**MASCOT-NUM 2022**
**Clermont Ferrand, France**
**7-9 June 2022**

# Multifidelity surrogate modelling with noisy grey-box models

A. Giannoukou
*ETH Zürich*

**Supervisor(s):** B. Sudret (ETH Zürich) and S. Marelli (ETH Zürich)

**PhD expected duration:** Sep. 2021 - Aug. 2025

**Address:** Institute of Structural Engineering
Chair of Risk, Safety and Uncertainty Quantification
ETH Zürich - HIL E 21.3
Stefano-Franscini-Platz 5
CH-8093 Zürich, Switzerland

**E-mail:** katerina.giannoukou@ibk.baug.ethz.ch

**Abstract:**

Modelling the response of complex systems has always been a challenge in engineering. Predictions for an engineering system can be obtained from different sources, such as mathematical models, computer simulations, physical experiments or expert opinions. Generally, it is possible to identify two classes of predictive models: high-fidelity (HF) models, that produce accurate estimations but are associated with high computational or financial costs, and low-fidelity models (LF) that are less accurate, but also less expensive to run.

Within the context of the Horizon 2020 GREYDIENT project [1], we consider the special case of grey-box systems for which very few and expensive-to-acquire experimental data are available, in addition to expensive computational models that describe the underlying physics. The two models are expensive to evaluate, and as a result, the data that can be obtained from each of them are limited. In this setting, the question that arises is how to perform accurate uncertainty quantification, statistical inference or optimisation, which typically involve a large amount of system analyses [3].

Multifidelity surrogate modelling (MFSM) approaches can provide the solution to this challenge. Surrogate models (SM) or metamodels, are inexpensive models that provide an approximation of expensive computational models based on a limited number of observations, and allow an accurate estimation of the output statistics of the approximated expensive models. Moreover, multifidelity (MF) models combine multiple models of varying fidelity, typically, supplementing few HF data with a larger number of lower-fidelity data. In MFSM, the different fidelities are combined inside a single surrogate model [1].

In our MFSM setting, the experimental data are considered to be single realisations of a black-box high-fidelity model. On the other hand, the available white-box computational models are considered as their low-fidelity counterparts. The MF model obtained by combining the two is a so-called grey-box model (Figure 1), aimed at improving the modelling accuracy and the predictive power of white-box models by integrating the data-driven black-box model, also including uncertainty measures in its predictions.

Moreover, real-world experiments produce data contaminated by measurement noise, and the low-fidelity physics-based model can in general be very expensive, and is itself often represented by a surrogate model. If we denote the experimental input samples as $\mathcal{X}_H = (\mathbf{x}_H^{(1)}, \mathbf{x}_H^{(2)}, ..., \mathbf{x}_H^{(N_H)})$, and the corresponding response vector as $\mathcal{Y}_H = (\mathbf{y}_H^{(1)}, \mathbf{y}_H^{(2)}, ..., \mathbf{y}_H^{(N_H)})$, then the experimental data $(\mathcal{X}_H, \mathcal{Y}_H)$ are assumed to be expensive and noisy evaluations of a black-box model. In other words, the high-fidelity model is seen as a mapping $M_H$: $X \rightarrow Y$, where $y_H^{(i)} = M_H(x_H^{(i)}) + \varepsilon_H$, and $\varepsilon_H$ is an additive noise term. Similarly, the white-box computer simulations, denoted as $M_L$, produce data contaminated by a random noise $\varepsilon_L$, i.e. $y_L^{(i)} = M_L(x_L^{(i)}) + \varepsilon_L$.

[1] `https://www.greydient.eu/`.

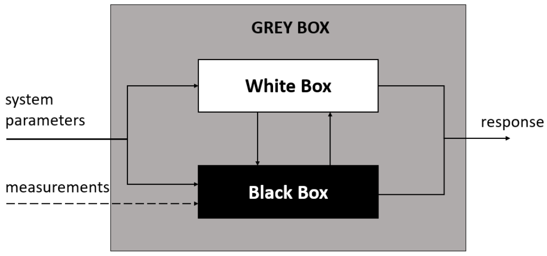


Figure 1: A grey-box model is a combination of a physics-driven white box and a data-driven black-box model.

In a first phase, we aim at developing and benchmarking different MF methodologies, initially under the assumption of noise-free models. Figure 2 showcases a preliminary investigation on the performance of MF modelling on a well-known benchmark, the borehole function [2]. Subsequently, we will investigate different techniques to fuse the noise-contaminated high- and low-fidelity information in MF models that are accurate and robust to noise.

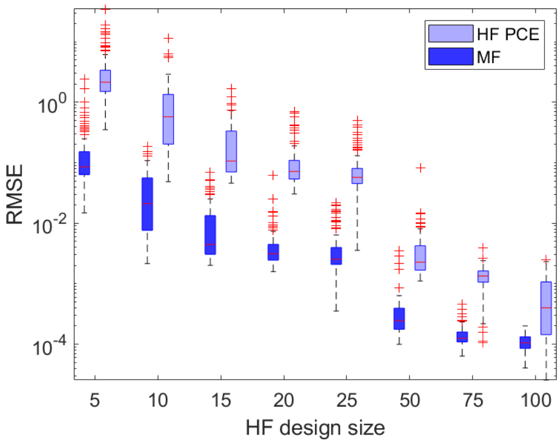


Figure 2: Comparison of a multifidelity (MF) model and a high-fidelity polynomial chaos expansion model (HF PCE) on the borehole function case study. The root mean square error is calculated for increasing high-fidelity experimental design size.

## References

[1] M. G. Fernández-Godino, C. Park, N. Kim, and R. T. Haftka. Review of multi-fidelity models. *arXiv: Applications*, 2016.

[2] 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*, 305:579–606, jun 2016.

[3] B. Peherstorfer, K. Willcox, and M. Gunzburger. Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization. *SIAM Review*, 60(3):550–591, jan 2018.

**Short biography** – Aikaterini (Katerina) Giannoukou has studied Electrical and Computer Engineering at the National Technical University of Athens. In September 2021, she started her PhD at the Chair of Risk, Safety and Uncertainty Quantification, at ETH Zurich. Her project is part of the Marie Skłodowska-Curie Innovative Training Network, GREYDIENT, and her work focuses on multifidelity modelling.