Exploring multi-fidelity noisy data: Methods and real-world examples

K Giannoukou, P Ascia, S Marelli, F Duddeck, B Sudret

ETH zürich

Exploring multi-fidelity noisy data: Methods and real-world examples

K. Giannoukou, P. Ascia, S. Marelli, F. Duddeck and B. Sudret

Chair of Risk, Safety and Uncertainty Quantification
Stefano-Franscini-Platz 5
CH-8093 Zürich

Risk, Safety & Uncertainty Quantification

Data Sheet

Journal:	Proc. 31st International Conference on Noise and Vibration Engineering (ISMA2024) — 10th International Conference on Uncertainty in Structural Dynamics (USD2024), Leuven (Belgium), September 9 - 11, 2024
Report Ref.:	RSUQ-2024-008
Arxiv Ref.:	-
DOI:	-
Date submitted:	June 28, 2024
Date accepted:	September 11, 2024

Exploring multi-fidelity noisy data: Methods and real-world examples

K. Giannoukou [1], **P. Ascia** [2], **S. Marelli** [1], **F. Duddeck** [2], **B. Sudret** [1]
[1] ETH Zurich, Department of Civil, Environmental and Geomatic Engineering,
Stefano-Franscini-Platz 5, 8093, Zurich, Switzerland

[2] Technical University of Munich, TUM School of Engineering and Design,
Theresienstraße 90, 80333, Munich, Germany

Abstract
Multi-fidelity surrogate models (MFSMs) are a well-established tool to combine information from sources with diverse computational fidelities into a single surrogate model. The sources of higher or lower fidelity can be, for example, computer simulations or physical experiments. MFSMs can exhibit enhanced predictive accuracy and reduced costs in emulating the response of complex systems, outperforming their single-fidelity surrogate model counterparts at comparable training costs. In real-world applications, uncertainty is present in the data, regardless of their fidelity. This uncertainty can be due to measurement noise, numerical noise, or unobserved/latent variables, and adds a layer of complexity by introducing non-deterministic behavior in the system response. In this work, we provide a framework to address the uncertainty in MFSM scenarios. The effectiveness of our approach is demonstrated through a transfer learning application in crashworthiness and a real-world wind turbine application, showcasing the applicability and versatility of our proposed methods.

1 Introduction

Engineering systems and models are often highly complex and costly to evaluate, while applications requiring numerous model evaluations, such as uncertainty quantification and optimization, are becoming increasingly popular. Consequently, being able to predict the response of complex systems in a cost-effective way is more crucial than ever.

In many cases, multiple computational models or experiments with varying levels of fidelity are available for making these predictions. High-fidelity (HF) models provide accurate results, but tend to be computationally and/or financially costly. Conversely, low-fidelity (LF) models are less costly to run, but they are also less accurate. In such scenarios, multi-fidelity surrogate models (MFSMs) can be employed to integrate these varying fidelity models into a single surrogate model (SM). MFSMs typically augment a small HF dataset with one or more larger lower-fidelity ones. In general, the HF and LF data to construct a MFSM can be obtained from experiments or computer simulations. When both kinds of data co-exist, experiments are usually considered as the HF, while computer simulations are treated as the LF [1, 2].

In the real world, both high- and low-fidelity models are commonly affected by various types of uncertainty. Depending on the nature of the uncertainty, we can distinguish two cases of models affected by latent variability, which can be treated differently. The first case is deterministic models contaminated by external noise, such as measurement noise in experimental data or numerical noise in computer simulators. In this case, the noise can be considered homoscedastic over the deterministic model response, and the objective of a SM is typically to predict the noise-free model response, effectively eliminating the external noise. Regression-based approaches are commonly employed for this purpose, thanks to their well-known denoising capabilities.

The second class of models are inherently stochastic models, also known as stochastic simulators. The response of these models at each input is a random variable. Recent methods for emulating stochastic

simulators in a single-fidelity setting aim to predict the entire response distribution at each point [3, 4].

In this work, we focus on the first case and consider it in a multi-fidelity (MF) setting. This means that both the HF and LF models are deterministic, and data from these models are possibly contaminated by additive homoscedastic noise. We consider this noise irreducible, assuming that once the corresponding datasets are provided for training surrogate models, there is no opportunity to repeat the experiments with more precise measuring instruments or repeat computations with increased precision. Moreover, due to the limited training data for both the low- and the high-fidelity models, MFSM predictions entail epistemic uncertainty. In this context, we are interested in assessing the underlying noise-free deterministic HF response, by combining information from all the different fidelity sources in a MFSM. Additionally, our goal is to provide estimates of the different kinds of uncertainty in our MFSM predictions, expressed through confidence and prediction intervals. Confidence intervals (CIs) reflect the uncertainty about where the underlying noise-free HF model lies, while prediction intervals (PIs) represent the uncertainty about unseen noise-contaminated HF observations [5].

The adoption of MFSMs has been consistently increasing across many fields of applied science over the past three decades. Most works consider an ideal noise-free setting where both the HF and LF models are computer simulations. Several studies are based on Gaussian process (GP) modeling [6, 7, 8, 9], and more recently deep Gaussian processes [10, 11]. Other approaches include, among others, polynomial chaos expansion-based based MFSMs [12, 13] and neural network-based approaches [14, 15, 16, 17]. A comparatively smaller body of works has addressed noisy models in a multi-fidelity setting, including the GP regression MFSMs in [18, 19], as well as the linear regression-based framework of [20]. Recently, multi-fidelity surrogate modeling has found applications in the gray-box modeling paradigm. [2] utilized MFSMs to construct gray-box surrogate models, by combining experimental data and computer simulations, with the former considered as high-fidelity and the latter as low-fidelity.

Our work is based on the comprehensive framework introduced in [2]. In addition to constructing MFSMs with noise contaminated HF and LF data, and providing the corresponding confidence and prediction intervals, we explore here the applicability of MF surrogate modeling in a transfer learning scenario. This scenario involves crashworthiness simulations, where knowledge from an earlier crash box design needs to be transferred to a newer design, reducing the need for a large number of simulations. Moreover, we demonstrate the applicability of the examined MF surrogate modeling framework through a real-world application involving wind turbine simulations of different fidelities.

2 Multi-fidelity surrogate modeling

2.1 Theory and methodology

Let us consider two information sources producing data at different fidelity levels. With the input space being $\boldsymbol{X} \in \mathbb{R}^M$, and the output space $Y \in \mathbb{R}$, let $(\boldsymbol{\mathcal{X}}_{\mathrm{H}}, \mathcal{Y}_{\mathrm{H}})$ be a HF dataset of size N_{H} with noise-contaminated observations. The HF model response y_{H} at input point $\boldsymbol{x}$ is expressed as:

$$y_{\mathrm{H}}(\boldsymbol{x}, \varepsilon_{\mathrm{H}}) = \psi_{\mathrm{H}}(\boldsymbol{x}) + \varepsilon_{\mathrm{H}}, \tag{1}$$

where $\boldsymbol{x}$ is a realization of $\boldsymbol{X}$, $\psi_{\mathrm{H}}(\boldsymbol{x})$ is the unknown HF deterministic function, and ε_{H} is an additive noise term, independent of $\boldsymbol{x}$ and modeled as a random variable following a zero-mean distribution:

$$\varepsilon_{\mathrm{H}} \sim f_{\varepsilon_{\mathrm{H}}}(\varepsilon_{\mathrm{H}}), \qquad \mathbb{E}\left[\varepsilon_{\mathrm{H}}\right] = 0. \tag{2}$$

Each observation $y_{\mathrm{H}}(\boldsymbol{x}_{\mathrm{H}}^{(i)}, \varepsilon_{\mathrm{H}}^{(i)})$ in $(\boldsymbol{\mathcal{X}}_{\mathrm{H}}, \mathcal{Y}_{\mathrm{H}})$ is denoted as $y_{\mathrm{H}}^{(i)}$.

In addition, consider another information source of lower fidelity providing the dataset $(\boldsymbol{\mathcal{X}}_{\mathrm{L}}, \mathcal{Y}_{\mathrm{L}})$:

$$y_{\mathrm{L}}(\boldsymbol{x}, \varepsilon_{\mathrm{L}}) = \psi_{\mathrm{L}}(\boldsymbol{x}) + \varepsilon_{\mathrm{L}}, \tag{3}$$

with $\varepsilon_{\mathrm{L}} \sim f_{\varepsilon_{\mathrm{L}}}(\varepsilon_{\mathrm{L}})$ and $\mathbb{E}[\varepsilon_{\mathrm{L}}] = 0$. Each observation $y_{\mathrm{H}}(\boldsymbol{x}_{\mathrm{L}}^{(i)}, \varepsilon_{\mathrm{L}}^{(i)})$ in $(\boldsymbol{\mathcal{X}}_{\mathrm{L}}, \mathcal{Y}_{\mathrm{L}})$, where $i = 1, \ldots, N_{\mathrm{L}}$, with $N_{\mathrm{L}} > N_{\mathrm{H}}$ is denoted as $y_{\mathrm{L}}^{(i)}$.

A MFSM aims to estimate the underlying deterministic HF function $\psi_{\mathrm{H}}(\boldsymbol{x})$ with $\hat{\psi}_{\mathrm{H}}(\boldsymbol{x})$ by leveraging all available variable-fidelity data. The low-fidelity (LF) response can be approximated by a surrogate model $\hat{\psi}_{\mathrm{L}}(\boldsymbol{x}) \approx \psi_{\mathrm{L}}(\boldsymbol{x})$. Then, the MFSM can be expressed as a linear combination between the LF surrogate and a discrepancy function:

$$\hat{\psi}_{\mathrm{H}}(\boldsymbol{x}) = \rho \cdot \hat{\psi}_{\mathrm{L}}(\boldsymbol{x}) + \delta(\boldsymbol{x}), \tag{4}$$

where ρ is a constant scaling factor and $\delta(\boldsymbol{x})$ is a discrepancy function. The choice of SMs for the LF model and the discrepancy function can vary, including polynomial chaos expansion, GP regression, or neural networks.

2.2 Implementation

Starting from a high fidelity dataset $(\boldsymbol{\mathcal{X}}_{\mathrm{H}}, \mathcal{Y}_{\mathrm{H}})$ and a lower-fidelity one $(\boldsymbol{\mathcal{X}}_{\mathrm{L}}, \mathcal{Y}_{\mathrm{L}})$, we use the approach introduced in [2] to construct of a MFSM. This approach employs polynomial chaos expansion (PCE) as a surrogate model for the MFSM, expressed as described in Eq. (4).

PCE approximates a model with finite variance using its spectral representation on a polynomial basis [21]. Consider a random vector $\boldsymbol{X} \in \mathbb{R}^M$ with independent components and a joint probability density function (PDF) $f_{\boldsymbol{X}}(\boldsymbol{x})$. Then, the truncated PCE of a model $\mathcal{M}(\boldsymbol{x})$ is given by:

$$\tilde{\mathcal{M}}_{\mathrm{PCE}}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}), \tag{5}$$

where $c_{\boldsymbol{\alpha}} \in \mathbb{R}$ are the coefficients of the multivariate polynomials $\{\Psi_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathcal{A}\}$. All polynomials $\Psi_{\boldsymbol{\alpha}}$ are orthogonal with respect to the PDF $f_{\boldsymbol{X}}$. The set $\mathcal{A} \subset \mathbb{N}^M$ consists of multi-indices of the polynomials and can be determined using various truncation schemes [22].

For the construction of the PCEs involved in this methodology, the spectral coefficients are computed using least angle regression. For the choice of the polynomial basis, degree adaptivity is used [22]. The construction of the multi-fidelity surrogate model involves the following steps:

1. Train a PCE model $\hat{\psi}_{\mathrm{L}}(\boldsymbol{x})$ the LF data $(\boldsymbol{\mathcal{X}}_{\mathrm{L}}, \mathcal{Y}_{\mathrm{L}})$;
2. Evaluate the LF PCE model at the HF input samples. These responses are denoted as $\hat{\psi}_{\mathrm{L}}(\boldsymbol{x}_{\mathrm{H}}^{(i)})$;
3. Estimate the scaling factor ρ as $\hat{\rho} \approx \frac{1}{N_H} \sum_{i=1}^{N_H} \frac{y_{\mathrm{H}}^{(i)}}{\hat{\psi}_{\mathrm{L}}(\boldsymbol{x}_{\mathrm{H}}^{(i)})}$;
4. Obtain a training set $(\boldsymbol{\mathcal{X}}_{\mathrm{H}}, \{y_{\mathrm{H}}^{(i)} - \hat{\rho}\hat{\psi}_{\mathrm{L}}(\boldsymbol{x}_{\mathrm{H}}^{(i)}), i = 1, ..., N_{\mathrm{H}}\})$ for the discrepancy function δ in Eq. (4), and use it to train a PCE model $\hat{\delta}(\boldsymbol{x})$;
5. Keep the sparse PCE basis functions of the PCE $\hat{\delta}(\boldsymbol{x})$, and using ordinary least squares (OLS), refit the corresponding coefficients jointly with ρ, considering $\hat{\psi}_{\mathrm{L}}(\boldsymbol{x})$ as another basis function;
6. Plug $\hat{\psi}_{\mathrm{L}}(\boldsymbol{x}), \hat{\rho},$ and $\hat{\delta}(\boldsymbol{x})$ in Eq. (4) to obtain the MFSM predictor.

For a more detailed discussion and explanation of the aforementioned steps, the reader is referred to [2].

3 Confidence and prediction intervals

3.1 Theory and methodology

Confidence and prediction intervals are used to express different kinds of uncertainty in predictive models trained using noise-contaminated data. Here, we examine both in a multi-fidelity setting.

Confidence intervals express the uncertainty about the true underlying HF model, or in other words, about the error between the true noise-free HF function and the MFSM regression model output:

$$\psi_{\mathrm{H}}(\boldsymbol{x}) - \hat{\psi}_{\mathrm{H}}(\boldsymbol{x}). \tag{6}$$

This uncertainty stems from the incomplete information available due to the finite size of the HF and LF experimental designs (EDs). The $(1-2\alpha)$ CI at $\boldsymbol{x}_0$ is an interval $[\psi_{\mathrm{l},\alpha}(\boldsymbol{x}_0), \psi_{\mathrm{u},\alpha}(\boldsymbol{x}_0)]$, such that:

$$\mathbb{P}\left[\psi_{\mathrm{l},\alpha}(\boldsymbol{x}_0) < \psi_{\mathrm{H}}(\boldsymbol{x}_0) < \psi_{\mathrm{u},\alpha}(\boldsymbol{x}_0)\right] = 1 - 2\alpha. \tag{7}$$

On the other hand, prediction intervals express the uncertainty about an unseen noise-contaminated HF observation, or in other words, about the error between a noise-contaminated HF observation and the MFSM regression model output:

$$y_{\mathrm{H}}(\boldsymbol{x}, \varepsilon_{\mathrm{H}}) - \hat{\psi}_{\mathrm{H}}(\boldsymbol{x}) \overset{\text{Eq. (1)}}{=} \psi_{\mathrm{H}}(\boldsymbol{x}) - \hat{\psi}_{\mathrm{H}}(\boldsymbol{x}) + \varepsilon_{\mathrm{H}}. \tag{8}$$

This means that the PIs are influenced not only by the reducible model error $\psi_{\mathrm{H}}(\boldsymbol{x}) - \hat{\psi}_{\mathrm{H}}(\boldsymbol{x})$, but also by the noise affecting the data ε_{H} on the HF data. From Eqs. (6), (8), we see that a PI is wider than the corresponding CI, since the former includes the latter.

The $(1-2\alpha)$ PI at $\boldsymbol{x}_0$ is an interval $[y_{\mathrm{l},\alpha}(\boldsymbol{x}_0), y_{\mathrm{u},\alpha}(\boldsymbol{x}_0)]$, such that:

$$\mathbb{P}\left[y_{\mathrm{l},\alpha}(\boldsymbol{x}_0, \varepsilon_{\mathrm{H}}) < y_{\mathrm{H}}(\boldsymbol{x}_0, \varepsilon_{\mathrm{H}}) < y_{\mathrm{u},\alpha}(\boldsymbol{x}_0, \varepsilon_{\mathrm{H}})\right] = 1 - 2\alpha. \tag{9}$$

3.2 Implementation

For the construction of confidence and prediction intervals, we follow the approach based on bootstrap [23] introduced in [2]. To construct a CI at a given input $\boldsymbol{x}_0$, we first obtain N_{B} multi-fidelity experimental designs by independently resampling with replacement N_{B} HF and N_{B} LF experimental designs and matching them one-to-one. Then, we train the N_{B} MFSMs as described in Section 2.2, and evaluate them at $\boldsymbol{x}_0$. We denote the bootstrap MFSM evaluations as $\hat{\psi}^*_{\mathrm{H},i}(\boldsymbol{x}_0)$, $i = 1, ..., N_{\mathrm{B}}$.
Then, we can obtain the $(1-2\alpha)$ CI about the underlying HF model as:

$$[\psi_{\mathrm{l},\alpha}(\boldsymbol{x}_0), \psi_{\mathrm{u},\alpha}(\boldsymbol{x}_0)] = \left[\hat{\psi}^{*[\alpha]}_{\mathrm{H}}(\boldsymbol{x}_0), \hat{\psi}^{*[1-\alpha]}_{\mathrm{H}}(\boldsymbol{x}_0)\right], \tag{10}$$

where $\hat{\psi}^{*[\alpha]}_{\mathrm{H}}(\boldsymbol{x}_0)$ and $\hat{\psi}^{*[1-\alpha]}_{\mathrm{H}}(\boldsymbol{x}_0)$ are the α- and $(1-\alpha)$-quantiles of the empirical quantile function of $\hat{\psi}^*_{\mathrm{H}}(\boldsymbol{x}_0)$.

To construct a PI at a given input $\boldsymbol{x}_0$, we initially follow the same procedure used for the corresponding CI. Then, we add an additional step to account for the irreducible noise ε_{H} in the HF observations. To this end, we first perform statistical inference for ε_{H} using the noise samples $\varepsilon^{(i)}_{\mathrm{H}}$ obtained as:

$$\varepsilon^{(i)}_{\mathrm{H}} = y^{(i)}_{\mathrm{H}} - \mu^*(\boldsymbol{x}^{(i)}_{\mathrm{H}}), \text{ for } i = 1, ..., N_{\mathrm{H}}, \tag{11}$$

where $y^{(i)}_{\mathrm{H}}$ is a HF model response and $\mu^*(\boldsymbol{x}^{(i)}_{\mathrm{H}})$ is the mean of the bootstrap models evaluations at $\boldsymbol{x}^{(i)}_{\mathrm{H}}$. Having these noise samples, the best-fitting distribution is selected among various zero-mean distributions using the Bayesian information criterion [24]. Finally, we add a new noise sample $\hat{\varepsilon}_{\mathrm{H},i}$ to each bootstrap model evaluation $\hat{\psi}^*_{\mathrm{H},i}(\boldsymbol{x}_0)$:

$$\hat{y}^*_{\mathrm{H},i}(\boldsymbol{x}_0, \hat{\varepsilon}_{\mathrm{H}}) = \hat{\psi}^*_{\mathrm{H},i}(\boldsymbol{x}_0) + \hat{\varepsilon}_{\mathrm{H},i}, \text{ for } i = 1, ..., N_{\mathrm{b}}. \tag{12}$$

Then the $(1-2\alpha)$ PI about a new unseen noisy observation at $\boldsymbol{x}_0$ can be obtained as:

$$[y_{\mathrm{l},\alpha}(\boldsymbol{x}_0, \varepsilon_{\mathrm{H}}), y_{\mathrm{u},\alpha}(\boldsymbol{x}_0, \varepsilon_{\mathrm{H}})] = \left[\hat{y}^{*[\alpha]}_{\mathrm{H}}(\boldsymbol{x}_0, \hat{\varepsilon}_{\mathrm{H}}), \hat{y}^{*[1-\alpha]}_{\mathrm{H}}(\boldsymbol{x}_0, \hat{\varepsilon}_{\mathrm{H}})\right], \tag{13}$$

where $\hat{y}_{\mathrm{H}}^{*[\alpha]}(\boldsymbol{x}_0, \hat{\varepsilon}_{\mathrm{H}})$ and $\hat{y}_{\mathrm{H}}^{*[1-\alpha]}(\boldsymbol{x}_0, \hat{\varepsilon}_{\mathrm{H}})$ are the α- and $(1-\alpha)$-quantiles of the empirical quantile function of $\hat{y}_{\mathrm{H}}^{*}(\boldsymbol{x}_0, \hat{\varepsilon}_{\mathrm{H}})$.

4 Applications

In this section, we examine the performance of the framework proposed in [2] and described in Sections 2, 3 in two applications. The first is a transfer learning application in crashworthiness, where negligible to minor numerical noise is expected. The second is a real-world application involving two aero-servo-elastic simulators for a wind turbine, where residual stochasticity is present instead.

The validation process entails assessing the predictive performance of the MFSM compared to the PCE surrogate model trained only on the corresponding HF data. We perform this comparison for increasing HF experimental design sizes. To assess the performance of the examined surrogate models, we use the normalized validation error ϵ_{val}, computed on a test set consisting of N_{t} unseen data points $\boldsymbol{x}_{\mathrm{t}}^{(i)}$, not previously used for training. This error is computed as follows:

$$\epsilon_{\text{val}} = \frac{\sum_{i=1}^{N_{\mathrm{t}}} (y_{\mathrm{t}}^{(i)} - \hat{\psi}_{\mathrm{H}}(\boldsymbol{x}_{\mathrm{t}}^{(i)}))^2}{\sum_{i=1}^{N_{\mathrm{t}}} (y_{\mathrm{t}}^{(i)} - \mu_y)^2}, \tag{14}$$

where $y_{\mathrm{t}}^{(i)}$ is the HF model response at the test point $\boldsymbol{x}_{\mathrm{t}}^{(i)}$, and μ_y is the mean value of the HF response. Finally, we construct and interpret the confidence and prediction intervals for the MFSM predictions.

In various stages of implementing the following applications, including the implementation of PCEs, sampling experimental designs, and other tasks, we used UQLab [25], a general-purpose uncertainty quantification software.

4.1 Transfer learning in crashworthiness simulations

Our first application of the multi-fidelity surrogate modeling framework is in the field of crashworthiness. In recent years, the rise in vehicle usage has driven the automotive industry to enhance crashworthiness for improved safety and environmental sustainability. A common design criterion for crashworthiness is the usage ratio (UR) of the energy absorber [26], defined as follows:

$$\mathrm{UR} = \frac{d}{l}, \tag{15}$$

where d is the total deformation measured after an impact and l is the total length of the undeformed crashworthy structure, i.e., of the crash box. The UR indicates the extent to which the material of the crash box is utilized to dissipate energy. A higher UR signifies better performance.

A common method for measuring the UR of a crash box is through a drop-tower test. In such test, the crash box is crushed by a rigid mass, known as the impactor, which is accelerated to a predefined speed to impact the crash box with a specific kinetic energy. The crash box is securely fixed to the test bed to prevent lateral movements during the impact. While the test can be conducted physically, it is also standard practice to evaluate a crash box virtually using explicit finite element simulations. Simulated testing is efficient and allows for easy modification of design parameters.

In this application, we focus on two crash box designs: an early-stage simplified design and a later-stage design that is closer to production, as shown in Figure 1. We have conducted numerous simulations for the early-stage design, examining how the thickness of the front part of the crash box (blue part in Figure 1(a)), the thickness of the rear part of the crash box (green part in Figure 1(a)), and the mass of the impactor affect the variation in the usage ratio. These three inputs are modeled as random variables, following uniform distributions, and their respective ranges are detailed in Table 1. The goal in this application is to examine how these three variables affect the usage ratio for the later-stage design crash box, while keeping the number

of expensive new simulations to a minimum. This transfer learning scenario is not typical for multi-fidelity modeling, as the simulation times for both designs are similar. By approaching it as a MF modeling problem, we aim to save computational resources and time by reducing the number of simulations needed for the new design. To this end, we consider the old design as the low fidelity and the new design as the high fidelity model. In both simulations little to no numerical noise is expected. We first evaluate the performance of

(a) Early-stage crash box design

(b) Later-stage crash box design

Figure 1: Illustration of the two crash box designs considered for transfer learning. The goal is to transfer knowledge from the early-stage crash box to the later-stage crash box, focusing on the energy absorbed under specified crash conditions.

Table 1: Crashworthiness application – Input variables and their distributions

Variable	Distribution	Parameters
Thickness of the front part (t_1) [mm]	Uniform	$[0.7, 4]$
Thickness of the front part (t_2) [mm]	Uniform	$[0.7, 4]$
Impactor mass (m) [kg]	Uniform	$[0.04, 0.35]$

our MFSM as we increase the size of the high-fidelity (HF) experimental design, ranging from 5 to 160 data points. The total number of available low-fidelity (LF) simulations is 400, with the LF experimental design fixed at these data points. The validation error is computed on an external validation set consisting of 225 unseen HF data points.

In Figure 2, we notice that for small HF experimental designs, the multi-fidelity surrogate model consistently outperforms the corresponding single-fidelity PCE model trained only on the corresponding HF data. This performance disparity becomes more pronounced as the HF experimental design size decreases. Once the number of HF data points exceed approximately 40, the performance of the MFSM aligns closely with that of the single-fidelity SM. For larger EDs, the single-fidelity SMs begin to outperform the MFSMs, likely due to the bias introduced by the low-fidelity model. Moreover, unlike the single-fidelity surrogate trained only on HF data, the MFSM always outperforms the single-fidelity PCE trained solely on LF data, as indicated by the black dashed line used as a baseline.

In conclusion, employing an MFSM in this context offers a significant advantage over single high-fidelity surrogate models when the number of available HF data points is limited, specifically when there are fewer than 40 new design (i.e HF) model evaluations. Moreover, the MFSM always offers an advantage over using a surrogate model trained solely on data for the old design (i.e LF).

We now showcase the CIs and PIs for our MFSM prediction. Following the results in the previous section, we selected a HF training set size of 20 samples, because more HF data does not offer a significant advantage over a single-fidelity approach for this specific example. The LF experimental design is kept to the original 400 samples. The number of bootstrap replications performed is $N_{\mathrm{B}} = 1,000$.

To illustrate the constructed CIs and PIs in this application, we select the thickness t_2 of the rear part of the crash box, and the mass m of the impactor as the most important input random variables. We selected them by performing a PCE-based Sobol' indices sensitivity analysis [27], where all the available HF data were used.

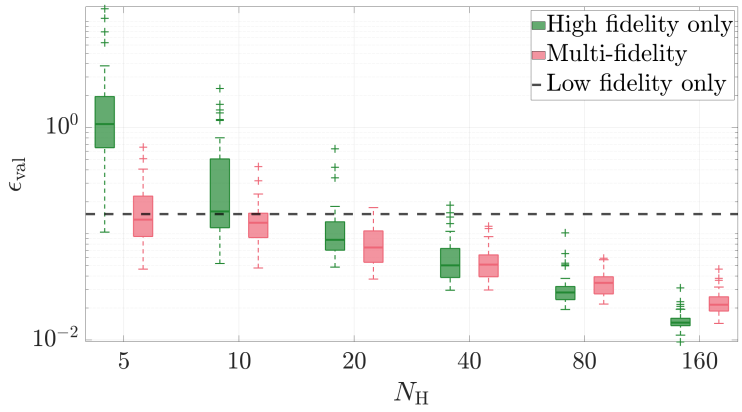
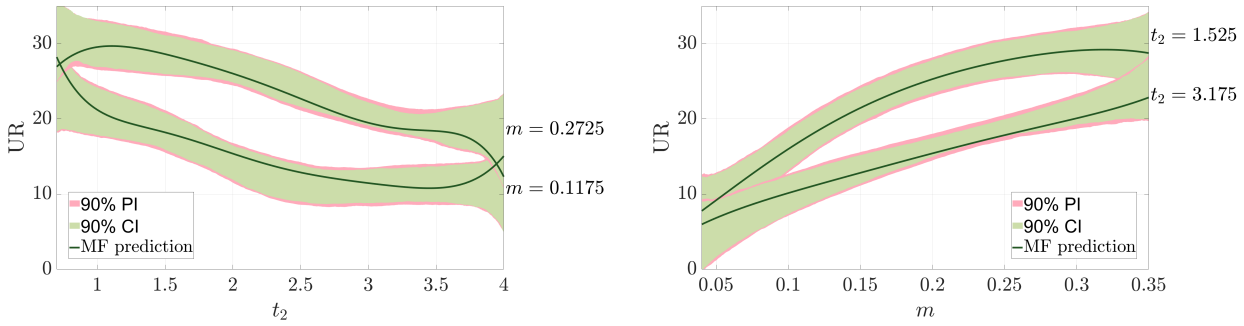


Figure 2: Crashworthiness application – Convergence of the validation error ϵ_{val} with increasing HF training data. Comparison between our MFSM and a PCE model trained solely on HF data. The dashed line indicates the error of a PCE model trained only on LF data.

Figure 3 illustrates the 90% CIs (green area) and PIs (pink area) along slices in the two selected dimensions with the other input parameter t_1 kept at its mean value. The green curve indicates the MFSM response. In Figures 3(a), 3(b), the two slices correspond to the 0.25- and 0.75-quantiles of t_2 and m respectively.

We observe that the prediction intervals indeed enclose the corresponding confidence intervals. However, the former are only marginally wider than the latter. This indicates that our approach correctly identifies the low-level of noise, demonstrating its applicability even in scenarios without noisy data. Moreover, the epistemic uncertainty, as indicated by the confidence intervals, appears to be quite large almost everywhere, likely due to the very small HF training set size used. With a larger HF training set, the confidence intervals are expected to narrow down significantly.



(a) Selected dimension: thickness of rear part t_2 [mm]

(b) Selected dimension: impactor mass m [kg]

Figure 3: Crashworthiness application – 90% confidence and prediction intervals for the MFSM along slices in the two selected dimensions.

4.2 Aero-servo-elastic simulation of a wind turbine

In our second multi-fidelity surrogate modeling application, we examine a real-world scenario, involving wind turbine simulations using two different simulators, as conducted by [28]. This study focuses on an onshore wind turbine with a 90-meter tower, a rotor diameter of 110 meters, and a rated power of 2 MW.

The objective is to assess how variations in wind speed, turbulence intensity, and shear profile affect the fore-aft extreme bending moment at the tower bottom. The quantity of interest is the maximum fore-aft bending moment at the tower bottom. The wind speed, turbulence intensity, and shear exponent are treated as random variables following uniform distributions, with their respective ranges presented in Table 2.

Table 2: Wind turbine application – Input variables and their distributions

Variable	Distribution	Parameters
Wind speed (U) [m/s]	Uniform	$[4, 25]$
Turbulence intensity (σ_{U}) [m/s]	Uniform	$[0.1, 6]$
Wind shear exponent (α) [-]	Uniform	$[-1, 1.5]$

The data used in this study is generated by [28], and consists in 10-minute time series simulations using two distinct numerical aeroservo-elastic simulators: Bladed [29] and FAST [30]. A 10-minute time series simulation in FAST runs in about 5 minutes, whereas the same simulation in Bladed takes around 30 minutes in real time. Data from Bladed are treated as high-fidelity, while data from FAST are considered low-fidelity. Detailed technical characteristics of these simulators are available in [28, 29, 30].

The three specified input random variables were used to generate wind fields. Both wind turbine simulators received a wind field as input and produced a 10-minute time series simulation of the structural response as output. Since wind fields are stochastic with respect to the three inputs, multiple wind field realizations were generated for each three-dimensional input sample using different random seeds. For each output time series, the maximum fore-aft extreme bending moment at the tower bottom was computed and averaged over the random seeds corresponding to each input. For each input, 12 different stochastic seeds were considered for Bladed and 24 for FAST, resulting in a total of $4,344$ simulations for Bladed and $33,480$ for FAST. This eventually produced 362 HF and $1,395$ LF data points.

Both wind turbine simulators are deterministic with respect to the input wind fields. However, the wind fields themselves are stochastic with respect to the input random variables, resulting in wind turbine simulator responses that contain significant residual stochasticity, which we treat here as additive homoscedastic noise.

Figure 4 shows the convergence of the validation error of the MFSM compared to the single-fidelity PCE trained only on HF data, as the size of the HF training dataset increases from 10 to 320 data points. In each case, for the training of the MFSM, all the available LF data is used each time. Also, the validation error of a PCE model trained exclusively on the LF data is represented by the black dashed line. As the validation data consists of unseen HF data containing noise, the validation error is not expected to reach zero but to approach a value that reflects the irreducible noise in the HF data.

As expected following [28], we once again observe that the MFSM consistently outperforms the single-fidelity HF surrogate model, with the performance difference being more pronounced for smaller HF training set sizes. Additionally, the MFSM always performs better than the single-fidelity model trained solely on the LF data.

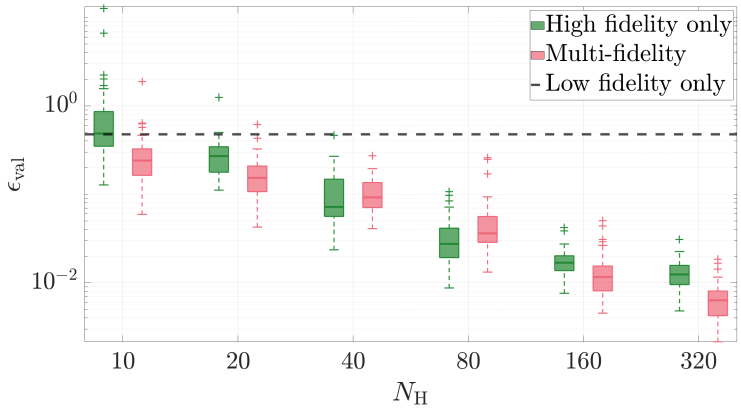


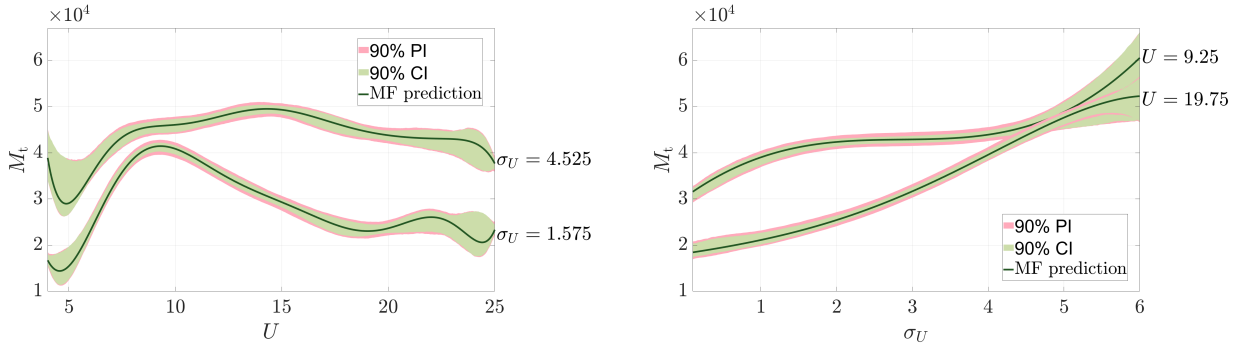
Figure 4: Wind turbine application – Convergence of the validation error ϵ_{val} with increasing HF training data. Comparison between our MFSM and a PCE model trained solely on HF data. The dashed line indicates the error of a PCE model trained only on LF data.

To showcase the CIs and PIs in this application, the wind speed U and the turbulence intensity σ_{U} are selected

as the most important input random variables. We select the HF training set size to be 70% of the available HF data, as the relative validation error for this training set size is already satisfactory, around 1%. The remaining HF data can be reserved for the appraisal of the constructed intervals, which is however not shown here. Furthermore, the LF experimental design consists of all the available $1,395$ LF samples. The number of bootstrap replications performed is $N_{\mathrm{B}} = 1,000$.

Figure 5 illustrates the 90% CIs (green area) and PIs (pink area) along slices in the two selected dimensions with the wind shear exponent set at its mean value. The green curve indicates the MFSM response. In Figures 5(a), 5(b), the two slices correspond to the 0.25- and 0.75-quantiles of σ_{U} and U respectively.

We observe that the epistemic uncertainty, as indicated by the confidence intervals, is generally low in both the U and σ_{U} dimensions, except occasionally near the bounds. In the U dimension, the total uncertainty for an unseen HF observation appears to be dominated by epistemic uncertainty, as the prediction intervals are only marginally wider than the corresponding CIs. Conversely, in the σ_{U} dimension, the epistemic uncertainty is very small, allowing the irreducible noise in the HF data to be more clearly observed.



(a) Selected dimension: wind speed U [m/s]

(b) Selected dimension: turbulence intensity σ_{U} [m/s]

Figure 5: Wind turbine application – 90% confidence and prediction intervals for the MFSM along slices in the two selected dimensions.

5 Conclusions and outlook

In this paper, we reviewed the multi-fidelity surrogate modeling framework introduced by [2] and examined its performance in two applications from different engineering contexts.

The first involved a transfer learning scenario, where information about the usage ratio of the energy absorber was transferred between simulations of an early-stage crash box prototype design and a later-stage crash box design. In this context, the early-stage crash box simulations were considered low-fidelity, while the later-stage crash box simulations were treated as high-fidelity. These simulations did not exhibit significant stochasticity.

The second application was a real-world scenario involving wind turbine simulations using two different simulators. Here, the stochasticity in the predictions of the fore-aft extreme bending moment at the tower bottom was clearly affecting the system responses.

In both cases, we observed that the multi-fidelity surrogate model provided significant benefits over single-fidelity surrogate models. The confidence intervals successfully conveyed information about the uncertainty regarding the true noise-free high-fidelity response. Furthermore, the prediction intervals successfully showed the uncertainty in new predictions, accounting for the irreducible noise, where it was present. Notably, the examined framework effectively handled the scenario with no noise.

In our future work, we plan to extend the methodology of [2] to accommodate inherently noisy models, a.k.a., stochastic simulators. This involves removing the current assumption of additive homoscedastic noise. Our goal is to combine information from multiple fidelity models into a multi-fidelity stochastic emulator,

capable of predicting the high-fidelity response distribution at each point. We also intend to apply this enhanced framework to appropriate stochastic simulator applications.

Acknowledgements

The application of multi-fidelity surrogate modeling in crashworthiness presented in this paper was performed during Katerina Giannoukou's stay at the group of Computational Solid Mechanics at the Technical University of Munich as a visiting student, under the supervision of Prof. Fabian Duddeck and in collaboration with Paolo Ascia, whose support is gratefully acknowledged.

The work in this paper is part of the GREYDIENT project, funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 955393.

References

[1] M. G. Fernández-Godino, "Review of multi-fidelity models," *Advances in Computational Science and Engineering*, vol. 1, no. 4, pp. 0–50, 2023.

[2] K. Giannoukou, S. Marelli, and B. Sudret, "A comprehensive framework for multi-fidelity surrogate modeling with noisy data: a gray-box perspective," *Submitted*, 2024.

[3] V. Moutoussamy, S. Nanty, and B. Pauwels, "Emulators for stochastic simulation codes," *ESAIM: Proceedings and Surveys*, vol. 48, pp. 116–155, 2015.

[4] X. Zhu and B. Sudret, "Stochastic polynomial chaos expansions to emulate stochastic simulators," *International Journal for Uncertainty Quantification*, vol. 13, no. 2, pp. 31–52, 2023.

[5] M. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied linear statistical models*, 5th ed. New York: McGraw-Hill, 2005.

[6] M. C. Kennedy and A. O'Hagan, "Predicting the output from a complex computer code when fast approximations are available," *Biometrika*, vol. 87, no. 1, pp. 1–13, 2000.

[7] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001.

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

[9] L. Le Gratiet and C. Cannamela, "Cokriging-based sequential design strategies using fast cross-validation techniques for multi-fidelity computer codes," *Technometrics*, vol. 57, no. 3, pp. 418–427, 2015.

[10] K. Cutajar, M. Pullin, A. Damianou, N. Lawrence, and J. González, "Deep Gaussian processes for multi-fidelity modeling," in *32nd Neural Information Processing Systems Conference*, Montreal, Canada, 2018.

[11] A. Hebbal, L. Brevault, M. Balesdent, E. Talbi, and N. Melab, "Multi-fidelity modeling with different input domain definitions using deep Gaussian processes," *Structural and Multidisciplinary Optimization*, vol. 63, no. 5, pp. 2267–2288, 2021.

[12] L. W.-T. Ng and M. S. Eldred, "Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation," in *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Honolulu, Hawaii*, 2012, p. 1852.

[13] P. S. Palar, T. Tsuchiya, and G. T. Parks, "Multi-fidelity non-intrusive polynomial chaos based on regression," *Computer Methods in Applied Mechanics and Engineering*, vol. 305, pp. 579–606, 2016.

[14] X. Meng and G. Karniadakis, "A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems," *Journal of Computational Physics*, vol. 401, p. 109020, 2020.

[15] X. Meng, H. Babaee, and G. E. Karniadakis, "Multi-fidelity Bayesian neural networks: Algorithms and applications," *Journal of Computational Physics*, vol. 438, p. 110361, 2021.

[16] C. Zhang, L. Liu, H. Wang, X. Song, and D. Tao, "SCGAN: stacking-based generative adversarial networks for multi-fidelity surrogate modeling," *Structural and Multidisciplinary Optimization*, vol. 65, no. 6, pp. 1–16, 2022.

[17] P. Conti, M. Guo, A. Manzoni, and J. S. Hesthaven, "Multi-fidelity surrogate modeling using long short-term memory networks," *Computer Methods in Applied Mechanics and Engineering*, vol. 404, p. 115811, 2023.

[18] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Inferring solutions of differential equations using noisy multi-fidelity data," *Journal of Computational Physics*, vol. 335, pp. 736–746, 2017.

[19] S. Ficini, U. Iemma, R. Pellegrini, A. Serani, and M. Diez, "Assessing the performance of an adaptive multi-fidelity Gaussian process with noisy training data: A statistical analysis," in *AIAA AVIATION 2021 FORUM*, 2021.

[20] Y. Zhang, N. H. Kim, C. Park, and R. T. Haftka, "Multifidelity surrogate based on single linear regression," *AIAA Journal*, vol. 56, no. 12, pp. 4944–4952, 2018.

[21] D. Xiu and G. E. Karniadakis, "The Wiener-Askey polynomial chaos for stochastic differential equations," *SIAM Journal on Scientific Computing*, vol. 24, no. 2, pp. 619–644, 2002.

[22] N. Lüthen, S. Marelli, and B. Sudret, "Automatic selection of basis-adaptive sparse polynomial chaos expansions for engineering applications," *International Journal for Uncertainty Quantification*, vol. 12, no. 3, pp. 49–74, 2022.

[23] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. CRC press, 1994.

[24] G. Schwartz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[25] S. Marelli and B. Sudret, "UQLab: A framework for uncertainty quantification in Matlab," in *Vulnerability, Uncertainty, and Risk*. American Society of Civil Engineers, 2014.

[26] J. Fang, G. Sun, N. Qiu, N. H. Kim, and Q. Li, "On design optimization for structural crashworthiness and its state of the art," *Structural and Multidisciplinary Optimization*, vol. 55, no. 3, pp. 1091–1119, 2016.

[27] B. Sudret, "Global sensitivity analysis using polynomial chaos expansions," *Reliability Engineering & System Safety*, vol. 93, no. 7, pp. 964–979, 2008.

[28] I. Abdallah, C. Lataniotis, and B. Sudret, "Parametric hierarchical Kriging for multi-fidelity aero-servo-elastic simulators – Application to extreme loads on wind turbines," *Probabilistic Engineering Mechanics*, vol. 55, pp. 67–77, 2019.

[29] E. A. Bossanyi, "GH Bladed theory manual," Garrad Hassan & Partners Ltd, Tech. Rep. 282/BR/009, 2003.

[30] J. Jonkman and M. Buhl, "FAST user's guide," National Renewable Energy Laboratory Golden, CO, USA, Tech. Rep. NREL/EL-500-38230, 2005.