Improving Bayesian networks multifidelity surrogate construction with basis adaptation

Xiaoshu Zeng*
University of Southern California, Los Angeles, CA 90089, USA.

Gianluca Geraci†
Sandia National Laboratories, Albuquerque, NM 87185, USA.

Alex A. Gorodetsky‡
University of Michigan, Ann Arbor, MI 48109, USA.

John D. Jakeman§ and Michael S. Eldred¶
Sandia National Laboratories, Albuquerque, NM 87185, USA

Roger Ghanem‖
University of Southern California, Los Angeles, CA 90089, USA.

Surrogate construction is an essential component for all non-deterministic analyses in science and engineering. The efficient construction of easy and cheaper-to-run alternatives to a computationally expensive code paves the way for outer loop workflows for forward and inverse uncertainty quantification and optimization. Unfortunately, the accurate construction of a surrogate still remains a task that often requires a prohibitive number of computations, making the approach unattainable for large-scale and high-fidelity applications. Multifidelity approaches offer the possibility to lower the computational expense requirement on the high-fidelity code by fusing data from additional sources. In this context, we have demonstrated that multifidelity Bayesian Networks (MFNets) can efficiently fuse information derived from models with an underlying complex dependency structure. In this contribution, we expand on our previous work by adopting a basis adaptation procedure for the selection of the linear model representing each data source. Our numerical results demonstrate that this procedure is computationally advantageous because it can maximize the use of limited data to learn and exploit the important structures shared among models. Two examples are considered to demonstrate the benefits of the proposed approach: an analytical problem and a nuclear fuel finite element assembly. From these two applications, a lower dependency of MFnets on the model graph has been also observed.

I. Introduction

Accurate uncertainty quantification (UQ) analysis of complex engineering systems is challenging due to the large amount of data, *i.e.*, high-fidelity (HF) numerical simulations, being required. For complex applications, the cost of a single numerical evaluation is such that single fidelity studies cannot be performed to a satisfying accuracy. Multifidelity approaches have been introduced to alleviate this computational burden by fusing data/information from several sources. For instance, in the context of the solution of partial differential equations (PDEs), it is not unusual to define resolution

*Postdoctoral Scholar, Civil and Environmental Engineering, University of Southern California, Los Angeles, CA 90089, USA.

†Principal Member of Technical Staff, Optimization and Uncertainty Estimation, Sandia National Laboratories, MS1318, Albuquerque, NM 87185, AIAA Member.

‡Assistant Professor, Department of Aerospace Engineering, University of Michigan, AIAA Member

§Principal Member of Technical Staff, Optimization and Uncertainty Estimation, Sandia National Laboratories, Albuquerque, NM, USA.

¶Distinguished Member of Technical Staff, Optimization and Uncertainty Estimation, Sandia National Laboratories, MS1318, Albuquerque, NM 87185, USA. AIAA Associate Fellow.

‖Professor, Civil and Environmental Engineering and Aerospace and Mechanical Engineering, University of Southern California, Los Angeles, CA 90089, USA.

levels associated to, either or both, varying spatial and temporal discretizations. Several MF strategies have been proposed in the UQ literature, based on sampling [1–5], surrogates [6–11], or hybrid approaches [12].

In this work, we focus on the so-called Multi Fidelity Bayesian Networks (MFNets) approach that has been recently proposed in [13, 14]. MFNets is a strategy that relies on latent variables to link models with an arbitrary complex dependency graph (as long as it can be represented by a Direct Acyclic Graph (DAG)). In [13], we have demonstrated how MFNets can significantly increase the quality of predictions, whenever the HF data are limited. However, in this work, we want to propose a strategy to improve its computational efficiency by combining MFNets with a model reduction strategy, namely the Basis Adaptation (BA) [15, 16]. BA provides a rigorous mechanism to limit and adapt a model representation to its important directions/variables. Our proposed approach consists in adopting the important directions/variables, discovered by BA, as the latent variables of the MFNets approach. This approach, by reducing the computational requirements due to a more compact models representation, is expected to improve the method accuracy.

The remainder of this manuscript is organized as it follows. In Section II, we briefly summarize the MFNets approach (originally presented in [13]) that serves as basis for our multifidelity data fusion. In Section III, we introduce the BA algorithm that allows for the identification of the reduced set of latent variables to be used in the multifidelity network. Numerical results are discussed in Section V, while conclusions end the manuscript in Section VI.

II. Multifidelity Bayesian Networks

In this section we briefly summarize the framework introduced in [13]. The reader should refer to the above reference for additional details. Let us consider a d-dimensional random variable vector $X : \Omega \to \mathbb{R}^d$, where the triplet $(\Omega, \mathcal{F}, \mathbb{P})$ denote a complete probability space. We consider M models, denoted via $h_i : \mathcal{X}_i \to \mathcal{Y}_i$ and a scalar quantity of interest (QoI). The input of the models are a random variable $X_i \in \mathcal{X}_i \subset \mathbb{R}^{d_i}$, where we do not assume a constant d_i for each model*.

The goal of the approach is to construct, in an efficient manner, a surrogate for

$$Y_i = h_i(X_i), \quad i = 1, \dots, M, \tag{1}$$

from which additional statistics can be evaluated, *e.g.*, the expected value. This problem is particularly challenging in the presence of HF simulations, for which the number of available realizations is very limited. If we indicate the HF model with Y_M, the goal of our approach is to fuse realizations, coming from the remaining low-fidelity (LF) models ($i = 1, \dots, M-1$), to limited data from Y_M.

As described in [13], we can represent each model, *i.e.*, the generic input-output map, with a linear expansion in term of latent variables $\theta_i \in \mathbb{R}^{\mathcal{P}_i}$

$$Y_i = \widehat{h}_i\left(X_i, \theta_i\right). \tag{2}$$

In this scenario, the surrogate construction problem corresponds to inferring the variables θ_i from available data $\mathcal{D}_i = \left(x_i^{(j)},\ y_i^{(j)}\right)_{j=1}^{m_i}$. In any multifidelity approach, it is possible to benefit from the presence of multiple data sources, if the data are correlated; however, in this approach, each model is solely defined by θ_i and, therefore, the correlation between models is controlled by that. More formally, we can assume a joint distribution $\mathbb{P}_\theta\left(\theta_1, \dots, \theta_M\right)$ among these variables to reflect the relationship among information sources. In the Bayesian context, the distribution $\mathbb{P}_\theta\left(\theta_1, \dots, \theta_M\right)$ can be interpreted as a prior, while the goal is to obtain a posterior distribution $\mathbb{P}_\theta\left(\theta_1, \dots, \theta_M | \mathcal{D}_1, \dots, \mathcal{D}_M\right)$.

The BNs are DAGs whose nodes represent random variables and whose edges represent conditional probability distributions [17]. Thanks to this feature, BNs allow to factorize $\mathbb{P}_\theta$ as a product of conditional probability distributions

$$\mathbb{P}_\theta(\theta_1, \dots \theta_M) = \prod_{i=1}^{M} \mathbb{P}(\theta_i | \theta_{\mathrm{pa}(i)}), \tag{3}$$

where $\theta_{\mathrm{pa}(i)}$ represents the set of parents of the random variable θ_i. Several multi-fidelity structures can be naturally encoded in this framework, *i.e.*, hierarchical and peer representations. For all of them, we can resort to a linear-Gaussian formulation for the inference on the BN. Given a set of M functions (features) $\phi = \{\phi_1, \dots, \phi_M : \phi_i : \mathcal{X}_i \to \mathbb{R}^{\mathcal{P}_i}\}$, we seek approximations of each model h_i of the form

$$Y_i = \widehat{h}_i(X_i, \theta_i) = \phi_i^T\left(X_i\right)\theta_i \equiv \sum_{k=1}^{\mathcal{P}_i} \phi_{ik}(X_i)\theta_{ik} \tag{4}$$

*By doing so, one could potentially consider models with dissimilar parametrization

Latent variables are usually modeled as Gaussian variables

$$\theta = (\theta_1, \theta_2, \ldots, \theta_M) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_M \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \cdots & \boldsymbol{\Sigma}_{1M} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{M1} & \cdots & \boldsymbol{\Sigma}_{MM} \end{bmatrix}, \tag{5}$$

where the means of each parameter are $\boldsymbol{\mu}_i \in \mathbb{R}^{\mathcal{P}_i}$ and covariances between parameters are $\boldsymbol{\Sigma}_{ij} \in \mathbb{R}^{\mathcal{P}_i \times \mathcal{P}_j}$. As done in [13], an additional constraint can be added to make the inference tractable: the mean of the conditional distribution of a parameter θ_i can be obtained as a linear combination of its parents values plus an offset

$$\theta_i | \theta_{\mathrm{pa}(i)} \sim \mathcal{N}\left(\boldsymbol{A}_{i|\mathrm{pa}(i)} \theta_{\mathrm{pa}(i)} + \boldsymbol{b}_{i|\mathrm{pa}(i)}, \boldsymbol{\Gamma}_{i|\mathrm{pa}(i)}\right), \tag{6}$$

where $\boldsymbol{A}_{i|\mathrm{pa}(i)} \in \mathbb{R}^{\mathcal{P}_i \times \sum_{j \in \mathrm{pa}(i)} \mathcal{P}_j}$, $\boldsymbol{b}_{i|\mathrm{pa}(i)} \in \mathbb{R}^{\mathcal{P}_i}$, and $\boldsymbol{\Gamma}_{i|\mathrm{pa}(i)} \in \mathbb{R}^{\mathcal{P}_i \times \mathcal{P}_i}$.

The last step, before an inference can be performed, is to assume a statistical model from the latent variables to the model output

$$\boldsymbol{y}_i = \boldsymbol{\Phi}_i \theta_i + \xi, \quad \xi \sim \mathcal{N}(0, \sigma_i^2 I_{m_i}), \quad \text{for } i = 1, \ldots, M, \tag{7}$$

where $\boldsymbol{\Phi}_i \in \mathbb{R}^{m_i \times \mathcal{P}_i}$ are Vandermonde-like matrices. The marginal likelihood of the output, given the inputs, is (under the assumption of independent data sources)

$$\boldsymbol{y}_1, \ldots, \boldsymbol{y}_M | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_M \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{C})$$

where

$$\boldsymbol{m} = \begin{bmatrix} \boldsymbol{\Phi}_1 \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\Phi}_M \boldsymbol{\mu}_M \end{bmatrix} \quad \boldsymbol{C} = \begin{bmatrix} \boldsymbol{\Phi}_1 \boldsymbol{\Sigma}_{11} \boldsymbol{\Phi}_1^T + \sigma_1^2 \boldsymbol{I}_{m_1} & \cdots & \boldsymbol{\Phi}_1 \boldsymbol{\Sigma}_{1M} \boldsymbol{\Phi}_M^T \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}_M \boldsymbol{\Sigma}_{M1} \boldsymbol{\Phi}_1^T & \cdots & \boldsymbol{\Phi}_M \boldsymbol{\Sigma}_{MM} \boldsymbol{\Phi}_M^T + \sigma_M^2 \boldsymbol{I}_{m_M} \end{bmatrix}. \tag{8}$$

If block diagonal matrices are built from the Vandermonde and covariance matrices as

$$\boldsymbol{\Phi} = \text{block-diag}\left[\boldsymbol{\Phi}_1 \ \cdots \ \boldsymbol{\Phi}_M\right] \qquad \text{and} \qquad \Sigma_M = \text{block-diag}\left[\sigma_1^2 I_{m_1} \ \cdots \ \sigma_M^2 I_{m_M}\right],$$

then $\boldsymbol{C} = \boldsymbol{\Phi} \boldsymbol{\Sigma} \boldsymbol{\Phi}^T + \Sigma_M$, and the joint density of the latent variables θ and the data $\boldsymbol{y}$ is a Gaussian random variable with mean and covariance

$$\theta, \boldsymbol{y} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{m} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma} \boldsymbol{\Phi}^T \\ \boldsymbol{\Phi} \boldsymbol{\Sigma} & \boldsymbol{C} \end{bmatrix}\right). \tag{9}$$

Finally, the posterior conditional distributions, given the observations $\theta \mid \boldsymbol{y} \sim \mathcal{N}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$, can be obtained as

$$\bar{\boldsymbol{\Sigma}} = \left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Phi}^T \Sigma_n^{-1} \boldsymbol{\Phi}\right)^{-1} \tag{10}$$

$$\bar{\boldsymbol{\mu}} = \bar{\boldsymbol{\Sigma}} \boldsymbol{\Phi}^T \Sigma_n^{-1} \boldsymbol{y} + \bar{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}. \tag{11}$$

One of the difficulties to face in this approach is the scalability issue related to the number of latent variables for each model, and the consequent need to acquire enough data to reduce the uncertainty on these variables. BNs helps in reducing the data requirement by enforcing a structure to the conditional probability, however, in this work, we explore the possibility to further reduce the number of latent variables by adopting the important directions used in the Basis Adaptation (BA) method. The introduction of latent variables, obtained from BA, has the potential to drastically reduce the data requirement by making the inference computationally more efficient. BA is introduced in the next section.

III. Basis Adaptation

Basis adaptation is in the framework of polynomial chaos expansions (PCE). For any model in the previous section, the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is considered. Let the random input variables, X_i, be independent standard Gaussian variables on $\boldsymbol{\Omega}$. Let $\mathcal{H}$ be a Hilbert space generated by the span of X_i and $\mathcal{F}(\mathcal{H})$ be the σ-algebra generated by $\mathcal{H}$. Then, $Y_i \in L^2(\Omega, \mathcal{F}(\mathcal{H}), \mathbb{P})$ can be approximated by the following truncated PCE of order up to p_i [18]

$$Y_i(X_i) = \sum_{k=0}^{\mathcal{P}_i - 1} \theta_{ik} \psi_{ik}(X_i), \tag{12}$$

where $\{\psi_{ik}\}_{k=0}^{\mathcal{P}_i-1}$ are Hermite polynomial basis functions that are orthonormal under probability measure of X_i and θ_{ik} are the associated PCE coefficients (to be computed). The truncated PCE converges to the physical model of Y_i in a mean-squared sense as the maximum PCE order $p_i \to \infty$. The PCE coefficients θ_{ik} can be found, *e.g.*, by projecting $Y_i(X_i)$ to the space spanned by $\{\psi_{ik}\}_{k=0}^{\mathcal{P}_i-1}$ as

$$Y_{ik} = \frac{\langle Y_i, \psi_{ik}\rangle}{\langle \psi_{ik}^2\rangle} = \langle Y_i, \psi_{ik}\rangle, \quad k = 0, \ldots, \mathcal{P}_i - 1\,, \tag{13}$$

where $\langle a, b\rangle$ indicates the L^2 inner product between a and b and $\langle \psi_{ik}^2\rangle = 1$, by construction. For a total order expansion, the number of PCE terms is

$$\mathcal{P}_i = \binom{d_i + p_i}{p_i} = \frac{(d_i + p_i)!}{d_i! p_i!}\,, \tag{14}$$

which grows factorially with d_i and p_i. Eq (13) involves the calculation of a d-dimensional integrals, however PCE coefficients can be also obtained with other strategies, *e.g.*, via ordinary least squares (OLS). For stable and accurate estimation, the number of required samples for OLS is in the order of $N \sim \mathcal{O}(\mathcal{P}_i \ln \mathcal{P}_i)$ and, in some cases, $N \sim \mathcal{O}(\mathcal{P}_i^2 \ln \mathcal{P}_i)$ [19]. When the dimension is high, the computation becomes prohibited. The curse-of-dimensionality remains, even with sparse quadrature rules where the model is evaluated on selecting pre-selected quadrature points.

Basis adaptation is a dimension reduction technique that intends to discover a low-dimensional manifold to represent the quantity of interest (QoI), Y_i. The basic idea is to employ a rotation matrix to transform the Gaussian input variables such that the QoI has concentrated information on the first several rotated variables. The polynomial span of these rotated variables is the low-dimensional space the method seeks. Assume $R_i \in \mathbb{R}^{d_i \times d_i}$ is a rotation matrix such that $R_i R_i^T = \boldsymbol{I}$, then, the rotated variables can be obtained as

$$\boldsymbol{\eta}_i = \boldsymbol{\eta}_i(X_i) = R_i X_i\,. \tag{15}$$

It follows that Y_i can be equivalently expanded with respect to $\boldsymbol{\eta}_i$ as

$$Y_i^{R_i}(\boldsymbol{\eta}_i) = \sum_{k=0}^{\mathcal{P}_i-1} \theta_{ik}^{R_i} \psi_{ik}(\boldsymbol{\eta}_i)\,. \tag{16}$$

Since the goal of basis adaptation is to find a low-dimensional manifold to represent Y_i, the construction of the rotation matrix is crucial.

In a classical Gaussian adaptation[†], the first row of the rotation matrix is constructed such that

$$\eta_{i1} = \sum_{k=1}^{d} \theta_{ik} X_{ik}\,, \tag{17}$$

where $\{\theta_{ik}\}_{k=1}^{d}$ is the set of first-order or Gaussian coefficients. The first rotated variable by Eq (17) captures the complete Gaussian components of the QoI, thus the name "Gaussian adaptation". The second to last rows of the rotation matrix is defined based on the sensitivity of X_i, quantified by the absolute value of the Gaussian coefficients. Specifically, the second row is defined such that η_{i2} is the most sensitive variable of X_i; the third row is defined such that η_{i3} is the second most sensitive variable of X_i; and so on. Since the Gaussian components are typically the most critical in many applications, the first rotated variable η_{i1} has the key information of the QoI. In addition, the other rotated variables also have descending order of importance by construction. Therefore, Gaussian adaptation motivates seeking a reduced representation of the QoI in a low-dimensional space.

Suppose the first r_i rotated variables are adequate to represent Y_i, we can partition $\boldsymbol{\eta}_i$ as

$$\boldsymbol{\eta}_i = \begin{bmatrix} R_{i_r} X_i \\ R_{i_{\neg r}} X_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{\eta}_{i_r} \\ \boldsymbol{\eta}_{i_{\neg r}} \end{bmatrix}\,. \tag{18}$$

Where R_{i_r} is the sub-matrix of the first r_i rows of R_i. Afterward, we can approximate the model Y_i with a PCE on $\boldsymbol{\eta}_{i_r}$

$$Y_i^{R_{i_r}}(\boldsymbol{\eta}_{i_r}) = \sum_{k=1}^{\mathcal{P}_i^r} \theta_{ik}^{R_{i_r}} \psi_{ik}(\boldsymbol{\eta}_{i_r})\,, \tag{19}$$

[†]Higher-order strategies are also available, see [15].

with its coefficients found by

$$\theta_{ik}^{R_{i_r}} = \left\langle Y_i^{R_{i_r}}, \psi_{ik} \right\rangle , \qquad k = 1, \ldots, \mathcal{P}_i^r . \tag{20}$$

Eq (20), similarly to Eq (13), requires the evaluation of a multidimensional integral, however the former only requires to evaluate an r-dimensional integral, where r_i is in general smaller than d_i. Thanks to this reduction in dimension, BA has been demonstrated to be efficient in significantly reducing the number of evaluations required for the accurate construction of a surrogate [15, 16, 20–22]. The scope of this manuscript is to demonstrate that this dimension reduction can be effective in the context of MFNets, as illustrated in the next section. In the following, the super script on PCE coefficients $\theta_{ik}^{R_{i_r}}$ can be removed without ambiguity.

IV. Embedded Basis Adaptation for MFNets

As illustrated in Section II, MFNets requires a computational effort that scales with the expansion dimension of each model. Here, BA is integrated into MFNets to obtain two possible advantages. First, a more compact representation for each model that only needs to use r variables is going to reduce the dimension of the inference problem presented in Eqs. (9), (10) and (11). Second, as demonstrated in previous work [23–28], the use of shared space among models has the potential of significantly increasing the correlation among models, which, in the MFNets context, roughly speaking corresponds to maximizing the information provided by the LF models for the HF model construction.

In this paper, as done in [13], we are concerned with the construction of the models from an assigned set of data $\mathcal{D}_i$, for each model $i = 1, \ldots, M$, while the allocation of resources, *i.e.*, evaluations, among models is not considered. In Section II, we have considered, for each model Y_i, a linear expansion over features ϕ_i (see Eq (4)), whereas here we adopt the Hermite polynomials ψ_i as basis[‡] and, for each model expansion (we assume a potentially different polynomial order p_i for each model), the number of terms $\mathcal{P}_i$ is given by Eq (14). In this work, we assume that the user is able to determine the number of important variables r_i after the construction of the rotation matrix, but determining them in an automated way is also possible by constructing bias measures, as suggested in [26–28]. As evident, the use of the adapted variable r_i, instead of d_i, enable us to represent each model with only

$$\mathcal{P}_i^r = \frac{(r_i + p_i)!}{r_i!\, p_i!} \tag{21}$$

variables, where $\mathcal{P}_i^r << \mathcal{P}_i$.

An additional, and equally important, benefit of using BA within MFNets, is that the correlation among models can be enhanced when the models are represented on their important variables $\boldsymbol{\eta}_{i_r}$. We have demonstrated this in a series of contributions, see [23–28], so this aspect is not further described here.

From an algorithmic perspective, embedding BA within MFNets requires few modifications to an existing MFNets algorithm. The construction of the surrogates via MFNets and PCE follows Algorithm 1. With the posterior mean $\bar{\boldsymbol{\mu}}$ and covariance $\bar{\boldsymbol{\Sigma}}$ obtained, the HF prediction, indexed by M, can be obtained. For any $\bar{Y}_M$ associated with samples $\bar{X}_M$, its mean can be estimated as

$$\mu_{\bar{Y}_M} = \boldsymbol{\Phi}_M(\bar{X}_M)\bar{\boldsymbol{\mu}}_M \tag{22}$$

and its variance can be estimated as

$$\mathbb{V}\mathrm{ar}(\bar{Y}_M) = \mathrm{diag}\left(\boldsymbol{\Phi}_M(\bar{X}_M)\bar{\boldsymbol{\Sigma}}_{MM}\boldsymbol{\Phi}_M(\bar{X}_M)^T\right) . \tag{23}$$

Algorithm 1 contains the cases with and without basis adaptation. When basis adaptation is used, the rotation matrix is constructed for each model independently and, from it, the number of adapted variables r_i is determined. Then, the original samples are mapped to reduced spaces, on which the subsequent model evaluations are carried out.

V. Numerical results

This section illustrates the benefits of embedding BA in MFNets through two numerical examples. We first introduce an analytical test problem in Section V.B, which we developed to verify the algorithm and gain additional intuition about the method. In Section V.C, we consider a more challenging computational problem in which a finite element model of a nuclear fuel rod assembly is used. For all the numerical examples provided in this section we directly employed

[‡]In the case of non-Gaussian input distributions, we can map the physical variable to standard Gaussian variables and retain this exact setup.

Algorithm 1: Surrogate modeling by MFNets and PCE

Input: m_i, σ_i, p_i, `adaptation` (using basis adaptation if `True`)

Generate desired number dataset $\mathcal{D}_i = \left(x_i^{(j)},\ y_i^{(j)}\right)_{j=1}^{m_i}$, for $i = 1, \ldots, M$;
Map the input samples to standard normal samples by, for example, inverse transformation;
if *adaptation=True* **then**
    for *i=1,. . ., M* **do**
        Construct first-order PCEs (12) with $p_i = 1$ through regression;
        Compute the rotation matrix R_i;
        Estimate/Assign the required adapted variables r_i;
        Map X_i to $\boldsymbol{\eta}_i$ by $\boldsymbol{\eta}_i = R_{i_{r_i}} X_i$;
        Evaluate the Hermite polynomial basis on $\boldsymbol{\eta}_i$ and construct the Vandermonde-like matrices $\boldsymbol{\Phi}_i$;
else
    Evaluate the Hermite polynomial basis on X_i and construct the Vandermonde-like matrices $\boldsymbol{\Phi}_i$, for $i = 1, \ldots, M$;
Let the prior mean of θ, $\boldsymbol{\mu}$, be zeros;
Compute the prior covariance matrix $\boldsymbol{\Sigma}$ based on the graph structures (details in the next section);
Construct the covariance matrix of the noise, Σ_M, by standard deviation of the noises σ_i, for $i = 1, \ldots, M$;
Compute the posterior mean $\bar{\boldsymbol{\mu}}$ and covariance $\bar{\boldsymbol{\Sigma}}$ by Eq (11).

Output: $\bar{\boldsymbol{\mu}}$, $\bar{\boldsymbol{\Sigma}}$.

the Sandia National Laboratories open source UQ software PyApprox [29], which already implements the MFNets approach. Three fidelity models are considered for both numerical test problems, *i.e.*, one HF model and two LF models. Moreover, two graphs are studied: a peer and a hierarchical structure. These graphs are illustrated in Figure 1.

Fig. 1 Graph structure with three nodes with (a) a peer and (b) a hierarchical structure. The HF model is Y_3 and is described by variable θ_3, while the two low-fidelity models are Y_2 and Y_1, with variables θ_2 and θ_1, respectively.

As explained in Section II, once the graph is assigned, a model for the conditional probability of the model coefficients needs to be assumed. Also, additional assumptions need to be provided for the statistics of the coefficients, *e.g.*, to model the covariance among coefficients. For all the results reported in this manuscript, we did not use hyperparameter tuning. The graph-dependent hyperparameters used in each example are discussed in Section V.A, while other problem dependent choices (*e.g.* specification of priors) are discussed in the corresponding test case section. In this paper, we simply explored few allowable choices of the parameters that we use to compare MFNets with and without BA. The choices of the admissible parameters are discussed in the following section.

A. Graph modeling and additional assumptions

Since we consider both cases with and without basis adaptation, we use θ_i to represent the sets of the PCE coefficients that can serve both expansions. In this section we directly follow the presentation of the tutorials released within PyApprox [29].

Peer graph Let's start by considering the peer graph illustrated in Figure 1(a), namely with $\varepsilon = \{\theta_1 \to \theta_3,\ \theta_2 \to \theta_3\}$. This graph is useful whenever, at least nominally, the LF sources could provide the same level of information to the HF

model.

From the model in Equation (6) and assuming that the PCE coefficients of an information source are linearly related to the PCE coefficients of other information sources, we can obtain the following expression for the HF model

$$\theta_3 = A_{31}\theta_1 + A_{32}\theta_2 + v_3\,, \tag{24}$$

where v_3 is a Gaussian noise with mean zero and covariance Σ_{v_3}. Furthermore, by assuming $\mathrm{Cov}[\theta_\alpha, v3] = 0\ \forall\alpha$ and $\mathrm{Cov}[\theta_1, \theta_2] = 0$, we can compute the prior covariance of the PCE coefficients as [29]

$$\begin{aligned}
\mathrm{Cov}[\theta_1, \theta_3] &= \mathbb{E}\left[\theta_1\theta_3^T\right] = \mathbb{E}\left[\theta_1(A_{31}\theta_1 + A_{32}\theta_2 + v_3)^T\right] = \mathrm{Cov}[\theta_1, \theta_1]\, A_{31}^T \\
\mathrm{Cov}[\theta_2, \theta_3] &= \mathbb{E}\left[\theta_2\theta_3^T\right] = \mathbb{E}\left[\theta_2(A_{31}\theta_1 + A_{32}\theta_2 + v_3)^T\right] = \mathrm{Cov}[\theta_2, \theta_2]\, A_{32}^T\,.
\end{aligned}$$

For simplicity, and to obtain a parametrization of reasonable size, we assume $A_{31} = a_{31}I$, $A_{32} = a_{32}I$, $\mathbf{\Sigma}_{11} = s_{11}I$, $\mathbf{\Sigma}_{22} = s_{22}I$, $\mathbf{\Sigma}_{33} = s_{33}I$, and $\Sigma_{v_3} = \mathbf{\Sigma}_{33} - a_{31}^2\mathbf{\Sigma}_{11} - a_{32}^2\mathbf{\Sigma}_{22}$, where the last assumption is to let the prior covariance of each information source be the same. Finally, the prior covariance matrix of the PCE coefficients can be written as

$$\mathbf{\Sigma} = \begin{bmatrix} s_{11}I & 0 & a_{31}s_{11}I \\ 0 & s_{22}I & a_{32}s_{22}I \\ a_{31}s_{11}I & a_{32}s_{22}I & s_{33}I \end{bmatrix}, \tag{25}$$

whereas the prior means of the PCE coefficients are assumed to be zeros. Given all the previous assumptions, the prior model is parameterized by the following scalar coefficients $(s_{11}, s_{22}, s_{33}, a_{31}, a_{32})$. For this graph, we consider two sets of admissible parameters values with common parameters $s_{11} = s_{22} = s_{33} = s = 1.0$. In the first case, we consider no correlations between coefficients among the models, *i.e.*, $a_{31} = a_{32} = 0$, which is useful to represent cases in which the correlation among the HF and LF model is low, or nonexistent. In the second case, we consider the correlation among models coefficients to be $a_{31} = a_{32} \approx \rho$, where ρ is the estimated correlation of the information sources (not the correlation of the PCE coefficients). If $\rho > 0.7$, then let $a_{31} = a_{32} = 0.7$ as 0.7 is determined as the upper bound of the allowable space with $s = 1.0$.

Hierarchical graph The second graph considered in this work is the hierarchical structure illustrated Figure 1(b) with $\varepsilon = \{\theta_1 \to \theta_2 \to \theta_3\}$. This arrangement of models is useful if it is possible to determine a sequence with accuracy/predictive capability increasing in a one-dimensional sequence.

By linear dependency assumption of the PCE coefficients, we have

$$\begin{aligned}
\theta_2 &= A_{21}\theta_1 + v_2 \\
\theta_3 &= A_{32}\theta_2 + v_3\,,
\end{aligned} \tag{26}$$

where v_2 is a Gaussian noise with mean zero and covariance Σ_{v_2}. Furthermore, we assume $\mathrm{Cov}[\theta_\alpha, v_2] = 0$ and $\mathrm{Cov}[\theta_\alpha, v_3] = 0\ \forall\alpha$. The prior covariance can be computed as [29]

$$\begin{aligned}
\mathrm{Cov}[\theta_1, \theta_2] &= \mathbb{E}\left[\theta_1\theta_2^T\right] = \mathbb{E}\left[\theta_1(A_{21}\theta_1 + v_2)^T\right] = \mathrm{Cov}[\theta_1, \theta_1]\, A_{21}^T \\
\mathrm{Cov}[\theta_2, \theta_3] &= \mathbb{E}\left[\theta_2\theta_3^T\right] = \mathbb{E}\left[\theta_2(A_{32}\theta_2 + v_3)^T\right] = \mathrm{Cov}[\theta_2, \theta_2]\, A_{32}^T \\
\mathrm{Cov}[\theta_1, \theta_3] &= \mathbb{E}\left[\theta_1\theta_3^T\right] = \mathbb{E}\left[\theta_1(A_{32}\theta_2 + v_3)^T\right] = \mathrm{Cov}[\theta_1, \theta_2]\, A_{32}^T = \mathrm{Cov}[\theta_1, \theta_1]\, A_{21}^T A_{32}^T\,.
\end{aligned}$$

Again, even in this case, in order to obtain a manageable parametrization we define $A_{21} = a_{21}I$, $A_{32} = a_{32}I$, $\Sigma_{v_2} = \mathbf{\Sigma}_{22} - a_{21}^2\mathbf{\Sigma}_{11}$, and $\Sigma_{v_3} = \mathbf{\Sigma}_{33} - a_{32}^2\mathbf{\Sigma}_{22}$, such that the prior covariance matrix of the coefficients can be derived as

$$\mathbf{\Sigma} = \begin{bmatrix} s_{11}I & a_{21}s_{11}I & a_{21}a_{32}s_{11}I \\ a_{21}s_{22}I & s_{22}I & a_{32}s_{22}I \\ a_{21}a_{32}s_{33}I & a_{32}s_{33}I & s_{33}I \end{bmatrix}, \tag{27}$$

while the prior means of the PCE coefficients are assumed to be zeros. The prior model in this graph is parameterized by the following scalar coefficients $(s_{11}, s_{22}, s_{33}, a_{21}, a_{32})$. For this graph, we will consider one case, where $s_{11} = s_{22} = s_{33} = s = 1.0$ and $a_{21} = a_{32} \approx \rho$, with ρ the estimated correlation among the information sources.

B. Analytical test problem

As the first numerical demonstration, we consider a simple test example with three information sources described by

$$
\begin{aligned}
f_1(\boldsymbol{x}) &= \exp\left(x_1 + 0.05x_2\right) + \exp 0.8x_3 + \exp\left(0.8x_4 + 0.05x_5 + 0.05x_6\right), \\
f_2(\boldsymbol{x}) &= \log\left(0.75x_1 + 0.05x_2 + 1\right) + \log\left(x_3 + 0.05x_5 + 1\right) + \log\left(0.5x_4 + 0.05x_6 + 1\right), \\
f_3(\boldsymbol{x}) &= \exp\left(0.1x_1 + 1.2x_2\right) + \exp 0.05x_3 + \exp\left(0.05x_4 + x_5 + x_6\right) \\
&\quad + \log(0.05x_1 + 0.8x_2 + 1) + \log(0.75x_5 + x_6 + 1),
\end{aligned} \tag{28}
$$

with f_1 and f_2 the LF models and f_3 the HF model. We see that the dimensions of these three models are all $d = 6$.

Suppose each information source is approximated by a third-order ($p = 3$) PCE. Then the total number of PCE terms for each information source is $\mathcal{P} = \binom{d+p}{p} = \binom{9}{3} = 84$. Let θ_1, θ_2, and θ_3 be the sets of PCE coefficients associated with these three models. Suppose we want to use MFNets to fuse the information sources and propose estimations of the HF model by a fixed number of samples. We consider a peer structure in the MFNets for this application.

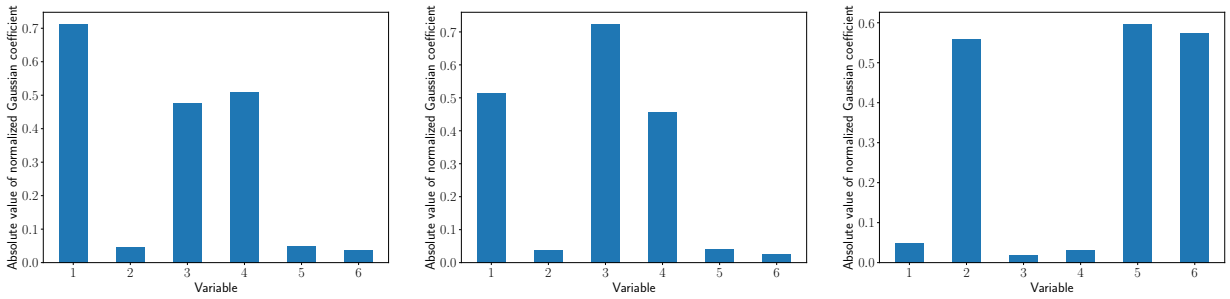
In the following, we report the results for the posterior prediction of the HF model in the original and adapted space. In all the cases, we assume $s_{11} = s_{22} = s_{33} = s = 1.0$. Section V.B.1 considers no correlation between coefficients among the models, *i.e.*, $a_{31} = a_{32} = 0$, and Section V.B.2 considers correlations in coefficients with $a_{31} = a_{32} = 0.7$ which is the upper bound of the allowable parameters.

1. Surrogate construction with peer graph ($a_{31} = a_{32} = 0$)

In this first case, the choice of the parameters, $a_{31} = a_{32} = 0$, reflects the absence of correlations between coefficients among the models. The parameter choice is reasonable in the original space, as correlations between the HF and LF models are almost zero by construction. Thus, the correlations between coefficients among the models can also be assumed to be zeros. As a comparison, the same choice of parameters are applied in the adapted space.

For the construction of the surrogate the original variables, *i.e.* without embedding BA, we let the number of samples be $m_1 = m_2 = 100$ and $m_3 = 20$. Further letting the PCE orders be $p_1 = p_2 = p_3 = 3$, the standard deviations of the models be $\sigma_1 = \sigma_2 = \sigma_3 = 0.01$, and `adaptation=False`. Then we can obtain the posterior mean and covariance of the coefficients by following Algorithm 1 with the prior covariance matrix in step 13 being computed via Eq (25).

Similarly, we can use Algorithm 1 with `adaptation=True` to obtain the posterior mean and covariance of the coefficients in the case based on BA. During the process, the rotation matrices for these three models are computed independently based on first-order PCEs from the available samples. The first rows of the rotation matrices are the normalized first-order PCE coefficients, which typically quantify the sensitivity of the QoI with respect to the parameters and determine the rotations uniquely. The absolute values of the normalized first-order PCE coefficients are reported in Figure 2. We can see from Figure 2 that the low-fidelity models depend on variables 1, 3, and 4, while for the



(a) Low-fidelity (f_1) **(b) Low-fidelity (f_2)** **(c) High-fidelity (f_3)**

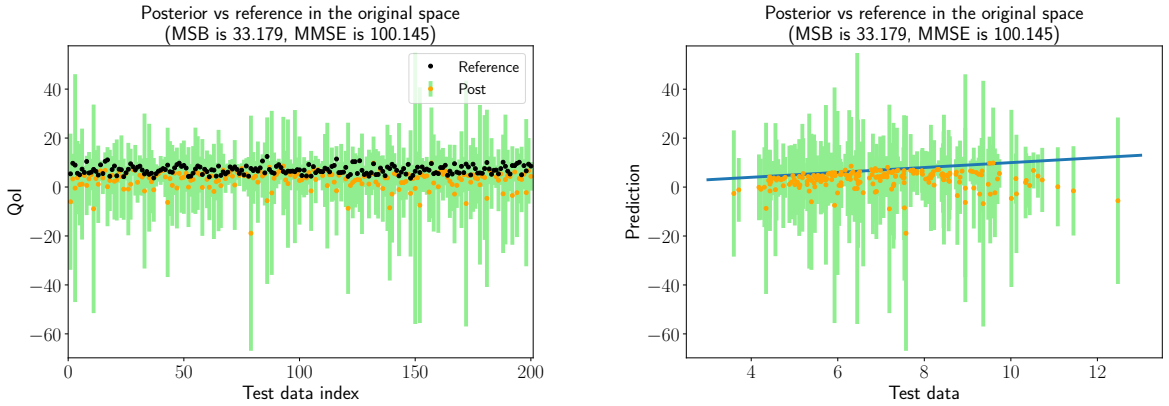
Fig. 2 Normalized first-order PCE coefficients in the original space. All the models are represented: low-fidelity f_1 (a), low-fidelity f_2 (b), and high-fidelity f_3 (c). We observe that variables 1, 3, and 4 are important for the low-fidelity models, whereas variables 2, 5, and 6 are very important for the high-fidelity model. The important variables identified here are the only ones used for the construction of the surrogate when embedding BA.

high-fidelity model the most important variables are 2, 5, and 6. Based on this information, we can estimate the required reduced dimensions for accurate representation as $r_1 = r_2 = r_3 = r = 3$. Then the total number of PCE terms for each

model in the adapted spaces are $\mathcal{P}_1 = \mathcal{P}_2 = \mathcal{P}_3 = \binom{r+p}{p} = \binom{6}{3} = 20$, which is much smaller than the number of terms for the same expansion in the original space.

We test the surrogate model for the HF model on 200 test points that were not included in the training data. The predicted responses can be obtained by evaluating Eqs (22) and (23). Several metrics are defined to quantitatively assess the quality of the MFNets surrogate. We consider the mean-squared bias (MSB) of the test set, which is the average squared bias of all the test data. Next, we define the mean-variance (MVAR) of the posterior over the test set, which is the average variance of all the test data. Finally, we define the average mean-squared error (MMSE) of the test set, where the mean-squared error of each test data equals the sum of its variance and squared bias. Note that MMSE = MSB + MVAR.

The MFNets surrogate performance, for the case built with the original variables, *i.e.* without BA, is shown in Figure 3. In Figure 3a, we plot all 200 test data, where the orange dots denote the posterior mean, and the green bars show



(a) Posterior predictions and test data **(b) Parity plot**

Fig. 3 Comparison between the posterior model (built in the original space with peer graph and $s = 1.0$, $a_{31} = a_{32} = 0$) and 200 reference HF test data. The green bars show the posterior mean plus and minus two standard deviations. In Figure a, both datasets, predictions and test data, are reported. In Figure b, the predictions are plotted against the corresponding test data (parity plot). In this case, the solid line (slope 1) indicates a perfect prediction.

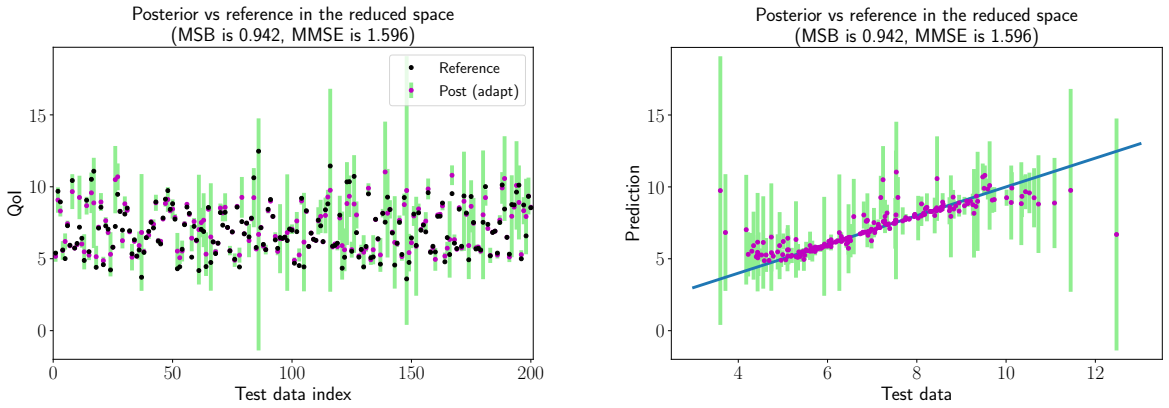
the posterior mean plus and minus two standard deviations of the posterior HF model. The predictions are compared directly against the test data (black dots). In Figure 3b, we report the parity plot, where the prediction of the posterior HF model are reported against the actual test data. Dots laying on the reference line, *i.e.*, slope 1, indicate a perfect prediction. We observe a noticeable bias of the posterior mean with MSB = 33.179. Also, the prediction variance is significant, leading to an MMSE = 100.145.

The performance for the MFNets surrogate of the HF model, based on the embedding of the BA method, is shown in Figure 4. Compared with Figure 3, the MSB of the posterior mean is reduced from 33.179 to 0.942, and the prediction variances are reduced significantly. The MMSE decreased from 100.145 to 1.596, which reflect almost two order of magnitude improvement with respect to the construction without BA. The biases and variances are reduced solely due to the reduced number of coefficients, while the number of datapoint we used for the training is the same. This example illustrates that the reduction of coefficients can increase the accuracy of the posterior estimation of the coefficients, and thus increase the accuracy of the posterior model.

2. Surrogate construction with peer graph ($a_{31} = a_{32} = 0.7$)

As discussed in the previous section, the correlation among models can be enhanced by representing the QoI in the adapted coordinates, rather than in their original models coordinates. This suggests that the increased correlation can be leveraged to further improve the accuracy of MFNets, by reducing the variance of the posterior model. For this case, we also assume the parameters $\sigma_1 = \sigma_2 = \sigma_3 = 0.01$.

The correlations between models 3 and 1 and 3 and 2 can be estimated by the adapted PCEs and they are 0.848 and 0.825, respectively . Although the correlations on the models are not necessarily the same as the correlation on the coefficients, we assume that they are related; based on this, we use the correlations on the models QoIs as the values of



(a) Posterior predictions and test data **(b) Parity plot**

Fig. 4 Comparison between the posterior model (built in the adapted space with peer graph and $s = 1.0$, $a_{31} = a_{32} = 0$) and 200 reference HF test data. The green bars show the posterior mean plus and minus two standard deviations. In Figure a, both datasets, predictions and test data, are reported. In Figure b, the predictions are plotted against the corresponding test data (parity plot). In this case, the solid line (slope 1) indicates a perfect prediction. These results show a significant improvement with respect to the same data reported in Figure 3b, which were obtained by constructing the surrogate without BA.

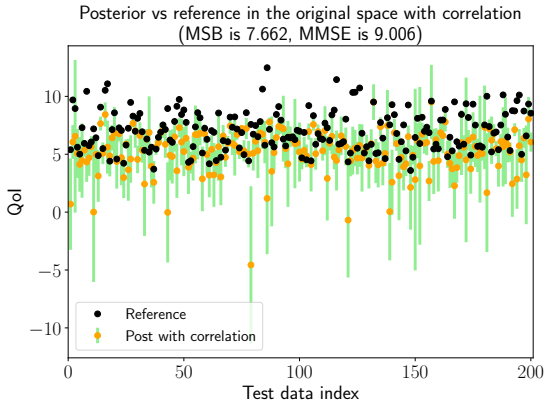
a_{31} and a_{32}. However, these values are greater than 0.7, which represents the upper bound of allowable parameters with $s = 1$, so we use $a_{31} = a_{32} = 0.7$. As done in the previous section, we discuss the accuracy of MFNets with and without the use of BA, by using the same 200 test data of the previous case.

For the MFNets training in the original space, the comparison between the reference and the posterior model of the HF model with $a_{31} = a_{32} = 0.7$ is shown in Figure 5. Compared to Figure 3, in which the surrogate was trained assuming no correlation among the coefficients, *i.e.*, $a_{31} = a_{32} = 0$, the prediction here shows a smaller MSB = 7.662 and a smaller MMSE = 9.006. Similarly, the prediction variance is also much smaller, which is reflected by the shorter green bars in the figure.

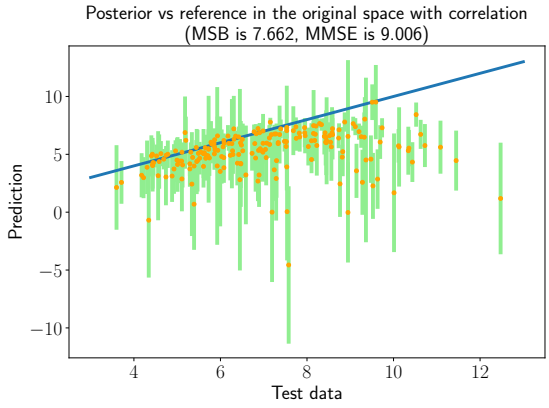
Despite the improvement in the surrogate construction that we obtained by assuming a non-zero correlation among the coefficients, the quality of the surrogate can be further improved by embedding BA in its construction. The performance of the surrogate constructed by embedding BA, with $a_{31} = a_{32} = 0.7$, is shown in Figure 6. Compared to Figure 4, we note that the posterior variance is much smaller and the MSB has decreased from 0.942 to 0.515, while MMSE decreased from 1.596 to 0.553.

3. Impact of the number of samples on the MFNets construction

As a final test case, we fixed the ratio of the LF samples to the HF samples to be 5 and explored the performance of the MFNets surrogate construction for an increasing number of HF samples. We considered the construction of the surrogate with the two sets of assumptions, namely $a_{31} = a_{32} = 0$ and $a_{31} = a_{32} = 0.7$, and with and without embedding BA. The results of these tests are summarized in Figure 7, where the MMSE is reported for each model. From Figure 7, we observe several interesting behaviors. First, the MMSE decreases more rapidly for the case with correlation among PCE coefficients. Second, we observe that the surrogate construction with BA is always more efficient than its counterpart, up to the point in which no distinction between the two methods can be observed. This behavior occurs when the surrogate construction without BA can already rely on enough data and the use of BA cannot further improve the accuracy of the model. As expected, the use of BA cannot arbitrarily improve the model accuracy, but rather train a model with reduced data requirements. For more complex applications, the threshold corresponding to the availability of enough data is expected to be significantly larger, therefore we expect the embedding of BA to become more effective with the increase in the problem dimensionality (both number of variables and total polynomial order).

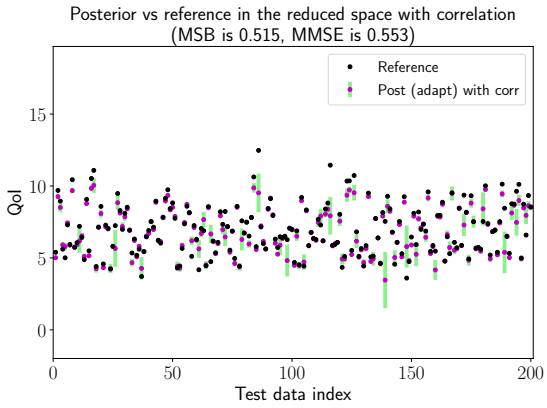


(a) Posterior predictions and test data

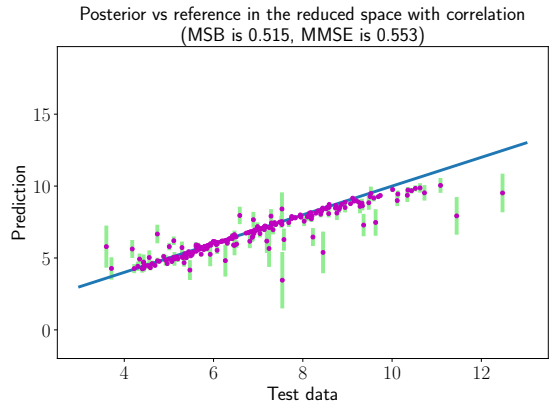


(b) Parity plot

Fig. 5 Comparison between the posterior model (built in the original space with peer graph and $s = 1.0$, $a_{31} = a_{32} = 0.7$) and 200 reference HF test data. The green bars show the posterior mean plus and minus two standard deviations. In Figure a, both datasets, predictions and test data, are reported. In Figure a, the predictions are plotted against the corresponding test data (parity plot). In this case, the solid line (slope 1) indicates a perfect prediction. These results show a significant improvement with respect to the same data reported in Figure 3b, which were obtained by assuming $a_{31} = a_{32} = 0$.



(a) Posterior predictions and test data



(b) Parity plot

Fig. 6 Comparison between the posterior model (built in the adapted space with peer graph and $s = 1.0$, $a_{31} = a_{32} = 0.7$) and 200 reference HF test data. The green bars show the posterior mean plus and minus two standard deviations. In Figure a, both datasets, predictions and test data, are reported. In Figure a, the predictions are plotted against the corresponding test data (parity plot). In this case, the solid line (slope 1) indicates a perfect prediction. These results show a significant improvement with respect to the same data reported in Figure 5, which were obtained by constructing the surrogate without BA.

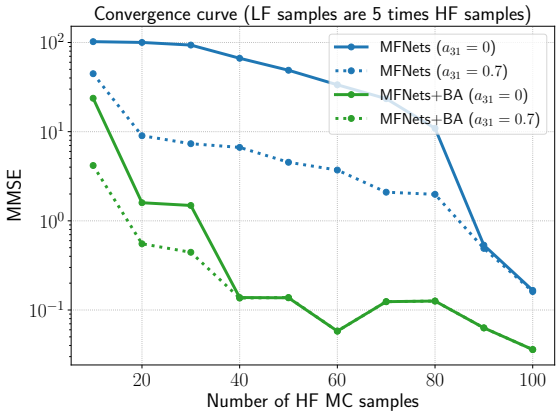
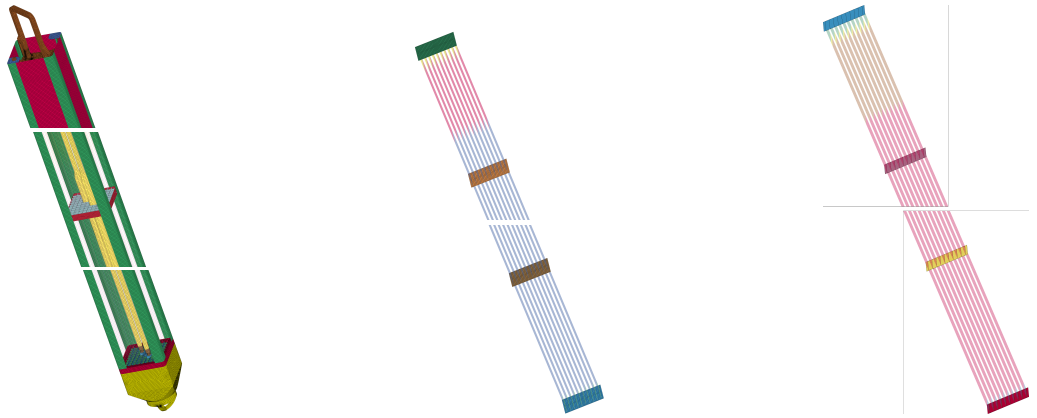


Fig. 7 MMSE for different surrogate constructions with MFNets with and without BA. For both cases we consider the set of assumptions $a_{31} = a_{32} = 0$ and $a_{31} = a_{32} = 0.7$, which correspond to the cases of zero and non-zero correlation among the PCE coefficients. The number of LF samples is fixed at 5 times the number of HF samples. For all methods, the MMSE decreases with increasing data, whereas the use of BA always improves the results, when compared to its counterpart without BA. When enough data are available, the difference among the surrogate construction with and without BA decreases to the point of being negligible. In this regime the number of data is sufficient for the training of the model and the remaining error is only driven by the availability of additional HF evaluations.

C. Fuel assembly

A more challenging numerical example is offered by a finite element (FE) model of a spent nuclear fuel assembly (FA), which we consider in this section. This model is based on a General Electric GE14 design used in boiling water reactors [30]. In more detail, the FA is a slender structure that holds 92 fuel rods and two water rods. The rods are then held in place by eight spacer grids along the long side. The water rods serve as guide tubes welded to the spacer grids, an upper handle, and a lower tie place. Finally, the spacer grids and rods are wrapped by a squared channel structure that is also welded to the upper handle and lower tie plate. The FE model is presented in Figure 8a. The model has three structural levels, the fuel rods, the spacer grids, and the skeleton, including the channel, upper handle, and lower tie plate.

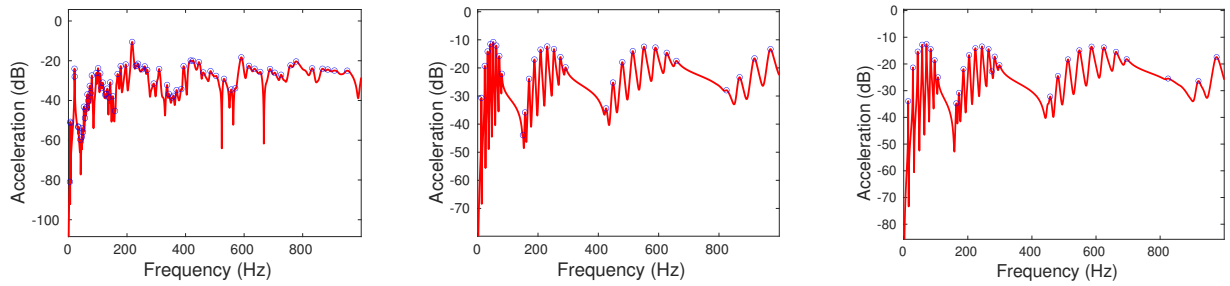


(a) Detailed fuel assembly (HF) **(b) First simplified fuel assembly (LF2)** **(c) Second simplified fuel assembly (LF1).**

Fig. 8 Finite element model of the fuel assembly models. Two simplified approximations are employed (b-c), in addition to the HF model (a).

The FE model has $N = 1,912,506$ degrees-of-freedom (DOFs), and we are interested in the vibration analysis in the frequency band $\mathcal{B} = 2\pi\times]0, 1000]$ rad/s. The boundary conditions are specified as it follows. The tip node of the lower tie plate is constrained, and two nodes on each of the four faces of the channel, at a height close to the handle structure, are constrained. Excitation is applied on a node of the lower tie plate (the red region in Figure 8a). As quantity of interest for the system, we considered the frequency response function (FRF) of a node on the adjacent face of the excitation node. The excitation and observation are both normal to the channel faces. Finally, the dynamic responses are computed with a model reduction method proposed in [31, 32], which could complete a computation by 50 nodes (16-core Intel(R) Xeon(R) CPU E5-2640 v3 @ 2.60 GHz) in about 5 minutes.

The connections of different structural levels, rod-to-grid, and grid-channel, are considered critical for the accurate evaluation of the system response; to account for their uncertainty, these connections are modeled as random variables with Beta distributions. Their mean values are 0.15 kN/mm, and 1.5 kN/mm, respectively, and the coefficients of variations for both is 20%. The random dimension of the system is $d = 8$, since there are 8 spacer grids. By choosing the parameters to be their mean values and solving the dynamic problem, we can obtain the FRF of acceleration, shown in Figure 9a. For the construction of the surrogate, we are interested in the average maximum acceleration experienced by the structure. We consider, as its proxy, the largest eight peaks along the frequency range, which are marked with blue circles in Figure 9. Two LF models are constructed. The first LF model (LF2) has an array of 10×10 fuel rods held by eight spacer grids, one upper-tie plate, and one low-tie plate; see the FE model in Figure 8b. Compared to the detailed model, the complex handle structure and the lower tie plate are simplified as two single plates; also, no channel or water rods are included in the LF model. Due to the simplification, the model has $N = 368,892$ DOFs. The upper and lower tie plates are constrained and the excitation is located at the first spacer grid from the bottom, which differs from the HF, where the excitation is introduced at the lower tie plate. The observation is also placed in a different location, with respect to the HF model: at the first space grid, but on the adjacent side of the excitation. Although the computational cost of the LF2 model is already significantly reduced by the mode superposition method, we further reduce its computational cost by applying the global reduce-order model (ROM) proposed in [33]. By applying this method, the cost to obtain the FRF is about 3 minutes, by using a single computer node. In this configuration, the cost ratio between HF and LF2 is about 80.



(a) FRF (detailed model - HF) **(b) FRF (first LF model - LF2)** **(c) FRF (second LF model - LF1)**

Fig. 9 Frequency response function (FRF) for the different approximations of the fuel assembly model. This response is associated to the mean values of the random variables describing the 8 structural connections.

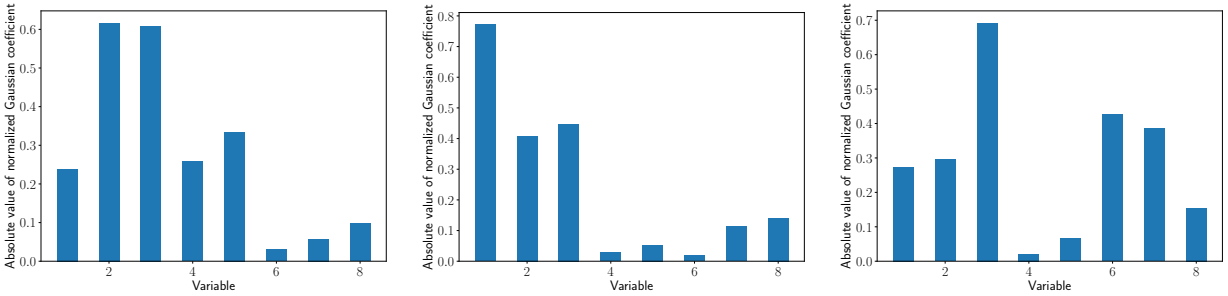
We use an even coarser mesh for the second LF model, named LF1, which leads to $N = 64,392$ DOFs. All other specifications for the second LF model are the same as the first LF model. The FE model and nominal FRF of the second LF are shown in Figure 8c and Figure 9c, respectively.

Due to the aggressive simplification of the LF models, their correlations to the HF model are small. For example, when estimated with 150 MC samples, the correlations of the first and second LF to the HF are about 0.09 and 0.015, respectively. Therefore, the two LF models are expected to only marginally help in increasing the accuracy of the HF prediction. We consider two cases with peer graph and $s_{11} = s_{22} = s_{33} = s = 1.0$. In the first case, we consider no correlations between coefficients among the models, *i.e.*, $a_{31} = a_{32} = 0$. In the second case, we consider the coefficients correlation (among the models) to be $a_{31} = a_{32} = 0.5$, which is approximately the correlations among the models in the adapted coordinates.

1. Surrogate construction with peer graph ($a_{31} = a_{32} = 0$)

Let consider a dataset of $m_1 = m_2 = 1000$, $m_3 = 150$, a polynomial degree $p_1 = p_2 = p_3 = 3$, and $\sigma_1 = \sigma_2 = \sigma_3 = 0.1$. For this application, we generated 100 test data to evaluate the performance of different surrogates construction.

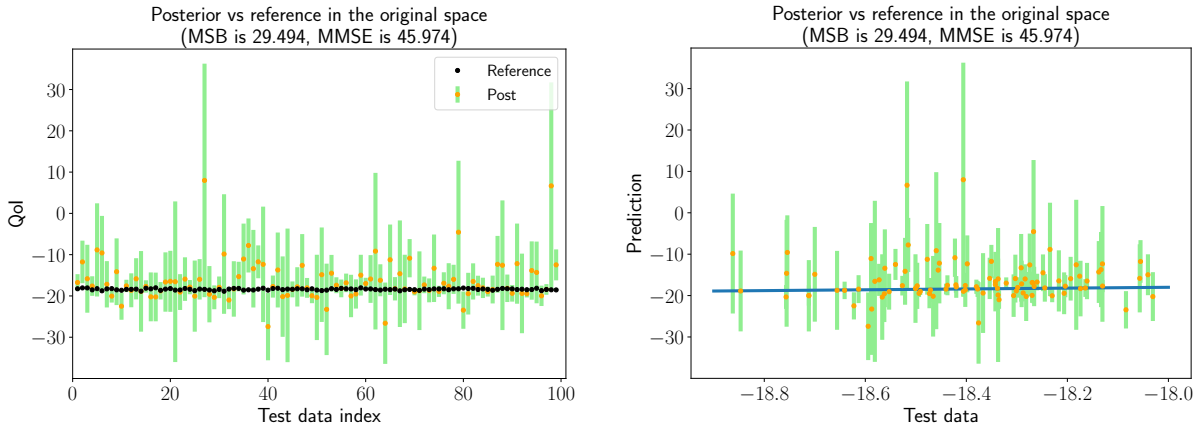
When `adaptation=True`, the rotation matrices for these three models are computed independently based on first-order PCEs. The absolute values of the normalized first-order PCE coefficients are presented in Figure 10.



(a) LF1 **(b) LF2** **(c) HF**

Fig. 10 Normalized first-order PCE coefficients in the original space of the FA. The figures from left to right correspond to models 1, 2, and 3.Normalized first-order PCE coefficients in the original space. All the models are represented: high-fidelity (c), low-fidelity LF2 (b), and the low-fidelity LF1 (a). We observe that variables 1 to 5 are important for LF1, variables 1 to 3 are important for LF2 and variables 1, 2, 3, 6 and 7 are important for the high-fidelity model. The important variables identified here are the only ones used for the construction of the surrogate when embedding BA.

The comparison between the reference data and the posterior models for the HF surrogate built with MFNets without BA is shown in Figure 11. Figure 11a presents 100 test data and the posterior predictions. Figure 11b shows the parity plot. For this case, we observe a noticeable bias of the posterior mean with MSB = 29.494. Also, the prediction



(a) Posterior predictions and test data **(b) Parity plot**

Fig. 11 Comparison between the posterior model (built in the original space with peer graph and $s = 1.0$, $a_{31} = a_{32} = 0$) and 100 reference HF test data for the fuel assembly finite element model. The green bars show the posterior mean plus and minus two standard deviations. In Figure a, both datasets, predictions and test data, are reported. In Figure b, the predictions are plotted against the corresponding test data (parity plot). In this case, the solid line (slope 1) indicates a perfect prediction.

variance is quite high, leading to MMSE = 45.974.

Similarly to the analytical test case, the prediction results for the construction of the surrogate with BA are shown in Figure 12. By comparing Figure 12 with Figure 11, we observe that the MSB of the posterior mean is reduced from 29.494 to 0.031, and the prediction variances are also significantly reduced. The MMSE decreased from 45.974 to 0.047.

2. Surrogate construction with peer structure ($a_{31} = a_{32} = 0.5$)

In this section, we still consider the construction of the surrogate based on a peer structure, but we assume a non-zero correlation between θ_3-θ_1 and θ_3-θ_2. Also, we assume $\sigma_1 = \sigma_2 = \sigma_3 = 0.1$. For these coefficients correlation, we use the same correlation estimated for the PCEs expansions in the reduced coordinates, *i.e.*, $a_{31} = a_{32} = 0.5$. Figure 13 reports the posterior of the HF surrogate constructed without BA. By comparing Figure 11 with Figure 13, we note that the surrogate construction with non-zero correlation among the PCE coefficient is more accurate, as expected. Both MSB = 2.939 and MMSE = 11.531 are significantly decreased. The prediction variance is also much smaller.

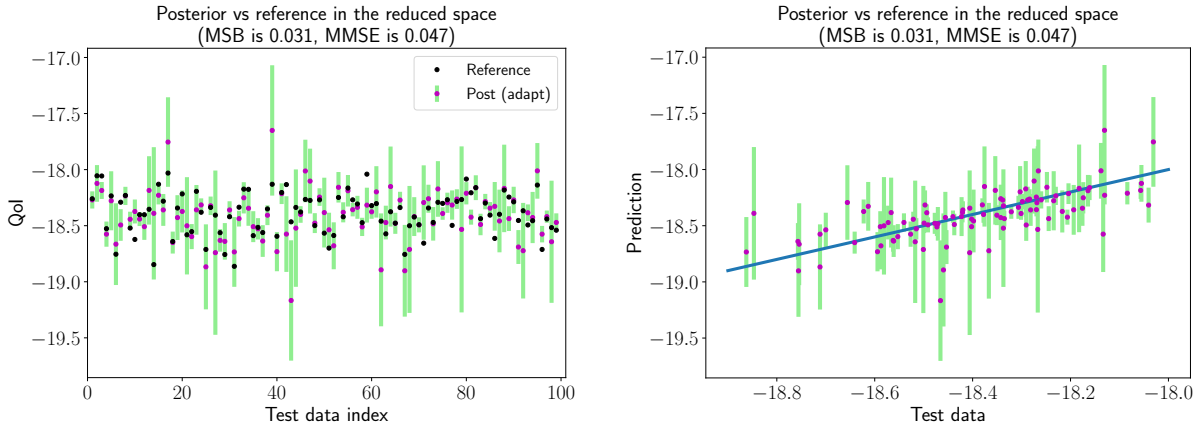
The posterior of the surrogate constructed by using MFNets with BA is shown in Figure 14. Interestingly, for this case, when training the surrogate with MFNets and embedded BA, there is little to no difference in using the correlation among the PCE coefficients. See Figure 14 compared to Figure 12; however, the use of BA is still beneficial (refer to Figures 14 and 12 compared with Figure 13).

3. Surrogate construction with hierarchical structure ($a_{21} = a_{32} = 0.5$)

For the case with correlation among the PCE coefficients, in this section we consider the use of a hierarchical graph, see Figure 1. Since the graph is changed, the prior covariance of the coefficients in step 13 of Algorithm 1 will use Eq (27). For comparison, we use the same parameters as in Section V.C.2, but with $a_{21} = a_{32} = 0.5$.

Figure 15 shows the posterior for the surrogate built with MFNets and BA on the hierarchical graph. By comparing Figure 14 with Figure 15, we note that the graph has no influence on the accuracy of the surrogate, since the results are almost identical. This result is somewhat expected following the observations made by comparing Figure 12 and Figure 14; since these figures are almost identical, we can infer that the effect of the covariance matrix of the PCE coefficient is negligible and, as a consequence, the graph, which only affects it, is expected to has a similar negligigle effect on the surrogate accuracy.

From the analytical problem, the convergence curves in Figure 7 indicate that incorporating the basis adaptation has much more profound effects, on the accuracy of the posterior model, than considering correlations in the PCE

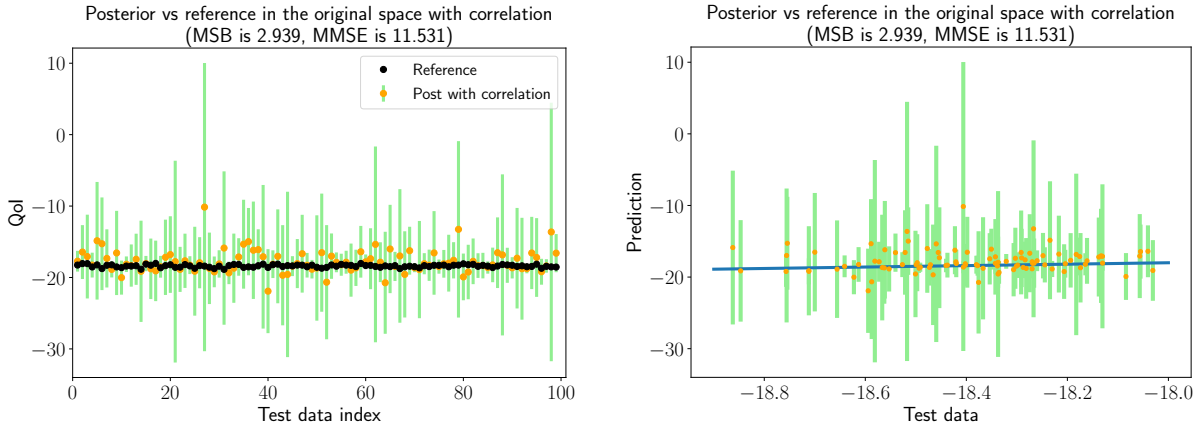


(a) Posterior predictions and test data

(b) Parity plot

Fig. 12 Comparison between the posterior model (built in the adapted space with peer graph and $s = 1.0$, $a_{31} = a_{32} = 0$) and 100 reference HF test data for the fuel assembly finite element model. The green bars show the posterior mean plus and minus two standard deviations. In Figure a, both datasets, predictions and test data, are reported. In Figure b, the predictions are plotted against the corresponding test data (parity plot). In this case, the solid line (slope 1) indicates a perfect prediction. A significant improvement in the surrogate accuracy is observed with respect to the surrogate constructed without BA, which is reported in Figure 11.

coefficients. if the number of HF samples is not significantly small, the basis adaptation can already take into account the correlation enhancement among the models. Considering the correlations in the PCE coefficients cannot further improve the accuracy. The FA application is an example where 150 MC samples of the HF model is large enough for the basis adaptation to introduce the correlation enhancement; see Figure 14. In such cases, the network structure has little to no effect on the accuracy of the posterior model, since its effect is to alter the conditional probability among coefficients. This could potentially provide a way to alleviate the computational cost associated with discovering the optimal models graph, however these results are only preliminary and should be confirmed with additional numerical experiments.



(a) Posterior predictions and test data **(b) Parity plot**

Fig. 13 Comparison between the posterior model (built in the original space with peer graph and $s = 1.0$, $a_{31} = a_{32} = 0.5$) and 100 reference HF test data for the fuel assembly finite element model. The green bars show the posterior mean plus and minus two standard deviations. In Figure a, both datasets, predictions and test data, are reported. In Figure b, the predictions are plotted against the corresponding test data (parity plot). In this case, the solid line (slope 1) indicates a perfect prediction. A higher accuracy for the surrogate is obtained in this case, when compared to the case with zero correlation among the PCE coefficients, reported in Figure 11.

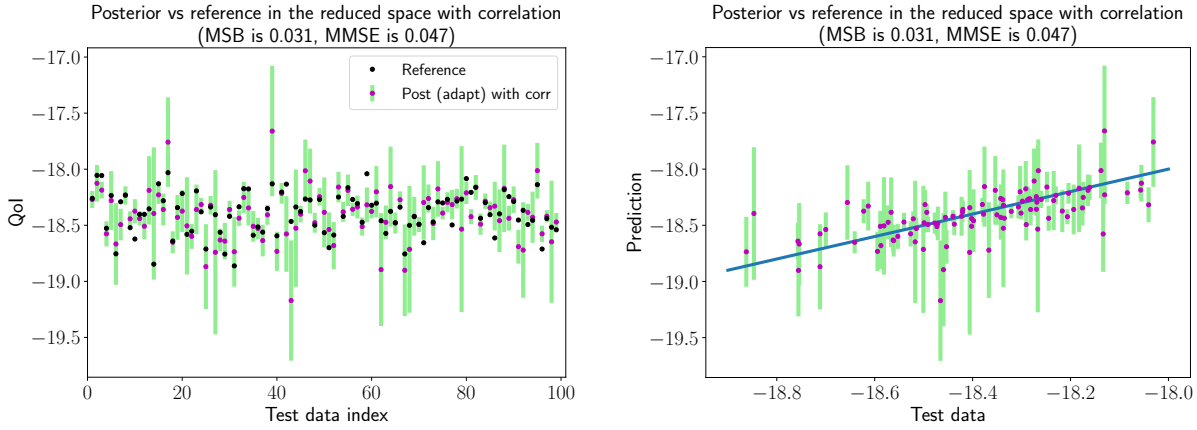
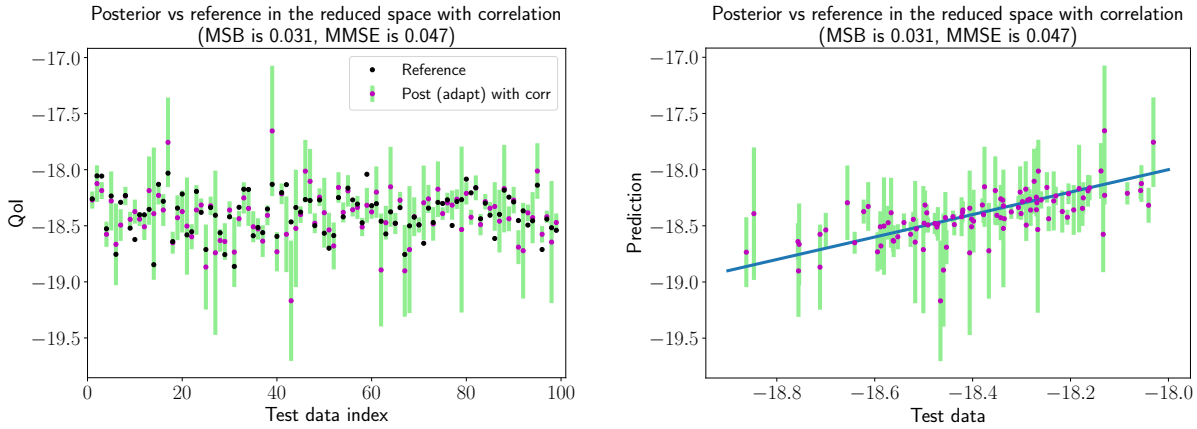


Fig. 14 Comparison between the posterior model (built in the adapted space with peer graph and $s = 1.0$, $a_{31} = a_{32} = 0.5$) and 100 reference HF test data for the fuel assembly finite element model. The green bars show the posterior mean plus and minus two standard deviations. In Figure a, both datasets, predictions and test data, are reported. In Figure b, the predictions are plotted against the corresponding test data (parity plot). In this case, the solid line (slope 1) indicates a perfect prediction. A higher accuracy for this surrogate is obtained, when compared to the same surrogate constructed with MFNets without BA, which is reported in Figure 13. However, the difference between this surrogate and the one constructed assuming zero correlation among the expansion coefficients, reported in Figure 12, is negligible.



(a) Posterior predictions and test data **(b) Parity plot**

Fig. 15 Comparison between the posterior model (built in the adapted space with $s = 1.0$, $a_{31} = a_{32} = 0.5$) and 100 reference HF test data on the hierarchical graph for the fuel assembly finite element model. The green bars show the posterior mean plus and minus two standard deviations. In Figure a, both datasets, predictions and test data, are reported. In Figure b, the predictions are plotted against the corresponding test data (parity plot). In this case, the solid line (slope 1) indicates a perfect prediction. The difference between this surrogate built on the hierarchical graph and the one obtained on the peer graph, reported in Figure 14, are almost identical.

VI. Conclusions

In this contribution, we explored the adoption of the Basis Adaptation (BA) dimension reduction strategy, within the multi-fidelity Bayesian Network strategy (MFNets), for enhancing the computational efficiency of the surrogate construction. By resorting to a more compact set of variables, discovered by the BA method, the inference problem of MFNets can be solved more efficiently, *i.e.*, relying on a smaller dataset for the high-fidelity model. We demonstrated that the use of BA within MFNets has two benefits. First, it allows to make the model representation more compact, thus reducing the latent variables to considr in the MFNets method. Second, it enhances the correlation among model, thus allowing for a more efficient use of lower-fidelity data sources. We demonstrated the idea and the efficiency of the proposed algorithm on a set of two problems: a simple analytical test problem, for verification, and a nuclear fuel assembly model. Although the adoption of BA within MFNets seems promising and can lead to orders of magnitude improvements, it is important to note here that, in our numerical experiments, MFNets has been used without an optimization of the hyperparameters. It is reasonable to assume that the performance of MFNets can be significantly improved by optimally selecting the hyperparameters first. Therefore, the numerical results presented in this manuscript should be considered as preliminary and should be extended before reaching a definitive assessment on the quantitative benefit that BA has when embedded within MFNets. Moreover, for the nuclear fuel assembly model, we observed that the use of BA within MFNets reduces its reliance of the graph structure linking the models. This observation is important because indicates that the computational effort needed to discover the graph could be potentially eliminated, although, even in this case, additional investigations are needed to corroborate this observation.

Acknowledgments

This article has been authored by employees of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA-0003525 with the U.S. Department of Energy (DOE). The employees co-own right, title and interest in and to the article and are responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan `https://www.energy.gov/downloads/doe-public-access-plan`.

References

[1] Giles, M. B., "Multilevel Monte Carlo path simulation," *Operations Research*, Vol. 56, No. 3, 2008, pp. 607–617.

[2] Haji-Ali, A.-L., Nobile, F., and Tempone, R., "Multi-index Monte Carlo: when sparsity meets sampling," *Numerische Mathematik*, Vol. 132, No. 4, 2016, pp. 767–806. doi:10.1007/s00211-015-0734-5.

[3] Peherstorfer, B., Willcox, K., and Gunzburger, M., "Survey of multifidelity methods in uncertainty propagation, inference, and optimization," *Preprint*, 2016, pp. 1–57.

[4] Geraci, G., Eldred, M. S., and Iaccarino, G., "A multifidelity multilevel Monte Carlo method for uncertainty propagation in aerospace applications," *19th AIAA Non-Deterministic Approaches Conference*, 2017, p. 1951.

[5] Gorodetsky, A. A., Geraci, G., Eldred, M. S., and Jakeman, J. D., "A generalized approximate control variate framework for multifidelity uncertainty quantification," *Journal of Computational Physics*, Vol. 408, 2020, p. 109257.

[6] Perdikaris, P., Venturi, D., Royset, J. O., and Karniadakis, G. E., "Multi-fidelity modelling via recursive co-kriging and Gaussian–Markov random fields," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Vol. 471, No. 2179, 2015. doi:10.1098/rspa.2015.0018.

[7] Kennedy, M. C., and O'Hagan, A., "Predicting the output from a complex computer code when fast approximations are available," *Biometrika*, Vol. 87, No. 1, 2000, pp. 1–13. doi:10.1093/biomet/87.1.1.

[8] Gratiet, L. L., and Garnier, J., "Recursive co-kriging model for design of experiments with multiple levels of fidelity," *International Journal for Uncertainty Quantification*, Vol. 4, No. 5, 2014, pp. 365–386.

[9] Narayan, A., Gittelson, C., and Xiu, D., "A Stochastic Collocation Algorithm with Multifidelity Models," *SIAM Journal on Scientific Computing*, Vol. 36, No. 2, 2014, pp. A495–A521. doi:10.1137/130929461.

[10] Doostan, A., Geraci, G., and Iaccarino, G., "A Bi-Fidelity Approach for Uncertainty Quantification of Heat Transfer in a Rectangular Ribbed Channel." *ASME. Turbo Expo: Power for Land, Sea, and Air, Volume 2C: Turbomachinery*, ASME, 2016.

[11] Jakeman, J. D., Eldred, M. S., Geraci, G., and Gorodetsky, A., "Adaptive multi-index collocation for uncertainty quantification and sensitivity analysis," *International Journal for Numerical Methods in Engineering*, Vol. 121, No. 6, 2020, pp. 1314–1343. doi:10.1002/nme.6268.

[12] Merritt, M., Geraci, G., Eldred, M., and Portone, T., "Hybrid multilevel Monte Carlo - polynomial chaos method for global sensitivity analysis." *Computer Science Research Institute Summer Proceedings 2020, A.A. Rushdi and M.L. Parks, eds. Technical Report SAND2020-12580R*, 2020, pp. 68–86.

[13] Gorodetsky, A., Jakeman, J., Geraci, G., and Eldred, M., "MFNets: multi-fidelity data-driven networks for Bayesian learning and prediction," *International Journal for Uncertainty Quantification*, Vol. 10, 2020, pp. 595–622.

[14] Gorodetsky, A., Jakeman, J., and Geraci, G., "MFNets: data efficient all-at-once learning of multifidelity surrogates as directed networks of information sources," *Computational Mechanics*, Vol. 68, 2021, pp. 741–758.

[15] Tipireddy, R., and Ghanem, R., "Basis adaptation in homogeneous chaos spaces," *Journal of Computational Physics*, Vol. 259, 2014, pp. 304–317.

[16] Zeng, X., Red-Horse, J., and Ghanem, R., "Accelerated basis adaptation in homogeneous chaos spaces," *Computer Methods in Applied Mechanics and Engineering*, Vol. 386, 2021, p. 114109.

[17] Koller, D., and Friedman, N., *Probabilistic graphical models: principles and techniques*, MIT press, 2009.

[18] Ghanem, R., and Spanos, P. D., *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, 1991.

[19] Cohen, A., and Migliorati, G., "Optimal weighted least-squares methods," *The SMAI journal of computational mathematics*, Vol. 3, 2017, pp. 181–203.

[20] Ghauch, Z. G., Aitharaju, V., Rodgers, W. R., Pasupuleti, P., Dereims, A., and Ghanem, R. G., "Integrated stochastic analysis of fiber composites manufacturing using adapted polynomial chaos expansions," *Composites Part A: Applied Science and Manufacturing*, Vol. 118, 2019, pp. 179–193.

[21] Thimmisetty, C., Tsilifis, P., and Ghanem, R., "Homogeneous chaos basis adaptation for design optimization under uncertainty: Application to the oil well placement problem," *Ai Edam*, Vol. 31, No. 3, 2017, pp. 265–276.

[22] Tsilifis, P., and Ghanem, R. G., "Reduced Wiener chaos representation of random fields via basis adaptation and projection," *Journal of Computational Physics*, Vol. 341, 2017, pp. 102–120.

[23] Geraci, G., and Eldred, M., "Leveraging Intrinsic Principal Directions for Multifidelity Uncertainty Quantification," *SAND2018-10817*, 2018.

[24] Geraci, G., Eldred, M., Gorodetsky, A., and Jakeman, J., "TOWARDS LEVERAGING ACTIVE DIRECTION FOR EFFICIENT MULTIFIDELITY UQ STRATEGIES," *6th European Conference on Computational Mechanics (ECCM 6)*, 2018, pp. **xxx–xxx**.

[25] Geraci, G., Eldred, M. S., Gorodetsky, A., and Jakeman, J., "Recent advancements in Multilevel-Multifidelity techniques for forward UQ in the DARPA Sequoia project," *AIAA Scitech 2019 Forum*, 2019, p. 0722.

[26] Zeng, X., Geraci, G., Eldred, M., and Ghanem, R., "Exploring Important Directions for Multifidelity Uncertainty Quantification by Basis Adaptation Method," *SIAM Conference on Computational Science and Engineering (CSE21)*, 2021.

[27] Zeng, X., Geraci, G., Eldred, M., Jakeman, J., Gorodetsky, A., and Ghanem, R., "Adaptive Basis for Multifidelity Uncertainty Quantification," *16th U.S. National Congress on Computational Mechanics (USNCCM16)*, 2021.

[28] Geraci, G., Zeng, X., Gorodetsky, A., Eldred, M., Jakeman, J., and Ghanem, R., "Uncertainty Quantification with Multifidelity Strategies Based on Models with Dissimilar Parametrizations," *14th WCCM & ECCOMAS Congress 2020*, 2021.

[29] Jakeman, J. D., "PyApprox: Approximation and probabilistic analysis of data." `https://sandialabs.github.io/pyapprox/index.html`, 2021.

[30] HITACHI, *GE14-BWR Nulcear Fuel*, 2020. `https://nuclear.gepower.com/fuel-a-plant/products/ge14`.

[31] Ezvan, O., Zeng, X., Ghanem, R., and Gencturk, B., "Multiscale modal analysis of fully-loaded spent nuclear fuel canisters," *Computer Methods in Applied Mechanics and Engineering*, Vol. 367, 2020, p. 113072.

[32] Ezvan, O., Zeng, X., Ghanem, R., and Gencturk, B., "Dominant substructural vibration modes for fully-loaded spent nuclear fuel canisters," *Computational Mechanics*, Vol. 67, No. 1, 2021, pp. 365–384.

[33] Ezvan, O., Zeng, X., Ghanem, R., and Gencturk, B., "Dominant vibration modes for broadband frequency analysis of multiscale structures with numerous local vibration modes," *International Journal for Numerical Methods in Engineering*, Vol. 117, No. 6, 2019, pp. 644–692.