Multifidelity surrogate modelling with noisy grey-box models

Global discrepancy multifidelity modelling

A. Giannoukou, S. Marelli, B. Sudret

ETH Zürich, Chair of Risk, Safety and Uncertainty Quantification, agiannoukou@ibk.baug.ethz.ch

Multifidelity grey-box modelling

The project

GREYDIENT

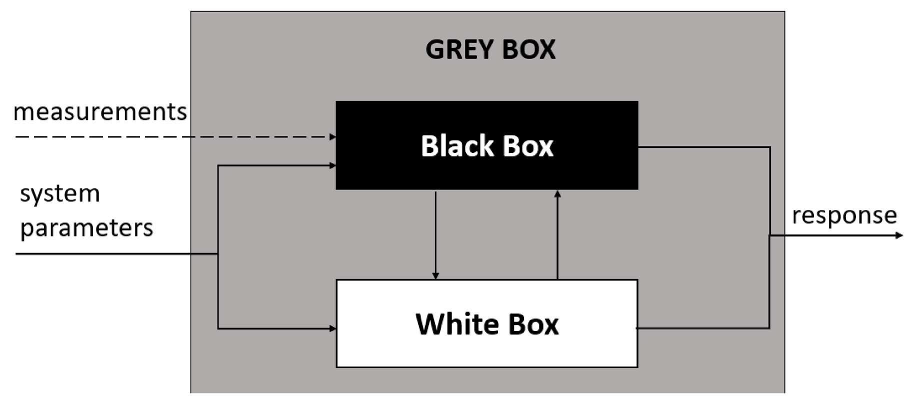
- **Goal**: Sustain the transition of European personal mobility towards safe and reliable systems
- **Idea**: Data from vehicle sensors + computational models of components and systems → grey-box model

Grey-box modelling

The output of engineering systems can often be obtained from multiple sources:

- Experimental data: **black-box models**
- Physics-based computational models: **white-box models**

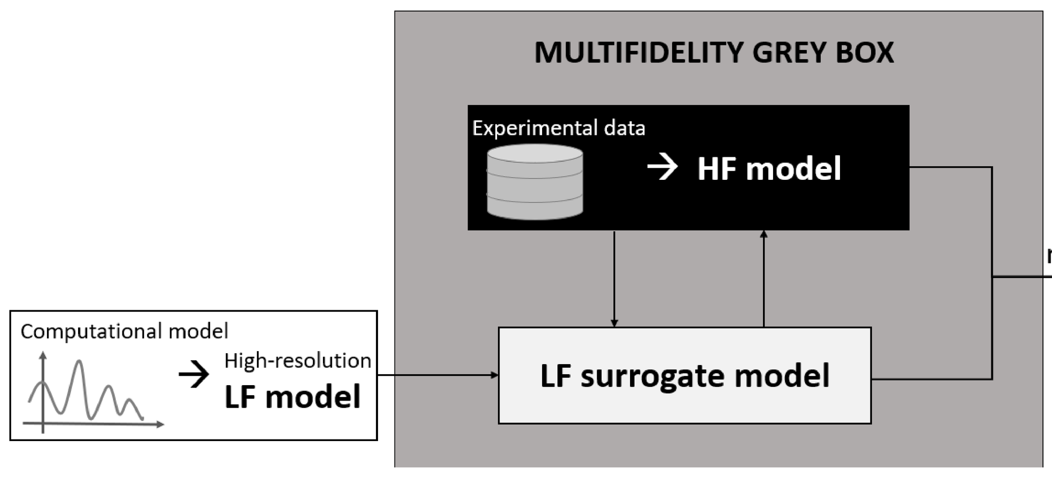
Grey-box models fuse information from white and black boxes



Multifidelity grey-box setting

- Few and expensive experimental data: **high fidelity** (HF)
- Expensive computational models: **low fidelity** (LF)

Goal: Combine information of multiple fidelities from black- and white-box models into a multifidelity grey-box model



Deterministic global discrepancy multifidelity modelling

Problem setup:

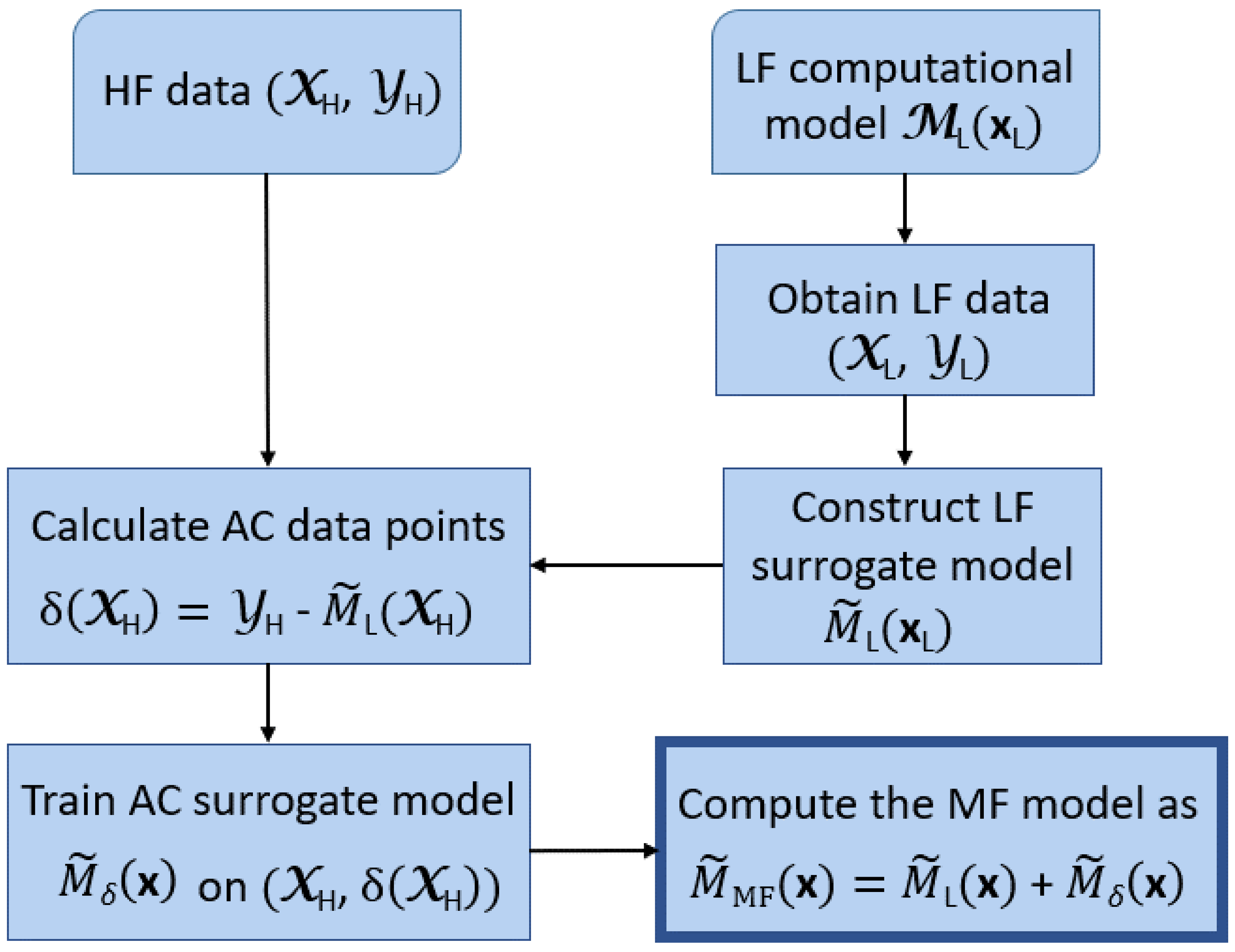
- HF experimental data: N_H input data $\mathcal{X}_H = (\boldsymbol{x}_H^{(1)}, \dots, \boldsymbol{x}_H^{(N_H)})$ with responses $\mathcal{Y}_H = (y_H^{(1)}, \dots, y_H^{(N_H)})$
- A LF computational model $\mathcal{M}_L$, and N_L LF model input data $\mathcal{X}_L = (\boldsymbol{x}_L^{(1)}, \dots, \boldsymbol{x}_L^{(N_L)})$ with model evaluations $\mathcal{Y}_L = (y_L^{(1)}, \dots, y_L^{(N_L)})$

Assumption: The HF response can in general be expressed as follows:

$$y_H = \rho(\boldsymbol{x}) \cdot y_L + \delta(\boldsymbol{x}) \quad (1)$$

An additive correction (AC) approach

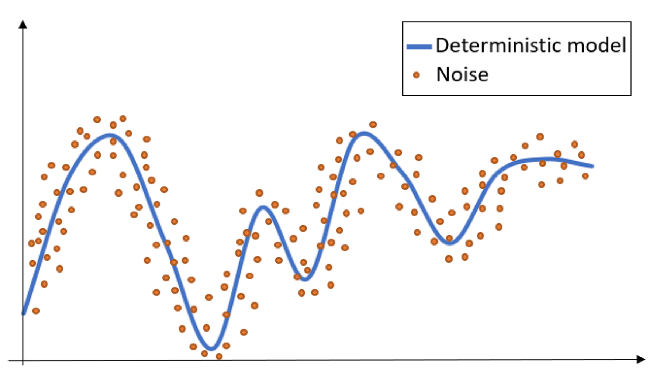
The method shown below is based on Berchier 2016; Ng, Eldred 2012.
It assumes that $\rho(\boldsymbol{x}) = 1$ in (1), so $y_H = y_L + \delta(\boldsymbol{x})$



Multifidelity modelling with noisy data

- Real-world data are contaminated by **measurement noise** ⇝ the HF unknown model can be written as: $\mathcal{M}_H \colon X \to Y$, where $y_H^{(i)} = \mathcal{M}_H(x_H^{(i)}) + \varepsilon_H$, and ε_H is random noise with $\mathbb{E}[\varepsilon_H] = 0$
- The LF physics-based model is represented by a surrogate model trained on a **finite experimental design** ⇝ the LF response can be written as: $y_L^{(i)} = \widetilde{\mathcal{M}}_L(x_L^{(i)}) + \varepsilon_L$, where ε_L is an approximation residual

The multifidelity model needs to denoise the noisy data.



Toy example: borehole function

Description of the example

- High-fidelity model:

$$f_H(x) = \frac{2\pi T_u (H_u - H_l)}{\ln(r/r_w)\left(1 + \frac{2LT_u}{\ln(r/r_w) r_w^2 K_w} + \frac{T_u}{T_l}\right)}$$

- Low-fidelity version:

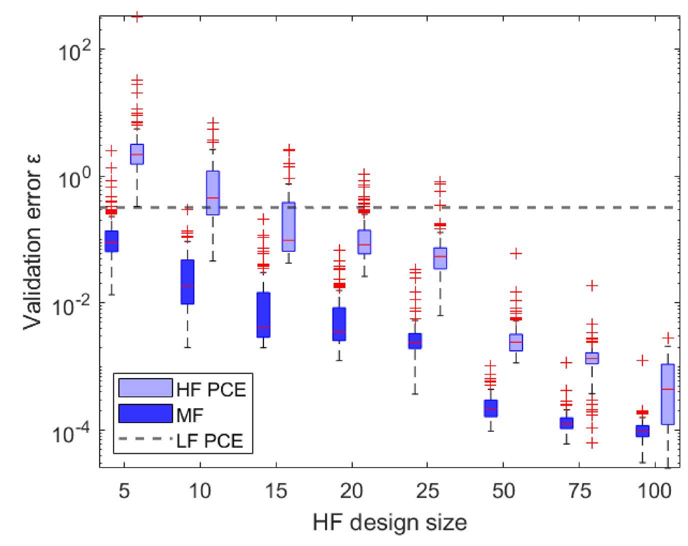
$$f_L(x) = \frac{5 T_u (H_u - H_l)}{\ln(r/r_w)\left(1.5 + \frac{2LT_u}{\ln(r/r_w) r_w^2 K_w} + \frac{T_u}{T_l}\right)}$$

Variable	Distribution
r_w	$\mathcal{N}(\mu = 0.10, \sigma = 0.0161812)$
r	$\mathcal{LN}(\mu = 7.71, \sigma = 1.0056)$
T_u	$\mathcal{U}(63070, 115600)$
H_u	$\mathcal{U}(990, 1110)$
T_l	$\mathcal{U}(63.1, 116)$
H_l	$\mathcal{U}(700, 820)$
L	$\mathcal{U}(1120, 1680)$
K_w	$\mathcal{U}(9855, 12045)$

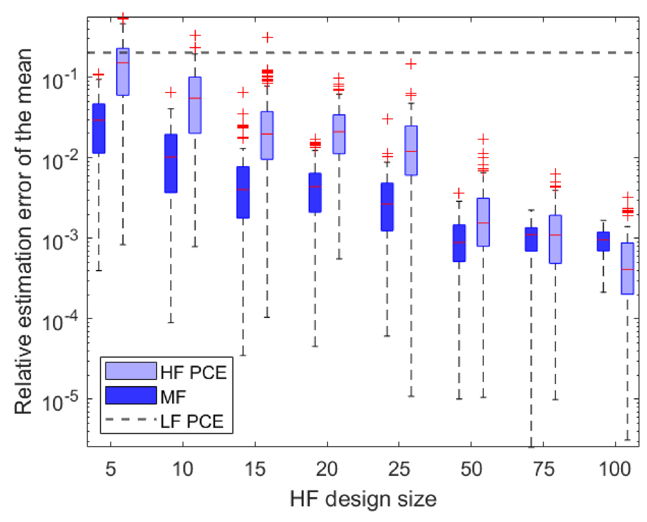
Data generation

- Surrogate for LF and MF: polynomial chaos expansions (PCE)
- HF experimental design (ED) sizes: 5, 10, 15, 20, 25, 50, 75, 100
- LF ED size (constant): 200
- HF ED not a subset of LF ED
- MF uses the HF and the LF samples
- Number of replications per HF ED size: 100
- Validation set size: $N_v = 10^5$ unseen data points

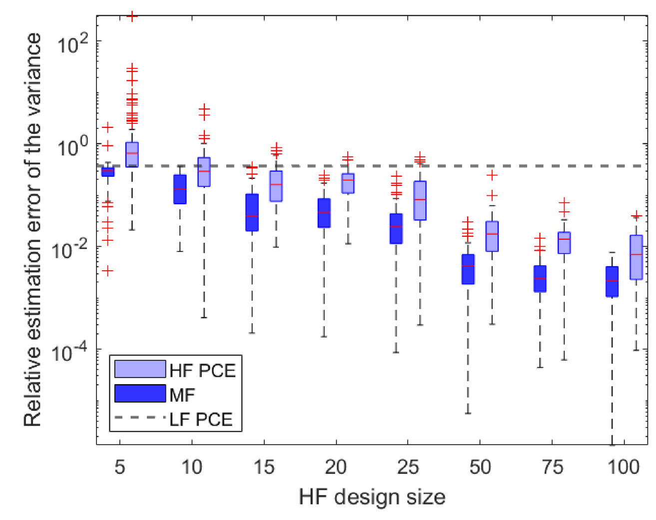
Results



Validation error of the MF model compared to the HF surrogate: $\varepsilon = \frac{\sum_{i=1}^{N_v}(y^{(i)} - \tilde{y}^{(i)})^2}{\sum_{i=1}^{N_v}(y^{(i)} - \mu)^2}$, where μ is the true model mean



Relative estimation error on the mean, given by $\epsilon_\mu = \left|\frac{\tilde{\mu} - \mu}{\mu}\right|$ where $\tilde{\mu}$ is the mean value of the surrogate model and μ is the true model mean



Relative estimation error on the variance, given by $\epsilon_{\sigma^2} = \left|\frac{\tilde{\sigma}^2 - \sigma^2}{\sigma^2}\right|$ where $\tilde{\sigma}^2$ is the variance of the surrogate model and σ^2 is the true model variance

Discussion and further work

- The presented MF method works well for analytical toy examples, but real data have not been used yet
- Global regression-based approaches will be investigated to deal with noisy data
- Multifidelity stochastic simulators will be investigated to deal with models with an aleatoric stochastic behaviour
- We used the UQLab software for the construction of the surrogate models in the example shown

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

Risk, Safety & Uncertainty Quantification

The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 955393