Uncertainty-Informed, Surrogate-Assisted Bayesian Optimized Experimental Design for Fluid Transport Process Monitoring

Nino Menzel[1], **Maria Fernanda Morales**[2], Wolfgang Nowak[2], Sergey Oladyshkin[2], Florian M. Wagner [1]

1) Geophysical Imaging and Monitoring, RWTH Aachen University, Wüllnerstraße 2, 52062 Aachen; **nino.menzel@gim.rwth-aachen.de**

2) Institute for Modelling Hydraulic and Environmental Systems - Stochastic Simulation and Safety Research for Hydrosystems (LS3)1, University of Stuttgart; **maria.morales@iws.uni-stuttgart.de**

1) Motivation

Enhancing the efficiency of geophysical surveys by performing Optimized Experimental Design (OED) is of **high importance** for field surveys, especially in highly **sensitive environments**.

However, one particular challenge in OED for fluid transport process monitoring is the **limitation of computational expenses** in both the underlying transport models as well as the OED step.

We present a **robust, surrogate-assisted Bayesian Optimized Experimental Design strategy** for geophysical field measurements, that…

- provides an efficient **stochastic alternative to deterministic OED approaches** for dynamic targets, such as the Compare-R approach (e.g., Wilkinson et al., 2015)
- aims at **considering parameter uncertainties** by introducing **uncertainty-informed** transport models for the optimization process
- uses surrogate models (emulators) to reduce the computational burden associated to the forward model

2) Bayesian optimization algorithm

The **Bayesian experimental design method** (Qiang et al., 2022) is based on the **Bayes formulation** of the geoelectrical inverse problem, written as:

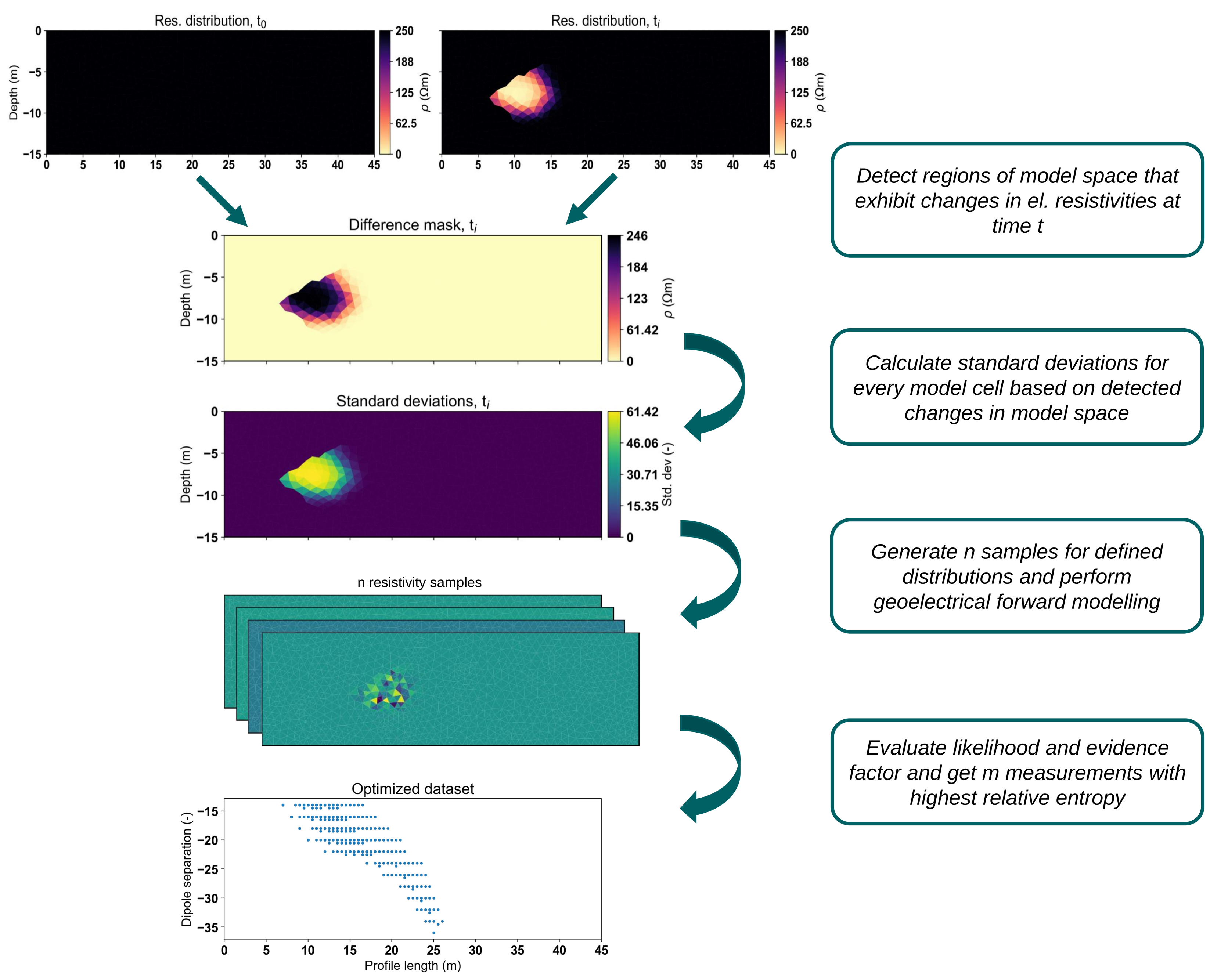
$$p(\boldsymbol{m}|\boldsymbol{d},\boldsymbol{s}) = \frac{p(\boldsymbol{m})\,p(\boldsymbol{d}|\boldsymbol{m},\boldsymbol{s})}{p(\boldsymbol{d}|\boldsymbol{s})}$$

where $p(\boldsymbol{m})$ is the **prior density** of model parameters ***m*** and $p(\boldsymbol{m}|\boldsymbol{d},\boldsymbol{s})$ is the **posterior density** of ***m***. $p(\boldsymbol{d}|\boldsymbol{m},\boldsymbol{s})$ is the **likelihood function**, which is a **joint probability distribution** of measurements ***d*** and model parameters ***m***. $p(\boldsymbol{d}|\boldsymbol{s})$ is the **Bayesian evidence factor**.

The **difference between prior and posterior is caused by the measurements *d*** and quantified by the Kullback-Leibler divergence *U(s)****:***

$$U(s) = \frac{1}{n}\sum_{i=1}^{n}\{\ln[p(\boldsymbol{d}^i|\boldsymbol{m}^i,\boldsymbol{s})] - \ln[p(\boldsymbol{d}^i|\boldsymbol{s})]\}$$

where m^i is the randomly generated MC samples with prior information of the model parameters m.

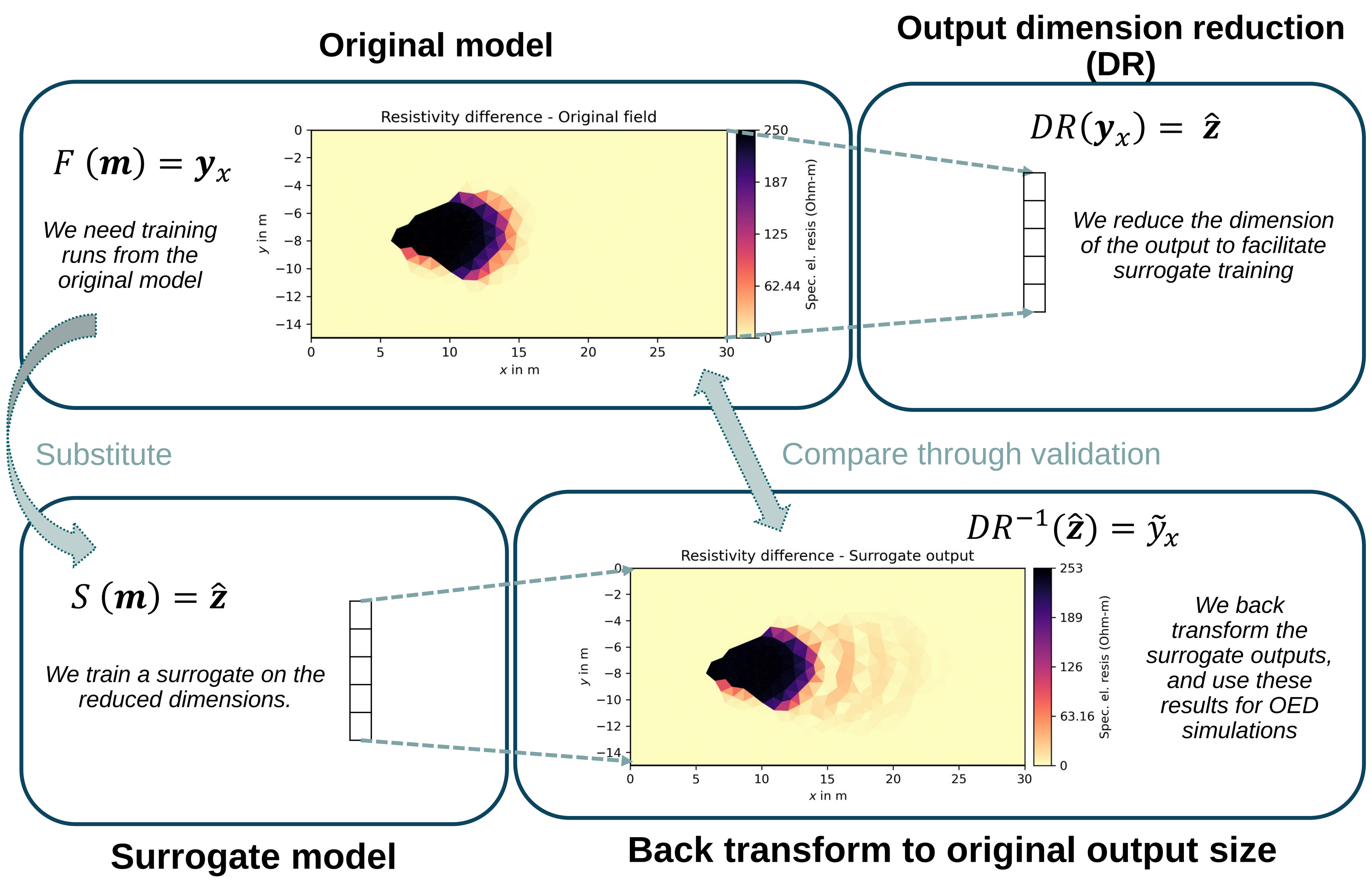


3) Surrogate-assisted approach

For a Bayesian OED approach, we need a **large number of model runs** to obtain representative masks and uncertainty estimators ⟶ **computationally expensive**

We aim to substitute the computationally expensive forward model with a surrogate model which

- is trained on **input-output data pairs** from the full model
- runs in a **fraction of the time**
- **approximates** the full model ⟶ has an **approximation error** associated to it, which must be considered



Research question: Which **surrogate + dimension reduction** approach reduces the associated error and fits the problem best? How to include the surrogate + DR error in OED algorithm?

Goal: use the surrogate to generate N_{MC} runs for Bayesian OED methodology, taking into consideration the error associated to surrogate output ($\tilde{\boldsymbol{y}}_x$)

4) Results of the synthetic study

Model Setup

- We consider **2 uncertain parameters**:
  - hydraulic conductivity (K): $U[1x10^{-4} - 3x10^{-4}]$
  - porosity (Φ): $U[0.2, 0.5]$
- Simulation of a **contaminant transport** in sedimentary rock (vague information on lithology / compaction)
- Contaminant is **released during a short timeframe** (quasi-instantaneous contaminant release)

OED considerations

- We use **N_{MC} model runs** to generate a mask of the area of interest based on the probability of each cell being affected: **stochastic approach**

Surrogate model considerations

- We use **Principal Component Analysis** (PCA) to reduce the output field to a smaller latent space ($\hat{\boldsymbol{z}}$)
- We use a **Gaussian Process Regression** (GPR) as a surrogate, trained with **50 model runs.**

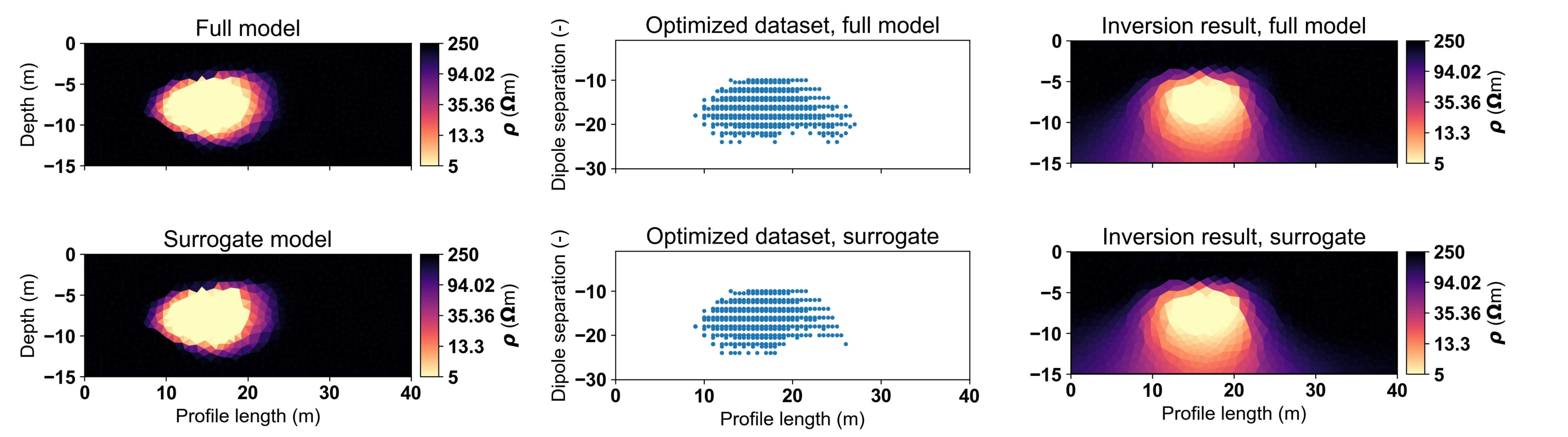


Fig. 1: *Comparison of the full model (upper row, left) and the surrogate (lower row, left) as well as the optimized electrical resistivity tomography (ERT) dataset using 1000 four-point configurations (middle plots in both rows). The left images show the inversion results of both the full model-based as well as the surrogate-based optimized datasets. The plot nicely visualizes the excellent performance of the surrogate in comparison to the full model, since both allow for precise focusing of the survey to the relevant model area and thus for detailed monitoring of the simulated transport process.*

5) Outlook

- Test and compare **different Bayesian OED criteria**
- Apply **cell-wise probabilities** instead of mask-averaged distributions to sample from **m**
- Implement **surrogate-assisted Bayesian inference** on available ERT measurements to reduce uncertainty associated with future time step simulations for OED purposes

References:

Rasmussen, C. E. (2006). *Gaussian processes for machine learning (Vol. 2, No. 3, p. 4).* Cambridge, MA: MIT Press: reliability of the surrogate predictions

Qiang, S., Shi, X., Kang, X., & Revil, A. (2022). Optimized arrays for electrical resistivity tomography survey using Bayesian experimental design. *Geophysics,* 87(4), E189–E203.

Wilkinson, P. B., Uhlemann, S., Meldrum, P. I., Chambers, J. E., Carrière, S., Oxby, L. S., & Loke, M. H. (2015). Adaptive time-lapse optimized survey design for electrical resistivity tomography monitoring. Geophysical Journal International, 203(1), 755–766.

Universität Stuttgart

Geophysical Imaging and Monitoring