Universität Stuttgart
Institute for Modelling Hydraulic and Environmental Systems
Stochastic Simulation and Safety Research for Hydrosystems (LS3)[1]

Training surrogate models using input dimension reduction for inverse modelling problems

Application to heterogeneous media problems

Maria Fernanda Morales Oreamuno[1], Sergey Oladyshkin[1] & Wolfgang Nowak[1]

Motivation

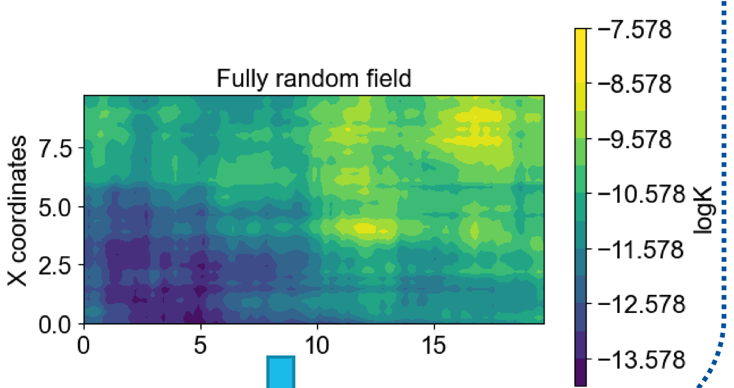
Surrogate model training can become computationally prohibitive for **high-dimensional, heterogeneous problems** given that **a)** more training points from computationally expensive models are needed and **b)** increased training time. Additionally, (Bayesian) inverse problems in high dimensions also suffer from the curse of dimensionality and convergence problems.

We implement surrogate training approaches with input dimension reduction (IDR) and apply them for forward and (Bayesian) inverse applications

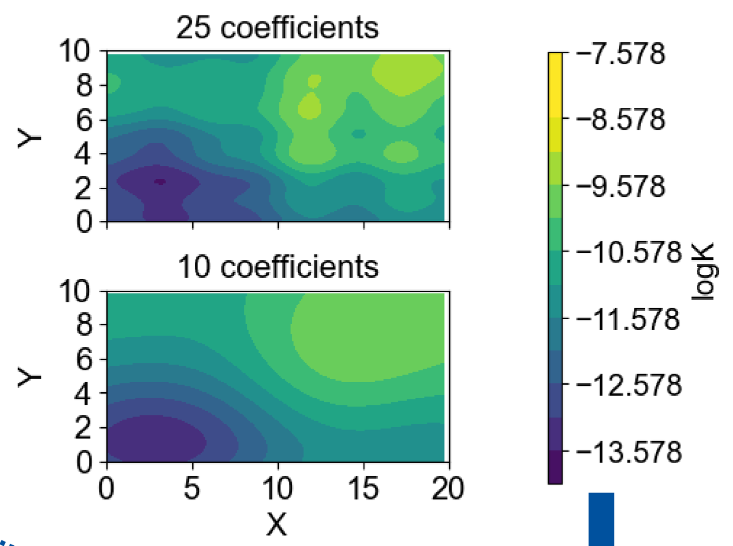
Our goal is to train surrogates for high dimensional problems, while optimizing the number of expensive model runs and accounting for any uncertainty induced by the IDR.

Karhunen-Loève decomposition (KLD) for IDR

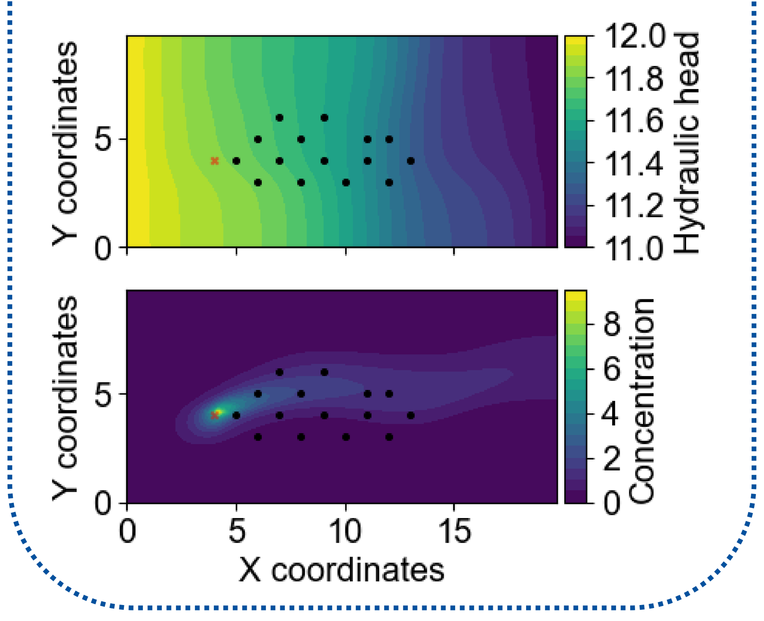
Full random field (ω)



Dimension reduction of parameters using KLD ($\hat{\omega}$)



Forward model outputs ($y(\omega)$)



Train surrogate, considering **IDR error**
$y(\omega) \sim \tilde{y} = \mathcal{S}(\hat{\omega}, \theta; y(\omega)) + \varepsilon(\hat{\omega})$

KLD truncation provides a smaller input training space, but smoothens the random field, and so provides less information to the surrogate when used as training inputs.

+ Fewer training points needed
- Higher uncertainty when using the surrogate for predictions

IDR-error „aware“ surrogate

- We focus on noisy-surrogates(GP), which consider the “noise” due to IDR.
- We focus on surrogates that provide **Gaussian distribution over the surrogate predictions.**
  - Allows to include a surrogate-induced error for a more representative UQ

Surrogate error „aware“ Bayesian inference

- We use available (borehole) observations, at time step $\mathbf{t_i}$, to update prior knowledge on uncertain parameters.
- Through Bayesian inference we can reduce the uncertainty associated to future time step ($\mathbf{t_{i+n}}$) simulations, for forward and optimal experimental design **(OED)** simulations

We consider the surrogate error in the likelihood estimation

$$p(\omega|y_o) = \frac{\overbrace{p(y_o|\omega)}^{\text{Likelihood}}p(\omega)}{p(y_o)} \qquad \overbrace{p(y_o|\omega)}^{\text{Likelihood}} = \left(\frac{1}{2\pi}\right)^{d/2} \left|\hat{R}\right|^{1/2} exp\left[-\frac{1}{2}(y-\tilde{y})^T \hat{R}^{-1}(y-\tilde{y})\right]$$

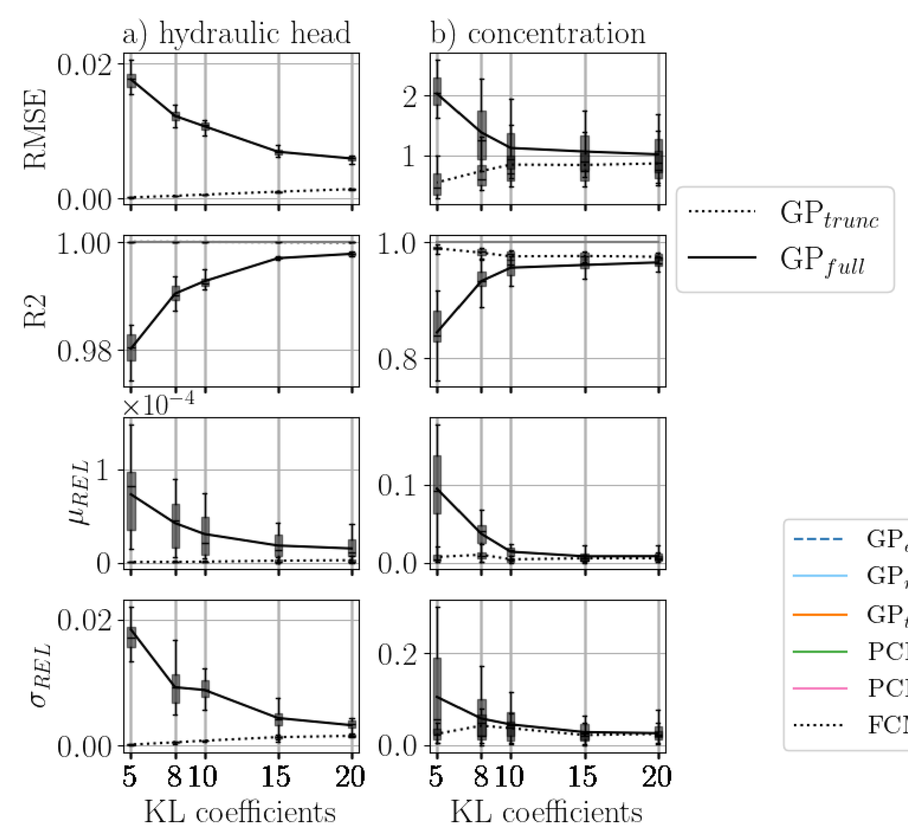
Augmented covariance $\hat{R} = (\varepsilon + \boldsymbol{K}(\cdot,\cdot))I$

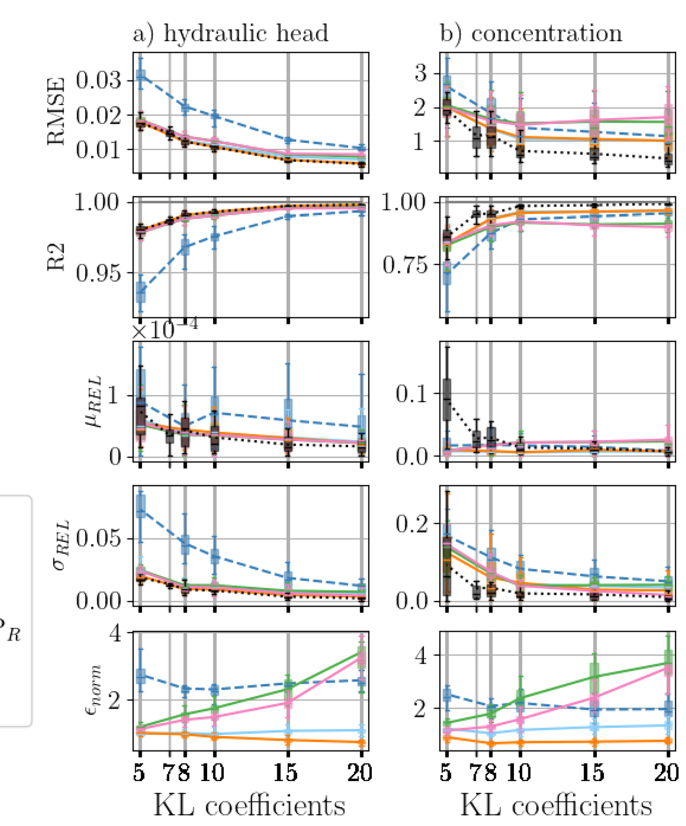
maria.morales@iws.uni-stuttgart.de

Application to GW transport model

Forward UQ: IDR error

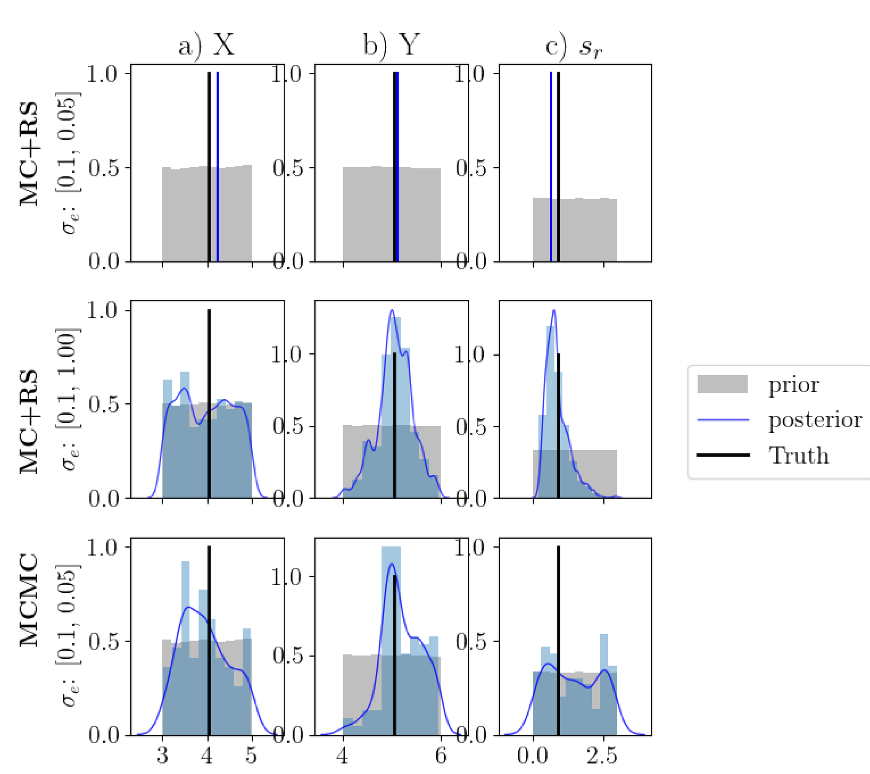
- We quantify and visualize the surrogate error compared to the surrogate approximation error.
- We train noisy/error-aware surrogate surrogates with different configurations, with different IDR levels, and compare them in validation



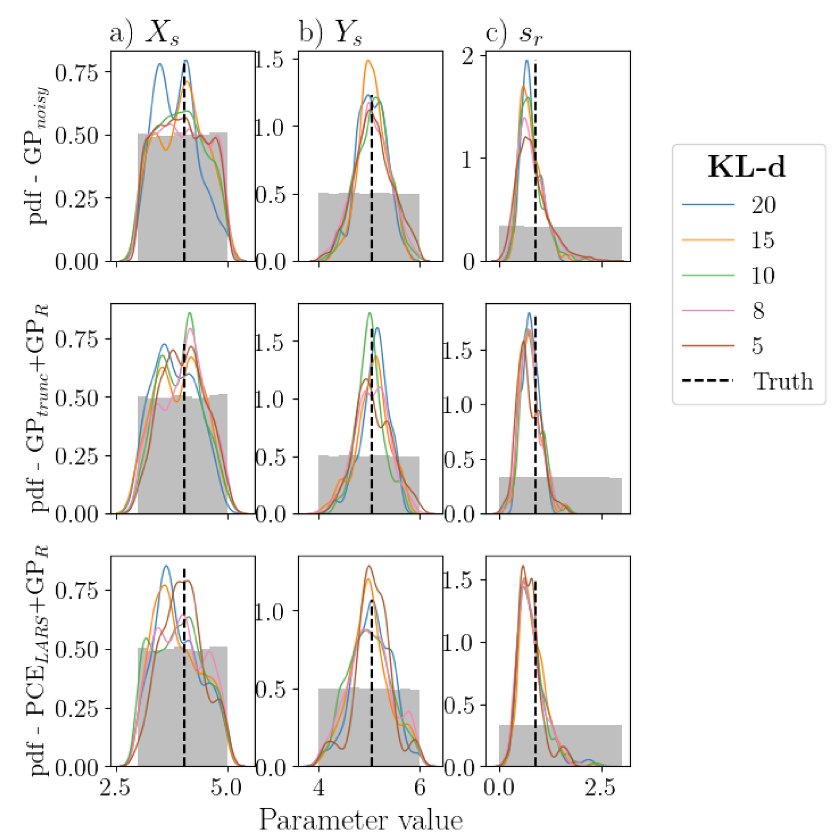


Bayesian inference

- IDR, up to a certain number of coefficients, did not affect the surrogate’s ability to infer the true posterior



Posterior distributions obtained with **full complexity model**, using 3 posterior sampling approaches



Posteriors using **surrogate models** with MC + rejection sampling, considering surrogate error

Sources

[1]Rasmussen, C. E. (2006). *Gaussian processes for machine learning* (Vol. 2, No. 3, p. 4). Cambridge, MA: MIT Press: reliability of the surrogate predictions
[2] Zhang, J., Zheng, Q., Chen, D., Wu, L., & Zeng, L. (2020). Surrogate-based Bayesian inverse modeling of the hydrological system: An adaptive approach considering surrogate approximation error. *Water Resources Research*, *56*(1), e2019WR025721.
[3] Schobi, R., Sudret, B., & Wiart, J. (2015). Polynomial-chaos-based Kriging. *International Journal for Uncertainty Quantification*, *5*(2).
[4] Kohlhaas, R., Kröker, I., Oladyshkin, S., & Nowak, W. (2023). Gaussian active learning on multi-resolution arbitrary polynomial chaos emulator: concept for bias correction, assessment of surrogate reliability and its application to the carbon dioxide benchmark. *Computational Geosciences*, *27*(3), 369-389.
[5] Oladyshkin, S., Mohammadi, F., Kroeker, I., & Nowak, W. (2020). Bayesian3 active learning for the Gaussian process emulator using information theory. *Entropy*, *22*(8), 890.

BVRox
BUNDESGESELLSCHAFT FÜR ENDLAGERUNG