

Coordinate Transformation and Polynomial Chaos for the Bayesian Inference of a Gaussian Process with Parametrized Prior Covariance Function

Ihab Sraj^{a,*}, Olivier P. Le Maître^b, Omar M. Knio^{c,d}, Ibrahim Hoteit^{a,d}

^a*Division of Physical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*

^b*LIMSI-CNRS, BP 133, Bt 508, 91403 Orsay Cedex, France*

^c*Department of Mechanical Engineering and Materials Science, Duke University, 144 Hudson Hall, Durham, North Carolina 27708, USA*

^d*Division of Computer, Electrical and Mathematical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*

Abstract

This paper addresses model dimensionality reduction for Bayesian inference based on prior Gaussian fields with uncertainty in the covariance function hyper-parameters. The dimensionality reduction is traditionally achieved using the Karhunen-Loève expansion of a prior Gaussian process assuming covariance function with fixed hyper-parameters, despite the fact that these are uncertain in nature. The posterior distribution of the Karhunen-Loève coordinates is then inferred using available observations. The resulting inferred field is therefore dependent on the assumed hyper-parameters. Here, we seek to efficiently estimate both the field and covariance hyper-parameters using Bayesian inference. To this end, a generalized Karhunen-Loève expansion is derived using a coordinate transformation to account for the dependence with respect to the covariance hyper-parameters. Polynomial Chaos expansions are employed for the acceleration of the Bayesian inference using similar coordinate transformations, enabling us to avoid expanding explicitly the solution dependence on the uncertain hyper-parameters. We demonstrate the feasibility of the proposed method on a transient diffusion equation by inferring spatially-varying log-diffusivity fields from noisy data. The inferred profiles were found closer to the true profiles when including the hyper-parameters' uncertainty in the inference formulation.

Keywords: Karhunen-Loève expansion, dimensionality reduction, Markov Chain Monte Carlo, polynomial chaos, Bayesian inference

1 Introduction

Inverse problems arise in many applications whenever we seek to find some information about a physical system based on some observations. From a computational point of view, a major challenge of inverse problems is their ill-posedness where there is no guarantee that a solution exists, multiple solutions may exist, or even the solution does not depend continuously on the observations. This can be significantly affected by measurement errors, and inferring a suitable solution from noisy observations is an important and challenging topic.

In this paper, we are only concerned with Bayesian approaches to inverse problems. This is motivated by their ability of providing complete posterior statistics and not just a single value for the quantity of interest. The multi-dimensional posterior can be directly explored via Markov Chain Monte Carlo (MCMC). This, however, requires repeated simulations (sometimes hundreds of thousands) of the forward model, once

*Corresponding author

Email address: `ihab.sraj@kaust.edu.sa` (Ihab Sraj)

for every proposed set of parameters of the Markov chain [1]. This practice renders Bayesian methods computationally prohibitive for large-scale applications. Acceleration techniques have been proposed in the literature in which a surrogate model is constructed that requires a much smaller ensemble of forward model runs which is then used in the sampling MCMC step instead at a significantly reduced computational cost. Marzouk *et al.* [2] for instance proposed a spectral projection method that uses spectral expansion of the prior model in Polynomial Chaos (PC) basis. The PC method has been extensively investigated in the literature, and its suitability for large-scale models has been demonstrated in various settings, including ocean [3, 4, 5, 6], tsunami [7, 8], climate modeling [9] and subsurface flow modeling [10].

The PC method has been shown to be efficient for inverse problems involving a limited number of stochastic parameters; yet in some cases the unknown quantity is a spatial or temporal field in which the number of stochastic parameters is quite large. Computational challenges in this case arise in the surrogate model construction as PC suffers from the curse of dimensionality [11]. In addition, convergence is hard to achieve using the Bayesian inference due to the high dimensionality of the posterior. To overcome this numerical issue, Marzouk *et al.* [12] introduced truncated Karhunen-Loève (KL) expansions to parametrize the stochastic field, endowed with a hierarchical Gaussian process prior. The idea is to transform the high-dimensional stochastic forward problem into a smaller problem whose solution captures that of the deterministic forward model over the support of the prior. Galerkin projection on a PC basis was used to seek the solution of the problem, and a reduced-dimensionality surrogate posterior density was constructed that is inexpensive to evaluate.

The Gaussian process prior assumed in Marzouk *et al.* [12] is associated with hyper-parameters that are rarely known in practice. Assuming otherwise renders the quantification of prior uncertainty unrealistic and incomplete. Hierarchical Bayesian inference is proposed in the literature for calibration in presence of uncertain hyper-parameters but is done *a priori* [13]. The method proposed by Marzouk *et al.* [12] does not explicitly consider the effect of length-scales and only includes one hyper-parameter accounting for prior variance. An attempt to extend the method proposed by Marzouk *et al.* [12] for priors with uncertain hyper-parameters has been recently proposed by Tagade and Choi [14]. In their work, a methodology is introduced to obtain a KL expansion of a stochastic process in terms of functions of the hyper-parameters. The prior uncertainty in these hyper-parameters was expanded in a PC basis, and Galerkin projection was used to evaluate PC coefficients of the surrogate model. The hyper-parameters hence become part of the inference problem and are estimated from the observations.

This paper proposes an extension of the method of Marzouk *et al.* [12] that is also an alternative to Tagade and Choi method [14]. Our proposed method explores the origin of the KL expansion where it is based on the eigen-functions and eigen-values of a given covariance function. These eigen-functions form a basis in a space dictated by the covariance hyper-parameters. Our method utilizes a change of basis methodology and therefore transforming the KL expansion based on one certain set of hyper-parameters into another. A fundamental distinction of the present work is that we avoid constructing a PC expansion for the uncertain hyper-parameters, and instead use the PC expansion constructed for a reference set of hyper-parameters and apply transformations to obtain PC expansion for any another set of hyper-parameters. The advantage of the proposed method is that the dimensionality of the PC expansion is not augmented by the number of hyper-parameters of the covariance function. Also, our method avoids cases when the hyper-parameters have complex distributions and PC bases may not even exist.

To outline the proposed developments, we start in Section 2 by providing a statistical formulation of the inverse problem based on Bayesian inference. Section 3 then presents the KL expansion and its generalization to account for uncertain hyper-parameters by means of change of basis. Section 4 describes the role of PC in Bayesian inference acceleration. In Section 5, numerical results for the calibration of a one-dimensional toy problem are presented and Section 6 concludes the paper with a summary of the results, discussion and conclusion.

2 Bayesian Inference

Bayesian inference is a statistical approach to inverse problems that has gained much interest in different applications including ocean [15, 16, 4], climate [17] and geophysical [1] modeling. We review the Bayesian

approach briefly below and discuss its implementation to our problem. Our objective is to infer a deterministic field $m(\mathbf{x})$, for some $\mathbf{x} \in D$, from a finite set of $N_o \geq 1$ observations $\mathbf{d} \in \mathbb{R}^{N_o}$. We consider situations where the observations \mathbf{d} are not direct measurements of $m(\mathbf{x})$, but are derived quantities that can be predicted using a model-problem (typically a set of partial differential equations), often called the forward model, relating the $m(\mathbf{x})$ to the model predictions: $m(\mathbf{x}) \mapsto \mathbf{u}(m) \in \mathbb{R}^{N_o}$. The Bayesian formula updates our prior knowledge of the m introducing an error model for the discrepancy between the model predictions $\mathbf{u}(m)$ and the observations \mathbf{d} ; the Bayes' rule is expressed as [18]:

$$p(m, \sigma_o^2 | \mathbf{d}) \propto p(\mathbf{d} | m, \sigma_o^2) p_m(m) p_o(\sigma_o^2), \quad (1)$$

where $p(\mathbf{d} | m, \sigma_o^2)$ is the likelihood of the observations, given m and σ_o^2 the error model hyper-parameter with prior $p_o(\sigma_o^2)$, and $p_m(m)$ is the field's prior. For simplicity, an unbiased additive Gaussian error model will be considered,

$$\boldsymbol{\epsilon} \doteq \mathbf{d} - \mathbf{u}(m), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_o^2 I_{N_o}), \quad (2)$$

where $\mathcal{N}(0, \sigma_o^2 I_{N_o})$ denotes the centered multivariate Gaussian distribution with diagonal covariance $\sigma_o^2 I_{N_o}$. In other words, the errors in the observations are assumed independent. For this choice, the likelihood becomes

$$p(\mathbf{d} | m, \sigma_o^2) = \prod_{i=1}^{N_o} p_\epsilon(d_i - u_i(m), \sigma_o^2), \quad p_\epsilon(x, \sigma_o^2) \doteq \frac{1}{\sqrt{2\pi\sigma_o^2}} \exp\left[-\frac{x^2}{2\sigma_o^2}\right]. \quad (3)$$

The main difficulties with the posterior above are the infinite dimensional character of $m(\mathbf{x})$ and its prior definition. A discretization of $m(\mathbf{x})$ is needed to perform the inference and setting a finite dimensional prior distribution. If $m(\mathbf{x})$ is endowed with a Gaussian prior, it is fully characterized by its second-order properties, namely its mean $\boldsymbol{\mu}(\mathbf{x})$ and covariance function $\mathcal{C}(\mathbf{x}, \mathbf{x}')$. From $\boldsymbol{\mu}$ and \mathcal{C} , one can rely on truncated Karhunen-Loève (KL) decomposition to represent $m(\mathbf{x})$ as a convergent series involving a finite set of KL coordinates (or expansion coefficients) η_k , $k = 1, \dots, K$ as discussed in Section 3. The inference problem can then be reformulated for the vector $\boldsymbol{\eta}$ of coordinates η_k , leading to

$$p(\boldsymbol{\eta}, \sigma_o^2 | \mathbf{d}) \propto p(\mathbf{d} | \boldsymbol{\eta}, \sigma_o^2) p_\eta(\boldsymbol{\eta}) p_o(\sigma_o^2), \quad (4)$$

where $p_\eta(\boldsymbol{\eta}) = \exp(-\boldsymbol{\eta}^T \boldsymbol{\eta} / 2) / (2\pi)^{K/2}$ is the Gaussian prior of the KL coordinates.

As discussed below, the covariance function $\mathcal{C}(\mathbf{x}, \mathbf{x}')$ is generally selected on the basis of limited knowledge and the inference of $m(\mathbf{x})$ can be improved by introducing additional hyper-parameters \mathbf{q} in the definition of \mathcal{C} i.e. $\mathcal{C}(\mathbf{x}, \mathbf{x}', \mathbf{q})$. This yields the generalized Bayes' formula,

$$p(\boldsymbol{\eta}, \mathbf{q}, \sigma_o^2 | \mathbf{d}) \propto p(\mathbf{d} | \boldsymbol{\eta}, \mathbf{q}, \sigma_o^2) p_\eta(\boldsymbol{\eta}) p_q(\mathbf{q}) p_o(\sigma_o^2), \quad (5)$$

where p_q is the prior distribution of the covariance parameters. For the case of covariance with hyper-parameters \mathbf{q} , the likelihood takes the following general form,

$$p(\mathbf{d} | \boldsymbol{\eta}, \mathbf{q}, \sigma_o^2) = \prod_{i=1}^{N_o} p_\epsilon(d_i - u_i(\boldsymbol{\eta}, \mathbf{q}), \sigma_o^2), \quad (6)$$

with p_ϵ defined in Eq. (3), and $u_i(\boldsymbol{\eta}, \mathbf{q})$ being a short-hand notation for the model prediction $u_i(m)$ with $m(\boldsymbol{\eta}, \mathbf{q})$ the reconstructed field.

Inferring the field then amounts to sampling the posterior of KL coordinates $\boldsymbol{\eta}$ and hyper-parameters \mathbf{q} . In general, the sample space is high-dimensional and suitable computational strategy is the Markov chain Monte Carlo (MCMC) method. In this work, we rely on an adaptive Metropolis-Hastings MCMC algorithm [19, 20] to accurately and efficiently sample the posterior distribution $p(\boldsymbol{\eta}, \mathbf{q}, \sigma_o^2 | \mathbf{d})$. This requires the evaluation of the posterior (up to its normalization constant) for multiple sample values of $(\boldsymbol{\eta}, \mathbf{q}, \sigma_o^2)$. The computational flow-chart for an evaluation of the posterior is schematically illustrated in Figure 1. Briefly,

given a sample value of \mathbf{q} , the dominant KL modes of $\mathcal{C}(\mathbf{x}, \mathbf{x}', \mathbf{q})$ are computed and the corresponding field $m(\mathbf{x})$ is constructed using the sampled value of $\boldsymbol{\eta}$. This field is fed into the solver to compute the model predictions $\mathbf{u}(\boldsymbol{\eta}, \mathbf{q})$ which are used, together with the sample value of the model error parameter σ_0^2 , to successively compute the likelihood and finally the posterior.

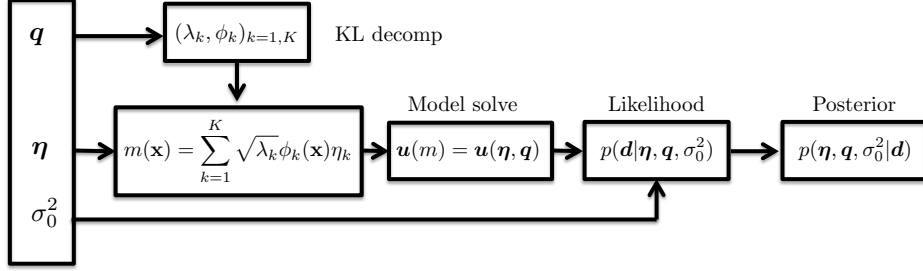


Figure 1: Flow-chart for the evaluation of the posterior distribution in the inference problem.

In general, the most computationally demanding part for sampling the posterior is the computation of the model predictions, given $(\boldsymbol{\eta}, \mathbf{q})$ (that is the realization of the field). This is particularly the case when the predictions involve the solution of partial differential equations. This computational cost motivates the substitution of $\mathbf{u}(\boldsymbol{\eta}, \mathbf{q})$ with a polynomial surrogate model $\tilde{\mathbf{u}}(\boldsymbol{\eta}, \mathbf{q})$, whose evaluation is inexpensive compared to the solution of the complete model. The surrogate is constructed offline and subsequently used on-line when running the MCMC algorithm. Specifically, the likelihood of the observations is approximated using

$$p(\mathbf{d}|\boldsymbol{\eta}, \mathbf{q}, \sigma_0^2) = \prod_{i=1}^{N_o} p_\epsilon(d_i - u_i(\boldsymbol{\eta}, \mathbf{q}), \sigma_0^2) \approx \prod_{i=1}^{N_o} p_\epsilon(d_i - \tilde{u}_i(\boldsymbol{\xi}(\boldsymbol{\eta}, \mathbf{q})), \sigma_0^2), \quad (7)$$

where, as mentioned previously, $\tilde{\mathbf{u}}(\boldsymbol{\xi})$ is a polynomial and $\boldsymbol{\xi} : (\boldsymbol{\eta}, \mathbf{q}) \mapsto \boldsymbol{\xi}(\boldsymbol{\eta}, \mathbf{q})$ is an explicit change of coordinates. Construction of the surrogate model for the predictions is detailed in the next two sections; Section 3 introduces the \mathbf{q} -dependent coordinate transformation, while the polynomial approximation $\tilde{\mathbf{u}}(\boldsymbol{\xi})$ is discussed in Section 4, together with the resulting surrogate-based sampling scheme.

3 Coordinate transformation for Uncertain Covariance Functions

3.1 Karhunen-Loève expansion

Let $D \subset \mathbb{R}^d$, $d \geq 1$, be a bounded domain, and denote $X \doteq L^2(D)$ equipped with inner product $(\cdot, \cdot)_X$ and norm $\|\cdot\|_X$:

$$u \in X \Leftrightarrow \|u\|_X < \infty, \quad \|u\|_X^2 = (u, u)_X = \int_D |u(\mathbf{x})|^2 d\mathbf{x}. \quad (8)$$

Consider a real-valued stochastic process $M(\mathbf{x}, \omega)$ with mean $\mu(\mathbf{x})$ and continuous covariance function $\mathcal{C}(\mathbf{x}, \mathbf{x}')$ on $D \times D$; ω is a random event belonging to a sample space Ω of a probability space (Ω, Σ, P) . The covariance function is defined as

$$\mathcal{C}(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(M(\mathbf{x}, \cdot) - \mu(\mathbf{x}))(M(\mathbf{x}', \cdot) - \mu(\mathbf{x}'))], \quad (9)$$

where \mathbb{E} denotes the expectation operator. The covariance function \mathcal{C} is symmetric positive semi-definite and thus by Mercer's theorem [21] it has the following spectral decomposition:

$$\mathcal{C}(\mathbf{x}, \mathbf{x}') = \sum_{k=1}^{\infty} \lambda_k \phi_k(\mathbf{x}) \phi_k(\mathbf{x}'), \quad (10)$$

where the λ_k and $\phi_k(\mathbf{x})$ are the eigen-values and associated (normalized) eigen-functions of the linear operator corresponding to the covariance function \mathcal{C} ; they satisfy the Fredholm equation of the second kind:

$$\int_D \mathcal{C}(\mathbf{x}, \mathbf{x}') \phi_k(\mathbf{x}') dx = \lambda_k \phi_k(\mathbf{x}), \quad \|\phi_k\|_X = 1. \quad (11)$$

The eigen-values λ_k are real and countable and the eigen-functions $\phi_k(\mathbf{x})$ are continuous and constitute an orthonormal basis in $L^2(D)$. Ordering the eigen-values in a decreasing sequence $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, the truncated Karhunen-Loève (KL) expansion M_K of M is given by [22]

$$M(\mathbf{x}, \omega) \approx M_K(\mathbf{x}, \omega) \doteq \mu(\mathbf{x}) + \sum_{k=1}^K \sqrt{\lambda_k} \phi_k(\mathbf{x}) \eta_k(\omega), \quad (12)$$

where K is the number of expansion terms retained in the spectral approximation. The stochastic coefficients

$$\eta_k(\omega) = (M(\mathbf{x}, \omega) - \mu(\mathbf{x}), \phi_k(\mathbf{x}))_X, \quad (13)$$

are mutually uncorrelated random variables with zero mean and unit variance, such that $\mathbb{E}[\eta_k \eta_{k'}] = \delta_{kk'}$. Under the assumption that M is a Gaussian Process (\mathcal{GP}) denoted by $M \sim \mathcal{GP}(\mu, \mathcal{C})$, the η_k 's are Gaussian and also independent. The truncated KL expansion is optimal in the mean square sense, meaning that of all possible K -term expansions, the M_K in Eq. (12) with λ_k and $\phi_k(\mathbf{x})$ satisfying Eq. (11) minimizes the mean-squared error in the approximation of M [22]. While it is known that the KL decomposition of M converges uniformly as $K \rightarrow \infty$ [23], the truncation error has implicit dependence on the covariance function \mathcal{C} .

The KL expansion is often employed to reduce the dimensionality in inverse problems, considering the expansion coefficients $\eta_{k=1, \dots, K}$ in Eq. (12) as reduced coordinates for the field $m(\mathbf{x})$ to be inferred from the collected observations [24]. In the Bayesian framework, this amounts to the determination of the posterior distribution of the expansion coefficients vector $\boldsymbol{\eta}$, which can be sampled or analyzed to estimate the characteristics of the field $m(\mathbf{x})$ (in particular, retrieving the median, MAP value, confidence intervals, ...). However, the posterior and so the inferred field m have implicit dependencies on the assumed prior covariance structure. This point has motivated the introduction of parametrized covariance families, as discussed in the following section, where the covariance parameters are treated as hyper-parameters in the inference procedure.

3.2 Covariance function with uncertain hyper-parameters

From now on, we assume the prior of $m(\mathbf{x})$ to be Gaussian and so completely characterized by its mean $\mu(\mathbf{x})$ and covariance function \mathcal{C} . However, in many applications, not all the aspects of the covariance function are well-known *a priori*. The stationarity of the covariance function can be easily determined and confirmed, yet, we have a large uncertainty in the other characteristics such as the values of the hyper-parameters. An example of a parametrized covariance function is

$$\mathcal{C}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^T \mathbf{M}(\mathbf{x} - \mathbf{x}')\right) + \sigma_d^2 \mathbf{x}^T \mathbf{x}' + \sigma_b^2 + \sigma_n^2 \delta_{pq} \quad (14)$$

where \mathbf{M} is a symmetric positive definite matrix. The covariance hyper-parameters \mathbf{M} , σ_f^2 , σ_b^2 , σ_d^2 , σ_n^2 are usually not exactly known *a priori* and should be treated as uncertain quantities. For many covariance functions it is easy to interpret the meaning of the hyper-parameters, which is of great importance when trying to understand the data. Traditionally, the hyper-parameters are estimated using Gaussian Process Regression (GPR) before inferring the model parameters [13]. To this end, a set of possibly noisy observations of the field $m(\mathbf{x})$ are used to perform stochastic interpolation of static data collected at few locations and maximize the marginal likelihood function using Bayesian inference or optimization techniques. Optimal values of the inferred hyper-parameters are then used in the covariance function and KL expansion is applied

as described in Eq. (12). The uncertainty bound can be estimated using GPR but is usually not considered in the expansion due to the complexity of the resulting model. This paper addresses the uncertainty in the hyper-parameters of covariance models. Specifically we develop a formulation that enables inferring the covariance function hyper-parameters along with the KL stochastic coordinates η_k . The formulation is based on basis transformations as described below.

3.3 Stochastic coordinate transformation

Without loss of generality, we assume that the stochastic prior process M is centered ($\mu(\mathbf{x}) = 0$) and has a parametrized covariance function $\mathcal{C}(\mathbf{x}, \mathbf{x}', \mathbf{q})$ defined by a random vector $\mathbf{q} \subset \mathbb{R}^h$ of hyper-parameters (h is the number of hyper-parameters, *e.g.* $\mathbf{q} = \{\mathbf{M}, \sigma_f^2, \sigma_b^2, \sigma_d^2, \sigma_n^2\}$ for the example in Eq. (14)), with joint density p_q . Because of the dependence of the covariance function on \mathbf{q} , the KL expansion of M in Eq. (12) becomes:

$$M_K(\mathbf{x}, \omega, \mathbf{q}) = \sum_{k=1}^K \sqrt{\lambda_k(\mathbf{q})} \phi_k(\mathbf{x}, \mathbf{q}) \eta_k(\omega), \quad \int_D \mathcal{C}(\mathbf{x}, \mathbf{x}', \mathbf{q}) \phi_k(\mathbf{x}', \mathbf{q}) dx = \lambda_k(\mathbf{q}) \phi_k(\mathbf{x}, \mathbf{q}). \quad (15)$$

To simplify the notation, we drop the \mathbf{x} and \mathbf{x}' dependence and introduce the scaled eigen-functions $\Phi_k(\mathbf{q})$:

$$\Phi_k(\mathbf{q}) \doteq \sqrt{\lambda_k(\mathbf{q})} \phi_k(\mathbf{q}), \quad \text{so} \quad M_K(\omega, \mathbf{q}) = \sum_{k=1}^K \Phi_k(\mathbf{q}) \eta_k(\omega). \quad (16)$$

We further assume the continuity of the scaled eigen-functions Φ_k with respect to \mathbf{q} , in the sense (see [25]) $\exists D_k(\mathbf{q}) > 0$, $\|\Phi_k(\mathbf{q}) - \Phi_k(\mathbf{q} + \delta\mathbf{q})\|_X^2 \leq D_k(\mathbf{q}) \|\delta\mathbf{q}\|_{\ell_h^2}^2$, and $\sum_{k=1}^K D_k(\mathbf{q}) \doteq D^{(K)}(\mathbf{q}) < \infty$ uniformly, so that

$$\mathbb{E} [\|M_K(\mathbf{q}) - M_K(\mathbf{q} + \delta\mathbf{q})\|_X^2] \leq D_K(\mathbf{q}) \|\delta\mathbf{q}\|_{\ell_h^2}^2. \quad (17)$$

In practice, when decomposing a covariance function $\mathcal{C}(\mathbf{q})$, the normalized eigen-functions are defined up to a factor of ± 1 . To ensure the \mathbf{q} -continuity of the $\phi_k(\mathbf{q})$'s, we have to select a consistent orientation of eigen-functions. A possibility, followed in this work, is to define the orientation of the eigen-functions with respect to a reference set of eigen-functions $\{\phi_k^r, k = 1, \dots, K\}$, *e.g.* the reference set defined below, such that $(\phi_k(\mathbf{q}), \phi_k^r)_X$ has a constant sign for all \mathbf{q} [26]. The dependence of the eigen-functions on the hyper-parameter \mathbf{q} is further illustrated in Section 3.4 below.

Let \mathcal{C}^r be a covariance function representative of the \mathbf{q} -dependent covariance function $\mathcal{C}(\mathbf{q})$ of M . As investigated below, a possible choice for \mathcal{C}^r can be

$$\mathcal{C}^r = \bar{\mathcal{C}} \doteq \int \mathcal{C}(\mathbf{q}) p_q(\mathbf{q}) d\mathbf{q}, \quad (18)$$

that is the \mathbf{q} -averaged of $\mathcal{C}(\mathbf{q})$, or a particular realization of \mathcal{C} corresponding to a deterministic value \mathbf{q}^r of the random parameters (*e.g.* nominal values obtained using GPR [13]). We denote ϕ_k^r the ordered and normalized eigen-vectors of \mathcal{C}^r . Note that $\{\phi_k^r, k = 1, 2, \dots, \infty\}$ is an orthonormal basis of X ; as a result, any scaled eigen-function $\Phi_k(\mathbf{q})$ can be expressed in this basis:

$$\Phi_k(\mathbf{q}) = \sum_{k'=1}^{\infty} b_{kk'}(\mathbf{q}) \phi_{k'}^r, \quad b_{kk'}(\mathbf{q}) = (\phi_k^r, \Phi_{k'}(\mathbf{q}))_X. \quad (19)$$

The continuity of the scaled eigen-functions implies the continuity of the projection coefficients $b_{kk'}(\mathbf{q})$. For computational purposes, the expansion in Eq. (19) needs to be truncated to the first K^r terms. Without loss of generality we shall use in the following $K^r = K$, allowing for convergence analysis with respect to a single parameter K .

Further, the change of basis gives:

$$M_K(\omega, \mathbf{q}) = \sum_{k=1}^K \Phi_k(\mathbf{q}) \eta_k(\omega) \approx \sum_{k=1}^K \left(\sum_{k'=1}^K b_{kk'}(\mathbf{q}) \phi_{k'}^r \right) \eta_k(\omega) = \sum_{k=1}^K \phi_k^r \hat{\eta}_k(\omega, \mathbf{q}), \quad (20)$$

where we have denoted

$$\hat{\eta}_k(\omega, \mathbf{q}) = \sum_{k'=1}^K b_{k'k}(\mathbf{q}) \eta_{k'}(\omega). \quad (21)$$

The transformation shows that the \mathbf{q} -dependence of \mathcal{C} can be translated into an expansion M_K with \mathbf{q} dependent scaled eigen-functions, see Eq. (16), or approximated by a \mathbf{q} -dependent linear transformation of the random variables in Eq. (20). Specifically, denoting the latter approximation \hat{M}_K we have the approximations

$$M(\omega, \mathbf{q}) \approx M_K(\omega, \mathbf{q}) = \sum_{k=1}^K \Phi_k(\mathbf{q}) \eta_k(\omega) \approx \hat{M}_K(\omega, \mathbf{q}) = \sum_{k=1}^K \phi_k^r \hat{\eta}_k(\omega, \mathbf{q}), \quad (22)$$

with $\hat{\eta}_k$ related to the η_k 's by Eq. (21).

We observe that $\hat{\eta}_k(\omega, \mathbf{q})$ is a linear combination of standard Gaussian random variables, so it is also Gaussian (with zero mean). However, the $\hat{\eta}_k(\omega, \mathbf{q})$ are generally correlated. The change of random coordinates in Eq.(21) can be cast in matrix form:

$$\hat{\boldsymbol{\eta}}(\omega, \mathbf{q}) = \mathcal{B}(\mathbf{q}) \boldsymbol{\eta}(\omega). \quad (23)$$

The covariance matrix for the random coefficients $\hat{\boldsymbol{\eta}}$, denoted $\Sigma^2(\mathbf{q})$, can be expressed as

$$\Sigma^2(\mathbf{q}) = \mathbb{E} [\hat{\boldsymbol{\eta}}(\mathbf{q}) \hat{\boldsymbol{\eta}}^t(\mathbf{q})] = \mathcal{B}(\mathbf{q}) \mathcal{B}^t(\mathbf{q}). \quad (24)$$

We shall assume that $\Sigma^2(\mathbf{q})$ is invertible (for almost every \mathbf{q}); a sufficient condition is that $\Phi_{1 \leq k \leq K}(\mathbf{q})$ is not orthogonal to $\text{span}\{\phi_1^r, \dots, \phi_K^r\}$. In addition, the conditional distribution of $\hat{\boldsymbol{\eta}}$, given \mathbf{q} , is

$$p_{\hat{\boldsymbol{\eta}}}(\hat{\boldsymbol{\eta}}|\mathbf{q}) = \frac{1}{\sqrt{2\pi^K |\Sigma^2(\mathbf{q})|}} \exp \left[-\frac{\hat{\boldsymbol{\eta}}^t (\Sigma^2)^{-1}(\mathbf{q}) \hat{\boldsymbol{\eta}}}{2} \right], \quad (25)$$

where $|\Sigma^2(\mathbf{q})|$ is the determinant of $\Sigma^2(\mathbf{q})$.

3.4 Example

We now provide a brief illustration of the convergence of the error in the approximation of $M(\mathbf{q})$. To this end, we consider $D = [0, 1]$ and a centered Gaussian process M with covariance function

$$\mathcal{C}(x, x', \mathbf{q}) = \sigma_f^2 \exp \left(-\frac{(x - x')^2}{2l^2} \right), \quad (26)$$

with hyper-parameter vector $\mathbf{q} = \{\sigma_f^2, l\}$. In this case, only the correlation length l affects the shape of the eigen-functions, while the process variance σ_f^2 simply scales the eigen-values. Therefore, we fix $\sigma_f^2 = 0.5$ through-out the section and assume uncertainty in l only, that is $\mathbf{q} = \{l\}$. Specifically, we assume the hyper-parameter l to have a uniform distribution in the range $[0.1, 1]$. It is important to note that the number of KL modes needed for convergence highly depends on the hyper-parameter l . In particular, if M has small-scale features (small l) a large number of KL modes will be needed.

For the selection of the reference covariance function, we contrast the choice $\mathcal{C}^r = \mathcal{C}(l^r)$, for several values $l^r \in [0.1, 1]$, with the case $\mathcal{C}^r = \bar{\mathcal{C}}$. The KL decompositions are numerically approximated with Galerkin piecewise constant modes over a uniform grid having $N = 128$ elements in space. Figure 2 compares in the left plot the considered reference covariance functions \mathcal{C}^r and in the right plot the respective decay rates with k of their eigen-values λ_k^r . When using $\mathcal{C}(l^r)$, it is seen that the smaller l^r the slowest the decay rate,

as expected, whereas for the \mathbf{q} -averaged covariance $\bar{\mathcal{C}}$ the decay rate is asymptotically similar (but with a lower magnitude) to the lowest l^r in the uncertainty range. Also, note that $\bar{\mathcal{C}}$ is evidently not Gaussian.

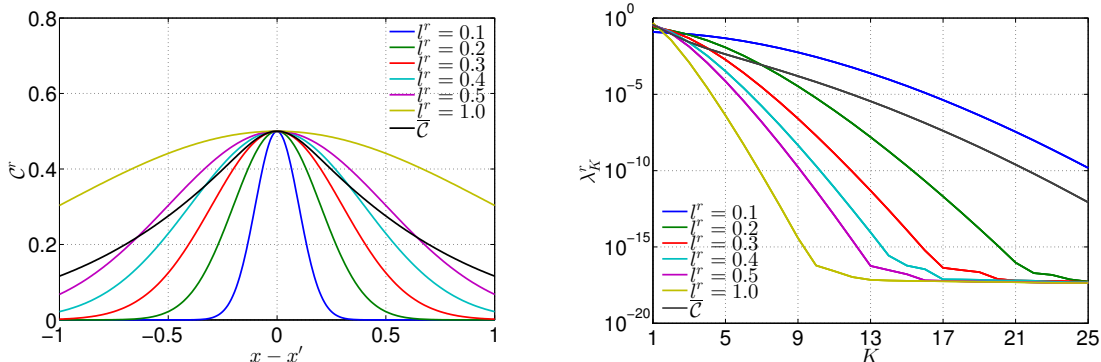


Figure 2: (Left) Reference covariance functions $\mathcal{C}^r = \mathcal{C}(l^r)$ for different values of l^r , as indicated. Also plotted is the \mathbf{q} -averaged covariance $\bar{\mathcal{C}}$. (Right) Spectra of the corresponding eigen-values decay with K .

To quantify the error in the approximation of $M(\omega, \mathbf{q})$ by the proposed transformation method, we introduce the following relative error measure

$$\epsilon_M(K, \mathbf{q}) = \frac{\|M(\mathbf{q}) - \hat{M}_K(\mathbf{q})\|_{L^2(\Omega, D)}}{\|M(\mathbf{q})\|_{L^2(\Omega, D)}}, \quad (27)$$

where

$$\|U\|_{L^2(\Omega, D)}^2 \doteq \mathbb{E}[(U, U)_X]. \quad (28)$$

The error $\epsilon_M(K, \mathbf{q})$ integrates both the truncation error in approximating M with M_K , and the subsequent projection error of M_K into the space of reference modes ϕ_k^r . The error ϵ_M is estimated by means of Monte Carlo sampling where realizations of M are generated given \mathbf{q} ; these realizations are projected on the K -dimensional dominant space of $\mathcal{C}(\mathbf{q})$ in order to compute the coordinates η_k (see Eq. (13)) which are transformed using Eq. (21) to obtain the corresponding realizations of \hat{M}_K . Observe also that $\|M(\mathbf{q})\|_{L^2(\Omega, D)} = \sigma_f$. Finally, the local (squared) error $\epsilon_M^2(K, \mathbf{q})$ can be averaged over \mathbf{q} to yield the averaged error, which we denote $E_M(K)$.

The mean square error $E_M(K)$ is shown in the left plot of Figure 3. Plotted are curves for different reference bases: using $\mathcal{C}^r = \mathcal{C}(l^r)$ with selected correlation lengths l^r within $[0.1, 1.0]$, and the \mathbf{q} -averaged covariance function $\bar{\mathcal{C}}$. A first comment from these curves is that the error decreases as K increases as expected. However, for $l^r > 0.1$, the error $E_M(K)$ stagnates as K increases when using $\mathcal{C}^r = \mathcal{C}(l^r)$. The stagnation occurs at lower K when l^r increases. This stagnation can be explained from the spectra reported in Figure 2 which shows that when using $l^r > 0.1$ the magnitude of λ_k^r quickly decays with k to zero machine precision, such that subsequent modes are not correctly estimated and cannot provide a suitable projection basis. To further illustrate the effect of finite numerical accuracy, we provide in Figure 4 plots of eigenfunctions $\phi_k(x, l)$ for selected k and $(x, l) \in D \times [0.1, 1]$. It is seen that for $k = 1, 4$ and 7 , the dependence on l of the numerical eigen-functions is smooth. In contrast, for $k = 10$ (resp. 13 and 19) the computed eigen-functions are seen to be noisy for $l \gtrsim 0.9$ (resp. $l \gtrsim 0.5$ and 0.25) because of finite numerical accuracy. Clearly, this indicates that under-resolved modes could be disregarded and that the reference basis should include only modes with indices k such that λ_k^r/λ_1^r remains in achievable accuracy ($\approx 10^{-16}$ for double precision). To keep the analysis simple, and because our approach is in fact robust to under-resolved modes, we continue in the following to compare for the same K the different choices of reference covariance functions. Note also that for the reference basis using the shortest correlation length, $l^r = 0.1$, and the \mathbf{q} -averaged covariance, this numerical issue has not yet emerged for the range of considered K , and the corresponding errors decay

monotonically up to $K = 25$. In addition, it is seen that the error curve corresponding to the \mathbf{q} -averaged reference covariance function $\bar{\mathcal{C}}$ has the lowest approximation error $E_M(K)$ for all K . This is not a surprise since by construction this choice uses eigen-functions ϕ_k^r spanning the optimal subspace to represent $M(\omega, \mathbf{q})$ when \mathbf{q} varies with law $p_q(\mathbf{q})$. In fact, finding the K -term expansion minimizing the \mathbf{q} -averaged mean square error approximation of $M \sim \mathcal{GP}(0, \mathcal{C}(\mathbf{q}))$ amounts precisely to the decomposition of $\bar{\mathcal{C}}$. In other words, if using the K dominant eigen-modes of $\mathcal{C}^r = \bar{\mathcal{C}}$ to construct the reference basis is non optimal to represent $M(\mathbf{q})$ for any value \mathbf{q} (obviously for each \mathbf{q} the optimal choice is the eigen-modes of $\mathcal{C}(\mathbf{q})$), there is no better choice on average over \mathbf{q} .

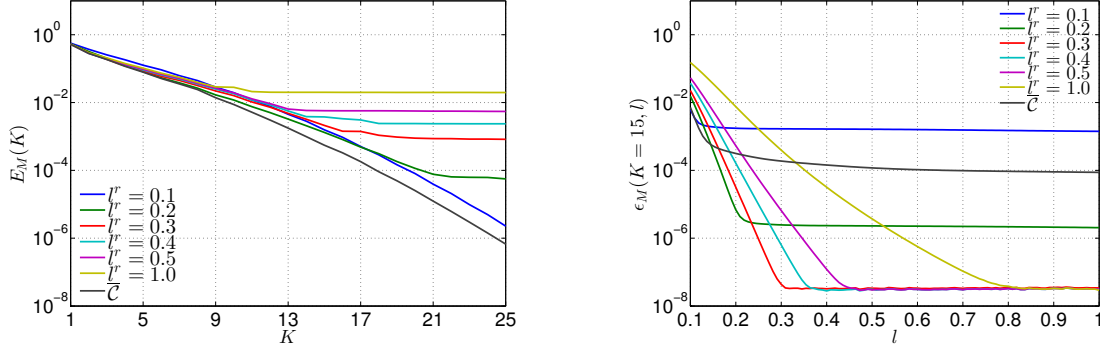


Figure 3: (Left) Error $E_M(K)$ in approximating the Gaussian Process M by \hat{M}_K for different reference covariance functions based on selected correlation lengths l^r as indicated. Also plotted are results obtained with $\bar{\mathcal{C}}$. (Right) Relative error $\epsilon_M(K = 15, l)$ for the same cases as in the left plot.

To better appreciate the behavior of the error with the hyper-parameter l in the present example, the right plot in Figure 3 reports the evolution of $\epsilon_M(K, l)$ for $K = 15$, using the same reference covariance functions considered previously. It is seen that for all reference covariance functions, the error $\epsilon_M(K = 15, l)$ increases when l decreases, reflecting the increasing truncation error for $K = 15$ when M involves smaller features. However, different behaviors are reported depending on the choice of \mathcal{C}^r when l increases. When using $\mathcal{C}^r = \mathcal{C}(l^r)$ with $l^r \geq 0.3$, the error converges to machine precision when $l \gtrsim l^r$, meaning that in this situation the 15-dimensional reference subspace span $\{\phi_k^r = \phi_k(l^r), k = 1, \dots, K\}$ essentially encompasses the 15-dimensional dominant subspace of $\mathcal{C}(l \gtrsim l^r)$. Further, this behavior highlights the robustness of the change of coordinates, even for situations where finite numerical accuracy prevents the correct determination of the whole set of eigen-functions. On the contrary, the choice $\mathcal{C}(l^r)$ with $l^r \leq 0.2$, while yielding a lower error at small correlation length $l \lesssim l^r$, exhibits a stagnating error for $l \gtrsim l^r$, denoting that the corresponding $K = 15$ -dimensional reference subspace is not rich enough to encompass the dominant subspaces at larger correlation lengths. Roughly speaking, the reference eigen-functions are too oscillating to properly represent processes with long-range correlations. Finally, the selection of $\bar{\mathcal{C}}$ for the reference covariance function provides the best compromise, by construction, maintaining a maximum error $\epsilon_M(K = 15, l)$ less than 10^{-2} over the whole range of l .

4 Polynomial Chaos Surrogate

A suitable Polynomial Chaos (PC) expansion for the model predictions is constructed to accelerate the Bayesian inference process. In Section 4.1 we briefly review the PC methodology and provide some details regarding the numerical methods used in the examples provided in Section 5. Then, in Section 4.2 we focus on exploiting the PC surrogates to efficiently handle uncertain hyper-parameter through the change of coordinates introduced previously in Section 3.3. Finally, Section 4.3 provides a brief analysis of the PC surrogate error.

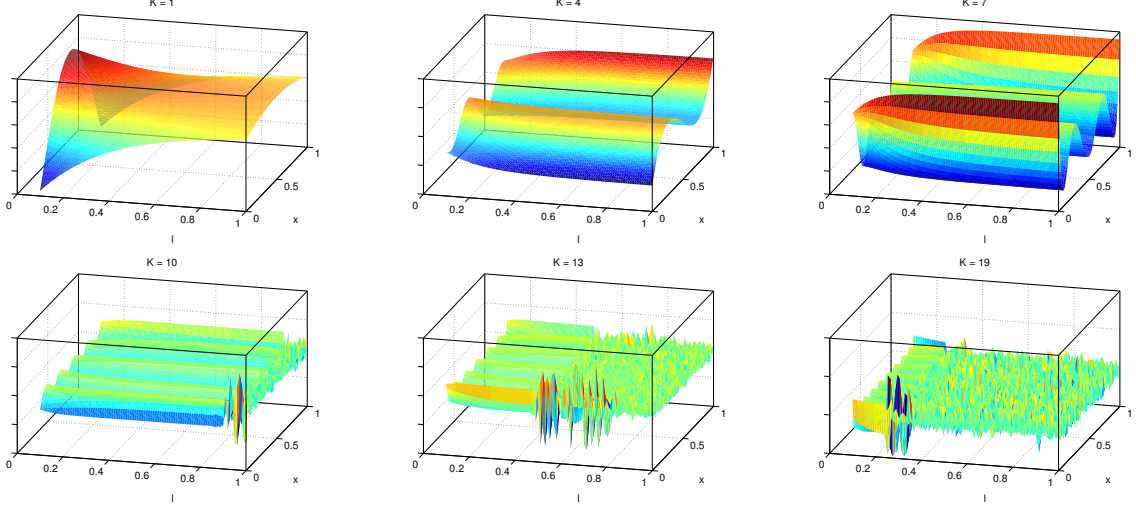


Figure 4: Dependence of eigen-functions $\phi_k(\mathbf{q})$ with the length-scale hyper-parameter l and selected k as indicated.

4.1 Polynomial Chaos expansion

Polynomial Chaos (PC) is a probabilistic methodology that expresses the dependencies of a model solution on some uncertain model inputs, through a truncated spectral polynomial expansion [22, 11]. Let $U \in Y$ be solution of a mathematical model \mathcal{L} (e.g. Partial Differential Equations), formally expressed as $\mathcal{L}U = 0$. We are interested in situations where the model \mathcal{L} is uncertain and parametrized with a finite set of independent second-order random variables $\boldsymbol{\xi} = (\xi_1, \dots, \xi_N)$ with known probability distribution. For simplicity, we shall restrict ourselves to the case of i.i.d. standard Gaussian random variables ξ_i , and will denote $p_{\boldsymbol{\xi}}$ the density function of $\boldsymbol{\xi}$, and $L_2(p_{\boldsymbol{\xi}})$ the space of second order random functionals in $\boldsymbol{\xi}$, that is

$$v(\boldsymbol{\xi}) \in L_2(p_{\boldsymbol{\xi}}) \Leftrightarrow \int \dots \int |v(\boldsymbol{\xi})|^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} < \infty. \quad (29)$$

Since the model depends on $\boldsymbol{\xi}$, its solution is also generally dependent on $\boldsymbol{\xi}$ and satisfies

$$\mathcal{L}(\boldsymbol{\xi})U(\boldsymbol{\xi}) = 0, \quad a.s. \quad (30)$$

Let $\{\Psi_{\alpha}, \alpha \in \mathbb{N}\}$ be a complete orthonormal set of $L_2(p_{\boldsymbol{\xi}})$, such that the model solution has an expansion of the form

$$U(\boldsymbol{\xi}) = \sum_{\alpha \in \mathbb{N}} U_{\alpha} \Psi_{\alpha}(\boldsymbol{\xi}), \quad \langle \Psi_{\alpha}, \Psi_{\beta} \rangle \doteq \int \dots \int \Psi_{\alpha}(\boldsymbol{\xi}) \Psi_{\beta}(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} = \delta_{\alpha, \beta}, \quad (31)$$

where the equality stands in the mean square sense and the expansion coefficients $U_{\alpha} \in Y$ are called the stochastic modes of U . A classical choice for the random functionals Ψ_{α} are orthonormal multi-variate polynomials in $\boldsymbol{\xi}$, leading to the so-called PC expansion of $U(\boldsymbol{\xi})$. The ξ_i being standard Gaussian random variables, the Ψ_{α} are in fact normalized multi-variate Hermite polynomials [27]. For practical purposes, the PC expansion of $U(\boldsymbol{\xi})$ needs to be truncated. When the basis is truncated to total order o the total number of terms in the PC expansion is given by $P+1 = (N+o)!/(N!o!)$ and therefore increases exponentially fast with both the expansion order o and the number N of random variables ξ_i . The series expansion approximating $U(\boldsymbol{\xi})$ is then finite and will be denoted $\tilde{U}(\boldsymbol{\xi})$ in the following:

$$U(\boldsymbol{\xi}) \approx \tilde{U}(\boldsymbol{\xi}) \doteq \sum_{\alpha=0}^P U_{\alpha} \Psi_{\alpha}(\boldsymbol{\xi}). \quad (32)$$

The existence and convergence of this series is asserted by the Cameron-Martin theorem [28] with the condition of U having a finite variance. The rate of convergence, and hence the number of terms in the series, depends on the smoothness of U with respect to $\boldsymbol{\xi}$. The series converges spectrally fast with P when U is infinitely smooth.

Various methods have been proposed for the determination of the PC coefficients U_α . They can be distinguished into the Non-intrusive and Galerkin methods. Non-intrusive methods rely on an ensemble of deterministic model evaluations of $U(\boldsymbol{\xi})$, for particular realizations of $\boldsymbol{\xi}$ selected either at random or deterministically. Non-Intrusive methods include Non-Intrusive Spectral and Pseudo-Spectral Projection [29, 30, 31], Least-Square-Fit and regularized variants [32, 33, 34], Collocation (interpolation) methods [35, 36, 37], that are often combined with Sparse-Grid algorithms to reduce computational complexity.

In the present paper, we instead rely on the Galerkin projection method [22, 11] for which the expansion coefficients U_α are defined through a reformulation of the model Eq. (30), using a weak form at the stochastic level. Specifically, Eq. (30) is projected on the PC basis, a procedure resulting in a set of $P + 1$ coupled problems,

$$\left\langle \mathcal{L}(\boldsymbol{\xi}) \sum_{\alpha=0}^P U_\alpha \Psi_\alpha(\boldsymbol{\xi}), \Psi_\beta(\boldsymbol{\xi}) \right\rangle = 0, \quad \beta = 0, \dots, P. \quad (33)$$

Numerical algorithms have been proposed to efficiently solve this set of coupled problems, both in the case of linear operators $\mathcal{L}(\boldsymbol{\xi})$ (see *e.g.* [38] for elliptic and parabolic problems) and non-linear operators (see *e.g.* [39, 40] and references in [11]).

4.2 PC surrogate for a parametrized covariance

Returning to the inference problem, we now want to construct a global PC surrogate for the model predictions, that accounts both for randomness of M_K , through its random coordinates $\boldsymbol{\eta}$, and the uncertainty in its covariance function, through the random hyper-parameter vector \boldsymbol{q} . We assume that the model problem amounts to solving for U a model depending on M_K . Using the notations above, it is written formally as

$$\mathcal{L}(\boldsymbol{\eta}, \boldsymbol{q})U(\boldsymbol{\eta}, \boldsymbol{q}) = 0. \quad (34)$$

The previous equation has motivated the idea of expanding the dependence of U with respect to the random vectors $\boldsymbol{\eta}$ and \boldsymbol{q} on a PC basis [12, 14], that is using $U(\boldsymbol{\eta}, \boldsymbol{q}) \approx \sum_{\alpha} U_\alpha \Psi_\alpha(\boldsymbol{\eta}, \boldsymbol{q})$. In the following, we consider an alternative approach, taking advantage of the change of coordinates discussed in Section 3. The change of coordinates allows us to approximate $M_K(\boldsymbol{q})$ on the fixed reference basis of KL modes $\{\phi_k^r, k = 0, \dots, K\}$, through the linear mapping $\boldsymbol{\eta} \mapsto \hat{\boldsymbol{\eta}}(\boldsymbol{q}) = \mathcal{B}(\boldsymbol{q})\boldsymbol{\eta}$. Eq. (25) provides the density of $\hat{\boldsymbol{\eta}}$ conditioned on \boldsymbol{q} . The model problem can therefore be recast as

$$\mathcal{L}(\hat{\boldsymbol{\eta}})U(\hat{\boldsymbol{\eta}}) = 0, \quad \text{where } \hat{\boldsymbol{\eta}} \sim p_{\hat{\boldsymbol{\eta}}}(\hat{\boldsymbol{\eta}}|\boldsymbol{q}). \quad (35)$$

The last expression shows that we only need to construct an approximation of the mapping $\hat{\boldsymbol{\eta}} \mapsto U(\hat{\boldsymbol{\eta}})$ which is accurate enough with respect to the conditional density $p_{\hat{\boldsymbol{\eta}}}(\hat{\boldsymbol{\eta}}|\boldsymbol{q})$ when \boldsymbol{q} varies. To get rid of the \boldsymbol{q} -dependence of the conditional density, we can consider averaging $p_{\hat{\boldsymbol{\eta}}}$ over \boldsymbol{q} . In the case of the reference covariance function $\mathcal{C}^r = \bar{\mathcal{C}}$, it can be shown that

$$\int \dots \int p_{\hat{\boldsymbol{\eta}}}(\hat{\boldsymbol{\eta}}|\boldsymbol{q})p_q(\boldsymbol{q})d\boldsymbol{q} = \frac{1}{\sqrt{2\pi^K|\Lambda^2|}} \exp \left[-\frac{\hat{\boldsymbol{\eta}}^t(\Lambda^2)^{-1}\hat{\boldsymbol{\eta}}}{2} \right], \quad (36)$$

where $\Lambda^2 = \text{diag}(\lambda_1^r, \dots, \lambda_K^r)$. In other words, the \boldsymbol{q} -marginal of the conditional density yields independent Gaussian random variables. This suggests constructing an approximate mapping of $\hat{\boldsymbol{\eta}} \mapsto U$, solving the model problem for a *reference* Gaussian field defined as

$$\hat{M}_K^{\text{PC}}(\boldsymbol{\xi}) = \sum_{k=1}^K \sqrt{\lambda_k^r} \phi_k^r \xi_k, \quad (37)$$

where the ξ_k 's are independent standard Gaussian random variables. It corresponds to a reference model problem $\hat{\mathcal{L}}(\boldsymbol{\xi})$ based on the reference Gaussian process $\hat{M}_K^{\text{PC}}(\boldsymbol{\xi})$. As before, we denote $\tilde{U}(\boldsymbol{\xi})$ the PC approximation of the reference model problem $\hat{\mathcal{L}}(\boldsymbol{\xi})U(\boldsymbol{\xi}) = 0$. From this PC approximation, we can approximate the model problem solution for couples $(\boldsymbol{\eta}, \mathbf{q})$ through

$$U(\boldsymbol{\eta}, \mathbf{q}) \approx \tilde{U}(\boldsymbol{\xi}(\boldsymbol{\eta}, \mathbf{q})) = \sum_{\alpha=0}^P U_\alpha \Psi_\alpha(\boldsymbol{\xi}(\boldsymbol{\eta}, \mathbf{q})), \quad \boldsymbol{\xi}(\boldsymbol{\eta}, \mathbf{q}) = \hat{\mathcal{B}}(\mathbf{q})\boldsymbol{\eta}, \quad (38)$$

where the \mathbf{q} -dependent matrix $\hat{\mathcal{B}}(\mathbf{q})$ expresses the change of coordinates $(\boldsymbol{\eta}, \mathbf{q}) \mapsto \boldsymbol{\xi}(\boldsymbol{\eta}, \mathbf{q})$. Based on Eq. (37), we propose to use

$$\hat{\mathcal{B}}_{kl}(\mathbf{q}) = \begin{cases} \frac{\mathcal{B}_{kl}(\mathbf{q})}{\sqrt{\lambda_k^r}}, & \lambda_k^r / \lambda_1^r > \kappa, \\ 0, & \text{otherwise,} \end{cases} \quad (39)$$

where $\kappa > 0$ is a small constant related to the numerical accuracy (typically $\kappa \sim 10^{-12}$) introduced to avoid ill-definition of the ξ_k 's associated to negligibly small λ_k^r . The κ -thresholding leads to transformed coordinates $\boldsymbol{\xi}(\boldsymbol{\eta}, \mathbf{q})$ where the first $K^{\text{PC}}(\kappa)$ components are non trivial, with $K^{\text{PC}}(\kappa) \leq K$. Note that the number of non-trivial components of $\boldsymbol{\xi}$ only depends on the reference covariance, \mathcal{C}^r , and not on \mathbf{q} . Also, the PC construction of $\tilde{U}(\boldsymbol{\xi})$ can in fact be reduced, considering $\hat{M}_{K^{\text{PC}}(\kappa)}^{\text{PC}}$ instead of \hat{M}_K^{PC} , with computational complexity reduction as a result when $K^{\text{PC}} < K$. However, we shall continue to report results as a function of K for simplicity.

When the reference covariance \mathcal{C}^r is not the \mathbf{q} -averaged of $\mathcal{C}(\mathbf{q})$, the \mathbf{q} -marginal conditional density $p_{\tilde{\eta}}$ remains Gaussian but introduces correlations between components. These correlations can be dealt with by introducing an additional change of basis in order to redefine a reference Gaussian process \hat{M}_K^{PC} in terms of independent standard Gaussian random variables ξ_k 's. In that case, Eq. (39) must be accordingly modified to account for the additional change of coordinates. Alternatively, when using $\mathcal{C}^r = \mathcal{C}(\mathbf{q}^r)$, we can continue to define the reference Gaussian process \hat{M}_K^{PC} by Eq. (37), which corresponds to solving the uncertain model problem assuming that $\tilde{\boldsymbol{\eta}}$ has for density $p_{\tilde{\eta}}$ conditioned on $\mathbf{q} = \mathbf{q}^r$. Although simpler, the approach is expected to yield a higher approximation error on average (over \mathbf{q}), as explained below.

4.3 Example

We consider the following model-problem consisting in the 1D transient diffusion equation,

$$\frac{\partial U}{\partial t} = \frac{\partial}{\partial x} \left(\nu \frac{\partial U}{\partial x} \right), \quad (40)$$

where the diffusivity ν is a stochastic field. Eq. (40) is solved for $t \in [0, T]$, in the unit domain $D = [0, 1]$, and with deterministic boundary conditions $U(x = 0, t) = -1$, $U(x = 1, t) = 1$, and homogeneous initial condition $U(x, t = 0) = 0$. We consider a log-normal stochastic diffusivity field of the form,

$$\nu = \nu_0 + \exp(M), \quad (41)$$

M is a (centered) Gaussian process with uncertain covariance function $\mathcal{C}(\mathbf{q})$. With $\nu_0 > 0$ the diffusivity is bounded away from 0 which ensures the well-posedness of the problem. In the computations we set $\nu_0 = 0.1$. In addition, we re-use the settings of Section 3.4 with Gaussian covariance function having an uncertain length-scale l with uniform distribution in $[0.1, 1]$ and fixed variance $\sigma_f^2 = 0.5$. For the solution of Eq. (40) we use a classical P1-finite element method (continuous piecewise linear approximation) for the spatial discretization, with a second order implicit time-integration scheme.

To investigate the error introduced by approximating $M \mapsto U$ by the PC map $\hat{M}_K \mapsto \tilde{U}$, we define the following error measures on the model problem solution. We first define the relative local error $\epsilon_U(o, K, \mathbf{q})$

as

$$\epsilon_U^2(o, K, \mathbf{q}) \doteq \frac{\|U(M(\mathbf{q})) - \tilde{U}(\boldsymbol{\xi}(\cdot, \mathbf{q}))\|_{L_2(\Omega, Y)}^2}{\|U(M(\mathbf{q}))\|_{L_2(\Omega, Y)}^2}, \quad (42)$$

where

$$\|V\|_{L_2(\Omega, Y)}^2 = \mathbb{E} \left[\int_0^T \|V(x, t)\|_{L_2(D)}^2 dt \right]. \quad (43)$$

This error measure incorporates the effects of several approximations: the approximation of $M(\mathbf{q})$ on the K -dimensional reference subspace, the truncation of the PC expansion to finite order o , and the spatial and time discretization errors inherent in the numerical resolution of the model problem. Because the PC surrogate will be used in place of solving numerically the model problem (given $\boldsymbol{\eta}$ and \mathbf{q}), we should not be concerned with the spatial and time discretization errors, and rather use for $U(M(\mathbf{q}))$ its discrete counterpart, provided that the same spatial and time discretizations are used. For the tests presented in this section, we use a uniform mesh with 56 elements and a fixed time-step $\Delta t = 10^{-4}$. These discretization parameters were selected to ensure that the error measurements reported below are dominated by the K and o -order truncation effects. Doing so, the local error ϵ_U^2 can be estimated by means of Monte Carlo average proceeding as follows. For a sample of \mathbf{q} , a) we generate a sample of the Gaussian process $M(\mathbf{q})$ on the finite-element mesh and solve the corresponding deterministic diffusion problem for the sample of $U(M(\mathbf{q}))$; b) we project $M(\mathbf{q})$ on the KL subspace, to obtain the KL coordinates $\boldsymbol{\eta}$ which are further translated to $\boldsymbol{\xi}(\boldsymbol{\eta}, \mathbf{q})$, and the PC approximation $\tilde{U}(\boldsymbol{\xi})$ is evaluated (see Eqs. 38-39); c) we compute $\|U(M(\mathbf{q})) - \tilde{U}(\boldsymbol{\xi})\|_{L_2(D \times T)}^2$ for the sample. Further, we set $T = 0.05$ in order to focus the error measure in the transient period.

Similarly, the local error can be \mathbf{q} -averaged to yield the relative global error counterpart:

$$E_U^2(o, K) \doteq \frac{\int \dots \int \|U(M(\mathbf{q})) - \tilde{U}(\boldsymbol{\xi}(\cdot, \mathbf{q}))\|_{L_2(\Omega, Y)}^2 p_{\mathbf{q}}(\mathbf{q}) d\mathbf{q}}{\int \dots \int \|U(M(\mathbf{q}))\|_{L_2(\Omega, Y)}^2 p_{\mathbf{q}}(\mathbf{q}) d\mathbf{q}}. \quad (44)$$

Figure 5 reports the global error for the present test problem. The left plot depicts $E_U(o, K)$ as a function of K and for a PC order $o = 10$. Errors are shown for some selection of reference covariance functions \mathcal{C}^r used in Section 3.4. We observe that for all the selected reference covariance functions, the global error on U stagnates for $K \gtrsim 9$. This indicates that the PC truncation becomes the dominant source of error for $K \gtrsim 9$. It is also seen that for all K shown, the error is the lowest when using $\bar{\mathcal{C}}$ for reference covariance, as expected.

Further, when using $\mathcal{C}(\mathbf{q}^r)$ as reference covariance function, the dependence of the global error on the reference length-scale l^r is non monotonic, but presents a minimum around $l^r = 0.4$. This minimum can be explained by the competition of two effects. On the one hand, we have seen that increasing l^r causes an increase in the approximation error of M , which translates in a larger approximation error on U . On the other hand, it can be shown that the lower l^r the more the \mathbf{q} -marginal of $p_{\hat{\eta}}$ departs from the K -variates standard Gaussian distribution assumed for the construction of \tilde{U} , with increasing averaged approximation error on U as a result. The right plot of Figure 5 also depicts the global error, but now for a fixed number of KL modes, $K = 15$, and increasing PC order $o \in [2, 10]$. Again, curves are shown for the different reference covariance functions. Similarly to the previous results, the global error is seen to stagnate for $o \gtrsim 8$, indicating here that for larger o the KL truncation error is dominant. In addition, for all shown o , using $\bar{\mathcal{C}}$ for reference covariance function appears to be superior to the choices $\mathcal{C}(l^r)$, while the later choice again exhibits a non-monotonic dependence of the error with respect to l^r .

Figure 6 presents in the left plot the normalized local error $\epsilon_U(o, K, l)$ for the case of $K = 15$ and PC order $o = 10$. As mentioned previously, the local error combines the effects of approximating M by \hat{M}_K , which has been reported in the right plot of Figure 3, and the PC truncation error. Focusing first on the cases where $\mathcal{C}(\mathbf{q}^r)$ is used as reference covariance, we observe a more complex behavior of the local error with l , depending on the selected reference length-scale l^r . Specifically, the local error at some l is always the lowest for the reference length-scale l^r the closest to l . This is expected, as using l^r is the optimal choice, given K and o , to achieve the lower error at $l = l^r$. For $\mathcal{C}^r = \bar{\mathcal{C}}$, which ensures by construction the best

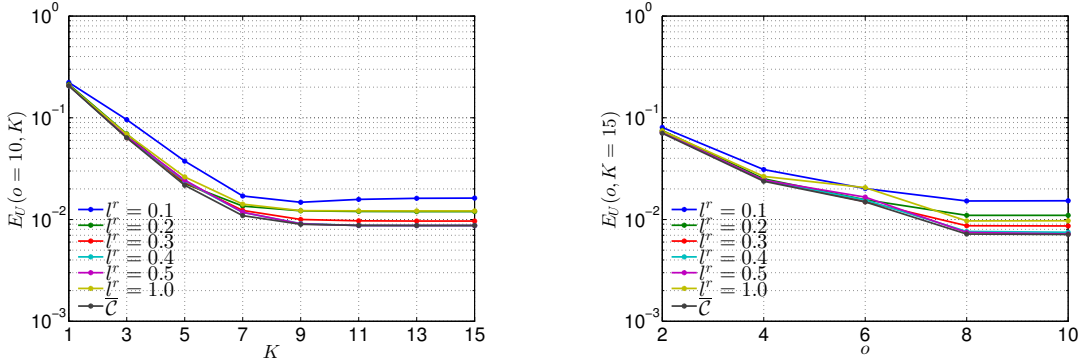


Figure 5: Global error $E_U(o, K)$ of the PC approximation \tilde{U} of the diffusion model problem solution. (Left) The plot shows the dependence of the error with K using a PC order $o = 10$. (Right) The plot is for different o and $K = 15$. The curves correspond to different definitions of the reference covariance function $C^r = \mathcal{C}(l^r)$ with l^r as indicated or the \mathbf{q} -averaged covariance function \bar{C} .

compromise over the \mathbf{q} -range, the local error remains below 2% over the whole range of hyper-parameters. Further, using $l^r > 0.2$, the local error first monotonically decreases with l and then stagnates (except for $l^r = 1$ where stagnation is not achieved).

Contrary to the local approximation error on M , the stagnation with $l \rightarrow 1$ occurs at an error level that strongly depends on l^r . This seems surprising as we have seen (right plot of Figure 3) that for $l^r > 0.2$ the approximation error on the process goes to zero as $l \rightarrow 1$, such that we could have expected an essentially constant local error ϵ_U for $0.2 < l^r \lesssim l$, depending only on the PC expansion order o . But one has to take into account the mapping from $\boldsymbol{\eta}$ to $\boldsymbol{\xi}$ to understand the behavior of the local error. Specifically, the PC approximation is constructed to minimize the approximation error for the reference model problem based on \hat{M}_K^{PC} (or $K^{\text{PC}}(\kappa)$) in Eq. (37), where the ξ_k 's are independent standard random variables. Therefore, the PC approximation aims at minimizing the error *with respect to* the standard K -variates Gaussian measure. When querying the PC approximation for some specific hyper-parameters value $\mathbf{q} \neq \mathbf{q}^r$, $\boldsymbol{\xi}$ follows a conditional Gaussian distribution $p_\xi(\boldsymbol{\xi}|\mathbf{q})$, induced by the transformation $\boldsymbol{\xi} = \hat{\mathcal{B}}(\mathbf{q})\boldsymbol{\eta}$. In general, this conditional distribution differs from the standard Gaussian one, affecting the quality of the approximation depending on \mathbf{q} . To get better insight into this effect, we remark that the conditional density $p_\xi(\boldsymbol{\xi}|\mathbf{q})$ is centered and Gaussian with covariance structure $\Sigma_\xi^2(\mathbf{q}) = \hat{\mathcal{B}}^t(\mathbf{q})\hat{\mathcal{B}}(\mathbf{q})$.

To measure the departure from the standard Gaussian multi-variables case, we present in the right plot of Figure 6 the largest eigen-value $\beta_{\max}(\mathbf{q})$ of $\Sigma_\xi^2(\mathbf{q})$, as a function of $\mathbf{q} = \{l\}$ and for the different reference covariance functions. $\sqrt{\beta_{\max}}$ measures of highest stretching rate induced by $\hat{\mathcal{B}}$. For the results reported in Figure 6, we used a thresholding parameter $\kappa = 10^{-12}$ in the definition of $\hat{\mathcal{B}}(\mathbf{q})$. It is seen that when using $\mathcal{C}(l^r)$ for reference, $\sqrt{\beta_{\max}}(l)$ increases exponentially fast with $l^r - l > 0$, denoting a more and more stretched distribution for $\boldsymbol{\xi}(\boldsymbol{\eta}, l)$, along some direction, as l decreases. Interestingly, although the maximal stretching rate can reach values as high as 10^6 , its impact on the PC approximation error is clearly much less important. The reason for the moderate sensitivity to coordinates stretching of the PC approximation error is that most of the stretching occurs along the directions associated with the lowest eigen-values λ_k^r , which have low to insignificant impacts on the model problem solution. In fact, our numerical experiments have demonstrated that the PC approximation error is essentially insensitive to κ , provided it is small enough. Indeed, a fast (exponential) decay of the successive KL modes' contributions to U is expected for elliptic and parabolic model problems, as the effects of short-scale fluctuations in the diffusivity field are filtered-out. However, coordinates stretching may yield robustness issues for other model types. Finally, it is seen that choosing l^r equal to the minimal length-scale ($l = 0.1$) yields a maximum stretching $\sqrt{\beta_{\max}} < 3$ which is controlled over the whole range of l , while the case of \bar{C} yields a significant stretching (picking to ≈ 10) around the minimal length-scale, but quickly decays with l and remains close to 1 (see left plot of Figure 6). These findings confirm the appropriateness of the reference covariance function for the construction of the

PC surrogate.

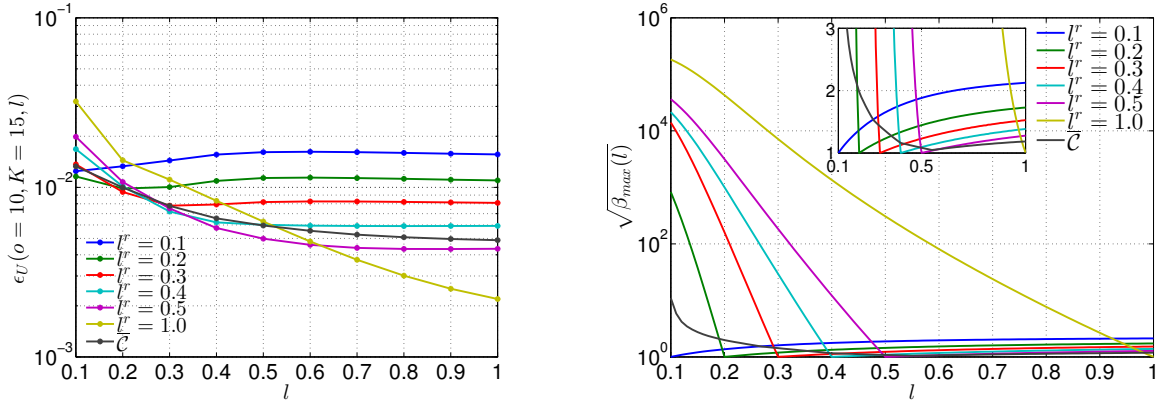


Figure 6: (Left) Local approximation error $\epsilon_U(o, K, l)$, for $o = 10$ and $K = 15$. (Right) (log-scale) and inset (linear-scale): dependence on l of the maximal stretching rate $\sqrt{\beta_{\max}(l)}$ induced by the coordinate transformation $\mathcal{B}(\mathbf{q})$. The curves correspond to different definitions of the reference covariance function $C^r = C(l^r)$ with l^r as indicated or the \mathbf{q} -averaged covariance function \bar{C} .

Concerning the PC approximation of the model-problem solution, we would like to stress the following points. First the approach can be readily extended to alternative and more elaborated PC constructions methods, including non-intrusive ones; in particular considering adaptive techniques where the set of polynomials used in the PC expansion is determined as to minimize the approximation error, instead of proceeding from PC basis with uniform truncation order o , would clearly be beneficial, especially for problems involving high numbers of KL modes K and requiring high polynomial order along certain ξ_k 's and not others. Second, the numerical tests have focused on length-scale uncertainty only, which is indeed the hardest source of uncertainty as it affects both the magnitude *and shape* of the KL modes. In contrast, uncertainty in the process variance σ_f^2 in the Gaussian covariance family only manifests itself in the magnitude of the eigenvalues. Therefore, uncertainty in the pre-exponential factor σ_f^2 of the Gaussian covariance can be handled through either an additional dimension to the PC expansion, as performed in [12, 14], or directly through our proposed change of coordinates approach based on the reference \bar{C} , which amounts to take the averaged variance as the reference one. Similar to the problem for uncertain length-scale l , numerical tests (not shown) have demonstrated that the \mathbf{q} -averaged definition of \bar{C} leads to globally lower errors in presence of variance uncertainty, compared to a definition of the reference based on some q^r . This has motivated the use of the \mathbf{q} -averaged definition of the reference covariance function in the remainder of the paper.

Finally, the PC expansion of the full model-problem solution has been considered here; there may be other situations where expansion of the full model-problem solution is not necessary. For instance, if the nature of the observations are known prior to constructing the PC expansion, the direct expansion of the model predictions $\mathbf{u}(\boldsymbol{\xi})$ could be considered. If in addition the measurements have been performed, considering the direct PC expansion of the measurements to model-predictions discrepancy, $\Delta_d(\boldsymbol{\xi}) = \sum_{i=1}^{N_o} |d_i - u_i(\boldsymbol{\xi})|^2$ (in the case of identically distributed additive noise) could be advantageous.

5 Application examples

In this section, we illustrate the benefit of considering prior Gaussian fields with parametrized covariance function in the inference of the diffusivity field in the transient diffusion problem introduced in Section 4.3. We first present in Section 5.1 the inference problem, and introduce 3 cases that will serve to investigate the proposed method. We also provide details on the exploitation the PC surrogate constructed in Section 4, and on the PC accelerated formulation of the inference problem. For comparison purposes, we first solve in Section 5.2 the Bayesian inference problem for a fixed covariance prior, that is without inferring the

covariance hyper-parameters, using instead preassigned values. Then, in Section 5.3, we considered the inference with hyper-parameters covariance and illustrate its advantage and behavior with respect to noise level, number of observations and surrogate polynomial order.

5.1 Set-up of the inference problem

The proposed method will be illustrated for the inference of a log-diffusivity field, using the transient diffusion model problem corresponding to Eq. (40). To test the proposed method we consider three different log-diffusivity fields $m(x) = \log(\nu - \nu_0)$, to be inferred:

- Sinusoidal profile: $m^{\text{sin}}(x) = \sin(2\pi x)$,
- Step function: $m^{\text{step}}(x) = \begin{cases} -1/2, & x < 0.5 \\ 1/2, & x \geq 0.5 \end{cases}$,
- Random profile: $m^{\text{ran}}(x)$ drawn at random from $\mathcal{GP}(0, \mathcal{C})$ where \mathcal{C} is the Gaussian covariance with length-scale $l = 0.25$ and variance $\sigma_f^2 = 0.65$.

The inferences are performed on sets of data, $\{d_i, i = 1, \dots, N_o\}$, consisting of noisy measurements of the solution to the diffusion equation for the three profiles. The measurements are taken at a set of n_x spatial locations x_i uniformly distributed inside $D = (0, 1)$, and for n_t times t_i uniformly distributed in $(0, T)$. The total number of observations is then $N_o = n_x \times n_t$. The observations are synthetically generated by perturbing the respective model solutions for the 3 fields tested with a measurement noise ϵ_i randomly and independently drawn from the Gaussian distribution $\mathcal{N}(0, \sigma_\epsilon^2)$. To avoid the so-called inverse crime [41], the solutions used to generate the observations are computed with a significantly finer spatial and temporal discretization than for the construction of the PC approximation. Unless otherwise specified, we use $n_x = 19$, $n_t = 13$ (so $N_o = 247$), with $T = 0.05$ and a Gaussian noise with $\sigma_\epsilon^2 = 0.01$. Figure 7 depicts the location of the observation points and the solution of the diffusion equation for m^{sin} at the different observation times.

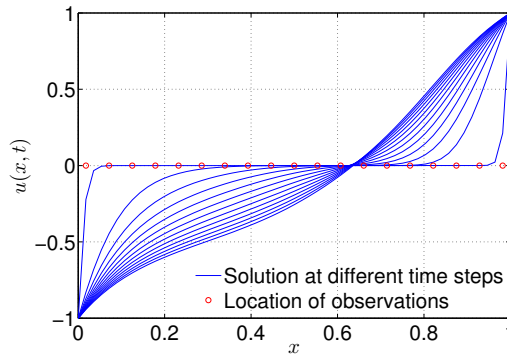


Figure 7: Illustration of inference problem for m^{sin} . Plotted are the $n_x = 19$ observation points and the solution of the diffusion equation with profile m^{sin} at the n_t observation times.

For the inference, we consider in all cases the Gaussian prior $M \sim \mathcal{GP}(0, \mathcal{C}(\mathbf{q}))$, where the covariance function $\mathcal{C}(\mathbf{q})$ has hyper-parameter $\mathbf{q} = \{l, \sigma_f^2\}$. The prior is then fully characterized once we have selected the prior of the hyper-parameters. We choose a uniform prior for l over the range $[l_{\min}, l_{\max}]$, with as previously $l_{\min} = 0.1$ and $l_{\max} = 1$, and an inverse Gamma prior [42, 43, 44] for σ_f^2 with parameters $\alpha = 3$ and $\beta = 1$. The prior of σ_f^2 thus has a long-tailed distribution with mean value $\beta/(\alpha - 1) = 0.5$ and variance $\beta^2/(\alpha - 1)^2(\alpha - 2) = 0.25$. Note that the existence of the first moment of σ_f^2 is enough to ensure the existence of the average covariance function, and that $M(\boldsymbol{\eta}, \mathbf{q}) \in L_2(D, p_\eta, p_q)$ (because the modes in its KL decomposition scales with $\sqrt{\sigma_f^2}$). In contrast, expanding the diffusion equation solution with respect to both

the KL coordinates $\boldsymbol{\eta}$ and hyper-parameter \mathbf{q} (as proposed in [12, 14]) could be problematic since, to our knowledge, there is no standard orthogonal polynomial family for the inverse Gamma distribution function and the solution U may not have second moment ($\exp(\sqrt{y})$ with $y \sim \text{Inv}\Gamma(3, 1)$ has unbounded second moment). Using the notation of Section 2, the prior of \mathbf{q} is then

$$p_{\mathbf{q}}(\mathbf{q}) = p_{\mathbf{q}}(l, \sigma_f^2) = \begin{cases} \frac{1}{|l_{\min} - l_{\max}| \Gamma(3)} (\sigma_f^2)^{-4} \exp\left(-\frac{1}{\sigma_f^2}\right), & l \in [l_{\min}, l_{\max}], \sigma_f^2 > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (45)$$

As for the noise hyper-parameter, we use the uninformative, improper, Jeffrey's prior

$$p_o(\sigma_o^2) \propto \frac{1}{\sigma_o^2}. \quad (46)$$

Having specified all priors, the determination of the Bayesian posterior $p(\boldsymbol{\eta}, \mathbf{q}, \sigma_o^2 | \mathbf{d})$ requires the evaluation of the likelihood of the data \mathbf{d} given $(\boldsymbol{\eta}, \mathbf{q}, \sigma_o^2)$. Instead of following the computational flow-chart presented in Figure 1, which would require the solution of a deterministic model problem for each new sample of $(\boldsymbol{\eta}, \mathbf{q})$, we rely on coordinate transformation and PC approximation as introduced in the previous sections. Following the findings of the previous section, the reference model problem is based on the stochastic process \hat{M}_K^{PC} corresponding to the \mathbf{q} -averaged covariance function $\bar{\mathcal{C}}$, whose KL decomposition is truncated to the first $K = 15$ dominant modes. Solving this reference problem, we obtain the approximation $\tilde{U}(\boldsymbol{\xi}) = \sum_{\alpha=0}^P U_{\alpha} \Psi_{\alpha}(\boldsymbol{\xi})$ of the reference model problem solution $U(\boldsymbol{\xi})$. Unless stated otherwise we use in the following results a PC order $o = 10$ with a spatial discretization involving 56 finite elements. From the approximate solution \tilde{U} , we can extract the PC approximations of the model predictions, $\tilde{\mathbf{u}}(\boldsymbol{\xi})$, whose components are

$$\tilde{u}_i(\boldsymbol{\xi}) = \tilde{U}(x_i, t_i, \boldsymbol{\xi}) = \sum_{\alpha=0}^P U_{\alpha}(x_i, t_i) \Psi_{\alpha}(\boldsymbol{\xi}), \quad i = 1, \dots, N_o. \quad (47)$$

This constitutes the offline step of the proposed PC-accelerated sampler. Once the PC approximation has been determined, one can use $\tilde{\mathbf{u}}(\boldsymbol{\xi}(\boldsymbol{\eta}, \mathbf{q}))$ as a surrogate of the model predictions $\mathbf{u}(\boldsymbol{\eta}, \mathbf{q})$ in the (online) computation of the likelihood:

$$p(\mathbf{d} | \boldsymbol{\eta}, \mathbf{q}, \sigma_o^2) \approx \tilde{p}(\mathbf{d} | \boldsymbol{\eta}, \mathbf{q}, \sigma_o^2) \doteq \prod_{i=1}^{N_o} p_{\epsilon}(d_i - \tilde{u}_i(\boldsymbol{\xi}(\boldsymbol{\eta}, \mathbf{q})), \sigma_o^2), \quad (48)$$

where $\boldsymbol{\xi}(\boldsymbol{\eta}, \mathbf{q})$ is given by Eq. (38) and p_{ϵ} is defined in Eq. (3). For the actual definition of the coordinate transformation $\hat{\mathcal{B}}(\mathbf{q})$ in Eq. (39), we set $\kappa = 0$ since $\lambda_{k \leq 15}^r / \lambda_1^r$ remains large enough for the present settings. Finally, multiplying by the hyper-parameter priors and prior of $\boldsymbol{\eta}$, one obtains (up to a constant normalization factor) the approximation $\tilde{p}(\boldsymbol{\eta}, \mathbf{q}, \sigma_o^2 | \mathbf{d})$ of the posterior distribution. The computational structure for the change of coordinates method and PC acceleration is schematically illustrated in Figure 8, distinguishing between offline and online steps. The online step is imbedded in an adaptive Metropolis-Hasting algorithm to generate samples of $(\boldsymbol{\eta}, \mathbf{q}, \sigma_o^2)$ following the posterior density.

5.2 Inference with fixed covariance parameters

For comparison purposes, the Bayesian inference problems are first solved for the case of Gaussian prior covariance function having pre-assigned parameters $l = 0.5$ and $\sigma_f^2 = 0.5$. The problem therefore consists in inferring only the 15 coordinates $\boldsymbol{\eta}$ and the noise hyper-parameter σ_o^2 . Note that in this case $\hat{\mathcal{B}}$ is the identity, as the PC approximation is based on the prior process with pre-assigned covariance function.

A total number of 2.5×10^5 MCMC steps were deemed necessary for the pre-assigned hyper-parameters case to properly explore the posterior. The resulting chain of the KL coordinates were observed to be well-mixed (not shown). The marginal posteriors, estimated using a standard Kernel Density Estimation

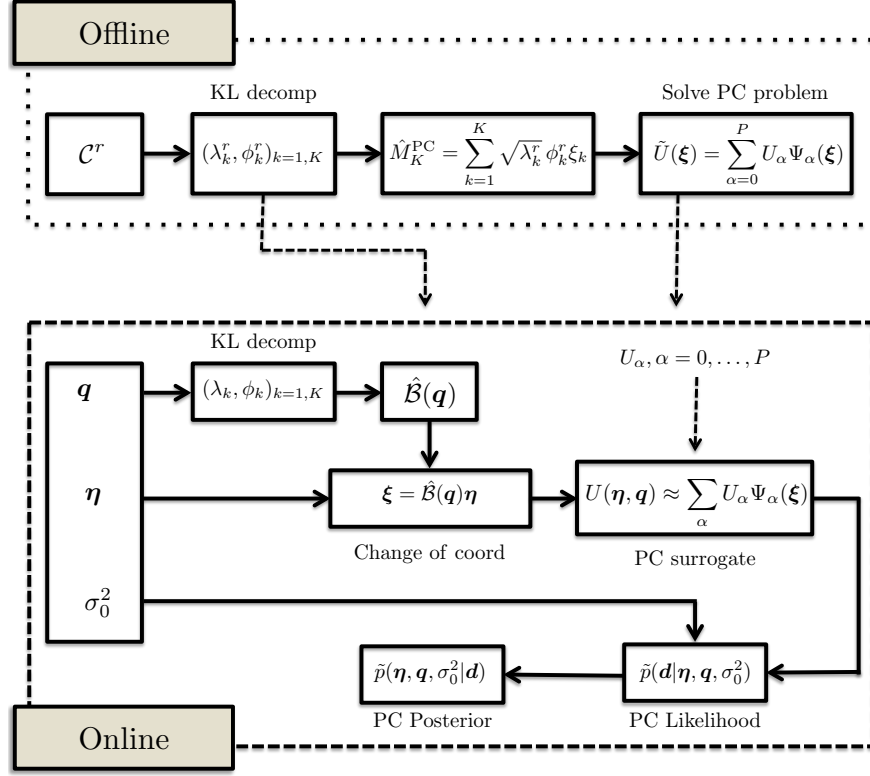


Figure 8: Offline step (surrogate construction) of the accelerated MCMC sampler and Online step of the PC surrogate based evaluation of the posterior.

(KDE) method [45, 46], of the first 8 KL coordinates for the inference of m^{sin} are shown in Figure 9. These posteriors are compared with their respective priors (standard Gaussian distributions). We notice that only the first 4 coordinates η_k show significant improvement in their posterior distributions. This improvement can be quantified using the Kullback-Leibler Divergence (KLD) which is a statistical measure that quantifies the distance between two probability distributions p and q [47], defined according to:

$$KLD(p, q) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} dx \quad (49)$$

Here we calculate the KLD between the prior and the (marginal) posterior of each KL coefficient η_k . The KLD is indicated on top of each plot and quantifies the information gain from the observations, which is found significant only for the first 4 KL coordinates. Figure 9 also shows the posterior of the noise variance hyper-parameter (bottom right plot), σ_0^2 , which exhibits a Maximum A Posteriori (MAP) value close to the value used to generate the data, $\sigma_\epsilon^2 = 0.01$. The similar findings are reported for the inferences of m^{step} and m^{ran} (results not shown for brevity).

To better analyze the quality of the inferred fields, we report in Figure 10, for the 3 test cases, the median, 5% and 95% quantiles values of the posteriors of the inferred field $m(x)$. These statistical characterizations of m are also compared with the true profiles. For the inference of m^{sin} we notice that the 5% to 95% quantiles range does not contain the true profile for a large set of x . This mismatch can be attributed to the pre-assigned hyper-parameter values that are not suitable. The same observation can be made for the case of m^{ran} . In contrast, m^{step} is nearly everywhere within the 5%-95% quantiles range of the inferred profile m .

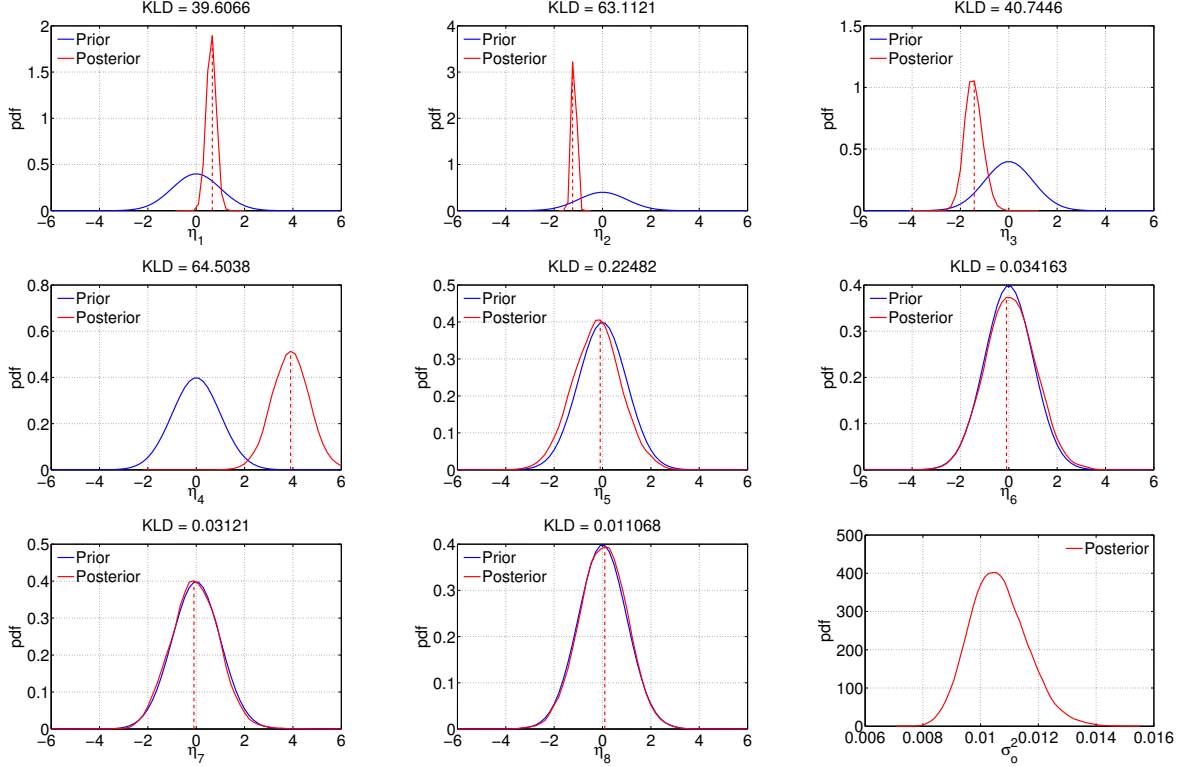


Figure 9: Comparison of the priors and (marginal) posteriors of the first 8 KL coordinates η_k and noise hyper-parameter σ_o^2 (posterior only) for the inference of m^{sin} without using covariance hyper-parameters (a Gaussian covariance with $l = 0.5$ and $\sigma_f^2 = 0.5$ is assumed). The corresponding Kullback-Leibler Divergences (KLD) for the KL coordinates are also indicated on top of each plot.

5.3 Inference with covariance hyper-parameters

Next, we repeat the previous inference problems but considering now the covariance hyper-parameters l and σ_f^2 in addition to the 15 KL modes and observation noise σ_o^2 . For the sampling of the posterior, a total of 2.5×10^5 MCMC steps was found also necessary to satisfactorily estimate the posterior statistics, same as for the case with pre-assigned parameters. The chains of all KL coordinates and hyper-parameters were observed to be well-mixed as illustrated in Figure 11.

The marginal posteriors of the first 8 KL coordinates η_k for m^{sin} are shown in Figure 12 together with their respective priors. The KLD values are also indicated on top of the plots. The results show a significant information gain for the first 7 KL coordinates, in contrast to only the first 4 KL coordinates when using pre-assigned parameters. In the same figure we show the marginal of the observation noise. The latter posterior has a MAP close to $\sigma_o^2 = 0.01$, corresponding to the value used to generate the observations. Similar conclusions can be made for the cases of $m^{\text{step,ran}}$ (results not shown for brevity).

The pdfs of the posterior of the hyper-parameters are shown in Figures 13 and compared with their priors for m^{sin} . The results show a significant difference between the prior and posterior of the covariance length scales l , with a MAP around $l = 0.2$, while the posterior probability of $l > 0.4$ is essentially zero. On the contrary, the posterior of the covariance variance σ_f^2 has a similar structure to that of its prior, with a shift of the expected (and MAP) value toward higher values.

5.3.1 Comparison with the inferences with and without hyper-parameters

To better appreciate the improvement resulting from the introduction of the covariance hyper-parameters, we first provide a comparison of the inferred median profiles, obtained by inferring covariance hyper-

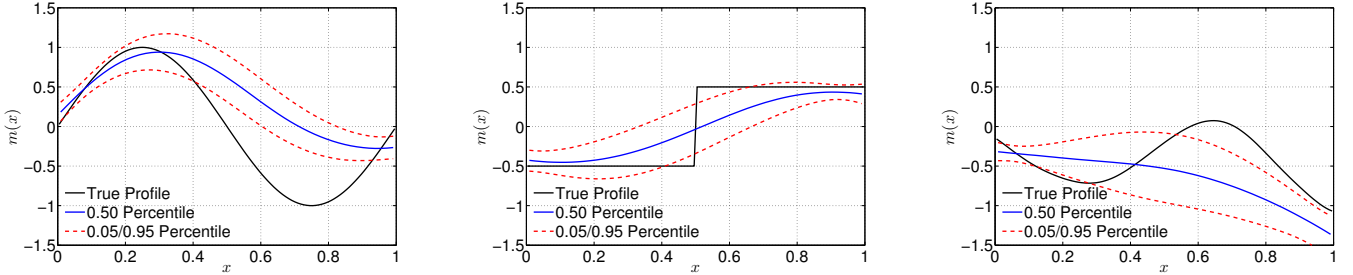


Figure 10: Comparison of the posterior of $m(x)$ with the true profile, for the cases of m^{sin} , m^{step} and m^{ran} (from left to right). The inferences use a fixed Gaussian covariance function with $l = 0.5$ and $\sigma_f^2 = 0.5$. Shown are the median, 5% and 95% quantiles of the posterior and true profile.

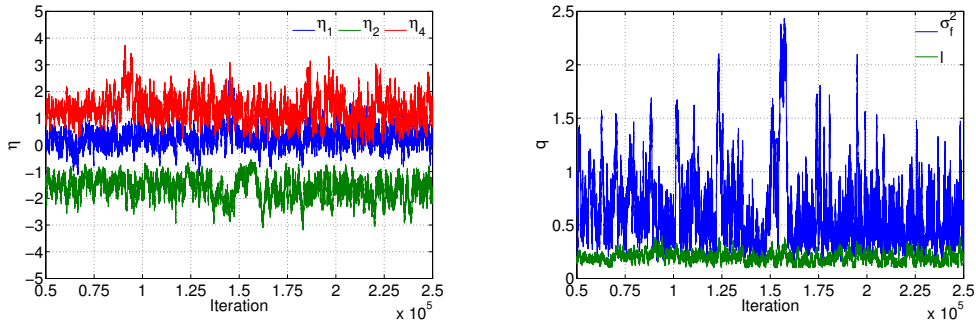


Figure 11: Illustration of the chain generated by MCMC using PC surrogate and coordinate transformation: successive samples of (Left) few KL coordinates and (Right) hyper-parameters \mathbf{q} . Case of the inference of m^{sin} .

parameters or by using pre-assigned values. The median profiles for all three cases are plotted in Figure 14, which also depicts the true profiles. It is seen that introducing the covariance hyper-parameter significantly reduces the distance between the median and true profiles in the smooth cases (m^{sin} and m^{ran}), while having no significant impact on the inference of the piecewise constant profile m^{step} . This behavior can be explained by the family of Gaussian processes considered, which is not well-suited for the inference of m^{step} , and so the introduction of hyper-parameters does not help improving the inference.

Second, the median, mean, MAP, 5% and 95% quantiles of the inferred log-diffusivity profiles are plotted in Figure 15 and compared with the respective true profiles $m^{\text{sin,step,ran}}$. These plots should be contrasted with the results shown in Figure 10, obtained with pre-assigned covariance. Consistent with the previous observations on the medians, we observe that in the case of the discontinuous profile, m^{step} , the inference of the hyper-parameters only affects slightly the 5% and 95% quantiles. On the contrary, for the smooth profiles $m^{\text{sin,ran}}$ the 5% and 95% quantiles bounds now contain the true profiles for nearly every x . This significant improvement is due partly to the better agreement between the true and median profiles, but also to a generally higher variability in the posterior when considering the hyper-parameters in the inference process. In other words, the inference of the covariance hyper-parameters appears to yield a more flexible approach than when using a fixed covariance assumption.

5.3.2 Effects of measurement noise and number of observations

To investigate the impact of the observations on the inference processes, with or without covariance hyper-parameters, we repeat the previous inference problems for different noise level σ_ϵ^2 in the observations and different number of spatial locations n_x . The results are reported in Figure 16 in terms of median profiles, for the three test profiles $m^{\text{sin,step,ran}}$ (from left to right) and the inference without (top row) and with covariance hyper-parameters (bottom row). As expected, the plots indicate an improvement of the inferred (median)

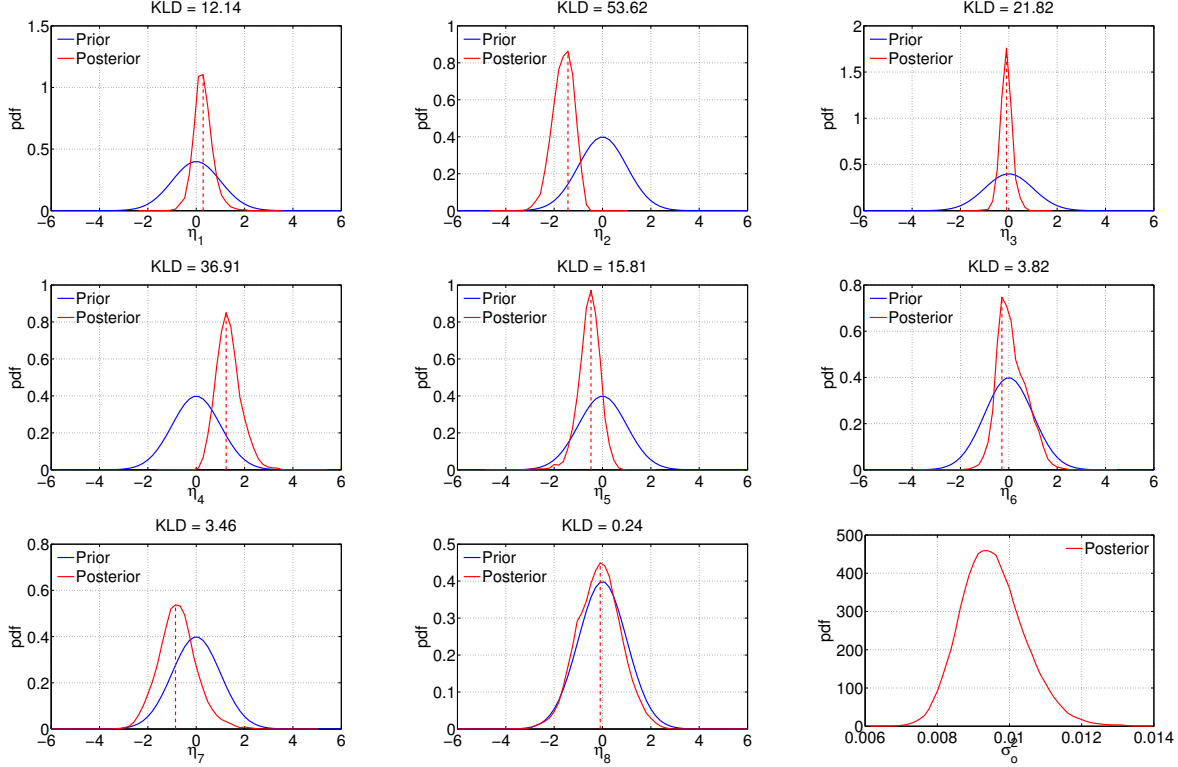


Figure 12: Comparison of the priors and (marginal) posteriors of the first 8 KL coordinates η_k and noise hyper-parameter σ_o^2 (posterior only) for the inference of m^{sim} with covariance hyper-parameters. The corresponding Kullback-Leibler Divergences (KLD) for the KL coordinates are also indicated on top of each plot.

profiles when the noise level is lowered, and when the number of observation increases. The improvements are more significant in the cases of the smooth profiles ($m^{\text{sim,ran}}$) than for the discontinuous one (m^{step}), a result consistent with the previous observations. In addition, for the smooth cases, the improvements carried by the introduction of the covariance hyper-parameters in the inference problem is seen to not only yield median profiles closer to the true ones, but also to significantly accelerate the convergence to the truth. The improvement of the convergence rate would require additional numerical experiments to be precisely measured, but it can already be safely asserted that more information is gained from the observations when considering the covariance hyper-parameters in the inference.

5.3.3 Convergence with the PC surrogate order

Finally, we illustrate in Figure 17 the dependence of the inferred median profiles on the selected order o for the PC surrogate model. The figure shows that, irrespective to the smoothness of the true profile, the inferred medians quickly converge as o increases, demonstrating that the L_2 convergence of the PC surrogate with coordinate transformation reported in Section 4.3 transfers to the inference problem. In fact, in view of the convergence curves shown in Figure 5, the differences in the inferred median profiles for $o = 8$ and $o = 10$ are more likely to come from sampling errors than from differences in the PC surrogates.

6 Discussion and Conclusion

This paper presented a Bayesian approach to infer a parameter field from prior \mathcal{GP} having a covariance function involving some hyper-parameters \mathbf{q} . The main contribution of the present work is the introduction of a coordinate transformation in order to represent the prior \mathcal{GP} using a unique reference basis of spatial

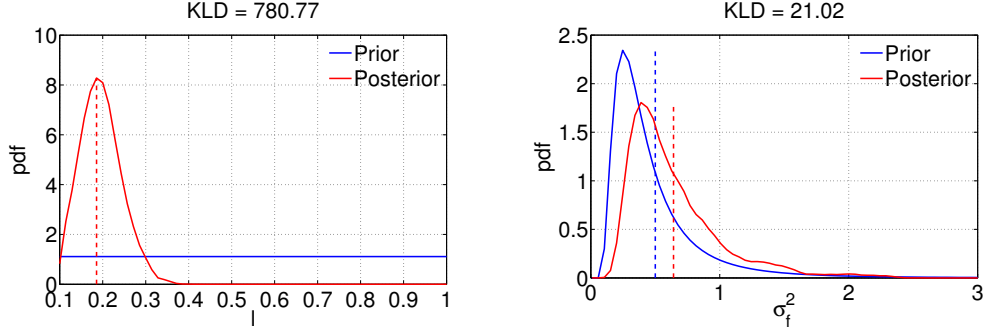


Figure 13: Priors and (marginal) posteriors of the covariance hyper-parameters l (left plot) and σ_f^2 (right plot) for the inference of m^{sin} . Also indicated are the corresponding Kullback-Leibler Divergences (KLD).

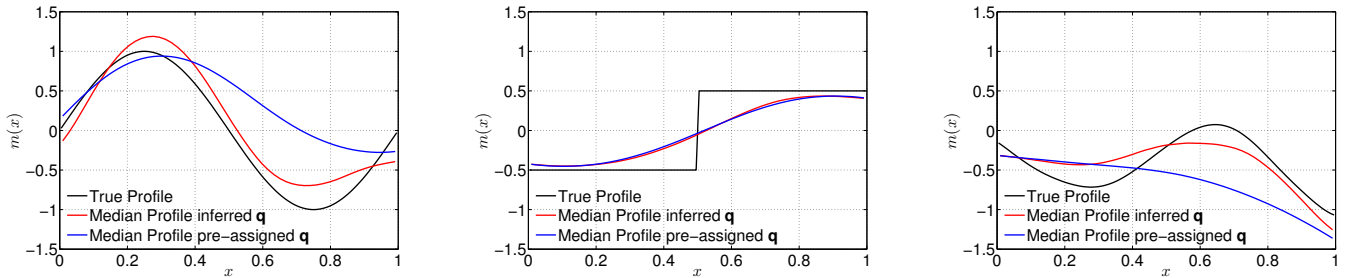


Figure 14: Comparison of the true log-diffusivity profiles with corresponding posterior medians for the inference with covariance hyper-parameters and preassigned covariance. Cases of m^{sin} , m^{step} and m^{ran} from left to right.

modes, while the effects of the covariance hyper-parameters is reflected by the (joint) prior probability density function of the random coordinates of the \mathcal{GP} that becomes conditioned on \mathbf{q} . The coordinate transformation naturally leads to the construction of a unique polynomial surrogate for the forward model predictions; this surrogate model accounts for the dependence of the model predictions on the coordinates of the \mathcal{GP} in the reference basis. For a Polynomial Chaos approximation, as considered in this paper, the construction of the surrogate amounts to solving a unique (stochastic) reference problem, assuming the independence of the \mathcal{GP} coordinates. The stochastic dimensionality of the surrogate model is therefore equal to the dimensionality of the (truncated) \mathcal{GP} representation, and is not augmented by the number of hyper-parameters intervening in the covariance function parametrization. This fact has to be contrasted with the alternative approaches proposed in [12, 14] where the PC expansion explicitly incorporates the dependencies on \mathbf{q} . Another advantage of selecting a reference problem for the construction of the PC surrogate, compared to the direct expansion with respect to the covariance hyper-parameters, is that it can overcome issues related to hyper-parameters with complex distributions, *e.g.* improper, no second-order moments, ... for which classical PC bases may not exist.

The surrogate model can then be substituted for the true model predictions in the definition of the likelihood of the observations appearing in Bayes' formula for the posterior of the \mathcal{GP} coordinates and covariance hyper-parameters. The resulting approximate likelihood can in turn be imbedded in a MCMC sampler to greatly accelerate the sampling of the posterior distribution, with significant computational savings. In its present form, the proposed method however introduces some overhead during the sampling stage, compared to other approaches relying on PC acceleration with explicit dependence on \mathbf{q} : for any new proposed values of the hyper-parameters the coordinate transformation must be determined. The determination of the transformation, given \mathbf{q} , requiring the computation of the dominant subspace of the covariance function (given \mathbf{q}) may constitute a severe limitation for large scale problems (for the simplified problems presented in Section 5, the CPU time of the inference with hyper-parameters was found roughly

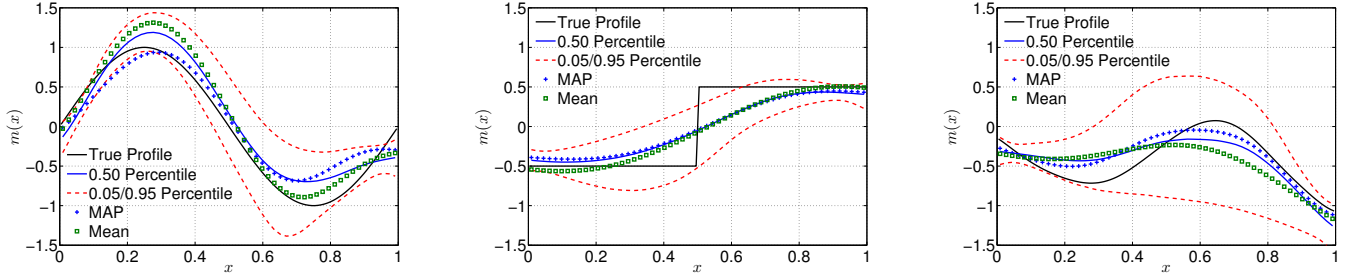


Figure 15: Comparison of the posteriors profiles with the true ones, for the cases of m^{sin} , m^{step} and m^{ran} (from left to right). The inferences use covariance function with hyper-parameters. Shown in each plots are the median, mean, MAP, 5% and 95% quantiles of the posterior and true profiles.

three time as large as for the case without hyper-parameters). To remedy this point in the future, we plan to approximate the dependence of the coordinate transformation, $\hat{\mathcal{B}}$, on \mathbf{q} using, again, a PC expansion. As for the construction of the PC surrogate of the model predictions, the approximation of the coordinate transformation will be computed off-line and subsequently used in-line within the sampler.

The numerical experiments presented in the paper, although based on a simple model, have highlighted the following points:

- Using for reference basis the truncated set of dominant modes of the \mathbf{q} -averaged covariance function is not only optimal (on average) for the representation of the processes with variable \mathbf{q} , but it also appears as the best choice in terms of averaged error for the PC surrogate of the model prediction in our example.
- The control of the stretching induced by the coordinate mapping $\hat{\mathcal{B}}(\mathbf{q})$ is crucial for the error control; while using the marginalized conditional density $p_{\bar{\boldsymbol{\eta}}}(\bar{\boldsymbol{\eta}}|\mathbf{q})$ appears to be an appropriate choice, other alternatives may be conceived. In particular, augmenting the variability of the reference process $M_K^{\text{PC}}(\boldsymbol{\xi})$ could improve the robustness of the surrogate PC model.
- The introduction of covariance functions with hyper-parameters clearly improved the inference results in the problems considered, particularly when inferring smooth profiles. In particular, information gain was observed for a larger set of coordinates. In addition, when covariance hyper-parameters was accounted for, the convergence rate of the inferred field with increased number of observations and reduced observation noise also seemed to improve.
- The convergence with the PC surrogate order seems quite fast for the presented problems, suggesting to possibility of using moderate PC orders, particularly to balance PC error and posterior sampling errors.

On the basis of the present findings, we plan for future work to develop the coordinate transformation approach to further exploit the posterior structure involving the conditional prior probability of the transformed coordinates $\bar{\boldsymbol{\eta}}$ and derive samplers adapted to this particular structure. Regarding the construction of the PC surrogate model, consideration of adaptive constructions would be beneficial to reduce the computational cost of the off-line step, to increase accuracy, and further accelerate the sampler. Further, the PC approximation of the coordinate transformation appears to be a key element to make the whole approach effective to handle large scale problems. In addition, the proposed method, in particular the construction of the PC approximation of the model prediction, would certainly benefit from fitting the procedure to the posterior distributions (of coordinates and hyper-parameters) rather than to the prior ones, especially when the observations are informative. Since these posterior distributions are not known *a priori*, iterative constructions are needed. Pursuit of these avenues is currently considered on a complex problem arising in subsurface geological models and earthquake model.

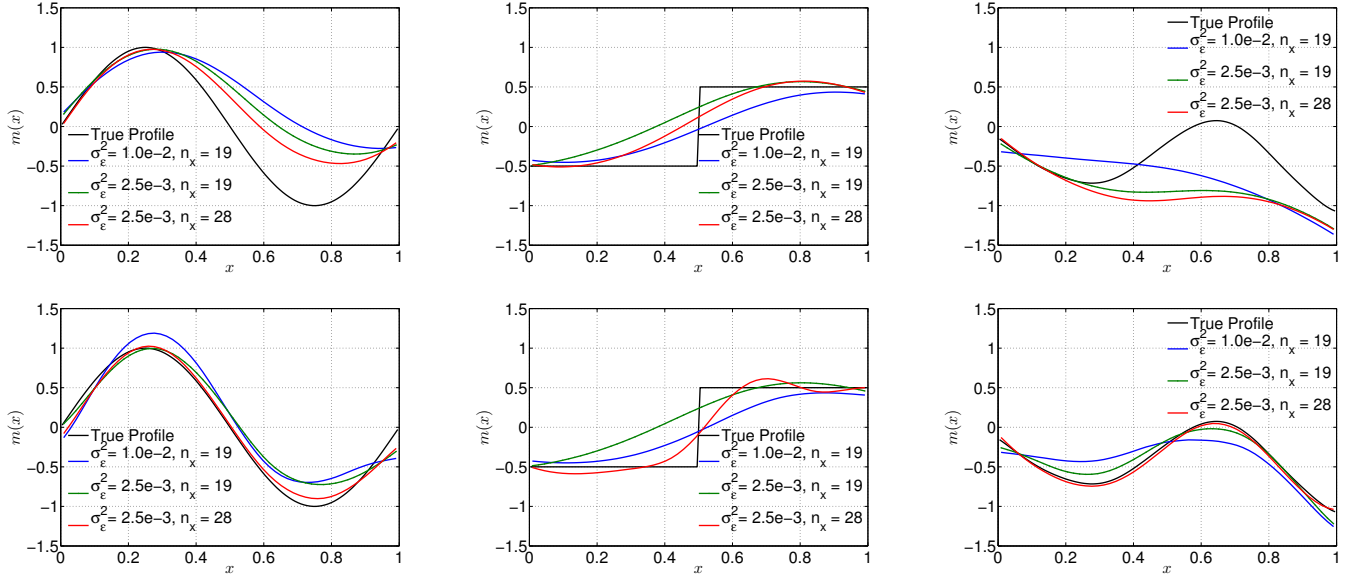


Figure 16: Effect of observations number and noise. Shown are the medians of the inferred profiles for the three test cases $m^{\text{sin,step,ran}}$ (from left to right), and inferences for a pre-assigned covariance function (top row) or with hyper-parameters (bottom row).

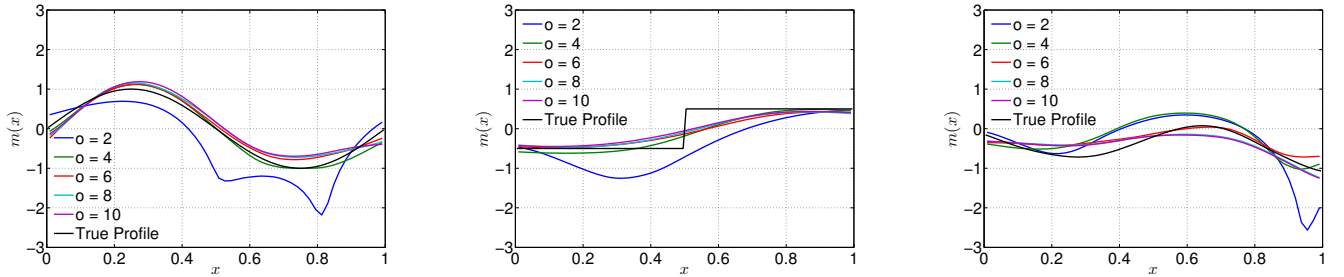


Figure 17: Effect of PC order o on the inferred median of the posterior: cases of (Left) m^{sin} , (Center) m^{step} and (Right) m^{ran} .

Acknowledgments

Research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST). OLM and OK also acknowledge partial support provided the US Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research, under Award Number DE-SC0008789.

References

- [1] A. Malinverno, Parsimonious bayesian markov chain monte carlo inversion in a nonlinear geophysical problem, *Geophysical Journal International* 151 (3) (2002) 675–688.
- [2] Y. M. Marzouk, H. N. Najm, L. A. Rahn, Stochastic spectral methods for efficient Bayesian solution of inverse problems, *Journal of Computational Physics* 224 (2) (2007) 560–586.

- [3] J. Winokur, P. Conrad, I. Sraj, O. Knio, A. Srinivasan, W. Thacker, Y. Marzouk, M. Iskandarani, A priori testing of sparse adaptive polynomial Chaos Expansions using an ocean general circulation model database, *Comput. Goesci.* 17 (2013) 899–911.
- [4] I. Sraj, M. Iskandarani, A. Srinivasan, W. C. Thacker, J. Winokur, A. Alexanderian, C.-Y. Lee, S. S. Chen, O. M. Knio, Bayesian inference of drag parameters using Fanapi AXBT data, *Monthly Weather Review* 141 (2013) 2347–2367.
- [5] I. Sraj, M. Iskandarani, A. Srinivasan, W. C. Thacker, O. M. Knio, Drag parameter estimation using gradients and hessian from a Polynomial Chaos model surrogate, *Monthly Weather Review* 142 (2013) 933–941.
- [6] P. Mattern, K. Fennel, M. Dowd, Estimating time-dependent parameters for a biological ocean model using an emulator approach, *Journal of Marine Systems* 96–97 (2012) 32–47.
- [7] I. Sraj, K. Mandli, O. M. Knio, I. Hoteit, Uncertainty quantification and inference of Manning’s friction coefficients using DART buoy data during the Tohoku Tsunami, *Ocean Modelling* 83 (2014) 82–97.
- [8] L. Ge, K. Cheung, Spectral sampling method for uncertainty propagation in long-wave runup modeling, *Journal of Hydraulic Engineering* 137 (3) (2011) 277–288.
- [9] K. Sargsyan, C. Safta, R. Berry, B. Debusschere, H. Najm, Uncertainty quantification in climate modeling, in: *AGU Fall Meeting Abstracts*, Vol. 1, 2011, p. 0899.
- [10] A. H. Elsheikh, I. Hoteit, M. F. Wheeler, Efficient bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates, *Computer Methods in Applied Mechanics and Engineering* 269 (0) (2014) 515 – 537.
- [11] O. P. Le Maître, O. M. Knio, *Spectral Methods for Uncertainty Quantification*, Springer Series in Scientific Computing, Springer-Verlag, 2010.
- [12] Y. M. Marzouk, H. N. Najm, Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems, *Journal of Computational Physics* 228 (2009) 1862–1902.
- [13] C. E. Rasmussen, C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*, The MIT Press, 2005.
- [14] P. M. Tagade, H.-L. Choi, A generalized polynomial chaos-based method for efficient bayesian calibration of uncertain computational models, *Inverse Problems in Science and Engineering* 22 (4) (2014) 602–624.
- [15] A. Alexanderian, O. Le Maître, H. Najm, M. Iskandarani, O. Knio, Multiscale stochastic preconditioners in non-intrusive spectral projection, *Journal of Scientific Computing* (2011) 1–35.
- [16] S. Zedler, G. Kanschat, R. Korty, I. Hoteit, A new approach for the determination of the drag coefficient from the upper ocean response to a tropical cyclone: a feasibility study., *Journal of Oceanography* 68 (2) (2012) 227 – 241.
- [17] R. Olson, R. Sriver, M. Goes, N. M. Urban, H. D. Matthews, M. Haran, K. Keller, A climate sensitivity estimate using Bayesian fusion of instrumental observations and an Earth System model, *Journal of Geophysical Research* 117 (2012) D04103.
- [18] D. S. Sivia, *Data Analysis - A Bayesian Tutorial*, Oxford Science Publications, Oxford, 2006.
- [19] H. Haario, E. Saksman, J. Tamminen, An adaptive Metropolis Algorithm, *Bernoulli* 7 (2) (2001) 223–242.
- [20] G. O. Roberts, J. S. Rosenthal, Examples of adaptive MCMC, *Journal of Computational and Graphical Statistics* 18 (2) (2009) 349–367.

- [21] G. Mircea, Stochastic calculus: Applications in science and engineering, Birkhauser, Boston, 2012.
- [22] R. Ghanem, P. Spanos, Stochastic Finite Elements: A Spectral Approach, Dover, 2002, 2nd edition.
- [23] R. J. Adler, J. E. Taylor, Random fields and geometry, Springer monographs in mathematics, Springer, New York, 2007.
- [24] A. H. Elsheikh, M. F. Wheeler, I. Hoteit, Hybrid nested sampling algorithm for bayesian model selection applied to inverse subsurface flow problems, *Journal of Computational Physics* 258 (0) (2014) 319 – 337.
- [25] P. Lax, Linear Algebra, Wiley-Interscience, 1996.
- [26] M. Salloum, A. Alexandrian, O. Le Maître, H. Najm, O. Knio, A simplified CSP analysis of a stiff stochastic ODE system, *Computer Methods in Applied Mechanics and Engineering* 217-229 (2012) 121–138.
- [27] N. Wiener, The homogeneous chaos, *American Journal of Mathematics* 60 (1938) 897–936.
- [28] R. H. Cameron, W. T. Martin, The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals, *Ann. Math.* 48 (1947) 385–392.
- [29] M. Reagan, H. Najm, R. Ghanem, O. Knio, Uncertainty quantification in reacting flow simulations through non-intrusive spectral projection, *Combustion and Flame* 132 (2003) 545–555.
- [30] P. Constantine, M. Eldred, E. Phipps, Sparse pseudospectral approximation method, *Computer Methods in Applied Mechanics and Engineering* 229-232 (2012) 1–12.
- [31] P. Conrad, Y. Marzouk, Adaptive smolyak pseudospectral approximations, *SIAM J. Sci. Comp.* 35 (6) (2013) 2643–2670.
- [32] M. Berveiller, B. Sudret, M. Lemaire, Stochastic finite element : a non intrusive approach by regression, *Eur. J. Comput. Mech.* 15 (2006) 81–92.
- [33] G. Blatman, B. Sudret, Adaptive sparse Polynomial Chaos expansion based on Least Angle Regression, *J. Comput. Phys.* 230 (6) (2011) 2345–2367.
- [34] J. Peng, J. Hampton, A. Doostan, A weighted ℓ_1 minimization approach for sparse polynomial chaos expansions, *Journal of Computational Physics* 267 (2014) 92–111.
- [35] I. Babuška, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data, *SIAM J. Numer. Anal.* 45 (3) (2007) 1005–1034.
- [36] D. Xiu, J. Hesthaven, High-order collocation methods for differential equations with random inputs, *SIAM J. Sci. Comput.* 27 (3) (2005) 1118–1139.
- [37] F. Nobile, R. Tempone, C. Webster, A sparse grid stochastic collocation method for partial differential equations with random input data, *SIAM J. Numer. Anal.* 46 (5) (2008) 2309–2345.
- [38] O. Le Maître, O. Knio, B. Debusschere, H. Najm, R. Ghanem, A multigrid solver for two-dimensional stochastic diffusion equations, *Computer Methods in Applied Mechanics and Engineering* 92 (41-42) (2003) 4723–4744.
- [39] O. Le Maître, O. Knio, H. Najm, R. Ghanem, A stochastic projection method for fluid flow. I. Basic formulation, *Journal of Computational Physics* 173 (2001) 481–511.
- [40] O. Le Maître, M. Reagan, B. Debusschere, H. Najm, R. Ghanem, O. Knio, Natural convection in a closed cavity under stochastic non-Boussinesq conditions, *SIAM J. Sci. Comput.* 26 (2004) 375–394.

- [41] J. Kaipio, E. Somersalo, Statistical inverse problems: Discretization, model reduction and inverse crimes, *Journal of Computational and Applied Mathematics* 198 (2) (2007) 493 – 504, special Issue: Applied Computational Inverse Problems.
- [42] J. Wang, N. Zabaras, Hierarchical bayesian models for inverse problems in heat conduction, *Inverse Problems* 21 (1) (2005) 183.
- [43] A. Gelman, J. B. Carlin, H. L. Stern, D. B. Rubin, *Bayesian Data Analysis*, 2nd Edition, Chapman and Hall/CRC, 2004.
- [44] A. Gelman, Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper), *Bayesian Analysis* 1 (3) (2006) 515–534.
- [45] B. W. Silverman, *Density estimation: for statistics and data analysis*, Chapman and Hall, 1986.
- [46] E. Parzen, On estimation of a probability density function and mode, *The Annals of Mathematical Statistics* 33 (3) (1962) 1065–1076.
- [47] S. Kullback, *Information theory and statistics*, John Wiley and Sons, NY, 1959.