Journal of Computational Physics 336 (2017) 36–50

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Multi-fidelity Gaussian process regression for prediction of random fields

L. Parussini [a], D. Venturi [b,*], P. Perdikaris [c], G.E. Karniadakis [d]

[a] Department of Engineering and Architecture, University of Trieste, Italy
[b] Department of Applied Mathematics and Statistics, University of California Santa Cruz, USA
[c] Department of Mechanical Engineering, Massachusetts Institute of Technology, USA
[d] Division of Applied Mathematics, Brown University, USA

ARTICLE INFO

Article history:
Received 17 July 2016
Received in revised form 10 November 2016
Accepted 23 January 2017
Available online 11 February 2017

Keywords:
Gaussian random fields
Multi-fidelity modeling
Recursive co-kriging
Uncertainty quantification

ABSTRACT

We propose a new multi-fidelity Gaussian process regression (GPR) approach for prediction of random fields based on observations of surrogate models or hierarchies of surrogate models. Our method builds upon recent work on recursive Bayesian techniques, in particular *recursive co-kriging*, and extends it to vector-valued fields and various types of covariances, including separable and non-separable ones. The framework we propose is general and can be used to perform uncertainty propagation and quantification in model-based simulations, multi-fidelity data fusion, and surrogate-based optimization. We demonstrate the effectiveness of the proposed recursive GPR techniques through various examples. Specifically, we study the stochastic Burgers equation and the stochastic Oberbeck–Boussinesq equations describing natural convection within a square enclosure. In both cases we find that the standard deviation of the Gaussian predictors as well as the absolute errors relative to benchmark stochastic solutions are very small, suggesting that the proposed multi-fidelity GPR approaches can yield highly accurate results.

1. Introduction

High-fidelity numerical simulations of complex stochastic dynamical systems require substantial computing time and data storage even in modern parallel architectures. This inherently limits the number of system states we can reliably simulate, thereby affecting accuracy when inferring statistical properties of any phase space function such as the performance of a certain engineering design. This basic observation has recently driven an explosive growth of fundamental and practical research at the interface of high-performance scientific computing, probability theory, and applied mathematics. One of the main features of such research is to replace expensive computational models with cheap surrogates or *hierarchies of surrogates*, and then come up with mathematical techniques leveraging on cross-correlations between the output of different surrogates to infer a quantity of interest. This yields a *multi-fidelity approach computational modeling*, which was first proposed by Kennedy and O'Hagan [19,20] in a Bayesian setting, and since then used in many different disciplines (see, e.g., [25,9]). For example, in [3] the authors presented a non-intrusive framework based on treed multi-output Gaussian processes, in which the response statistics are obtained through sampling a properly trained surrogate model of the physical system. The tree is built in a sequential way and its refinement depends on the observations through a global measure of

* Corresponding author.
E-mail address: venturi@ucsc.edu (D. Venturi).

http://dx.doi.org/10.1016/j.jcp.2017.01.047
0021-9991/

the uncertainty in the prediction, the inferred length scales, as well as the input probability distribution. Gaussian process regression has also been studied in the context of hierarchical multi-scale modeling of materials. For example, in [22] an adaptive moving kriging interpolation method was proposed to reduce the number of model evaluations at the fine-scale and inform the coarse-scale model with essential constitutive information. A multi-fidelity approach for minimizing the number of evaluations of expensive high-fidelity models was also proposed in [23], where the statistics of the high-fidelity model are computed based on realizations of a corrected low-fidelity surrogate. The correction function can be additive, multiplicative, or a combination of the two and it may be updated occasionally by high fidelity model evaluations.

In this paper we propose a new multi-fidelity Gaussian process regression (GPR) approach for predicting finite-dimensional random fields based on observations of surrogate models or hierarchies of surrogate models. The key idea relies on representing the map between the space of (random) Fourier or other spectral coefficients associated with any series expansion (relative to a spatial basis) and the physical space in which the random field develops. In this setting, the multi-fidelity inference problem for a random field reduces to an inference problem of a multivariate random vector of Fourier coefficients given data, i.e., vector samples produced by surrogate models at different levels of fidelity. To perform such inference, one can apply the recursive co-kriging technique recently developed by Le Gratiét et al. [15,13] (see also [25]) to each component of the vector of Fourier coefficients. From a Bayesian standpoint, this is equivalent to assuming independent priors for each model output, which may result in loss of information. To overcome this issue, we extend the recursive co-kriging technique to a multivariate setting (statistical models with vector outputs), to capture the cross-correlation structure among different vector components. We consider both separable and non-separable priors and quantify the advantages and trade-offs of each approach. We remark that although the main focus of this paper is revolving around uncertainty propagation and quantification in model-based computations, the proposed framework can be readily applied to more general parametric studies such as multivariate inverse problems and multi-objective surrogate-based optimization.

The paper is organized as follows. In Section 2 we introduce the general framework along with some theoretical background. In Section 3 we give a brief overview of Gaussian process regression (GPR) methods for vector-valued random fields and introduce our multi-fidelity recursive GPR technique. In Section 4 we evaluate the accuracy of the proposed multi-fidelity GPR approach by applying it to the stochastic Burgers equation and a stochastic thermal convection problem. Finally, in Section 5 we summarize our main findings.

2. Methodology

Consider a scalar random field $u(\boldsymbol{x}, \boldsymbol{\xi})$ depending on a set of coordinates (or design variables) $\boldsymbol{x} \in \mathbb{R}^n$, as well as on a set of random parameters $\boldsymbol{\xi} \in \mathbb{R}^d$. The field u could be, e.g., the solution to a partial differential equation in which the boundary conditions are set to be random and represented in terms of $\boldsymbol{\xi}$. Suppose that $u(\boldsymbol{x}, \boldsymbol{\xi})$ is in a separable Hilbert space. This allows us to write the series expansion

$$u(\boldsymbol{x}, \boldsymbol{\xi}) \cong \sum_{i=1}^{k} a_i(\boldsymbol{\xi}) L_i(\boldsymbol{x}), \tag{1}$$

where $L_i(\boldsymbol{x})$ are basis functions depending on the coordinates (or design variables) $\boldsymbol{x}$ while $a_i(\boldsymbol{\xi})$ are functions of random variables $\boldsymbol{\xi}$. If $u(\boldsymbol{x}, \boldsymbol{\xi})$ is the solution to a stochastic PDE model, then $L_i(\boldsymbol{x})$ are usually set a priori (spatial basis functions), while the functions $a_i(\boldsymbol{\xi})$ are determined by the PDE, e.g., by computing its solution at specific values of $\boldsymbol{\xi}$ through the probabilistic collocation method [1,8,4]. At this point we pose the following question: can we determine a model for the random vector field

$$\boldsymbol{a}(\boldsymbol{\xi}) = [a_1(\boldsymbol{\xi}) \cdots a_k(\boldsymbol{\xi})] \tag{2}$$

based on data collected at a specific nodes in the $\boldsymbol{\xi}$-space? This question is obviously not new and researchers have been working on it for decades. For instance, one can use polynomial interpolation of each $a_i(\boldsymbol{\xi})$ at Chebyshev sparse grids [2]. However, this implicitly assumes that $a_i(\boldsymbol{\xi})$ is a multivariate polynomial (which we do not know for sure), and also that we can predict the value of $a_i(\boldsymbol{\xi})$ at a non-observed location with probability 1, i.e., with no uncertainty. This is obviously not correct from a statistical standpoint. In a more robust setting, $a_i(\boldsymbol{\xi})$ should be considered as a random field with known values at observed points. Following the classical literature [18,19,6,29,12], we shall assume that the distribution of the vector $\boldsymbol{a}(\boldsymbol{\xi})$, conditional to the realization $\{\boldsymbol{a}_1 = \boldsymbol{a}(\boldsymbol{\xi}_1), ..., \boldsymbol{a}_n = \boldsymbol{a}(\boldsymbol{\xi}_n)\}$, is Gaussian with mean $\boldsymbol{m}(\boldsymbol{\xi})$ and (matrix-valued) covariance function $\boldsymbol{C}(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)$, i.e.,

$$\boldsymbol{a}(\boldsymbol{\xi}) | \boldsymbol{a}_1, ..., \boldsymbol{a}_n \sim \mathcal{GP}\left(\boldsymbol{m}(\boldsymbol{\xi}), \boldsymbol{C}(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)\right). \tag{3}$$

As we shall see in the subsequent sections, this setting allows us to build a multi-fidelity Gaussian process regression framework, in which observations of $\boldsymbol{a}(\boldsymbol{\xi})$ obtained from models with different levels of fidelity are combined in a seamless way to yield a highly accurate Gaussian predictor of $\boldsymbol{a}(\boldsymbol{\xi})$.

2.1. Calculation of statistical moments: measuring the uncertainty of uncertainty

In the traditional uncertainty quantification setting, $a_i(\boldsymbol{\xi})$ are considered as deterministic functions of the random variables $\boldsymbol{\xi}$, e.g., multivariate orthogonal polynomials selected from the Askey scheme [35,33,7,32], or generalized bi-orthogonal series expansion [5,30]. This allows us to evaluate the statistical moments of $u(\boldsymbol{x}, \boldsymbol{\xi})$ in terms of integrals with respect to the probability density function of $\boldsymbol{\xi}$ (assuming it exists), i.e.,

$$\mathbb{E}_{\boldsymbol{\xi}}\left[u(\boldsymbol{x}, \boldsymbol{\xi})^n\right] = \int u(\boldsymbol{x}, \boldsymbol{\xi})^n p(\boldsymbol{\xi}) d\boldsymbol{\xi}. \tag{4}$$

A substitution of (1) into (4), yields the following expressions for the first and second moment

$$\mathbb{E}_{\boldsymbol{\xi}}\left[u(\boldsymbol{x}, \boldsymbol{\xi})\right] = \sum_{i=1}^{k} L_i(\boldsymbol{x}) \int a_i(\boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi}, \tag{5}$$

$$\mathbb{E}_{\boldsymbol{\xi}}\left[u(\boldsymbol{x}, \boldsymbol{\xi})^2\right] = \sum_{i,j=1}^{k} L_i(\boldsymbol{x}) L_j(\boldsymbol{x}) \int a_i(\boldsymbol{\xi}) a_j(\boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi}. \tag{6}$$

Next, suppose that $a_i(\boldsymbol{\xi})$ are Gaussian random fields, with mean $\boldsymbol{m}$ and cross covariance $\boldsymbol{C}$ (see Eq. (3)). Under this assumption, the first- and second-order moments (5)–(6) are no longer deterministic quantities, but they are rather random fields. In fact, the expectation $\mathbb{E}_{\boldsymbol{\xi}}[\cdot]$ becomes a *conditional expectation*, i.e., $\mathbb{E}_{\boldsymbol{\xi}|\boldsymbol{a}}[\cdot]$ (integral with respect to $\boldsymbol{\xi}$ given a realization of the Gaussian random field $\boldsymbol{a}(\boldsymbol{\xi})$). This yields a Gaussian distribution for the mean field (5) (infinite sum of Gaussian random variables), while the second-order moment has the distribution of an indefinite quadratic form in Gaussian random variables [26,28]. With such distributions we can measure the *uncertainty of uncertainty*, i.e., how accurate we can be when we compute the statistical properties of the random field $u(\boldsymbol{x}, \boldsymbol{\xi})$ given that we are using a stochastic (Gaussian) model for the functions $a_i(\boldsymbol{\xi})$, which depend on the random variables $\boldsymbol{\xi}$. These observations can be generalized, and new ways of approximating integrals and derivatives of functions based on *stochastic interpolants* can be built (see, e.g., [16]). If we take the average of (1) relative to the conditional distribution (3), then we immediately obtain

$$\mathbb{E}_{\boldsymbol{\xi}|\boldsymbol{a}}\left[u(\boldsymbol{x}, \boldsymbol{\xi})\right] = \sum_{i=1}^{k} m_i(\boldsymbol{\xi}) L_i(\boldsymbol{x}), \tag{7}$$

$$\mathbb{E}_{\boldsymbol{\xi}|\boldsymbol{a}}\left[u(\boldsymbol{x}, \boldsymbol{\xi})^2\right] = \sum_{i,j=1}^{k} C_{ij}(\boldsymbol{\xi}, \boldsymbol{\xi}) L_i(\boldsymbol{x}) L_j(\boldsymbol{x}), \tag{8}$$

where C_{ij} is the cross-covariance of the random fields $a_i(\boldsymbol{\xi})$ and $a_j(\boldsymbol{\xi})$. In the next section, we discuss different Gaussian process regression (GPR) methods – including the proposed multivariate recursive co-kriging approach – to infer the conditional distribution (3) in a *multi-fidelity* setting.

3. Gaussian process regression (GPR)

In this section, we give a brief overview of Gaussian process regression methods (GPR) for scalar and vector-valued fields depending on an arbitrary number of variables. We also introduce our multi-fidelity recursive GPR method.

3.1. Kriging and recursive co-kriging for scalar fields

Consider a deterministic real-valued scalar field $A(\boldsymbol{\xi})$ depending on k variables $\boldsymbol{\xi} = (\xi_1, ..., \xi_k)$. We think of $A(\boldsymbol{\xi})$ as a realization of a Gaussian random field $a(\boldsymbol{\xi})$ in the form

$$a(\boldsymbol{\xi}) = m(\boldsymbol{\xi}) + y(\boldsymbol{\xi}), \tag{9}$$

where $m(\boldsymbol{\xi})$ is a deterministic function, usually called *trend function*, representing the mean of $a(\boldsymbol{\xi})$, while $y(\boldsymbol{\xi})$ is a zero-mean Gaussian random field with covariance function

$$C(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\theta}) = \sigma^2 r(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\theta}). \tag{10}$$

Here σ^2 is a scale parameter – the *process variance* – and r is a symmetric and positive definite kernel depending on a set of hyperparameters $\boldsymbol{\theta}$. The function $m(\boldsymbol{\xi})$ in (9) can have different forms. In simple kriging we assume $m(\boldsymbol{\xi}) = \mu$, where μ is a known constant. A more popular choice is $m(\boldsymbol{\xi}) = \beta$, where β is a regression parameter learned from data (ordinary kriging). More generally, we can model $m(\boldsymbol{\xi})$ in terms of series expansions relative to known basis functions $h_j(\boldsymbol{\xi})$ (universal kriging) as [13]

$$m(\boldsymbol{\xi}) = \sum_{j=1}^{p} \beta_j h_j(\boldsymbol{\xi}). \tag{11}$$

The basis functions $\{h_1(\boldsymbol{\xi}), ..., h_p(\boldsymbol{\xi})\}$ are often chosen to be polynomials,[1] while the expansion coefficients $\boldsymbol{\beta} = [\beta_1 \cdots \beta_p]^T$ depend on data and they are usually learned though maximum likelihood estimates. Now, suppose we have available observations of $A(\boldsymbol{\xi})$ at n distinct nodes $(\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_n)$, i.e., observations of the random field (9) at $\boldsymbol{\xi}_i$ $(i = 1, ..., n)$. It is well-known that the (conditional) posterior distribution of $a(\boldsymbol{\xi})$ given data $\boldsymbol{A} = [A(\boldsymbol{\xi}_1) \cdots A(\boldsymbol{\xi}_n)]^T$ and the regression parameters $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and σ^2 is Gaussian with mean

$$m_a(\boldsymbol{\xi}) = \boldsymbol{h}(\boldsymbol{\xi})\boldsymbol{\beta} + \boldsymbol{r}^T(\boldsymbol{\xi})\boldsymbol{R}^{-1}(\boldsymbol{A} - \boldsymbol{H}\boldsymbol{\beta}), \tag{12}$$

and variance

$$s_a^2(\boldsymbol{\xi}) = \sigma^2 \left[1 - \boldsymbol{r}^T(\boldsymbol{\xi})\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{\xi}) + \boldsymbol{q}(\boldsymbol{\xi}) \left(\boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{H} \right)^{-1} \boldsymbol{q}(\boldsymbol{\xi})^T \right]. \tag{13}$$

In equations (12) and (13) we defined $\boldsymbol{h}(\boldsymbol{\xi}) = [h_1(\boldsymbol{\xi}) \cdots h_p(\boldsymbol{\xi})]$,

$$\boldsymbol{q}(\boldsymbol{\xi}) = \boldsymbol{h}(\boldsymbol{\xi}) - \boldsymbol{r}^T(\boldsymbol{\xi})\boldsymbol{R}^{-1}\boldsymbol{H}, \tag{14}$$

and

$$\boldsymbol{r}(\boldsymbol{\xi}) = \begin{bmatrix} r(\boldsymbol{\xi}, \boldsymbol{\xi}_1; \boldsymbol{\theta}) \\ \vdots \\ r(\boldsymbol{\xi}, \boldsymbol{\xi}_n; \boldsymbol{\theta}) \end{bmatrix}, \qquad \boldsymbol{R} = \begin{bmatrix} \boldsymbol{r}(\boldsymbol{\xi}_1)^T \\ \vdots \\ \boldsymbol{r}(\boldsymbol{\xi}_n)^T \end{bmatrix} \qquad \boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}(\boldsymbol{\xi}_1) \\ \vdots \\ \boldsymbol{h}(\boldsymbol{\xi}_n) \end{bmatrix} \tag{15}$$

All parameters $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and σ^2 are learned from data by classical maximum likelihood estimation. We remark that equations (12)–(13) hold for the universal kriging model, i.e., when $m(\boldsymbol{\xi})$ is given in (11). Similar equations can be obtained for the simple kriging by simply integrating out $\boldsymbol{\beta}$ from the conditional posterior of the universal kriging.

3.1.1. Multi-fidelity recursive co-kriging for scalar fields

Co-kriging was originally proposed for enhancing the accuracy of a high-fidelity surrogate using supplementary observations of low-fidelity models. In essence, it aims at exploiting the cross-correlation between two or more Gaussian process surrogates through a stochastic auto-regressive scheme, and it yields a predictive posterior distribution for the high-fidelity model output that encodes the contribution of lower fidelity levels with quantified uncertainty. Co-kriging was first proposed by Kennedy and O'Hagan in the landmark paper [19], and since then it has been used in many different applications, such as surrogate-based optimization [9]. Unfortunately, the algorithm described in [19] has an unfavorable cubic scaling with the number of observed data and the number of surrogate models. To overcome this issue, Le Gratiét et al. [15,13] recently proposed a recursive version of the co-kriging algorithm, which allows us to break up the computation of posterior distribution into a sequence of distinct inferences of smaller dimensions. The advantage of this approach is clear: the inversion of the small covariance matrices associated with the sequence of inferences is less expensive and better conditioned than the inversion of the large covariance matrix associated with the full inference problem.

To describe the recursive co-kriging approach, suppose we have available multiple models, say s-models, providing an estimate of the real-valued scalar field

$$\begin{aligned} A &: \mathbb{R}^d \to \mathbb{R} \\ &\boldsymbol{\xi} \to A(\boldsymbol{\xi}) \end{aligned}$$

We denote by $A^{(1)}(\boldsymbol{\xi}), ..., A^{(s)}(\boldsymbol{\xi})$ the output of such models, ranked in terms of increasing fidelity,[2] $A^{(1)}(\boldsymbol{\xi})$ being the model with lowest level of fidelity. Usually, the computational cost of a model is proportional to its fidelity, i.e., the higher the fidelity the higher the computational cost. This means that we cannot usually afford sampling $A^{(s)}(\boldsymbol{\xi})$ extensively. On the other hand, we can usually afford many samples of $A^{(1)}(\boldsymbol{\xi})$, but such samples will not be accurate. The recursive co-kriging technique proposed in [14,13] aims at constructing an accurate estimate of $A(\boldsymbol{\xi})$ by leveraging on data from all models $A^{(j)}(\boldsymbol{\xi})$ $(j = 1, ..., s)$, in a *computationally efficient* way. The key idea is to represent each model output in terms of a Gaussian random field $a^{(j)}(\boldsymbol{\xi})$ and then make the hypothesis that such fields are related to each other by the autoregressive model

$$a^{(j)}(\boldsymbol{\xi}) = \rho^{(j-1)}(\boldsymbol{\xi})\tilde{a}^{(j-1)}(\boldsymbol{\xi}) + z^{(j)}(\boldsymbol{\xi}), \qquad j = 2, .., s \tag{16}$$

[1] For example, a linear basis in one dimension is $\boldsymbol{h}(\xi) = [1, \xi]$.

[2] Assessing the level of fidelity of mathematical model in the absence of prior information or data is a challenging problem.

where $\tilde{a}^{(j-1)}(\boldsymbol{\xi})$ and $z^{(j)}(\boldsymbol{\xi})$ are appropriate Gaussian random fields, while $\rho^{(j-1)}(\boldsymbol{\xi})$ are deterministic scaling fields (see [15,14,13] for further details). It can be shown that the posterior distribution of $a^{(j)}(\boldsymbol{\xi})$, conditional on the observations and parameters of all lower-fidelity models, is Gaussian with mean and variance given by

$$m^{(j)}(\boldsymbol{\xi}) = \rho^{(j-1)}(\boldsymbol{\xi}) m^{(j-1)}(\boldsymbol{\xi}) + Z^{(j)}(\boldsymbol{\xi}), \tag{17}$$

$$s^{(j)}(\boldsymbol{\xi})^2 = \rho^{(j-1)}(\boldsymbol{\xi})^2 s^{(j-1)}(\boldsymbol{\xi})^2 + W^{(j)}(\boldsymbol{\xi}), \tag{18}$$

where $Z^{(j)}(\boldsymbol{\xi})$ and $W^{(j)}(\boldsymbol{\xi})$ are functions that can be explicitly computed based on univariate GPR results at fidelity levels j and $j-1$ (see [15,14,13] for further details). Note that the mean (17) and the variance (18) satisfy a Markov property, which in turn allows us to break up the co-kriging problem as formulated in [19] into a sequence of univariate GPR (single-level kriging). This is very convenient from a computational viewpoint. In [13], a fully Bayesian formulation of recursive co-kriging which incorporates prior information in the maximum likelihood estimation of the model hyperparameters is given. This allows us to speed up the estimation process by leveraging on analytical expressions.

3.2. Multi-fidelity recursive co-kriging for vector-valued fields

In this section, we generalize the recursive co-kriging approach discussed in the previous section to systems with multiple (vector) outputs. To this end, consider the real-valued vector field

$$\begin{aligned} &\boldsymbol{A} : \mathbb{R}^d \to \mathbb{R}^k, \\ &\qquad \boldsymbol{\xi} \to \boldsymbol{A}(\boldsymbol{\xi}). \end{aligned}$$

Suppose we have available s different models of $\boldsymbol{A}$, i.e., $\boldsymbol{A}^{(1)}(\boldsymbol{\xi}), ..., \boldsymbol{A}^{(s)}(\boldsymbol{\xi})$, ranked in terms of increasing fidelity, with $\boldsymbol{A}^{(1)}(\boldsymbol{\xi})$ being the model with lowest fidelity. As before, we assume that the computational cost of a model is proportional to its fidelity, i.e., the higher the fidelity the higher the computational cost. We represent the output of each model in terms of a vector-valued Gaussian random field $\boldsymbol{a}^{(j)}(\boldsymbol{\xi})$ and make the hypothesis that such fields are related to each other by the autoregressive model

$$\boldsymbol{a}^{(j)}(\boldsymbol{\xi}) = \boldsymbol{\rho}^{(j-1)}(\boldsymbol{\xi}) \circ \tilde{\boldsymbol{a}}^{(j-1)}(\boldsymbol{\xi}) + \boldsymbol{z}^{(j)}(\boldsymbol{\xi}), \qquad j = 2, .., s \tag{19}$$

where $\circ$ denotes the componentwise product between two vectors or matrices, i.e., the Hadamard product,

$$\boldsymbol{\rho}^{(j-1)}(\boldsymbol{\xi}) = \left(\boldsymbol{I}_k \otimes \boldsymbol{g}_{j-1}^T(\boldsymbol{\xi}) \right) \boldsymbol{\alpha}_{j-1} \tag{20}$$

is a deterministic scaling vector field represented in terms of known basis functions $\boldsymbol{g}_{j-1}(\boldsymbol{\xi}) = [g_{j-1,1}(\boldsymbol{\xi}) \cdots g_{j-1,p}(\boldsymbol{\xi})]$ ($\otimes$ the tensor product operator, $\boldsymbol{I}_k$ is the $k \times k$ identity matrix and $\boldsymbol{\alpha}_{j-1}$ are regression coefficients). Also, we assume that

$$\boldsymbol{z}^{(j)}(\boldsymbol{\xi}) \sim \mathcal{GP}\left(\left(\boldsymbol{I}_k \otimes \boldsymbol{f}_j^T(\boldsymbol{\xi}) \right) \boldsymbol{\beta}_j, \boldsymbol{C}^{(j)}(\boldsymbol{\xi}, \boldsymbol{\xi}') \right), \tag{21}$$

$$\boldsymbol{a}^{(1)}(\boldsymbol{\xi}) \sim \boldsymbol{z}^{(1)}(\boldsymbol{\xi}), \tag{22}$$

and that $\tilde{\boldsymbol{a}}^{(j-1)}(\boldsymbol{\xi})$ has the conditional distribution

$$\tilde{\boldsymbol{a}}^{(j-1)}(\boldsymbol{\xi}) \sim \boldsymbol{a}^{(j-1)}(\boldsymbol{\xi}) \,\Big|\, \boldsymbol{z}^{(j-1)}(\boldsymbol{D}_{j-1}) = \boldsymbol{A}^{(j-1)}(\boldsymbol{D}_{j-1}),\, \boldsymbol{\beta}_{j-1},\, \boldsymbol{\alpha}_{j-2},\, \boldsymbol{C}_{j-1}(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\theta}_{j-1}). \tag{23}$$

In these equations, $\boldsymbol{f}_j(\boldsymbol{\xi}) = [f_{j,1}(\boldsymbol{\xi}) \cdots f_{j,q}(\boldsymbol{\xi})]$ is a vector of known basis functions whose linear combination represents the mean of $\boldsymbol{z}^{(j)}$, $\boldsymbol{C}_j$ is the covariance of $\boldsymbol{z}^{(j)}$ and $\boldsymbol{D}_{j-1}$ is a set of nodes $\{\boldsymbol{\xi}_n\}$ in which we evaluate the output of the model $\boldsymbol{A}^{(j-1)}(\boldsymbol{\xi})$. We assume that such sets are nested, i.e.,

$$\boldsymbol{D}_s \subseteq \boldsymbol{D}_{s-1} \cdots \subseteq \boldsymbol{D}_1. \tag{24}$$

The hypothesis (23) allows us to break up the full multi-fidelity GPR regression for random vector fields into a sequence of independent GPRs at each level of fidelity. The posterior distribution of $\boldsymbol{a}^{(j)}(\boldsymbol{\xi})$, conditional to data at level j and model parameters is Gaussian, i.e.,

$$\boldsymbol{a}^{(j)}(\boldsymbol{\xi}) \,\Big|\, \boldsymbol{z}^{(j)}(\boldsymbol{D}_j) = \boldsymbol{A}^{(j)}(\boldsymbol{D}_j),\, \boldsymbol{\beta}_j,\, \boldsymbol{\alpha}_{j-1},\, \boldsymbol{C}_j(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\theta}_j) \sim \mathcal{GP}\left(\boldsymbol{m}^{(j)}(\boldsymbol{\xi}), \boldsymbol{C}^{(j)}(\boldsymbol{\xi}, \boldsymbol{\xi}') \right). \tag{25}$$

The mean and variance have expressions

$$\boldsymbol{m}^{(j)}(\boldsymbol{\xi}) = \boldsymbol{\rho}^{(j-1)}(\boldsymbol{\xi}) \circ \boldsymbol{m}^{(j-1)}(\boldsymbol{\xi}) + \boldsymbol{M}^{(j)}(\boldsymbol{\xi}), \tag{26}$$

$$\boldsymbol{C}^{(j)}(\boldsymbol{\xi}, \boldsymbol{\xi}') = \boldsymbol{\rho}^{(j-1)}(\boldsymbol{\xi})^2 \circ \boldsymbol{C}^{(j-1)}(\boldsymbol{\xi}, \boldsymbol{\xi}') + \boldsymbol{V}^{(j)}(\boldsymbol{\xi}, \boldsymbol{\xi}'). \tag{27}$$

Note that the mean and the variance of the predictive distribution have a Markovian structure, which allows us to perform GPR across different levels of fidelity independently. Regarding $\boldsymbol{M}^{(j)}(\boldsymbol{\xi})$ and $\boldsymbol{V}^{(j)}(\boldsymbol{\xi}, \boldsymbol{\xi}')$, a lengthy calculation shows that,

$$\boldsymbol{M}^{(j)}(\boldsymbol{\xi}) = \hat{\boldsymbol{f}}_j^T(\boldsymbol{\xi})\boldsymbol{\beta}_j + \boldsymbol{c}_j^T(\boldsymbol{\xi})\hat{\boldsymbol{C}}_j^{-1}\left[\boldsymbol{A}^{(j)}(\boldsymbol{D}_j) - \boldsymbol{P}_{j-1} \circ \boldsymbol{A}^{(j-1)}(\boldsymbol{D}_j) - \boldsymbol{f}_j^T(\boldsymbol{D}_j)\boldsymbol{\beta}_j\right], \tag{28}$$

$$\boldsymbol{V}^{(j)}(\boldsymbol{\xi}, \boldsymbol{\xi}') = \boldsymbol{C}_j(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\theta}_j) - \boldsymbol{c}_j^T(\boldsymbol{\xi})\hat{\boldsymbol{C}}_j\boldsymbol{c}_j(\boldsymbol{\xi}'), \tag{29}$$

where

$$\hat{\boldsymbol{f}}_j(\boldsymbol{\xi}) = \boldsymbol{I}_k \otimes \boldsymbol{f}_j(\boldsymbol{x}), \qquad \boldsymbol{c}_j(\boldsymbol{\xi}) = \boldsymbol{C}_j(\boldsymbol{\xi}, \boldsymbol{D}_j; \boldsymbol{\theta}_j), \qquad \hat{\boldsymbol{C}}_j = \boldsymbol{C}_j(\boldsymbol{D}_j, \boldsymbol{D}_j; \boldsymbol{\theta}_j), \tag{30}$$

and

$$\boldsymbol{P}_j = \left(\boldsymbol{I}_k \otimes \boldsymbol{g}_j(\boldsymbol{D}_j)\right)^T \boldsymbol{\alpha}_j. \tag{31}$$

We remark that equation (29) holds for simple kriging. We have also obtained the analytical expression of $V^{(j)}(\boldsymbol{\xi}, \boldsymbol{\xi}')$ corresponding to *universal kriging* (not shown here). So far, we have made no assumption on the covariance model $\boldsymbol{C}_j(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\theta}_j)$ we use for GPR at each level of fidelity. In the next subsection we discuss two important cases.

3.2.1. Separable covariance function

Suppose that covariance model $\boldsymbol{C}_j(\boldsymbol{\xi}, \boldsymbol{\xi}')$ is separable, in the sense that

$$\boldsymbol{C}_j(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\theta}_j) = r_j(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\theta}_j)\boldsymbol{\Sigma}_j, \tag{32}$$

where r_j is a correlation function and $\boldsymbol{\Sigma}_j$ is a matrix with fixed entries. The assumption (32) implies that there is only one spatial correlation function representing the cross covariance between all components of the model $\boldsymbol{a}^{(j)}(\boldsymbol{\xi})$. The separable covariance has a conjugate prior for $\boldsymbol{\Sigma}_j$, which allows us to integrate out $\boldsymbol{\Sigma}_j$ analytically from the posterior (25). This yields a multivariate Student-t conditional posterior [11,24] with mean and variance that can be computed analytically.

3.2.2. Non-separable covariance function

Unlike the separable case, a non-separable covariance function can have a different spatial correlation function for each component of the model output. In particular, we consider here the linear model of coregionalization (LMC), where

$$\boldsymbol{C}_j(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\theta}_j) = \boldsymbol{B}\left[\mathbf{diag}\left(r_1(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\theta}_1), ..., r_k(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\theta}_k)\right)\right]\boldsymbol{B}^T. \tag{33}$$

As is well known [21], depending on the matrix $\boldsymbol{B}$ we can have different models of coregionalization. For example, if the choose $\boldsymbol{B}$ to be diagonal (with positive entries), then the LMC is called *independent*; on the other hand, if we assume that $\boldsymbol{B}$ is symmetric and positive definite then the LMC is called *dependent*. In the latter case we can use the spectral decomposition of $\boldsymbol{B}\boldsymbol{B}^T$ to represent $\boldsymbol{B}$ efficiently. In both cases, the coefficients of the matrix $\boldsymbol{B}$ become additional parameters that have to be estimated when performing GPR at each level of fidelity.

4. Numerical results

In this Section we provide numerical results and study the accuracy of the multi-fidelity GPR approach we presented in this paper. To this end, we first study a simple pedagogical example, i.e., a real function in $[0, 2\pi]$ depending on one random parameter ξ. This allows us to validate our methods and assess their accuracy and computational efficiency. Subsequently, we apply our multi-fidelity GPR approach to the stochastic Burgers equation in 1D and to a stochastic thermal convection problem in 2D.

4.1. A pedagogical example

Consider a real function $u(x, \xi)$, periodic in $x \in [0, 2\pi]$, and depending on one random parameter ξ, which we assume to be uniformly distributed in $[-1, 1]$. Such function can be approximated by the Fourier series

$$u(x, \xi) = \sum_{i=1}^{k} a_i(\xi)L_i(x), \tag{34}$$

where $L_i(x)$ are trigonometric polynomials (odd expansion) ([17], p. 29). The number of modes k here defines the "fidelity" of the model, i.e., higher k corresponds to higher fidelity. Our goal is to predict $u(x, \xi)$ based on samples in the ξ-space obtained from series expansions with different number of modes, i.e., different fidelity. In particular, we consider the following prototype low-fidelity model

$$u_L(x, \xi) = \sum_{i=1}^{k_L} a_i^{(L)}(\xi)L_i(x), \qquad \text{(low-fidelity model)} \tag{35}$$

where $k_L = 5$ and

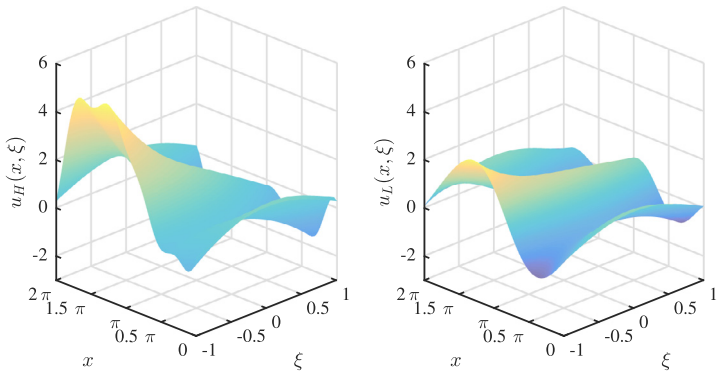

Fig. 1. Prototype high-fidelity (left) and low-fidelity (right) models defined in equations (36) and (35), respectively.

$$a_1^{(L)} = \cos(1.5\xi), \quad a_2^{(L)} = -\xi^2, \quad a_3^{(L)} = \left(0.5\xi^2 + 1\right)^{-1}, \quad a_4^{(L)} = e^{-\xi}, \quad a_5^{(L)} = \xi^2 - \xi.$$

On the other hand, the high-fidelity model is defined as

$$u_H(x,\xi) = \sum_{i=1}^{k_H} a_i^{(H)}(\xi) L_i(x), \qquad \text{(high-fidelity model)} \tag{36}$$

where $k_H = 9$ and

$$a_j^{(H)} = \sum_{i=1}^{k_L} a_i^{(L)}(\xi) L_i(x_j) - 0.5\xi x_j + 0.2 \qquad j = 1, \ldots, k_H. \tag{37}$$

Note that the coefficients $a_j^{(H)}$ are defined in terms of $a_j^{(L)}$. This makes the low-fidelity and high-fidelity models (35)–(36) correlated in both x and ξ. In Fig. 1 we plot $u_L(x,\xi)$ and $u_H(x,\xi)$ versus x and ξ. To determine the accuracy of the proposed multi-fidelity GPR[3] methods, we have computed the absolute error of both the predicted coefficients $a_j(\xi)$ and the predicted random function $u(x,\xi)$ relative to $a_j^{(H)}(\xi)$ and $u_H(x,\xi)$, respectively. In particular, we have used the root-mean-square errors (RMSE) defined as

$$RMSE(a_j) = \sqrt{\frac{1}{N_p}\sum_{p=1}^{N_p}\left(a_j^{(s)}(\xi_p) - a_j^{(H)}(\xi_p)\right)^2} \qquad j = 1, \ldots, k_H, \tag{38}$$

and

$$RMSE(u) = \sqrt{\frac{1}{N_p N_q}\sum_{p=1}^{N_p}\sum_{q=1}^{N_q}\left(u_s(x_q,\xi_p) - u_H(x_q,\xi_p)\right)^2}. \tag{39}$$

In the last two equations, $\{x_q\}_{q=1,\ldots,N_q}$ and $\{\xi_p\}_{p=1,\ldots,N_p}$ are evenly spaced nodes in $[0, 2\pi]$ and $[-1, 1]$, respectively, while $N_p = N_q = 51$. $a_j^{(s)}$ indicates the surrogate of the j-th coefficient $a_j^{(H)}$ and u_s the surrogate of high-level function u_H. Next, we compare the performance of different GPR approaches in reconstructing the high-fidelity model u_H based on samples of u_L and u_H. Specifically, we consider

1. Univariate co-kriging (Section 3.1);
2. Multivariate co-kriging with separable or non-separable covariance function (Section 3.2).

In each approach we studied the effects of different covariance models, i.e., Gaussian, 5/2-Matérn and Wendland [27,34], as well as the effects of different orders in the regressors of the trend function (see Section 3.1). Such effects are shown in Fig. 2 and Fig. 3 in the case of univariate kriging and co-kriging, respectively. It is seen that a high-order regressor of the trend function yield better accuracy, i.e., predictors with smaller standard deviation (see Fig. 2). Regarding the choice of the covariance model, our numerical results suggest that the Gauss and the Wendland models yield the most accurate

[3] In this example we have only two levels of fidelity defined by the models u_L and u_H.

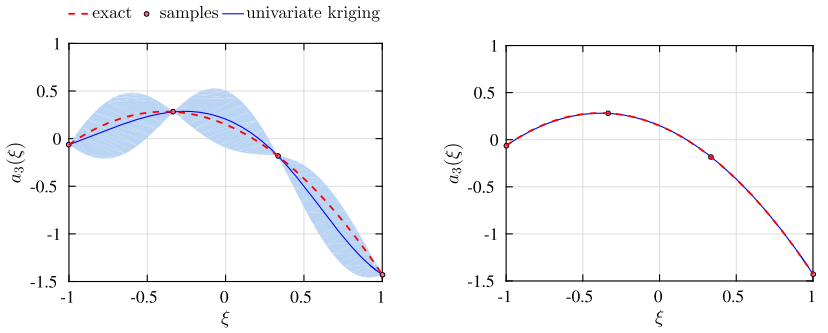

Fig. 2. Effect of the trend function in the univariate kriging model. We plot the mean and the standard deviation of the Gaussian predictor of $a_3(\xi)$ obtained with 4 evenly-spaced samples, 5/2-Matérn covariance and different polynomial order in the trend function (11): zero-order universal kriging (left), second-order universal kriging (right).

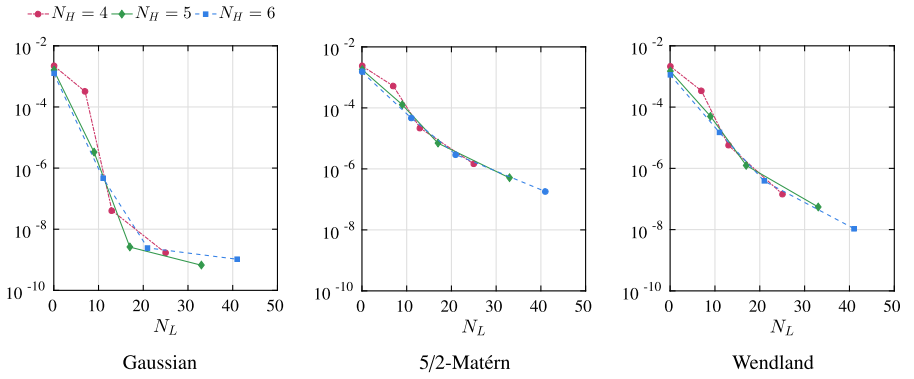

Fig. 3. Convergence of the univariate co-kriging model with respect to the number of low-fidelity samples for different covariance kernels. We plot Root-mean-square errors (38) of the coefficient $a_3(\xi)$ versus the number of low-fidelity samples (N_L). We show results obtained by using a different number of high-fidelity samples (N_H) and different covariance kernels: Gaussian, 5/2-Matérn and Wendland.

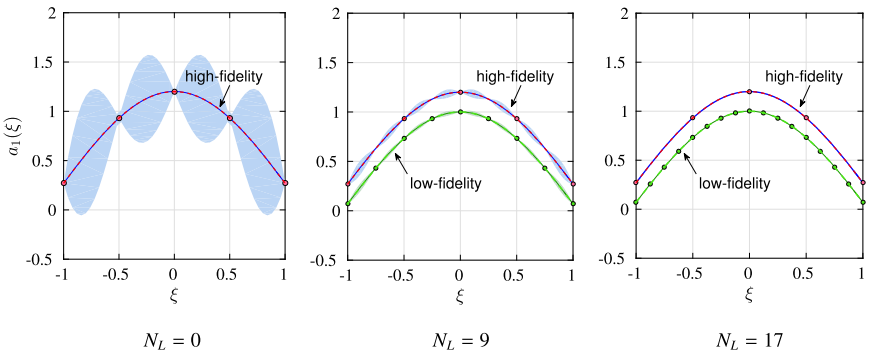

Fig. 4. Multivariate recursive co-kriging with non-separable covariance. Shown are results for the coefficient $a_1(\xi)$ obtained with $N_H = 5$ high-fidelity samples and a different number of low-fidelity samples. It is seen that as we increase N_L the co-kriging predictor becomes more and more accurate (the standard deviation of the high-fidelity predictor decreases).

predictions, but in terms of computational time 5/2-Matérn is more efficient. Also, the 5/2-Matérn and Wendland covariance models are better conditioned than the Gaussian models. As is well known, the accuracy of GPR depends on the functions we aim to predict as well as on the location of the sample points. Therefore, the conclusions we have just drawn about accuracy and efficiency may not be extended to other cases.

Next, we compare the accuracy of different GPR models in reconstructing the high-fidelity model u_H based on samples of u_L and u_H. We assume a 5/2-Matérn covariance and an order 2 for the regressors of the trend functions. A typical plot we obtained when inferring the modes $a_i^{(H)}(\xi)$ from multi-fidelity data is shown in Fig. 4. We see that for a fixed value of high-fidelity samples, the multi-variate recursive GPR with non-separable covariance converges to the high-fidelity model as we increase the number of low-fidelity runs. In Fig. 5 we compare the accuracy of the proposed GPR approaches. It is seen that univariate co-kriging and multivariate co-kriging with non-separable covariance functions yield similar convergence

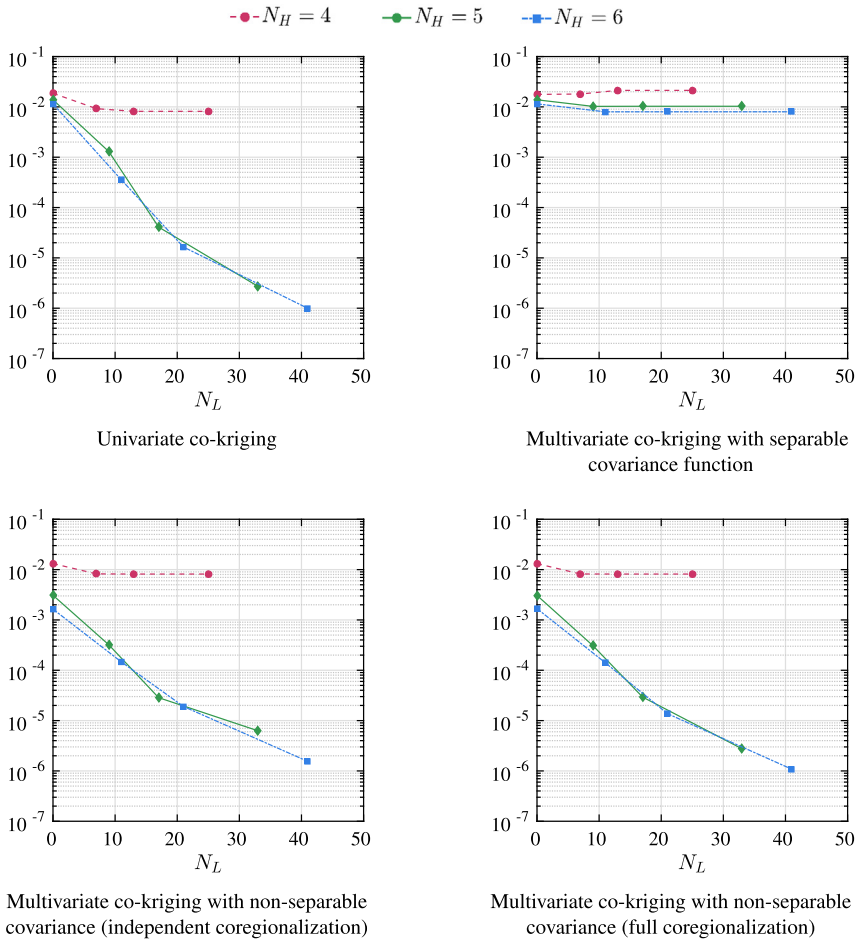

Fig. 5. Accuracy of different GPR approaches in reconstructing the high-fidelity model u_H based on samples of u_L and u_H. Shown are RMSEs (39) versus the number of low fidelity samples N_L. Each curve corresponds to a different number of high-fidelity samples.

rates. The multivariate co-kriging with separable covariance function in this case yields a plateau in the error slopes. This is due to the fact that the coefficients $a_i(\xi)$ have different correlation lengths, which cannot be effectively captured by a single correlation function (see Section 3.2.1). Regarding the computational time, univariate GPR is usually faster than multivariate GPR with non-separable covariances. The main bottleneck of the multivariate GPR is the optimization of the likelihood function, which depends on the model parameters of all functions $a_i(\xi)$. On the other hand, multivariate GPR with separable covariance function can be very efficient in situations where there is a strong statistical correlation between different $a_i(\xi)$, i.e., when all modes $a_i(\xi)$ can be modeled by a single correlation function at each fidelity level.

4.2. Stochastic Burgers equation

Consider the following initial/boundary value problem for the Burgers equation

$$\begin{cases} \dfrac{\partial u}{\partial t} + u\dfrac{\partial u}{\partial x} = \dfrac{1}{2}\dfrac{\partial^2 u}{\partial x^2} + f(x,t) \qquad x \in [0, 2\pi] \quad t \geq 0 \\ \text{Periodic B.C.} \\ u(x, 0, \boldsymbol{\xi}) = u_0(x; \boldsymbol{\xi}) \end{cases} \tag{40}$$

where $u_0(x, \boldsymbol{\xi})$ is a random initial condition depending on two uniformly distributed random variables (ξ_1, ξ_2) (uniform in $[-\sqrt{3}, \sqrt{3}]$)

$$u_0(x, \boldsymbol{\xi}) = 1 + \xi_1(\omega)\sin(x) + \xi_2(\omega)\cos(x) \tag{41}$$

and $f(x,t)$ is a deterministic forcing term defined as

$$f(x,t) = 8\sin(10x)\sin(5t) + 8\cos(7x)e^{-\sin(t)}. \tag{42}$$

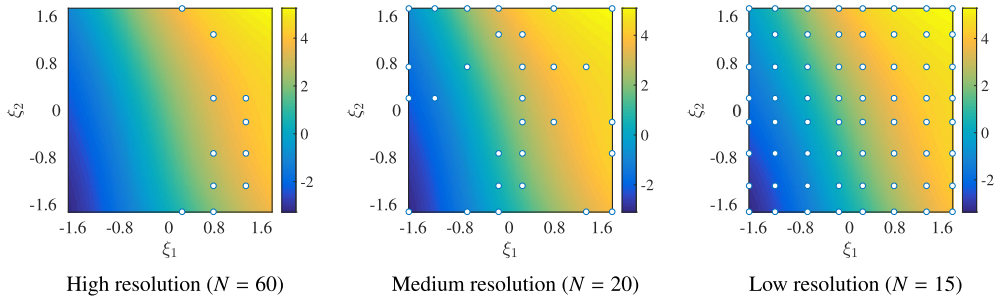

Fig. 6. Stochastic Burgers equation: Sample locations in (ξ_1, ξ_2)-space for each level of fidelity superimposed to the contour plots of the field $a_{35}(1, \xi_1, \xi_2)$ (see Eq. (43)).

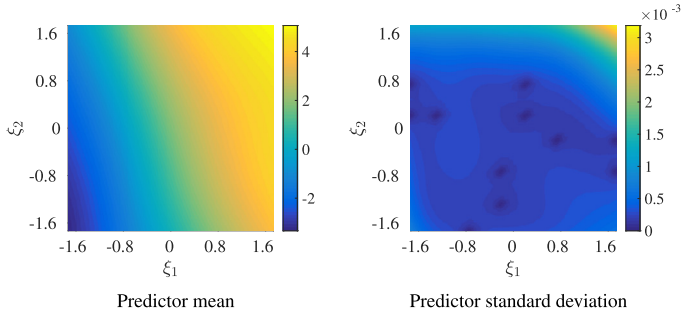

Fig. 7. Stochastic Burgers equation: Univariate recursive co-kriging. Predictor mean and standard deviation of the coefficient a_{35} at time $t = 1$. To determine the recursive co-kriging model, we computed 11 high-fidelity, 29 medium-fidelity and 64 low-fidelity solution samples of (40)–(42). The sample locations in the (ξ_1, ξ_2)-space are shown in Fig. 6.

We would like to use multi-fidelity GPR to infer the statistics of the random field $u(x, t, \boldsymbol{\xi})$ at time $t = 1$. To this end, we represent u in terms of the Fourier series

$$u(x, t, \xi_1, \xi_2) = \sum_{q=-N/2}^{N/2} a_q(t, \xi_1, \xi_2) e^{iqx} \tag{43}$$

and run stochastic simulations of (40) by using the probabilistic collocation method with a different number of spatial modes N, i.e., a different resolution in space. Specifically, we considered $N = 15$ (low-fidelity), $N = 20$ (medium-fidelity) and $N = 60$ (high-fidelity). We study the accuracy and the computational cost of multi-fidelity GPR in modeling the solution to (40)–(42) at $t = 1$. In particular, we consider the univariate recursive co-kriging approach,[4] in which we determine (in a multi-fidelity setting) each function $a_q(1, \xi_1, \xi_2)$ $(q = -N/2, ..., N/2)$ appearing in (43) independently of the others. We remark that in principle one can derive the analytical solution to the initial/boundary value problem (40) and determine $a_q(\xi_1, \xi_2, t)$. In this study, however, we rely on the accurate numerical stochastic solution we obtained on a tensor product Gauss–Legendre Lobatto grid of 40×40 points in (ξ_1, ξ_2) (see also [25]).

To determine the recursive co-kriging model of each $a_q(1, \xi_1, \xi_2)$, we computed 11 high-fidelity ($N = 60$), 29 medium fidelity ($N = 20$) and 64 low fidelity ($N = 15$) solution samples of equation (40). In Fig. 6 we show the sample locations for each level of fidelity superimposed to the contour plots of the field $a_{35}(1, \xi_1, \xi_2)$ we obtained from high-fidelity runs. In Fig. 7 we plot the mean and standard deviation of the Gaussian random field representing the coefficient $a_{35}(1, \xi_1, \xi_2)$ we obtained form the recursive co-kriging approach. It is seen that univariate GPR yields a predictor with good approximation properties, with relatively small standard deviation. Similar results are obtained for other coefficients $a_q(1, \xi_1, \xi_2)$. In Fig. 8 we plot the mean and standard deviation of the velocity field $u(x, t, \xi_1, \xi_2)$ at $t = 1$ we obtained from GPR modeling.

4.3. Stochastic Rayleigh–Bénard convection

In this section we study stochastic Rayleigh–Bénard convection of a Newtonian incompressible fluid within two-dimensional square enclosures. In particular we consider steady state solutions corresponding to a random (uniformly

[4] In this case multivariate recursive co-kriging is not practical. In fact, if we set $N = 60$ (high-fidelity simulation), we obtain a set of 121 functions $a_q(1, \xi_1, \xi_2)$, which yields a multivariate covariance model represented by a matrix 7381 times bigger than the univariate case (if we sample each a_q at the same set of nodes).

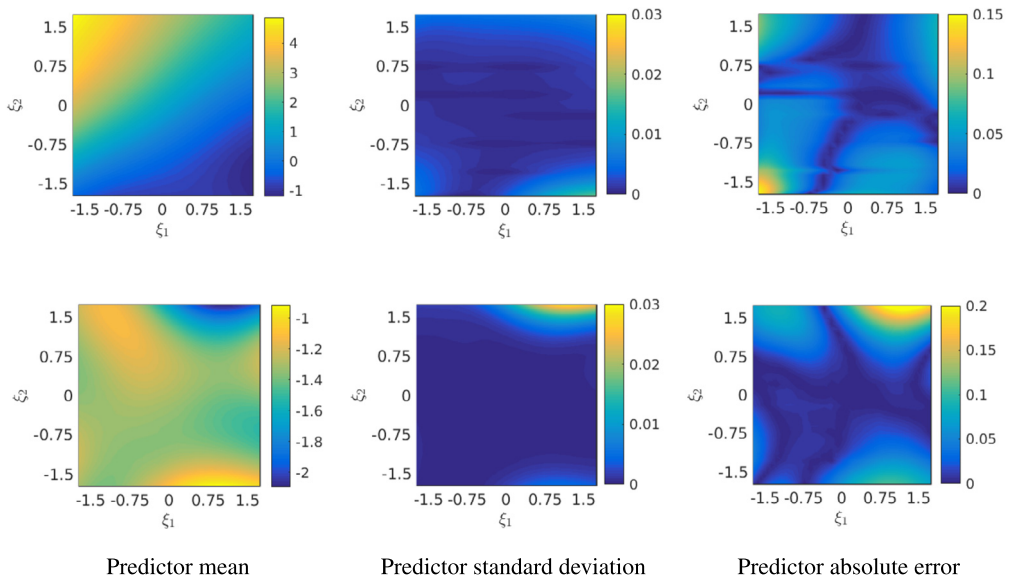

Fig. 8. Stochastic Burgers equation: Univariate recursive co-kriging. Shown are the predictor mean, the predictor standard deviation and the absolute error of the predictor mean relative to the exact solution to (40)–(42) at $t = 1$ and for two different sections: $x = 0$ (first row), $x = \pi$ (second row). To determine the recursive co-kriging model, we computed 11 high-fidelity, 29 medium-fidelity and 64 low-fidelity solution samples of (40)–(42). The sample locations in the (ξ_1, ξ_2)-space are shown in Fig. 6.

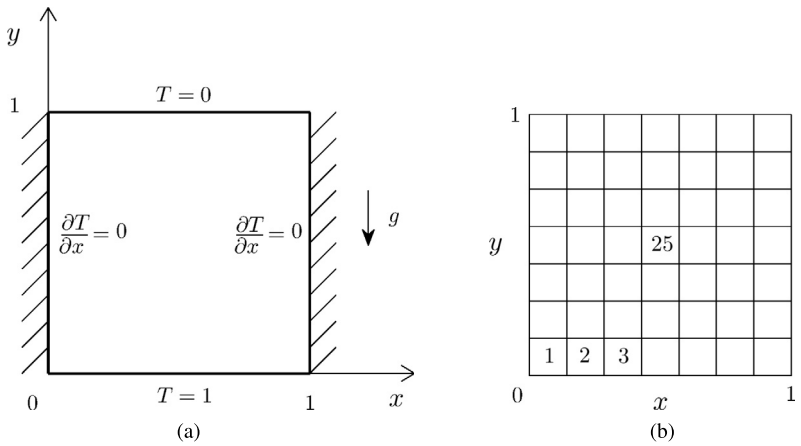

Fig. 9. (a) Schematic of dimensionless geometry and temperature boundary conditions. The sidewalls of the cavity are assumed to be adiabatic while the horizontal walls are kept at constant temperature. The velocity boundary conditions are of no-slip type, i.e. $\partial\psi/\partial x = 0$ and $\partial\psi/\partial y = 0$ at the walls. (b) Spectral element mesh we used for the local modeling of the temperature field.

distributed) Rayleigh number [31]. Such solutions are governed by the Oberbeck–Boussinesq equations

$$\frac{\partial\psi}{\partial y}\frac{\partial\left(\nabla^2\psi\right)}{\partial x} + \frac{\partial\psi}{\partial x}\frac{\partial\left(\nabla^2\psi\right)}{\partial y} = -Pr\nabla^4\psi + Ra(\omega)Pr\frac{\partial T}{\partial x}, \tag{44}$$

$$\frac{\partial\psi}{\partial y}\frac{\partial T}{\partial x} - \frac{\partial\psi}{\partial x}\frac{\partial T}{\partial y} = \nabla^2 T, \tag{45}$$

where ψ denotes the dimensionless stream function, T the dimensionless temperature field, while Ra and Pr are the Rayleigh and the Prandtl numbers, respectively. A sketch of the geometry together with the boundary conditions associated with the system (44)–(45) is shown in Fig. 9. We represent the temperature and velocity fields in terms of series expansions relative to globally defined eigenfunctions $\psi_n(x, y)$ and $\Gamma_m(x, y)$ as (see [31])

$$\psi(x, y) = \sum_{n=1}^{N_v} A_n(Ra)\,\psi_n(x, y), \qquad T(x, y) = \sum_{m=1}^{N_T} B_m(Ra)\,\Gamma_m(x, y). \tag{46}$$

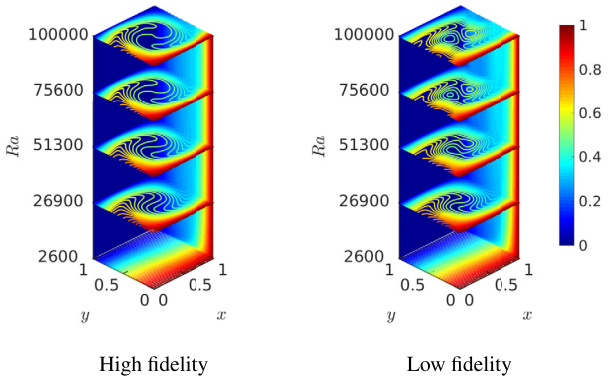

Fig. 10. Stochastic Rayleigh–Bénard convection. Steady-state temperature field corresponding to one-roll convection patterns within a wide range of Rayleigh numbers, i.e., between 2.6×10^3 and 10^5. Shown are simulations results with different resolutions in physical space.

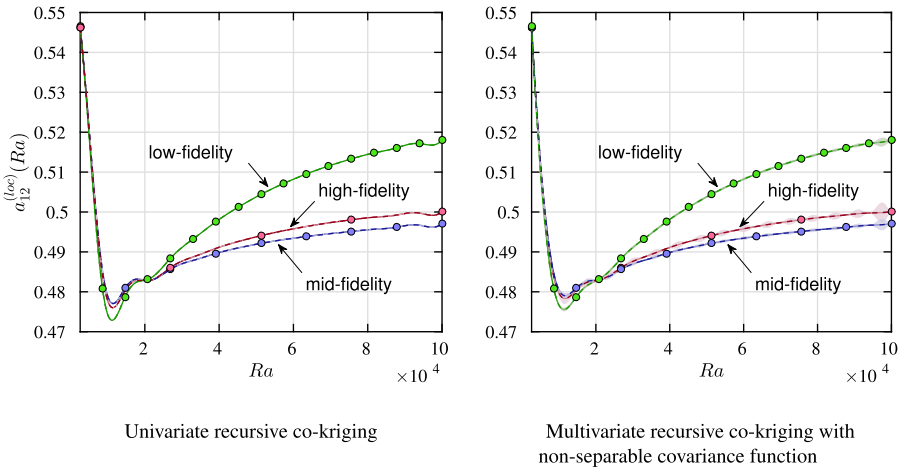

Fig. 11. Stochastic Rayleigh–Bénard convection. GP representation of the coefficient $a_{12}^{(loc)}(Ra)$ in the expansion of the local temperature field (47) within the spectral element 25 shown in Fig. 9. We plot the predictor means (lines) and standard deviations (bands) obtained with univariate recursive co-kriging and multivariate recursive co-kriging with non-separable covariance model.

Our goal is to represent the steady-state temperature field corresponding to the one-roll flow pattern by using GPR within a wide range of Rayleigh numbers, i.e., between 2.6×10^3 and 10^5. To this end, we performed variable fidelity simulations of the system (44)–(45) both in physical space (i.e., by varying N_v and N_T in (46)) as well as in probability space, (i.e., by sampling Ra at a different number of collocation points). In particular, we considered three levels of fidelity in physical space: the low-fidelity solution is obtained by setting $N_v = N_T = 20$ in (46), the medium fidelity solution is obtained with $N_v = N_T = 50$ and the high-fidelity solution is obtained with order $N_v = N_T = 100$ (see Fig. 10).

To model the temperature field within the GPR framework we first decompose the spatial domain $[0, 1]^2$ into 49 square spectral elements of equal size. In each element we approximate the temperature field by using a fourth-order Legendre polynomial expansion. This yields $N_{T_l} = 25$ local degrees of freedom, i.e. 25 functions of the Rayleigh number representing the local temperature as

$$T_{loc}(x, y) = \sum_{j=1}^{N_{T_l}} a_j^{(loc)}(Ra)\phi_j(x, y), \tag{47}$$

where $\phi_i(x, y)$ are polynomial basis functions (tensor product of Legendre polynomials).

To perform GPR in multi-fidelity setting we sampled the solution of (44)–(45) at different Rayleigh numbers within the range $[2.6 \times 10^3, 10^5]$. In particular, we computed 5 equally-spaced samples of the high-fidelity model ($N_v = N_T = 100$), 9 equally-spaced samples of medium-fidelity model ($N_v = N_T = 50$) and 17 equally-spaced samples of the low-fidelity model. In Fig. 11 we plot one coefficient in the expansion of the local temperature field (47), i.e., $a_{12}^{(loc)}(Ra)$ in the spectral element 25 (see Fig. 9), versus the Rayleigh number. Other coefficients have similar trend. We also compare such exact results with GPR models obtained by using univariate recursive co-kriging and multivariate recursive co-kriging with non-separable (independent coregionalization) covariance. It is seen that both these GPR approaches yield relatively accurate

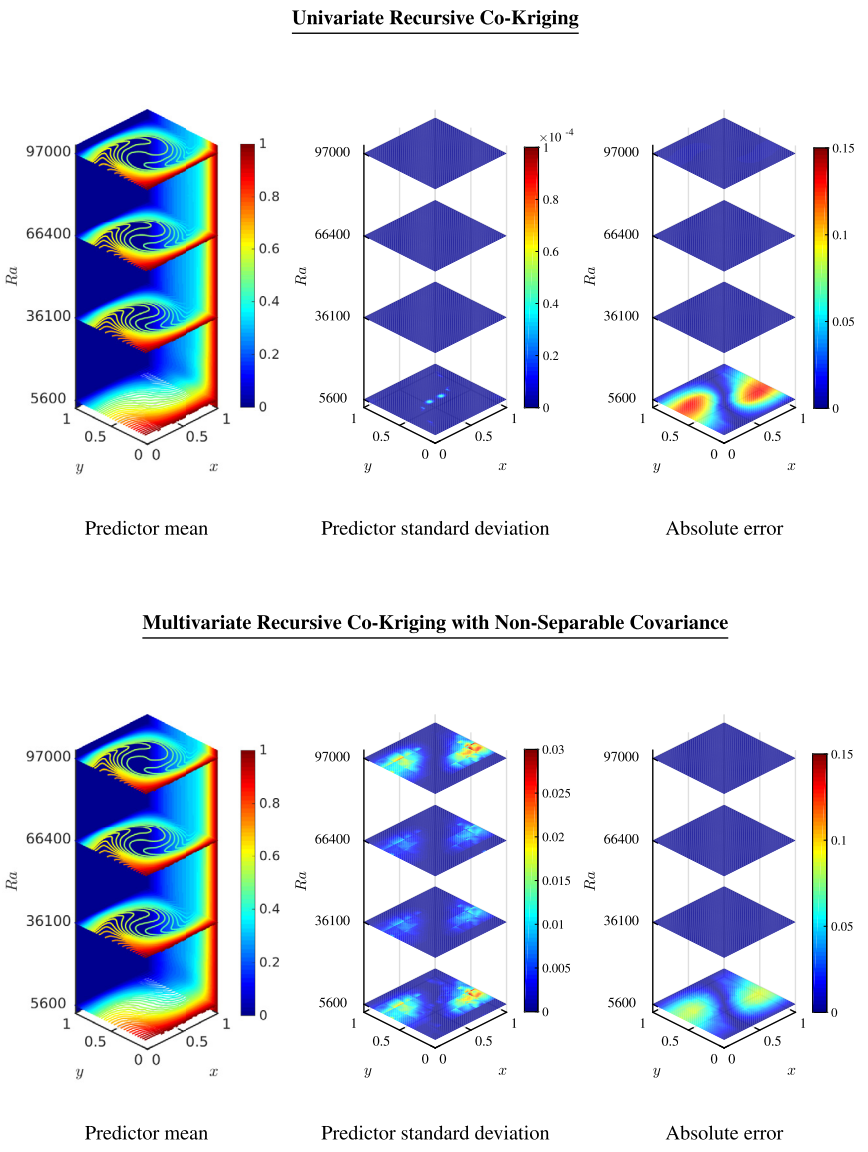

Fig. 12. Stochastic Rayleigh–Bénard convection. Mean and standard deviation of the Gaussian process representing the temperature field as a function of the Rayleigh number. Shown are results obtained with univariate recursive co-kriging and multivariate recursive co-kriging with non-separable covariance.

results. By taking the linear combination of all GPs representing the coefficients $a_j^{(loc)}(Ra)$ we can represent the local temperature field (47) as well as the global one in terms of GPs. This is done in Fig. 12 for GPR models obtained by using univariate recursive co-kriging and multivariate recursive co-kriging with non-separable covariance functions. It is seen that in both cases the maximum standard deviation of the predictor is within 3% of the maximum mean temperature while the absolute errors relative to the high-fidelity benchmark solution are smaller than 1.5e-1 (univariate co-kriging) and 8.6e-2 (multivariate co-kriging with non-separable covariance), respectively. The absolute errors here are computed by taking the absolute value of difference between the high fidelity solution and the mean temperature field obtained from GPR. The standard deviation of univariate co-kriging predictor, however, is very small compared with the absolute error, especially for $Ra \simeq 5600$. On the other hand, the multivariate predictor produces uncertainty estimates which are in much better agreement with the absolute error.

5. Summary

In this paper we proposed a new multivariate recursive co-kriging approach for prediction of random fields. Our method builds upon recent work on recursive co-kriging [12] and extends it to vector valued functions and various types of covariances, including separable and non-separable covariances. The framework we proposed is general and it can be employed to

perform multi-fidelity data fusion, i.e., prediction of vector-valued fields based on variable-fidelity models. We applied the proposed multivariate recursive GPR approach to the stochastic Burgers equation and the stochastic Oberbeck–Boussinesq equations. In both cases the predictor mean and standard deviation as well as the absolute error relative to the benchmark stochastic solution are very small, suggesting that the proposed methods can yield highly accurate results. Regarding the computational cost, the multivariate GPR approach with non-separable covariance functions is more expensive than the univariate GPR approach applied to each component of the vector output. The main reason is that in the multivariate approach we model all cross-correlations between different vector components, which yields a large matrix of cross-covariances that needs to be trained on data. There is no consensus on whether multivariate GPR can achieve better accuracy than univariate GPR applied independently to each component of the vector output (see, e.g., [21]). The numerical results presented in this paper indeed show that they have comparable accuracy. On the other hand, if the components of the vector output are highly correlated then the multivariate GPR approach with separable covariance function can outperform univariate GPR as well as multivariate GPR with non-separable covariances (see [10]), both in terms of accuracy and computational cost. The main reason is that in the separable approach the same covariance model is applied to all components of the vector output, yielding significant computational savings when learning the model parameters from data (it has the same cost of a single univariate GPR). This approach, however, is accurate only if the components of the vector output are highly correlated.

Acknowledgements

This work was supported by DARPA EQUiPS grant N66001-15-2-4055 and by the Army Research Office (ARO) grant W911NF-14-1-0425.

References

[1] I. Babuska, 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.
[2] V. Barthelmann, E. Novak, K. Ritter, High dimensional polynomial interpolation on sparse grids, Adv. Comput. Math. 12 (2000) 273–288.
[3] I. Bilionis, N. Zabaras, Multi-output local Gaussian process regression: applications to uncertainty quantification, J. Comput. Phys. 231 (17) (2012) 5718–5746.
[4] H.J. Bungartz, M. Griebel, Sparse grids, Acta Numer. 13 (2004) 147–269.
[5] H. Cho, D. Venturi, G.E. Karniadakis, Karhunen–Loéve expansion for multi-correlated stochastic processes, Probab. Eng. Mech. 34 (2013) 157–167.
[6] S. Conti, A. O'Hagan, Bayesian emulation of complex multi-output and dynamic computer models, J. Stat. Plan. Inference 140 (3) (2010) 640–651.
[7] O.G. Ernst, A. Mugler, H.-J. Starkloff, E. Ullmann, On the convergence of generalized polynomial chaos expansions, ESAIM: Math. Model. Numer. Anal. 46 (2) (2012) 317–339.
[8] J. Foo, G.E. Karniadakis, Multi-element probabilistic collocation method in high dimensions, J. Comput. Phys. 229 (2010) 1536–1557.
[9] A.I.J. Forrester, A. Sóbester, A.J. Keane, Multi-fidelity optimization via surrogate modelling, Proc. R. Soc. A 463 (2088) (2007) 3251–3269.
[10] T.E. Fricker, J.E. Oakley, N.M. Urban, Multivariate Gaussian process emulators with nonseparable covariance structures, Technometrics 55 (1) (2012) 47–56.
[11] T.E. Fricker, J.E. Oakley, N.M. Urban, Multivariate Gaussian process emulators with nonseparable covariance structures, Technometrics 55 (1) (2013) 47–56.
[12] L. Le Gratiét, Building kriging models using hierarchical codes with different levels of accuracy, in: 11th Annual Meeting of the European Network for Business and Industrial Statistics, 2011.
[13] L. Le Gratiét, Multi-fidelity Gaussian process regression for computer experiments, Ph.d. thesis, 2013.
[14] L. Le Gratiét, Bayesian analysis of hierarchical multi-fidelity codes, SIAM/ASA J. Uncertain. Quantificat. 1 (2013) 244–269.
[15] L. Le Gratiét, J. Garnier, Recursive co-kriging model for design of computer experiments with multiple levels of fidelity, Int. J. Uncertain. Quantificat. 4 (5) (2014) 365–386.
[16] P. Hennig, M.A. Osborne, M. Girolami, Probabilistic numerics and uncertainty in computations, Proc. R. Soc. A 471 (2179) (2015) 20150142.
[17] J.S. Hesthaven, S. Gottlieb, D. Gottlieb, Spectral Methods for Time-Dependent Problems, Cambridge Univ. Press, 2007.
[18] D.R. Jones, A taxonomy of global optimization methods based on response surfaces, J. Glob. Optim. 21 (5) (2001) 345–383.
[19] M.C. Kennedy, A. O'Hagan, Predicting the output from a complex computer code when fast approximations are available, Biometrika 87 (1) (2000) 1–13.
[20] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, J. R. Stat. Soc., Ser. B, Stat. Methodol. 63 (3) (2001) 425–464.
[21] J. Kleijnen, E. Mehdad, Multivariate versus univariate kriging metamodels for multi-response simulation models, Eur. J. Oper. Res. 236 (2) (2014) 573–582.
[22] J. Knap, N.R. Barton, R.D. Hornung, A. Arsenlis, R. Becker, D.R. Jefferson, Adaptive sampling in hierarchical simulation, Int. J. Numer. Methods Eng. 76 (4) (2008) 572–600.
[23] L.W.T. Ng, M.S. Eldred, Multifidelity uncertainty quantification using nonintrusive polynomial chaos and stochastic collocation, in: 53rd AIAA/ASME/ASCE/AHS/ASC Conference on Structures, Structural Dynamics and Materials, Honolulu, Hawaii, 2012.
[24] J.K. Ord, Families of Frequency Distributions, Griffin's Statistical Monographs & Courses, Hafner Pub. Co., 1972.
[25] P. Perdikaris, D. Venturi, J. Royset, G.E. Karniadakis, Multi-fidelity modeling via recursive co-kriging and Gaussian Markov random fields, Proc. R. Soc. A 471 (2179) (2015) 1–22.
[26] S.B. Provost, H.-T. Ha, D. Sanjel, On approximating the distribution of indefinite quadratic forms, Statistics 43 (6) (2013) 597–609.
[27] C.E. Rasmussen, C.K.I. Williams, Covariance Functions, The MIT Press, 2006.
[28] S.O. Rice, Distribution of quadratic forms in normal random variables-evaluation by numerical integrations, SIAM J. Sci. Stat. Comput. 1 (4) (1980) 438–448.
[29] M.L. Stein, Interpolation of Spatial Data: Some Theory for Kriging, Series in Statistics, Springer, New York, Berlin, 1999.
[30] D. Venturi, A fully symmetric nonlinear biorthogonal decomposition theory for random fields, Physica D 240 (4–5) (2011) 415–425.
[31] D. Venturi, X. Wan, G.E. Karniadakis, Stochastic bifurcation analysis of Rayleigh–Bénard convection, J. Fluid Mech. 650 (2010) 391–413.

[32] D. Venturi, X. Wan, R. Mikulevicius, B.L. Rozovskii, G.E. Karniadakis, Wick–Malliavin approximation to nonlinear stochastic partial differential equations: analysis and simulations, Proc. R. Soc. A 469 (2158) (2013) 1–20.
[33] X. Wan, G.E. Karniadakis, Multi-element generalized polynomial chaos for arbitrary probability measures, SIAM J. Sci. Comput. 28 (3) (2006) 901–928.
[34] H. Wendland, Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree, Adv. Comput. Math. 4 (1) (1995) 389–396.
[35] D. Xiu, G.E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2) (2002) 619–644.