites and micro-structural defects such as dislocations, grain boundaries or vacancies [69,70]. Although this phenomenon has been extensively documented since the 19th century [71], the growing trend towards a hydrogen-based economy as a means of mitigating CO_2 emissions and fossil fuel dependency has generated unprecedented interest on HE research. Micro-structural defects in metals act as 'trap' sites, which sequester hydrogen and govern the susceptibility to HE [72,73]. Their characterization is thus of pivotal importance for the understanding of HE and the design of HE-resistant alloys, and this is generally achieved using TDS experiments [74]. The TDS test involves several stages [75]: charging a sample with hydrogen, heating the sample at a fixed rate, and detecting the flux of desorbing hydrogen as a function of temperature by using a mass spectrometer. The hydrogen flow curve of desorbing hydrogen as a function of temperature defines the TDS spectrum, whose peaks can be associated with the presence of diverse micro-structural defects. Nonetheless, the formation of peaks in the TDS spectrum may be induced by the combined action of manifold hydrogen traps, being necessary to use simulation models and inverse calibration for their identification. Previous investigations on the modelling of the TDS test evidenced the existence of non-smooth relationships between the flux curves and the parameters characterizing micro-structural defects (see e.g. [76]), making this application a formidable benchmark case study for the formulation presented in this work. In the remainder of this section, the PDE governing the hydrogen diffusion in materials tested by TDS is introduced in Section 3.2.1. The construction of the surrogate model and its performance evaluation is reported in Section 3.2.2 and, finally, Section 3.2.3 presents the Bayesian parameter identification results.

3.2.1. TDS governing diffusion equation

Consider a one-dimensional specimen of length L as sketched in Fig. 5(a). The specimen is subjected to increasing temperatures T, starting from T_o and increasing at a constant heating rate ϕ. Hydrogen atoms occupy normal interstitial lattice sites (NILS) and additionally can reside at trapping sites such as interfaces or dislocations. The kinetics of hydrogen trapping and detrapping in metals is commonly described with a two-level system as sketched in Fig. 5(b) for the case of a single trap. The potential landscape in this figure describes the diffusion path of hydrogen in metals, the trap binding energy ΔH being the difference between detrapping and trapping energies. Let us assume that the number of hydrogen traps in the specimen amounts to N_t. Then, let $C_L(x, t)$ and $C_{T,i}(x, t)$, $i = 1, \ldots, N_t$, denote the hydrogen concentration in the lattice and in the i-th trap, respectively, with $x \in [-L/2, L/2]$ and t respectively denoting space and time. On this basis, the Fickian diffusion equation needs to be enriched with source and sink terms as [76]:

$$\frac{\partial C_L}{\partial t} + \sum_{i=1}^{N_t} \frac{\partial C_{T,i}}{\partial t} = D_L \frac{\partial^2 C_L}{\partial x^2}, \tag{27}$$

with $D_L = D_o \exp(-Q/RT)$ being the lattice diffusion coefficient, which is expressed in terms of the lattice activation energy Q, diffusion pre-exponential factor D_o, and the universal gas constant R. It is convenient to introduce the lattice and

Table 4
Non-dimensional variables used in the hydrogen diffusion PDE employed for the TDS tests.

Spatial coordinate	$\bar{x} = x/L$	Lattice activation energy	$\bar{Q} = Q/(RT_o)$
Time coordinate	$\bar{t} = tD_o/L^2$	Trap binding energy	$\overline{\Delta H_i} = \Delta H_i/(RT_o)$
Heating rate	$\bar{\phi} = (\phi L^2)/(T_o D_o)$	Lattice diffusion coefficient	$\bar{D}_L = D_L/D_o$
Trap density	$\bar{N}_i = (\alpha N_{T,i})/(\beta N_L)$	Local equilibrium constant	$\bar{K} = \exp\left\{-\frac{\overline{\Delta H_i}}{\bar{T}}\right\}$
Temperature	$\bar{T} = T/T_o$	Fractional lattice occupancy	$\bar{\theta}_L = \theta_L/\theta_L^o$

trap occupancy fractions θ_L and $\theta_{T,i}(\{\theta_L, \theta_{T,i}\} \in [0,1])$, respectively, by rewriting the corresponding concentrations in the form $C_L = \theta_L \beta N_L$ and $C_{T,i} = \theta_{T,i} \alpha N_{T,i}$. Here, β is the number of NILS per unit volume, α is the number of atoms sites per trap, N_L is the number of lattice atoms per unit volume, and $N_{T,i}$ is the number of trap sites per unit volume. Therefore, Eq. (27) can be rewritten as:

$$\frac{\partial \theta_L}{\partial t} + \sum_{i=1}^{N_t} \left(\frac{\alpha N_{T,i}}{\beta N_L}\right) \frac{\partial \theta_{T,i}}{\partial t} = D_L \frac{\partial^2 \theta_L}{\partial x^2}. \tag{28}$$

The PDE in Eq. (28) needs to be complemented with trap kinetic equations. To this aim, the formulations by MacNabb and Foster [77] and Oriani [78] are commonly adopted. The latter represents a simplification of the former by assuming that a local equilibrium exists between the hydrogen atoms at the lattice sites and the i-th trap such that, for $\theta_L \ll 1$,

$$\theta_{T,i} = \frac{K_i \theta_L}{(1 + K_i \theta_L)}, \tag{29}$$

with K_i being the local equilibrium constant for the i-th trap:

$$K_i = \exp\left\{-\frac{\Delta H_i}{RT}\right\}. \tag{30}$$

Introducing Eq. (29) into (28), and following the non-dimensional formulation developed by Raina et al. [76], the governing PDE describing hydrogen diffusion in the TDS test can be recast in a compact form as:

$$\frac{\partial \bar{\theta}_L}{\partial \bar{t}}\left[1 + \sum_{i=1}^{N_t} \frac{\bar{K}_i \bar{N}_i}{\left(1 + \bar{K}_i \theta_L^0 \bar{\theta}_L\right)^2}\right] + \frac{\bar{\theta}_L}{\bar{T}^2} \sum_{i=1}^{N_t} \frac{\bar{K}_i \bar{N}_i \overline{\Delta H_i} \bar{\phi}}{\left(1 + \bar{K}_i \theta_L^0 \bar{\theta}_L\right)^2} = \bar{D}_L \frac{\partial^2 \bar{\theta}_L}{\partial \bar{x}^2}, \tag{31}$$

with θ_L^0 being the initial lattice occupancy. The non-dimensional variables employed are listed in Table 4.

The initial and boundary conditions of the PDE in Eq. (31) are schematically presented in Fig. 5(a). At $t = 0$, it is assumed an initial uniform lattice occupancy $\bar{\theta}_L(\bar{x}, \bar{t} = 0) = 1$. Thereafter, the hydrogen lattice occupancy is assumed zero at the boundaries, that is $\bar{\theta}_L(\bar{x} = \pm 1/2, \bar{t} > 0) = 0$. As temperature raises, the lattice occupancy evolves spatially and temporally as sketched in Fig. 5(c), and the flux of hydrogen atoms $J(t)$ diffusing out at boundaries is measured as presented in Fig. 5(d). This flux can be obtained in non-dimensional terms after solving Eq. (31) as [76]:

$$\bar{J} = -\bar{D}_L \theta_L^o \frac{\partial \bar{\theta}_L}{\partial \bar{x}}. \tag{32}$$

Generally, the magnitudes of Q, D_0 and θ_L^0 are known, and the heating rate ϕ is an input to the TDS system. Therefore, the TDS spectrum can be used to map the microstructural hydrogen traps, as characterised by their trap densities ($\bar{N}_i$) and binding energies ($\overline{\Delta H_i}$). These can be obtained for a given flux curve $\bar{J}$ by the inverse calibration of the PDE in Eq. (31).

The surrogate modelling of the flux curves obtained after solving Eq. (31) represents a formidable problem due to the strong nonlinearities of these curves. Specifically, depending upon the hydrogen trap configuration, several different regimes can be observed, as previously discussed by Raina *et al.* [76]. Specifically, their results for the case of metals containing a single trap showed that no peak flux is attained for low trap densities and binding energies. Alternatively, when a peak flux is found, those authors identified two distinct regimes (I and II) originated by two types of microstructural defects, referred to as shallow and deep traps. Shallow traps are characterized by large trap densities, and give origin to peak fluxes that are highly sensitive to both N and ΔH. On the other hand, deep traps are characterized by low trap densities, resulting in peak fluxes that are insensitive to the trap binding energy. The existence of these different regimes turns the construction of a surrogate model covering the whole domain of the traps into an notably challenging task. Note that a large number of high-order polynomials and a dense ED need to be included in the PCE to accurately represent the whole global behaviour of the hydrogen flux. Such large EDs may severely compromise the computational efficiency of the surrogate model since, as indicated above, the complexity of the Cholesky decomposition of the correlation matrix **R** in Eq. (13) is $\mathcal{O}(N^3)$. The metamodeling of TDS experiments thus represents an exceptional case study to justify the use of the domain partitioning approach presented in Section 2.1.3.

Table 5

Accuracy and computational efficiency analysis of PCK metamodels developed for the surrogate modelling of hydrogen diffusion flux curves obtained by TDS (VS = 36,000).

	$\hat{\mathcal{M}}_1$ 72 subdomains ED = 72,000	$\hat{\mathcal{M}}_2$ 72 subdomains ED = 108,000	$\hat{\mathcal{M}}_3$ 108 subdomains ED = 72,036	$\hat{\mathcal{M}}_4$ 108 subdomains ED = 108,000
t_e [ms]	1.97	4.86	0.83	2.02
NAAE	1.0637E−02	7.540E−03	1.115E−02	7.736E−03
NMAE	2.517E−05	2.239E−05	2.363E−05	1.656E−05
NRMSE	9.452E−04	5.259E−04	1.208E−03	5.229E−04
R^2	0.9991	0.9995	0.9988	0.9995

3.2.2. Surrogate modelling of TDS flux curves for metals with two traps

The PDE in Eq. (31) is solved numerically by using the *pdepe* solver in MATLAB. A space discretization of 201 elements along x was found to provide mesh-independent results. To illustrate the behaviour of a ferritic steel sample, representative model parameters from reference [76] have been adopted herein, including a lattice activation energy $Q = 6.7$ kJmol^{-1}, diffusion pre-exponential factor $D_o = 2 \times 10^{-7}$ m^{2}s^{-1}, heating rate $\overline{\phi} = 0.1$ and lattice density $N_L = 8.46 \times 10^{28}$ atoms m^{-3}, with $\alpha = \beta = 1$. The initial temperature and occupancy fraction are chosen as $T_o = 293$ K and $\theta_L^0 = 10^{-6}$, respectively, and the thickness of the specimen is chosen as $L = 5$ mm. The variation of trap binding energies and trap densities are selected as the physically meaningful ranges $-40 \leq \overline{\Delta H_i} \leq -10$ and $10^{-7} \leq \overline{N_i} \leq 10^{-2}$. In the present study, we limit to the modelling of metals with two hydrogen traps, i.e. $N_t = 2$. Therefore, in the surrogate modelling, temperature and the trap densities and binding energies are considered as input variables, which amounts to 5 design variables, i.e. $\mathcal{M}(\boldsymbol{x}) = \bar{J}$ with $\boldsymbol{x} = \left[\overline{T}, \overline{\Delta H_1}, \overline{\Delta H_2}, \log(\overline{N_1}), \log(\overline{N_2})\right]^{\mathrm{T}} \subset \mathbb{R}^5$.

After some preliminary sensitivity analyses, two partitions of $\mathcal{D}_{\mathbf{x}}$ have been considered, namely $\mathcal{P}^1$, $\mathcal{P}^2$. Partition $\mathcal{P}^1$ has been defined by splitting the temperature $\overline{T}$ and activation energy domains ($\overline{\Delta H_i}$, $i = 1, 2$) in two, while three segments were considered for the partition of the domain of the trap densities ($\log(\overline{N_i})$, $i = 1, 2$). On the other hand, the partitions in $\mathcal{P}^2$ remain identical except for the temperature domain which is divided in three sub-domains. In order to define the optimal surrogate model, EDs of 1000 and 1500 sampling points per-subdomain have been considered for $\mathcal{P}^1$, while EDs of 667 and 1000 points per sub-domain have been defined for $\mathcal{P}^2$. This amounts to four different surrogate models labelled with $\hat{\mathcal{M}}_i$, $i = 1, \ldots, 4$. In order make a fair comparison between the different proposals, models $\hat{\mathcal{M}}_1$ and $\hat{\mathcal{M}}_3$ are trained with EDs of 72,000 and 72,036 samples respectively, while $\hat{\mathcal{M}}_2$ and $\hat{\mathcal{M}}_4$ are trained with 108,000 samples. The comparison of the metamodels in terms of accuracy and computational efficiency is reported in Table 5 over a VS of 36,000 samples. Similarly to the results in the previous case study, the consideration of domain partitioning leads to considerable computational time reductions and moderate reductions in prediction accuracy. Note that the evaluation time of the forward model is about 280 ms, so all the metamodels achieve reductions between 98.3%–99.7%. The computation time of the metamodel depends upon the size of the ED in each region, which explains why models $\hat{\mathcal{M}}_3$ and $\hat{\mathcal{M}}_2$ are the fastest and slowest ones, respectively. On the other hand, the accuracy of the metamodel increases as so does the size of the ED. Indeed, models $\hat{\mathcal{M}}_2$ and $\hat{\mathcal{M}}_4$ exhibit significantly lower errors compared to models $\hat{\mathcal{M}}_1$ and $\hat{\mathcal{M}}_3$. Therefore, in view of these results, $\hat{\mathcal{M}}_4$ provides a good trade-off between computational efficiency and accuracy, and it is selected in the subsequent Bayesian model parameter inference. To illustrate the effectiveness of the surrogate model in representing the different stages observed in the TDS test, Fig. 7 shows the comparison of the forward model and the predictions by $\hat{\mathcal{M}}_4$ for a variety of combinations of traps, including the case of fluxes without peak, one single peak, and two peaks. It is observed that the proposed PCK model can accurately reproduce all the different regimes observable in the TDS test. Only some minor errors are observed in the no flux regime, given the imposed limitation on the order of the polynomials in the PCE for the sake of computational efficiency. Finally, in order to highlight the superior performance of the proposed multi-element PCK metamodel, Fig. 6 furnishes the comparison of the predictions by standard LAR-PCE (trained with 76,000 samples) and $\hat{\mathcal{M}}_4$. These results clearly evidence the superior performance of the proposed approach with respect to LAR-PCE, whose predictions versus the forward model exhibits a large scatter around the diagonal line with a low coefficient of determination of $R^2 = 0.46$.

3.2.3. Bayesian inference of the trapping sites from a TDS experiment

In this last subsection, the previous surrogate model $\hat{\mathcal{M}}_4$ is used to conduct Bayesian parameter inference following the MCMC algorithm in Section 2.2. The trap binding energies and densities of the two trap system are chosen as the inference parameters $\boldsymbol{\theta} = \left(\overline{\Delta H_1}, \overline{\Delta H_2}, \log(\overline{N_1}), \log(\overline{N_2})\right)$ in Eq. (20). With the purpose of assessing the performance of the implemented DRAM MCMC approach to infer the parameters of hydrogen traps covering the two different regions described in reference [76], two different trap configurations are considered to generate synthetic experimental data from the forward model. A two-trap system (EI) with properties $\boldsymbol{\theta} = (-25, -35, -3, -2.5)$ is considered first. The second one (EII) instead is defined by $\boldsymbol{\theta} = (-15, -30, -6, -3)$. The flux curves obtained in EI and EII correspond to those previously shown in Fig. 7(f) and (g), respectively. In addition, to evaluate the sensitivity of the model parameter inference to the presence of noise pollution in the experiment, a second analysis of the EII experiment was performed after affecting the flux curve with a zero-mean Gaussian white noise with a standard deviation equal to 0.4 times the mean value of the unpolluted flux curve
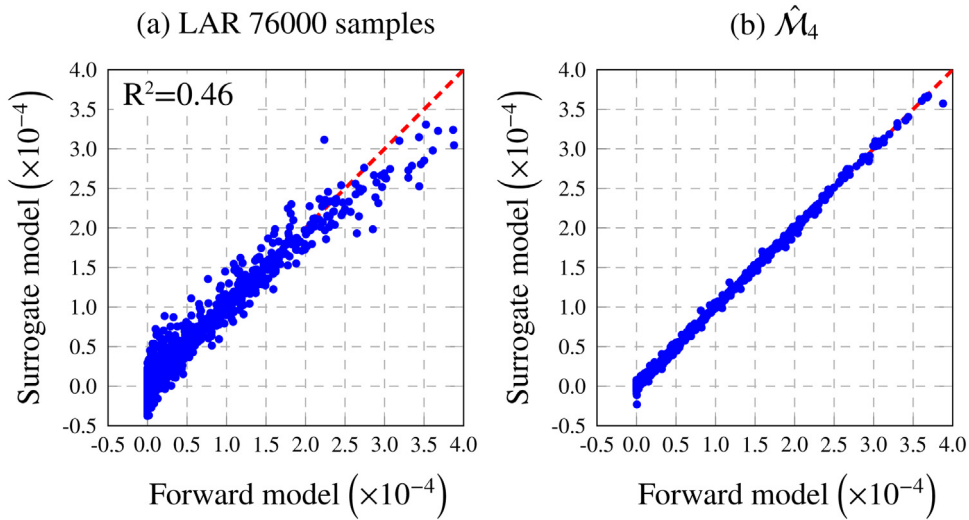
Fig. 6. Scatter plots of Hydrogen flux curves obtained by the forward solution of the PDF of the TDS test versus the predictions by standard LAR-PCE (a) and by the proposed multi-element PCK metamodel $\hat{\mathcal{M}}_4$ (b) (VS of 36,000 samples).
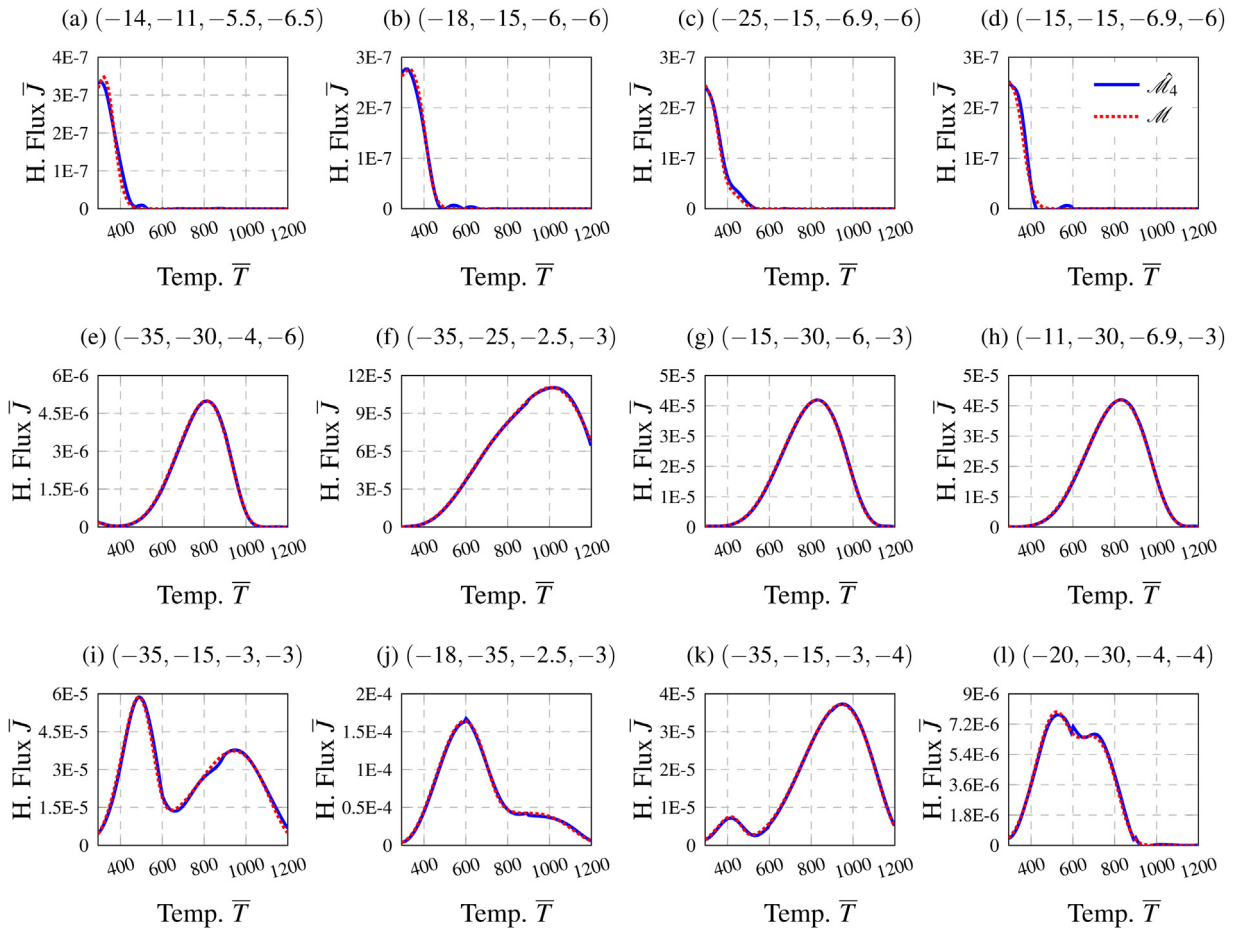

Fig. 7. Surrogate modelling of the Hydrogen flux curves obtained by TDS of metals with different values of trap binding energies and concentrations. Quantities in parenthesis represent the parameters of the traps $\left(\overline{\Delta H_1}, \overline{\Delta H_1}, \log(\overline{N}_1), \log(\overline{N}_2)\right)$.
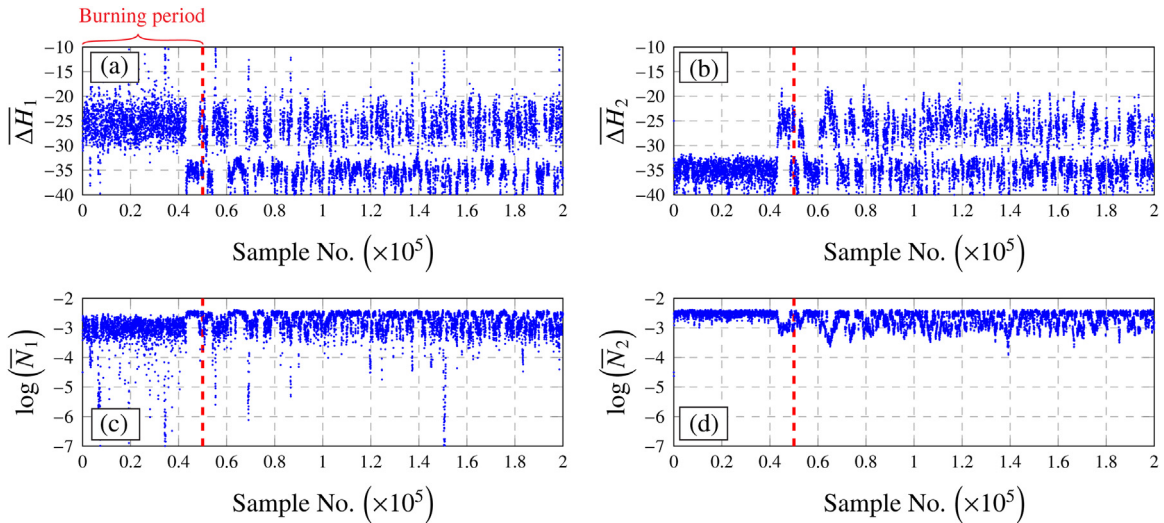
Fig. 8. Markov chains generated by DRAM MCMC of trap parameters $\overline{\Delta H_1}$, $\overline{\Delta H_2}$, $\log(\overline{N}_1)$, $\log(\overline{N}_2)$ for TDS Experiment EI.

(note later in Fig. 11 that such a noise level represents a considerably low signal-to-noise ratio). The experiment EI was defined to illustrate the potentials of the implemented DRAM MCMC to draw samples from a multi-modal distribution. Note that the PDE in Eq. (31) does not differentiate the order of the traps, thereby the problem is ill-posed and the posterior distribution is expected to exhibit two modes corresponding to two symmetric solutions. Instead, the experiment EII was designed to account for a trap ($\overline{\Delta H_1} = -15$, $\log(\overline{N}_1) = -6$) in the regime with no flux as identified by Raina and et al. [76], while the second trap ($\overline{\Delta H_2} = -30$, $\log(\overline{N}_2) = -3$) represents a deep trap. Therefore, the PDF in this case should be uni-modal.

In the inference analyses, uninformative uniform priors $\mathcal{U}(-40, -10)$ and $\mathcal{U}(-7, -2)$ are selected for $\overline{\Delta H_i}$ and $\log(\overline{N}_i)$ ($i = 1, 2$), respectively. A total number of 200 000 samples with a burning time of 50 000 samples were drawn by the previously introduced Bayesian inference approach for EI. The sampling of the posterior PDF in Experiment EII was more challenging given its uni-modal nature with large regions of low probability, requiring up to 480,000 samples with a burning period of 160,000 samples to achieve convergence. Interestingly, this phenomenon attenuates when the flux curve is affected by noise, only requiring a chain of 120,000 samples with a burning period of 30,000 to attain convergence. This is expectable since the noise-induced lower probability concentration around the exact true solutions makes it easier for the chain to span from one solution to the symmetric one. The initial location state was defined as $\theta_0 = (-25, -25, -4.5, -4.5)$, while the prediction error was set to $\sigma_{\varepsilon} = 1E - 9$ and $1.6E - 5$ for the noise unpolluted and polluted cases, respectively. After some initial calibration by visual inspection of the chain traces, a diagonal covariance matrix with entries equal $(0.05 \cdot \theta_0)^2$ was initially defined for the Gaussian proposal. In the AM step the proposal distribution was scaled by a factor $s_d = 2.4^2/d$ and the non-adaptation period n_0 was set to 500 and 4000 for the EI and EII experiments, respectively. On the other hand, in the DR step the proposal is scaled down by a factor of 0.2.

The Markov chain and the joint posterior PDF obtained for Experiment EI are presented in Figs. 8 and 9, respectively. As anticipated, the problem is ill-posed and there exist two potential solutions, namely $\theta = (-25, -35, -3, -2.5)$ and $\theta = (-35, -25, -2.5, -3)$. This manifests in the marginal PDFs in Fig. 9. Specifically, the PDFs corresponding to parameters $\overline{\Delta H_1}$ and $\overline{\Delta H_2}$ have two identical modes at -35 and -25, and parameters $\log(\overline{N}_1)$ and $\log(\overline{N}_2)$ have two modes at -3 and -2.5. It is observed in Fig. 9 that, indeed, the implemented DRAM algorithm is capable of exploring the two modes in the distribution, without getting stuck around one of them as it is usually the case when implementing standard MCMC methods. For validation purposes, the posterior PDF has been also computed by direct integration of the forward solution. To do so, the evidence of the model has been computed over a mesh of 60^4 elements. This required forty five hours of parallel computation on ten cores, while the MCMC approach only required about four hours on a single core. The Highest Density Regions (HDRs) at the 80% and 50% level of both distributions are reported in Table 6 The close fittings between the exact marginal PDFs and those predicted by the surrogate model-based Bayesian inference in Fig. 9 demonstrate the accuracy of the developed approach, as it is also evident from the computed HDRs in Table 6. Finally, the Markov chain, and the posterior PDF obtained for the TDS experiment EII are reported in Figs. 10 and 11, respectively, and the posterior HDR values are reported in Table 7. In this case, the PDFs exhibit one single mode as previously anticipated. This corresponds to the shallow trap ($\overline{\Delta H_1} = -30$, $\log(\overline{N}_1) = -3$), while the trap in the no-flux regime goes unnoticed. From a Bayesian perspective, this represents an observability limitation of the experiment, being the model of one single trap more likely to represent the material given the experimental evidence. Furthermore, it is noted that the presence of measurement noise does not substantially alter the inference outcome. The modes of the posteriors for the trap densities parameters $\overline{\Delta H_1}$ and
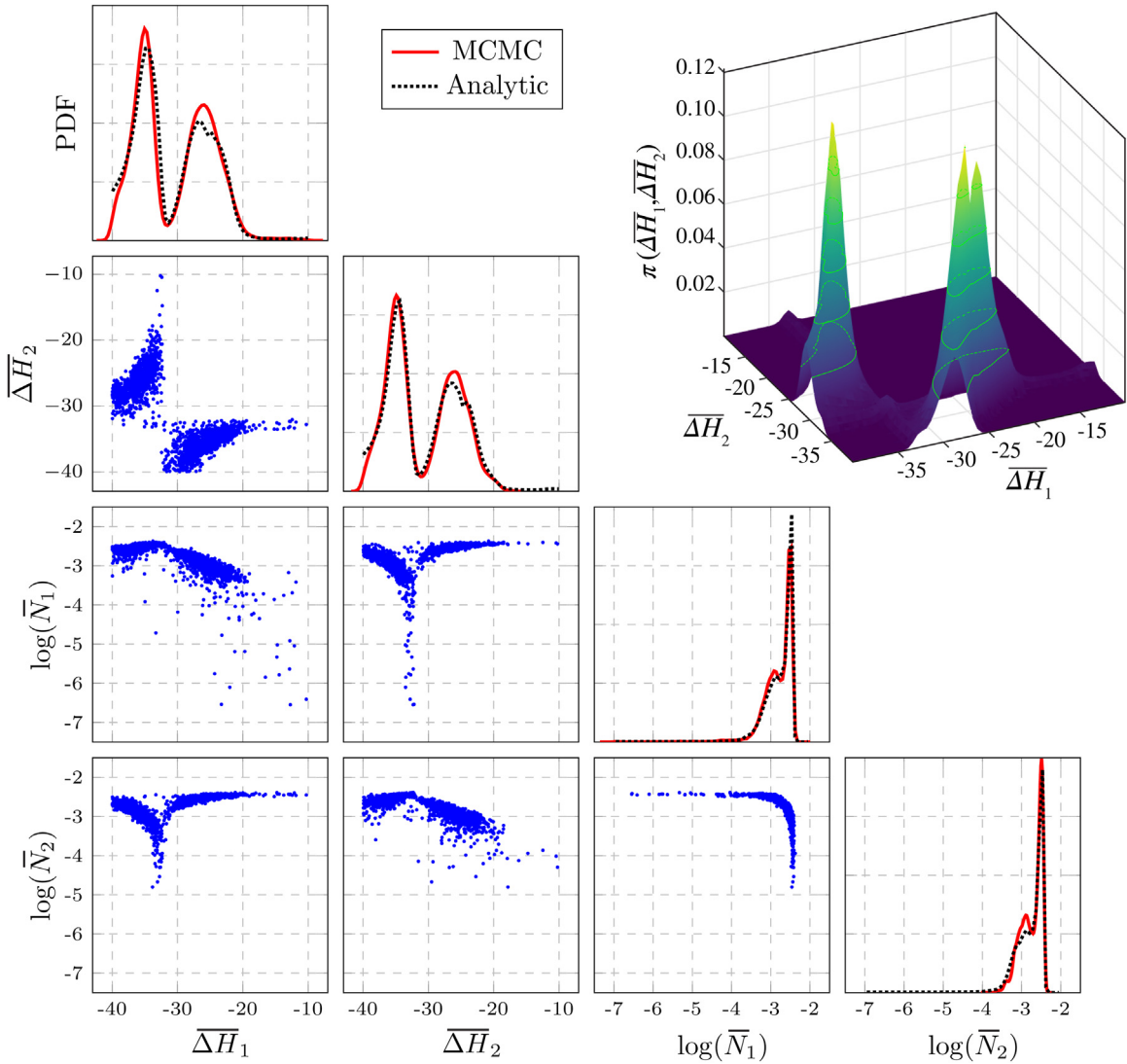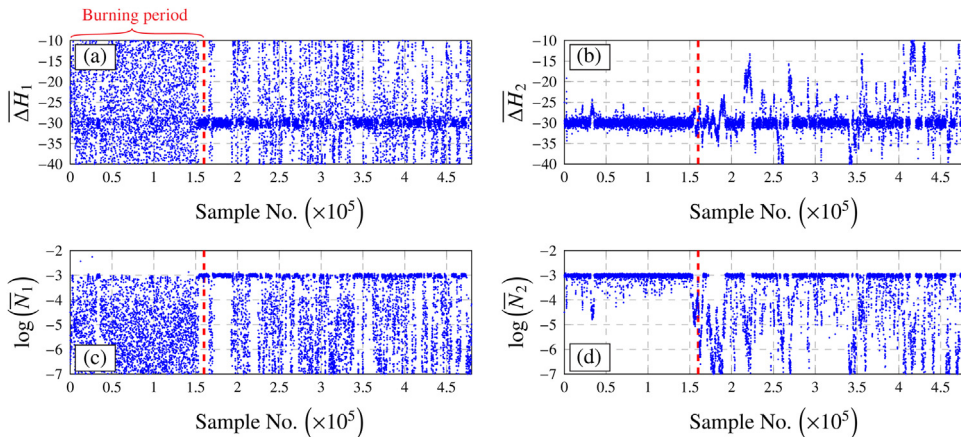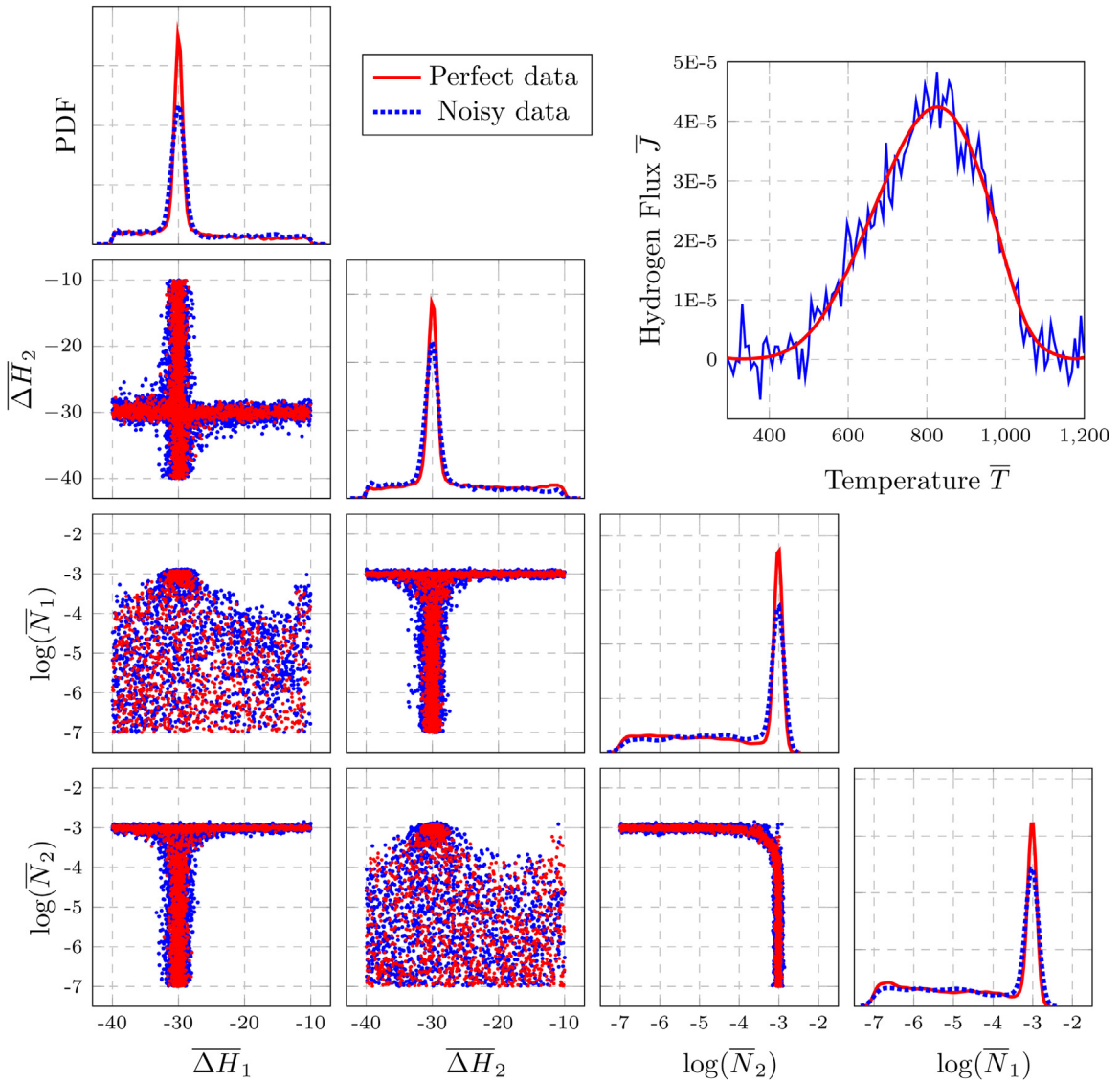
Fig. 9. Bayesian identification results of the trap parameters $\boldsymbol{\theta} = \left(\overline{\Delta H}_1, \overline{\Delta H}_2, \log(\overline{N}_1), \log(\overline{N}_2)\right)$ of TDS Experiment EI. The surface plot in the top right corner corresponds to the marginal PDF over $\left(\overline{\Delta H}_1, \overline{\Delta H}_2\right)$ obtained by numerical integration.

Fig. 10. Markov chains generated by DRAM MCMC of trap parameters $\overline{\Delta H}_1$, $\overline{\Delta H}_2$, $\log(\overline{N}_1)$, $\log(\overline{N}_2)$ for TDS Experiment EII under noise-free data conditions.

Table 6
HDR at 80% and 50% of the PDFs obtained by direct integration and by MCMC for experiment EI.

HDR	$\overline{\Delta H}_1$	$\overline{\Delta H}_2$	$\log(\overline{N}_1)$	$\log(\overline{N}_2)$
80% HDR (MCMC)	$(-38.004,-33.084) \cup (-28.931,-22.606)$	$(-37.833,-32.686) \cup (-28.628,-23.134)$	$(-3.075,-2.418)$	$(-3.059,-2.736) \cup (-2.683,-2.409)$
80% HDR (Analytical)	$(-38.515,-32.253) \cup (-29.121,-22.729)$	$(-38.179,-32.247) \cup (-29.184,-22.991)$	$(-3.069,-2.386)$	$(-3.099,-2.389)$
50% HDR (MCMC)	$(-36.726,-33.659) \cup (-27.398,-24.842)$	$(-36.546,-33.280) \cup (-27.143,-25.164)$	$(-2.925,-2.869) \cup (-2.653,-2.427)$	$(-2.898,-2.868) \cup (-2.637,-2.416)$
50% HDR (Analytical)	$(-36.884,-32.840) \cup (-27.295,-24.816)$	$(-36.680,-32.899) \cup (-27.424,-25.208)$	$(-2.702,-2.406)$	$(-2.703,-2.399)$

Fig. 11. Bayesian identification results of the trap parameters $\boldsymbol{\theta} = \left(\overline{\Delta H}_1, \overline{\Delta H}_2, \log(\overline{N}_1), \log(\overline{N}_2)\right)$ of experiment EII with noise unpolluted (red) and polluted data (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 7
HDR at 80% and 50% of the PDFs obtained from noisy and noise-free data by MCMC for experiment EII.

HDR	$\overline{\Delta H}_1$	$\overline{\Delta H}_2$	$\log(\overline{N}_1)$	$\log(\overline{N}_2)$
80% HDR (noise-free data)	$(-39.872, -22.145)$	$(-34.206, -25.128)$	$(-5.887, -2.927)$	$(-5.346, -2.920)$
80% HDR (noisy data)	$(-39.970, -26.221)$	$(-39.314, -24.378)$	$(-5.410, -2.900)$	$(-5.733, -2.889)$
50% HDR (noise-free data)	$(-31.526, -28.406)$	$(-30.841, -29.199)$	$(-3.886, -2.924)$	$(-3.088, -2.946)$
50% HDR (noisy data)	$(-31.534, -28.698)$	$(-31.576, -28.656)$	$(-3.307, -2.896)$	$(-3.286, -2.883)$

$\overline{\Delta H}_2$ of the noise-free scenario are -29.966 and -30.002, while for the noisy scenario the values -29.989 and -29.961 are obtained, which represents a difference of 0.077% and 0.137%, respectively. On the other hand, for parameters $\log(\overline{N}_1)$ and $\log(\overline{N}_2)$ the modes of the posteriors in the noise-free case are -3.009 and -3.016, whereas in the noisy scenario they take values -3.017 and -3.019, meaning a difference of 0.266% and 0.010%, respectively. This confirms that the proposed approach is robust to the presence of measurement noise. Overall, these results illustrate the potential of the developed approach for model selection and information gain analysis of TDS results, which are left for future developments.

4. Conclusions

This work presents the development of a multi-element PCK meta-model for surrogate model-based Bayesian parameter inference of highly nonlinear engineering models. The proposed metamodel combines adaptive sparse PCE and Kriging metamodelling to attain both global and local prediction capabilities. The optimal order of the polynomials in the PCE is automatically identified by the LAR algorithm. Then, the optimal PCE is inserted into a Kriging predictor as the trend term, while the stochastic term is fitted through GA optimization. With the aim of tackling non-smoothness in the forward model, a simple regular block partitioning approach has been implemented. On this basis, the space domain is split into a discrete number of subsets where local surrogate models are constructed. Then, the global model response is obtained by combining the local metamodels in a piecewise fashion. Finally, the surrogate model is used for Bayesian parameter estimation using a cost-efficient DRAM MCMC with DR and AM capabilities. The effectiveness of the proposed approach has been validated through two benchmark case studies: (i) the analytical Drop-Wave function; (ii) and a PDE for TDS tests. Key findings and contributions of this work include:

- Optimal surrogate models ought to be defined by preliminary parametric analyses accounting for prediction accuracy and computational cost. The latter is particularly critical when performing computationally intense applications such as Bayesian parameter estimation. To this aim, this work has presented a set of error metrics and a methodological discussion through two validation case studies.
- The analysis of the Salomon function has shown that the proposed multi-element PCK model with regular block partitioning provides similar accuracy ($R^2 > 0.99$, NMAE $< 10^{-4}$) as the (classical) PCK approach, while achieving 100 and 10 times shorter evaluation and construction times, respectively. Moreover, the presented results have shown that the proposed method outperforms the SSE technique for the analysis of such a highly nonlinear surface, requiring 20 times fewer samples to achieve a comparable accuracy.
- The size of the ED and the number of domain partitions critically determine the computational cost of the developed sparse PCE-Kriging metamodel. Specifically, the partition of non-smooth problems into a finite set of sub-domains allows the sparse adaptive PCE to eliminate a considerable number of high-order components through LAR, so achieving important savings in the construction of the Kriging model and the evaluation of the resulting metamodel.
- The developed surrogate model-based DRAM MCMC approach allows to conduct fast Bayesian parameter inference. In particular, the proposed approach has been applied to the identification of micro-structural traps in metallic alloys subject to TDS. The hydrogen fluxes obtained in TDS test represent a considerable challenge in surrogate modelling due to the presence of diverse regimes depending on the configuration of the hydrogen traps. In terms of R^2, the proposed approach is capable of reproducing more than 99.9% of the hydrogen diffusion TDS model with computational time savings of 99.3% with respect to the forward numerical model.
- The presented analyses evidence the potential of the developed approach for conducting inverse characterisation of hydrogen-metal interactions. The accuracy of the proposed PCK surrogate model in conjunction with DRAM MCMC opens vast possibilities for future applications in model selection, and information gain analysis of TDS hydrogen desorption tests.

Despite its simplicity, the adopted regular block partitioning model has demonstrated significant performance in terms of computational savings. In this respect, future research will involve the development of more efficient partitioning algorithms that would allow the sampling effort to be localised where the forward model presents greater non-linearities, thus achieving similar accuracies with smaller sample sizes. Another interesting goal for future work consists in the development of multielement surrogate PCK-based models capable of dealing with discontinuities in the response surface.

Data availability

Data will be made available on request.

Acknowledgements

References

[1] Q. Chen, C. Yang, Hybrid algorithm for multi-objective optimization design of parallel manipulators, Appl. Math. Model. 98 (2021) 245–265.
[2] X. Sun, Y.Y. Choi, J.I. Choi, Global sensitivity analysis for multivariate outputs using polynomial chaos-based surrogate models, Appl. Math. Model. 82 (2020) 867–887.
[3] E. García-Macías, I. Venanzi, F. Ubertini, Metamodel-based pattern recognition approach for real-time identification of earthquake-induced damage in historic masonry structures, Autom. Constr. 120 (2020) 103389.

[4] L. Hong, H. Li, K. Peng, A combined radial basis function and adaptive sequential sampling method for structural reliability analysis, Appl. Math. Model. 90 (2021) 375–393.

[5] Y. Yuan, A. Jones, R. Setchfield, C.W. Schwingshackl, Robust design optimisation of underplatform dampers for turbine applications using a surrogate model, J. Sound Vib. 494 (2021) 115528.

[6] J. Stork, M. Friese, M. Zaefferer, T. Bartz-Beielstein, A. Fischbach, B. Breiderhoff, B. Naujoks, T. Tušar, et al., Open issues in surrogate-assisted optimization, in: T. Bartz-Beielstein, et al. (Eds.), High-Performance Simulation-Based Optimization, Springer, 2020, pp. 225–244.

[7] J. Hao, W. Ye, L. Jia, G. Wang, J. Allen, Building surrogate models for engineering problems by integrating limited simulation data and monotonic engineering knowledge, Adv. Eng. Inform. 49 (2021) 101342.

[8] S. Chakraborty, S. Adhikari, R. Ganguli, The role of surrogate models in the development of digital twins of dynamic systems, Appl. Math. Model. 90 (2021) 662–681.

[9] H. Sun, H.V. Burton, H. Huang, Machine learning applications for building structural design and performance assessment: state-of-the-art review, J. Build. Eng. 33 (2021) 101816.

[10] M.J. Asher, B.F.W. Croke, A.J. Jakeman, L.J.M. Peeters, A review of surrogate models and their application to groundwater modeling, Water Resour. Res. 51 (8) (2015) 5957–5973.

[11] B.M. de Gooijer, J. Havinga, H.J.M. Geijselaers, A.H. van den Boogaard, Evaluation of pod based surrogate models of fields resulting from nonlinear FEM simulations, Adv. Model. Simul. Eng. Sci. 8 (1) (2021) 1–33.

[12] L. Zhang, Y. Wu, P. Jiang, S.K. Choi, Q. Zhou, A multi-fidelity surrogate modeling approach for incorporating multiple non-hierarchical low-fidelity data, Adv. Eng. Inform. 51 (2022) 101430.

[13] Y. Shi, Z. Lu, L. Xu, S. Chen, An adaptive multiple-Kriging-surrogate method for time-dependent reliability analysis, Appl. Math. Model. 70 (2019) 545–571.

[14] J.P.C. Kleijnen, Kriging metamodeling in simulation: a review, Eur. J. Oper. Res. 192 (3) (2009) 707–716.

[15] M.D. Buhmann, Radial basis functions, Acta Numer. 9 (2000) 1–38.

[16] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (3) (2004) 199–222.

[17] G.B. Kingston, M. Rajabalinejad, B.P. Gouldby, P.H.A.J.M. Van Gelder, Computational intelligence methods for the efficient reliability analysis of complex flood defence structures, Struct. Saf. 33 (1) (2011) 64–73.

[18] E. Schulz, M. Speekenbrink, A. Krause, A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions, J. Math. Psychol. 85 (2018) 1–16.

[19] R. Schobi, B. Sudret, J. Wiart, Polynomial-chaos-based Kriging, Int. J. Uncertain. Quantif. 5 (2) (2015) 59–63.

[20] R. Alizadeh, J.K. Allen, F. Mistree, Managing computational complexity using surrogate models: a critical review, Res. Eng. Des. 31 (3) (2020) 275–298.

[21] N.V. Queipo, R.T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, P.K. Tucker, Surrogate-based analysis and optimization, Prog. Aerosp. Sci. 41 (1) (2005) 1–28.

[22] J.N. Fuhg, A. Fau, U. Nackenhorst, State-of-the-art and comparative review of adaptive sampling methods for Kriging, Arch. Comput. Methods Eng. 28 (2021) 2689–2747.

[23] S.S. Garud, I.A. Karimi, M. Kraft, Design of computer experiments: a review, Comput. Chem. Eng. 106 (2017) 71–95.

[24] H. Liu, Y.S. Ong, J. Cai, A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design, Struct. Multidiscip. Optim. 57 (1) (2018) 393–416.

[25] O. Le Maître, H.N. Najm, R.G. Ghanem, O.M. Knio, Multi-resolution analysis of wiener-type uncertainty propagation schemes, J. Comput. Phys. 197 (2) (2004) 502–531.

[26] A. Menafoglio, G. Gaetani, P. Secchi, Random domain decompositions for object-oriented Kriging over complex domains, Stoch. Environ. Res. Risk Assess. 32 (12) (2018) 3421–3437.

[27] A. Resmini, J. Peter, D. Lucor, Sparse grids-based stochastic approximations with applications to aerodynamics sensitivity analysis, Int. J. Numer. Methods. Eng. 106 (1) (2016) 32–57.

[28] S.A. Mattis, T. Butler, Enhancing piecewise-defined surrogate response surfaces with adjoints on sets of unstructured samples to solve stochastic inverse problems, Int. J. Numer. Methods. Eng. 119 (10) (2019) 923–940.

[29] Y.S. Ong, P.B. Nair, A.J. Keane, Evolutionary optimization of computationally expensive problems via surrogate modeling, AIAA J. 41 (4) (2003) 687–696.

[30] R.P. Liem, C.A. Mader, J.R.R.A. Martins, Surrogate models and mixtures of experts in aerodynamic performance prediction for aircraft mission analysis, Aerosp. Sci. Technol. 43 (2015) 126–151.

[31] R. Teixeira, M. Nogal, A. O'Connor, Adaptive approaches in metamodel-based reliability analysis: a review, Struct. Saf. 89 (2021) 102019.

[32] J. Foo, X. Wan, G.E. Karniadakis, The multi-element probabilistic collocation method (ME-PCM): error analysis and applications, J. Comput. Phys. 227 (22) (2008) 9572–9595.

[33] C. Rasmussen, Z. Ghahramani, Infinite mixtures of gaussian process experts, Adv. Neural Inf. Process. Syst. 14 (2001) 881–888.

[34] B.A. Konomi, A.A. Hanandeh, M. Pulong, E.L. Kang, Computationally efficient nonstationary nearest-neighbor gaussian process models using data-driven techniques, Environmetrics 30 (8) (2019) e2571.

[35] S. Marelli, P.R. Wagner, C. Lataniotis, B. Sudret, Stochastic spectral embedding, Int. J. Uncertain. Quantif. 11 (2) (2021).

[36] P. Wagner, S. Marelli, B. Sudret, Bayesian model inversion using stochastic spectral embedding, J. Comput. Phys. 436 (2021) 110141.

[37] Y. Huang, C. Shao, B. Wu, J.L. Beck, H. Li, State-of-the-art review on Bayesian inference in structural system identification and damage assessment, Adv. Struct. Eng. 22 (6) (2019) 1329–1351.

[38] W. Liu, X. Luo, F. Huang, M. Fu, Prediction of soil water retention curve using Bayesian updating from limited measurement data, Appl. Math. Model. 76 (2019) 380–395.

[39] J.M. Emery, M.D. Grigoriu, R.V. Field Jr., Bayesian methods for characterizing unknown parameters of material models, Appl. Math. Model. 40 (13–14) (2016) 6395–6411.

[40] P. Hauseux, J.S. Hale, S. Cotin, S.P.A. Bordas, Quantifying the uncertainty in a hyperelastic soft tissue model with stochastic parameters, Appl. Math. Model. 62 (2018) 86–102.

[41] R. Dwivedi, Y. Chen, M.J. Wainwright, B. Yu, Log-concave sampling: metropolis-hastings algorithms are fast!, J. Mach. Learn. Res. 20 (183) (2019) 1–42.

[42] S.H. Cheung, S. Bansal, A new Gibbs sampling based algorithm for Bayesian model updating with incomplete complex modal data, Mech. Syst. Signal Process. 92 (2017) 156–172.

[43] Y. Che, X. Wu, G. Pastore, W. Li, K. Shirvan, Application of Kriging and variational Bayesian Monte Carlo method for improved prediction of doped UO2 fission gas release, Ann. Nucl. Energy 153 (2021) 108046.

[44] P. Del Moral, A. Doucet, A. Jasra, Sequential monte carlo samplers, J. R. Stat. Soc. Ser. B Methodol. 68 (3) (2006) 411–436.

[45] J. Ching, Y.C. Chen, Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging, J. Eng. Mech. 133 (7) (2007) 816–832.

[46] T. Yin, H. Zhu, An efficient algorithm for architecture design of Bayesian neural network in structural model updating, Computer-Aided Civ. Infrastruct. Eng. 35 (4) (2020) 354–372.

[47] F. Schneider, I. Papaioannou, D. Straub, C. Winter, G. Müller, Bayesian parameter updating in linear structural dynamics with frequency transformed data using rational surrogate models, Mech. Syst. Signal Process. 166 (2022) 108407.

[48] W.W. Xing, A.A. Shah, P. Wang, S. Zhe, Q. Fu, R.M. Kirby, Residual Gaussian process: a tractable nonparametric Bayesian emulator for multi-fidelity simulations, Appl. Math. Model. 97 (2021) 36–56.

[49] L. Ierimonti, N. Cavalagli, I. Venanzi, E. García-Macías, F. Ubertini, A transfer Bayesian learning methodology for structural health monitoring of monumental structures, Eng. Struct. 247 (2021) 113089.

[50] A. del Val, O.P. Le Maître, T.E. Magin, O. Chazot, P.M. Congedo, A surrogate-based optimal likelihood function for the Bayesian calibration of catalytic recombination in atmospheric entry protection materials, Appl. Math. Model. 101 (2022) 791–810.

[51] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, J. Comput. Phys. 230 (6) (2011) 2345–2367.

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

[53] H. Haario, M. Laine, A. Mira, E. Saksman, DRAM: efficient adaptive MCMC, Stat. Comput. 16 (4) (2006) 339–354.

[54] K.K. Phoon, J. Ching, Risk and Reliability in Geotechnical Engineering, CRC Press, Boca Raton, FL, USA, 2015.

[55] J. Sacks, W.J. Welch, T.J. Mitchell, H.P. Wynn, Design and analysis of computer experiments, Stat. Sci. 4 (4) (1989) 409–423.

[56] C. Rasmussen, C. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.

[57] H. Cham, E. Reshetnyak, B. Rosenfeld, W. Breitbart, Full information maximum likelihood estimation for latent variable interactions with incomplete indicators, Multivar. Behav. Res. 52 (1) (2008) 12–30.

[58] J. Kleffe, J.N.K. Rao, Estimation of mean square error of empirical best linear unbiased predictors under a random error variance linear model, J. Multivar. Anal. 43 (1) (1992) 1–15.

[59] G. Pellegrini, Polynomial Chaos Expansion with Applications to PDEs, University of Verona, 2014 Ph.D. thesis.

[60] G.K. W. Xiaoliang, Multi-element generalized polynomial chaos for arbitrary probability measures, SIAM J. Sci. Comput. 28 (3) (2006) 901–928.

[61] M. Moustapha, J.M. Bourinet, B. Guillaume, B. Sudret, Comparative study of Kriging and support vector regression for structural engineering applications, ASCE-ASME J. Risk Uncertain. Eng. Syst. A 4 (2) (2018) 04018005.

[62] K. Ritter, Average-Case Analysis of Numerical Problems, Springer, 2000.

[63] W. Wang, R. Tuo, C.F. Jeff Wu, On prediction properties of Kriging: uniform error bounds and robustness, J. Am. Stat. Assoc. 115 (530) (2020) 920–930.

[64] A. Mira, On metropolis-hastings algorithms with delayed rejection, Metron 59 (3–4) (2001) 231–241.

[65] H. Haario, E. Saksman, J. Tamminen, An adaptive metropolis algorithm, Bernoulli 7 (2) (2001) 223–242.

[66] D.C. Stenning, D.A. van Dyk, Bayesian Statistical Methods For Astronomy Part II: Markov Chain Monte Carlo, STATS Astrophysics. EDP Sciences, 2021.

[67] R. Salomon, Re-evaluating genetic algorithm performance under coordinate rotation of benchmark functions. a survey of some theoretical and practical aspects of genetic algorithms, Biosystems 39 (3) (1996) 263–278.

[68] S. Marelli, B. Sudret, Uqlab: a framework for uncertainty quantification in matlab, in: The 2nd International Conference on Vulnerability and Risk Analysis and Management, 2014, pp. 2554–2563.

[69] R.P. Gangloff, Hydrogen-assisted Cracking, in: I. Milne, R. Ritchie, B. Karihaloo (Eds.), Comprehensive Structural Integrity, vol. 6, Elsevier Science, New York, NY, 2003, pp. 31–101.

[70] S.K. Dwivedi, M. Vishwakarma, Hydrogen embrittlement in different materials: a review, Int. J. Hydrog. Energy. 43 (46) (2018) 21603–21616.

[71] W.H. Johnson, On some remarkable changes produced in iron and steel by the action of hydrogen and acids, Proc. R. Soc. Lond. 23 (1875) 168–179.

[72] R. Fernández-Sousa, C. Betegón, E. Martínez-Pañeda, Analysis of the influence of microstructural traps on hydrogen assisted fatigue, Acta Mater. 199 (2020) 253–263.

[73] M. Isfandbod, E. Martínez-Pañeda, A mechanism-based multi-trap phase field model for hydrogen assisted fracture, Int. J. Plast. 144 (2021) 103044.

[74] A. Zafra, Z. Harris, C. Sun, E. Martinez-Pañeda, Comparison of hydrogen diffusivities measured by electrochemical permeation and temperature-programmed desorption in cold-rolled pure iron, J. Nat. Gas Sci. Eng. 98 (2022) 104365.

[75] F.J. Castro, G. Meyer, Thermal desorption spectroscopy (TDS) method for hydrogen desorption characterization (I): theoretical aspects, J. Alloys Compd. 330 (2002) 59–63.

[76] A. Raina, V.S. Deshpande, N.A. Fleck, Analysis of thermal desorption of hydrogen in metallic alloys, Acta Mater. 144 (2018) 777–785.

[77] A. McNabb, P.K. Foster, A new analysis of diffusion of hydrogen in iron and ferrite, Trans. Met. Soc. 227 (3) (1963) 618–627.

[78] R.A. Oriani, The diffusion and trapping of hydrogen in steel, Acta Metall. 18 (1) (1970) 147–157.