Verbesserung der prädiktiven Güte endlagerrelevanter Simulationen durch optimale Datenakquise und Smart-Monitoring

Prof. Julia Kowalski

Prof. Florian Wagner

apl. Prof. Sergey Oladyshkin

Dr. Marc S. Boxberg

Prof. Wolfgang Nowak

Qian Chen, M.Sc

Nino Menzel, M.Sc

Maria Fernanda Morales, M.Sc

MBD | RWTH AACHEN UNIVERSITY

Chair of Methods for Model-based Development in Computational Engineering
Faculty of Mechanical Engineering

GIM | RWTH AACHEN UNIVERSITY

Geophysical Imaging and Monitoring Teaching and Research Unit, Faculty of Georesources and Materials Engineering

University of Stuttgart | LS³

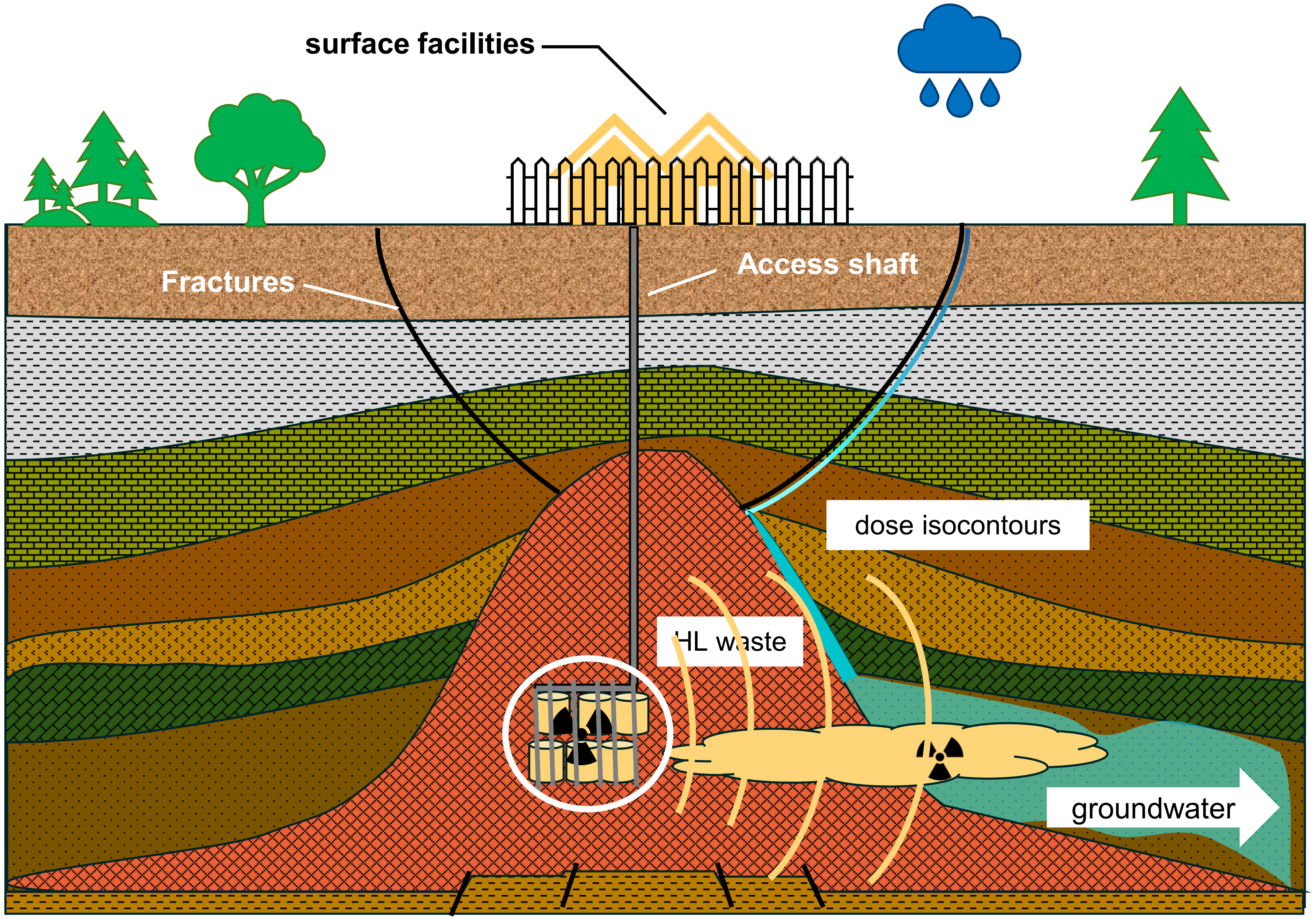
Department of Stochastic Simulation and Safety Research for Hydrosystems
Institute for Modelling Hydraulic and Environmental Systems
Stuttgart Center for Simulation Science

Table of Contents

University of Stuttgart

RWTH AACHEN UNIVERSITY

Research objectives



Our knowledge of the subsurface is governed by uncertainties!

Question: How can we disentangle uncertainties, hence manage reliability?

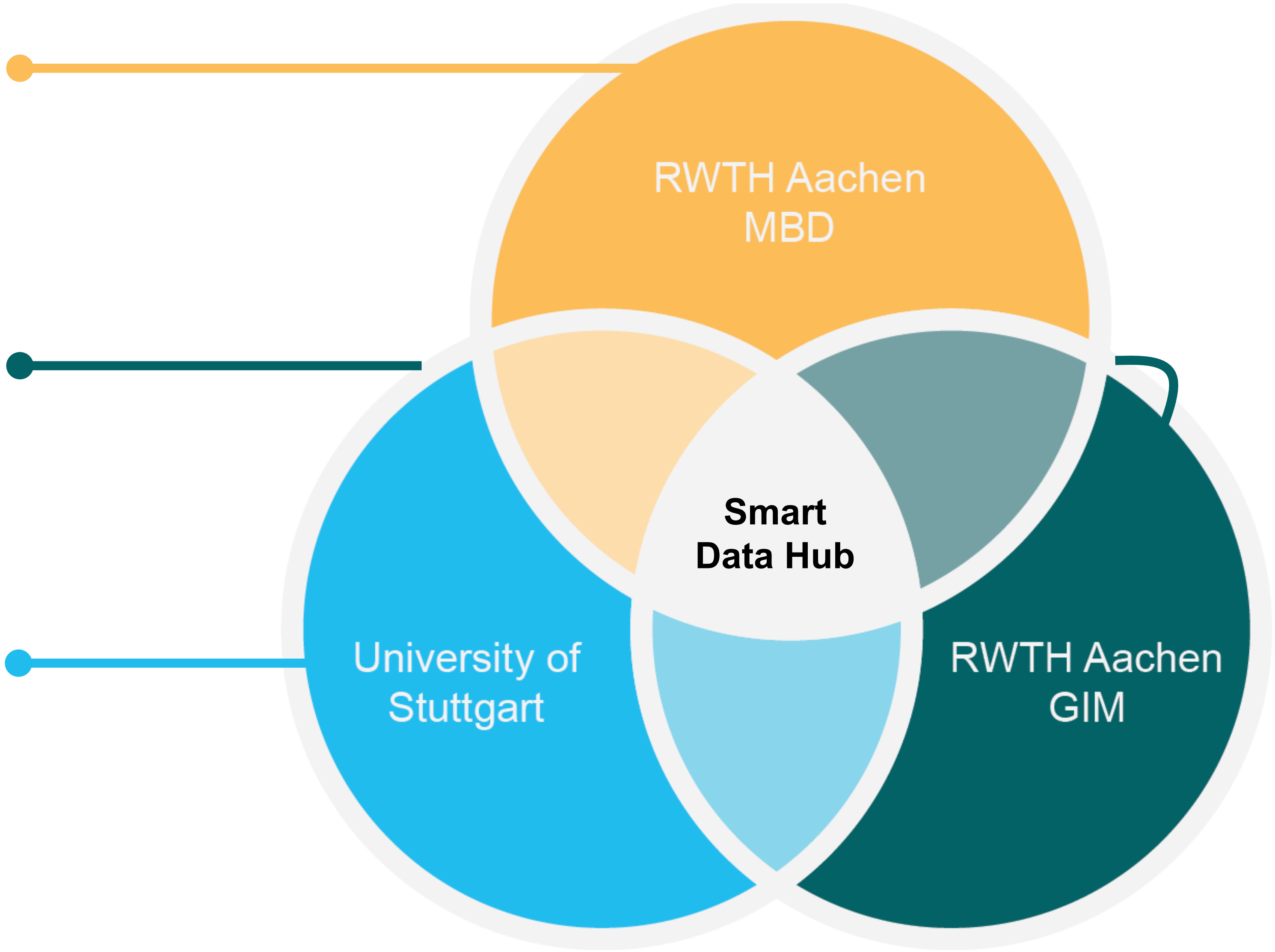
- **Facilitate transparency and reproducibility**
  Collect and integrate system uncertainty; analyze uncertainty impact
- **Decision support for data acquisition (overarching goal)**
  Assess which measurement would be most beneficial to reduce uncertainty
- **Decision support for monitoring radioactive waste repository**
  Assess which measurements would be most beneficial for reliable monitoring

Key methodological building blocks

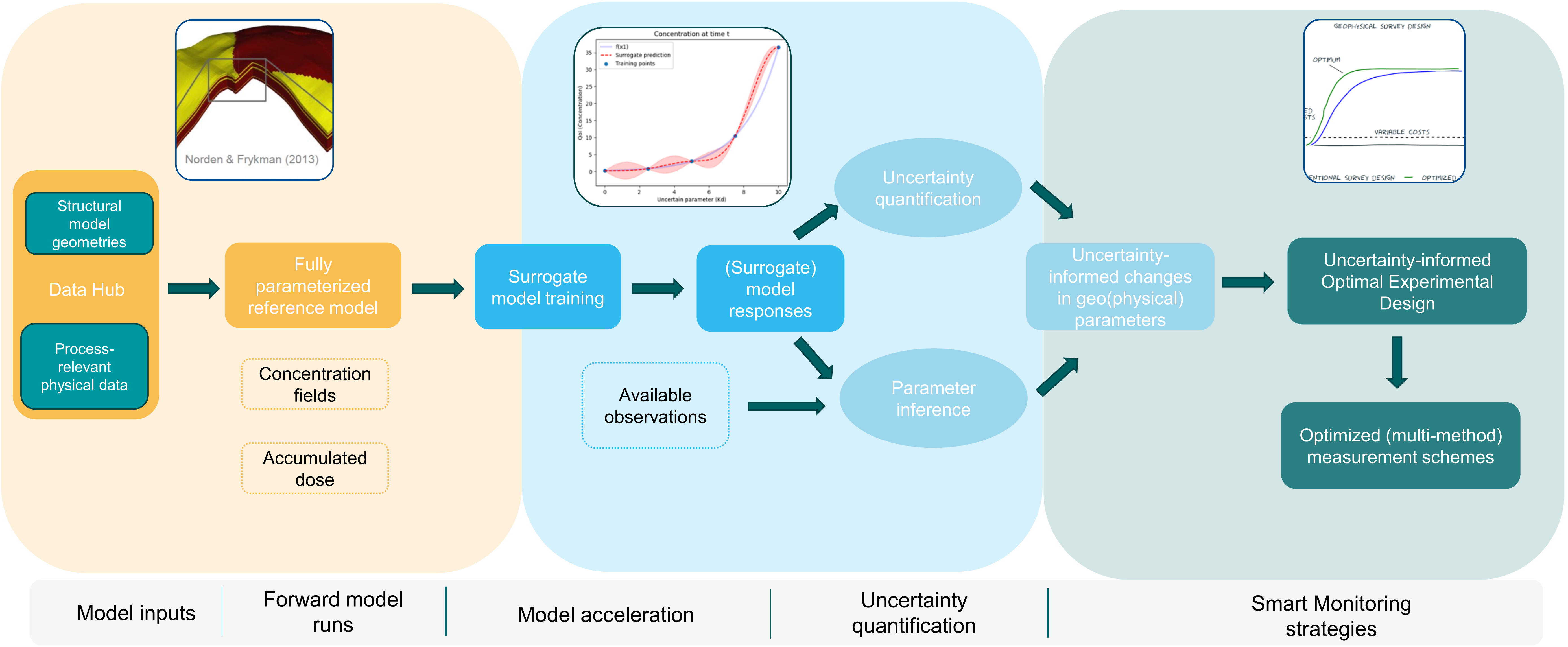
- **Impact modeling**
  orchestrates a workflow based on uncertainty informed geology, hydrothermal setting and impact scenario (nuclide transport)

- **Optimal experimental design**
  uses uncertainty-informed (impact) models to assess the value-add of surface probing and geophysical measurements

- **Surrogate modeling**
  constitutes an enabling technology for compute- intense tasks in impact modeling and optimal experimental design



Framework



University of Stuttgart
RWTH AACHEN UNIVERSITY

The Need for a Smarter Data Management Tool

We need material properties and uncertainties to feed into our models.

Common challenge: **Data findability**!

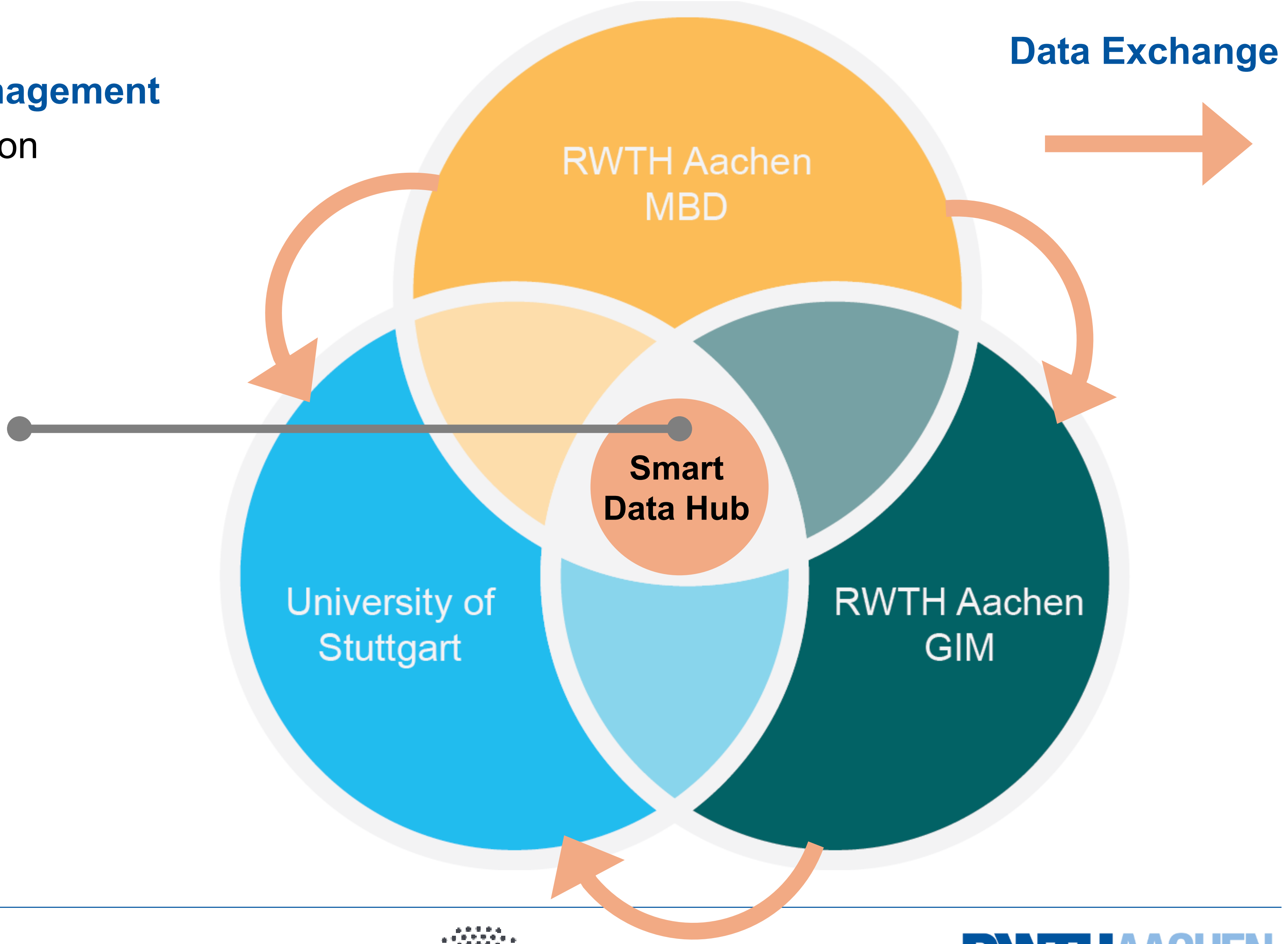
Are we managing it in the most efficient way?

What if we could **seamlessly integrate** this **data into simulation** workflows?

data is structured, **accessible**, and directly compatible with your simulations!

Core digital infrastructure

- **Smart Data Hub**
  A FAIR approach to input/output data management
  idea: provides central instance to collect info on
  - geological site,
  - material properties,
  - subsurface geological model,
  - …



starting point for data-integrated white-box impact model

provides uncertainty information as input for uncertainty management

facilitates traceability and reproducibility

Smart Data Hub: Database

- Rock properties

YAML-DB

density ρ:
- type
- unit
- value/relation
- Uncertainty
- source
- flag

conductivity κ:
- type
- unit
- value/relation
- Uncertainty
- source
- flag

heat capacity c_p:
- type
- unit
- value/relation
- Uncertainty
- source
- flag

porosity φ:
- type
- unit
- value/relation
- Uncertainty
- source
- flag

Sites

synthetic
- DE_Crystalline.yml
- DE_Rocksalt.yml
- ...

reference
- USA_Wipp_Rocksalt.yml
- CH_MontTerri_OpalinusClay.yml

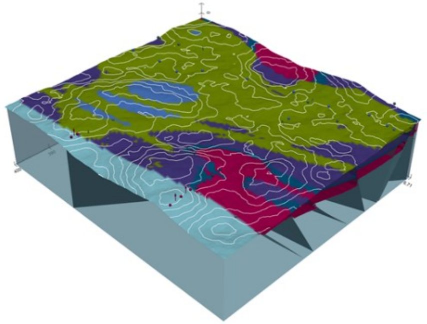
candidate

sensible defaults

age_lithology
- Cenozoic_Claystone.yml
- Cenozoic_Sandstone.yml
- Cenozoic_Limestone.yml
- Mesozoic_Sandstone.yml
- Mesozoic_Claystone.yml
- Mesozoic_Limestone.yml
- Paleozoic_Claystone.yml
- Paleozoic_Sandstone.yml
- Paleozoic_Sandstone.yml
- Paleozoic_Limestone.yml

- Geomodel

GemPy



+ PyVista

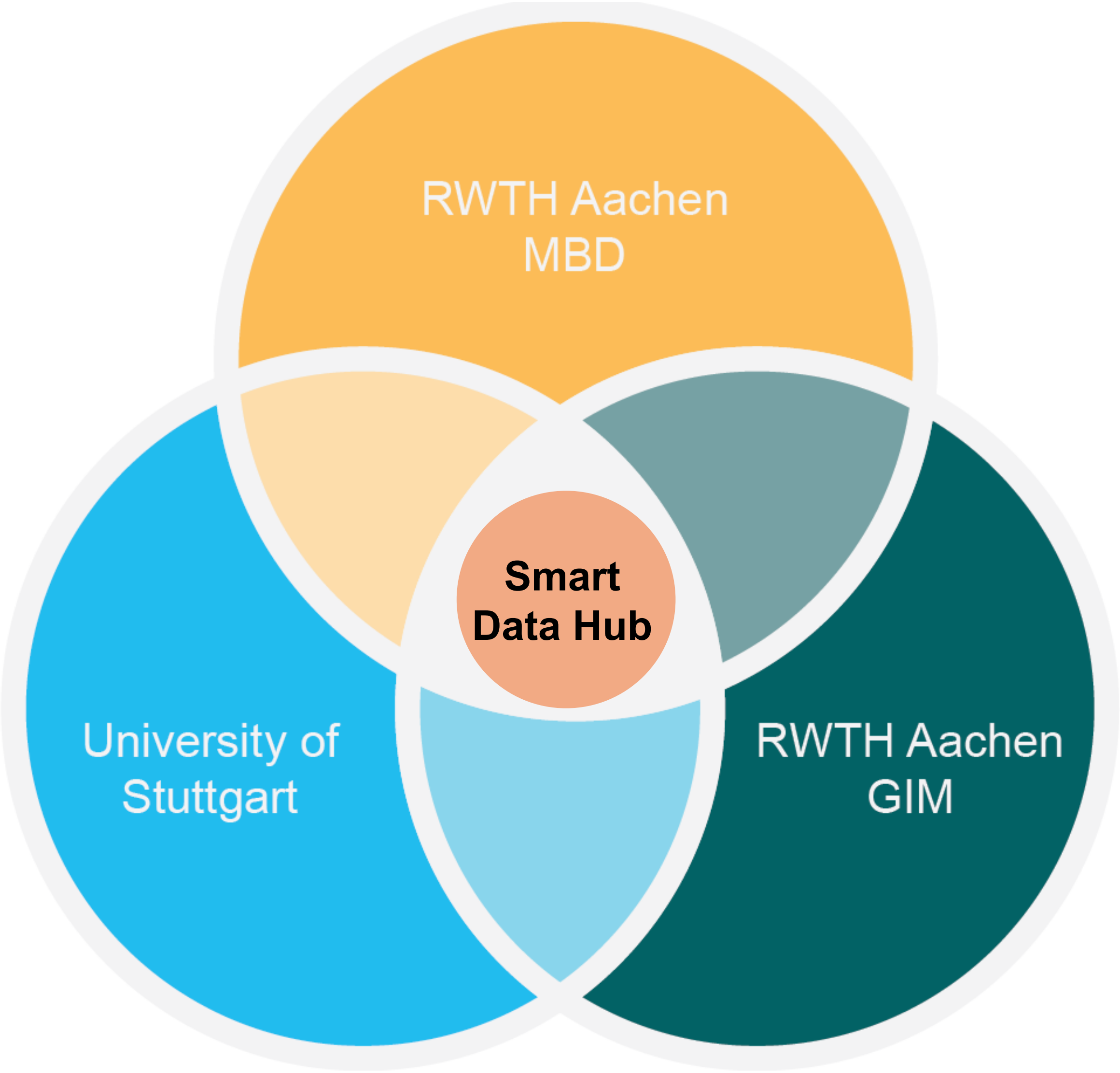
save as plydata for Dashboard visualization

Example:
geometry/USA_Wipp_Rocksalt

- Changhsingian.ply
- Tertiary.ply
- Wuchiapingian-Lower.ply
- Wuchiapingian-Middle.ply
- Wuchiapingian-Upper.ply

Live Demo

Smart Data Hub: Publicly accessible via GitHub



Smart Data Hub - A Data-Centric Approach for Integrated Simulation Workflow Management in Radioactive Waste Disposal

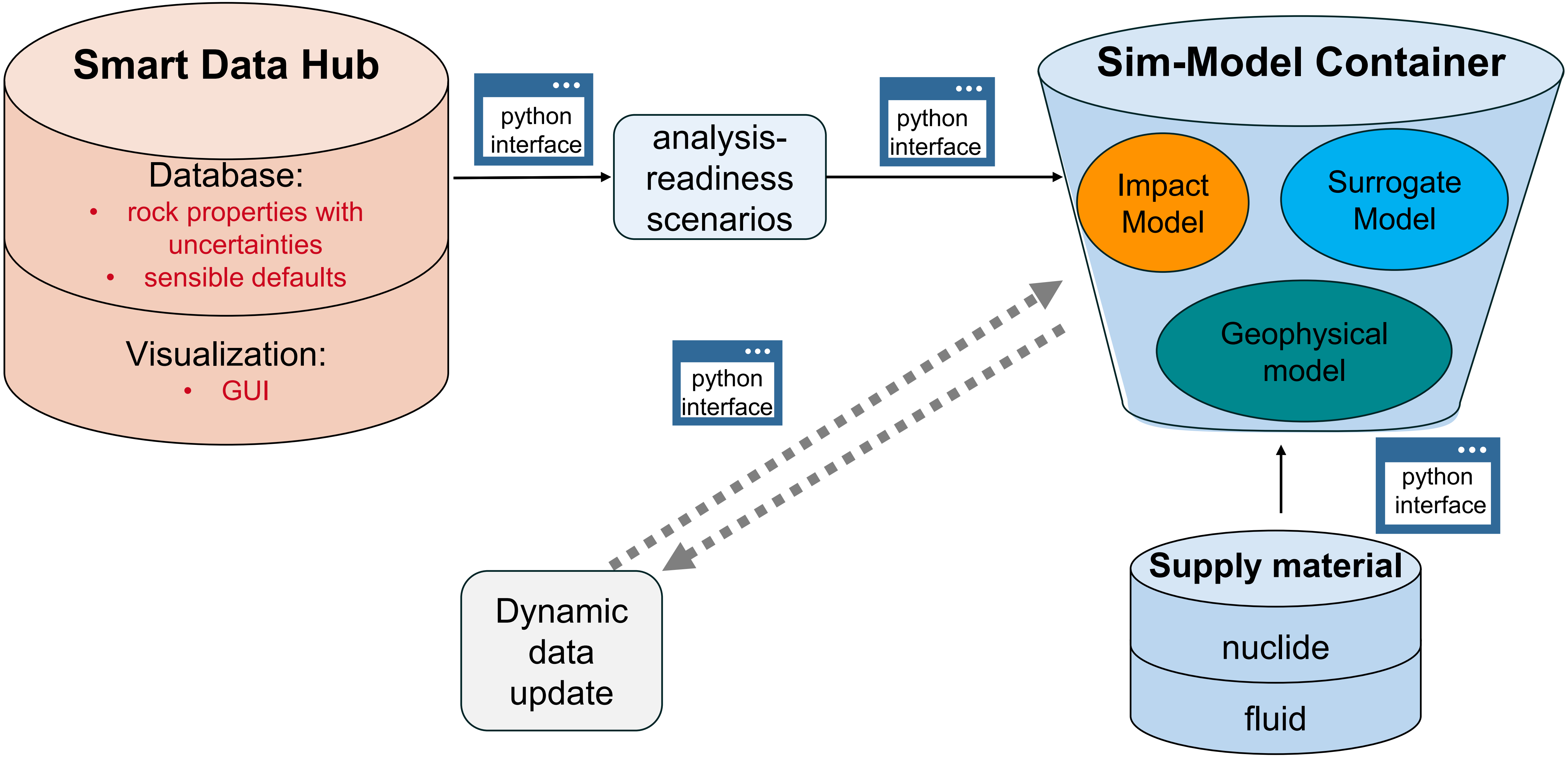
Contribution to a conference proceedings:

- **DECOVALEX 2023, Troyes France**
- **EGU2024, Vienna Austria**
- **Research Data Day2024, Julich Germany**

Available in GitHub

Paper in preparation

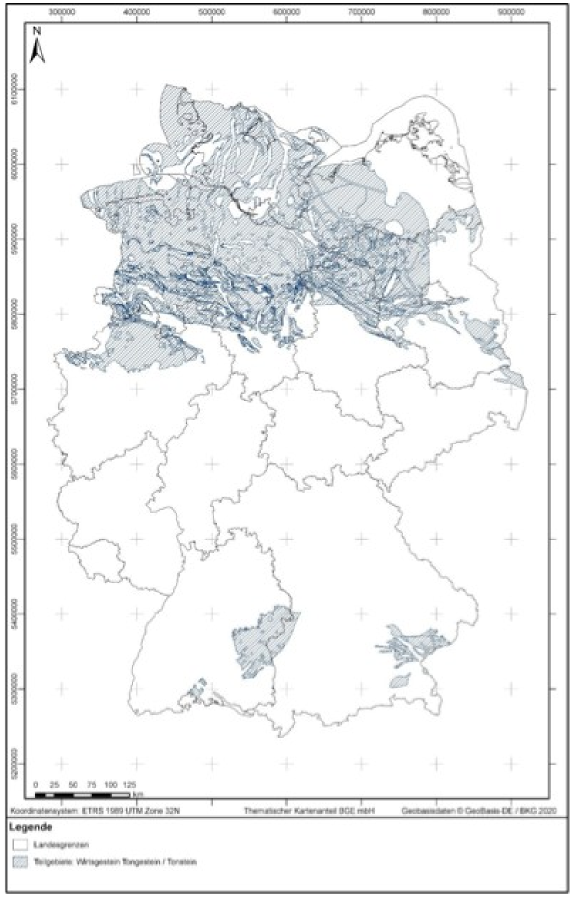
Data need to be readily accessible

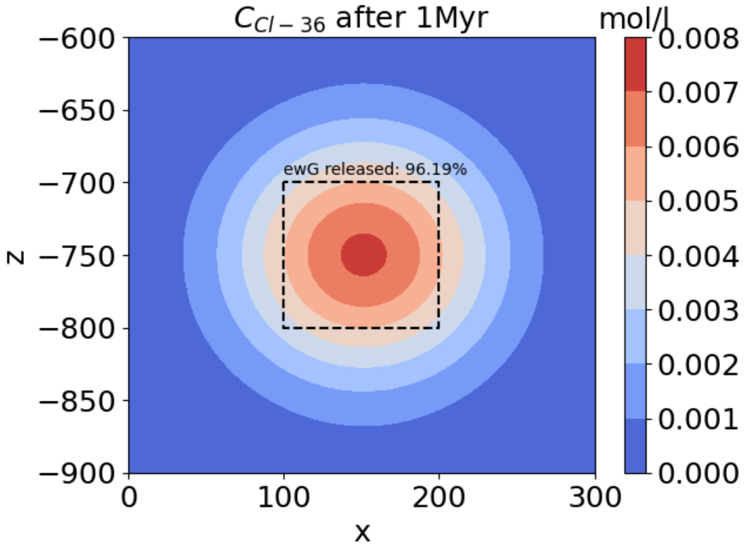


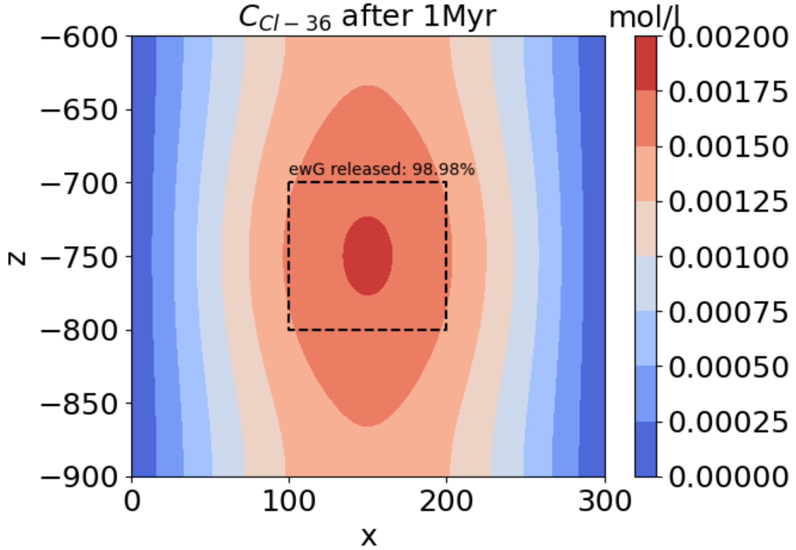
University of Stuttgart RWTH AACHEN UNIVERSITY

Process and Impact models

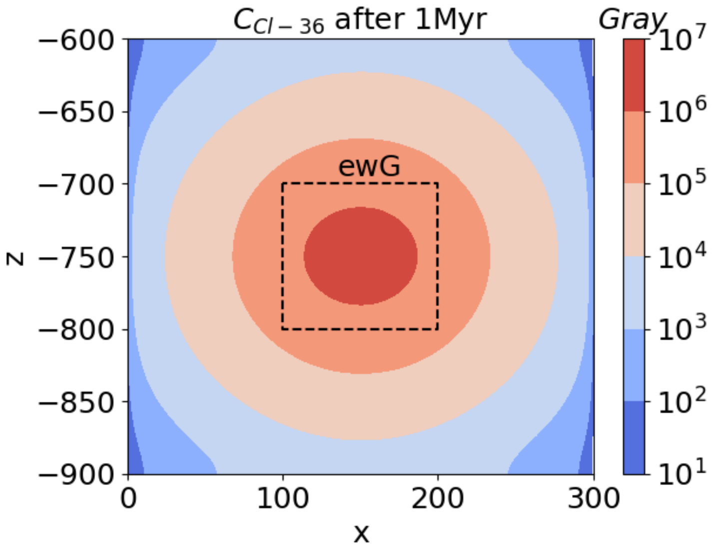
- Claystone

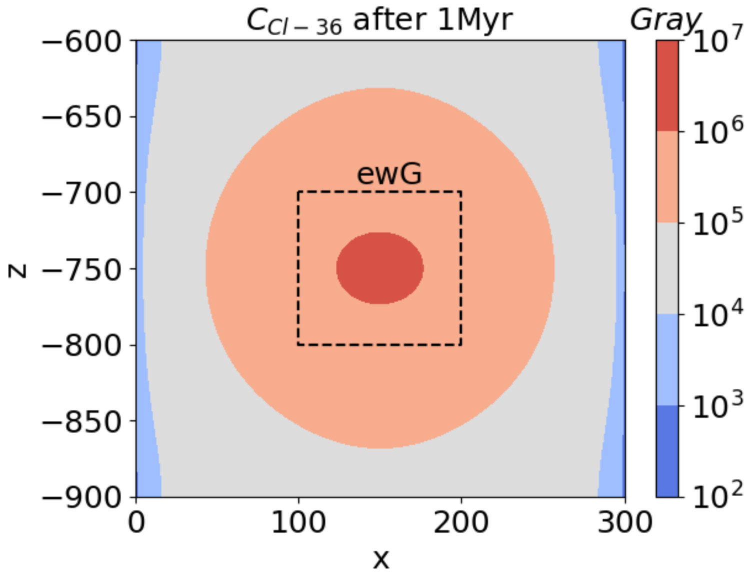


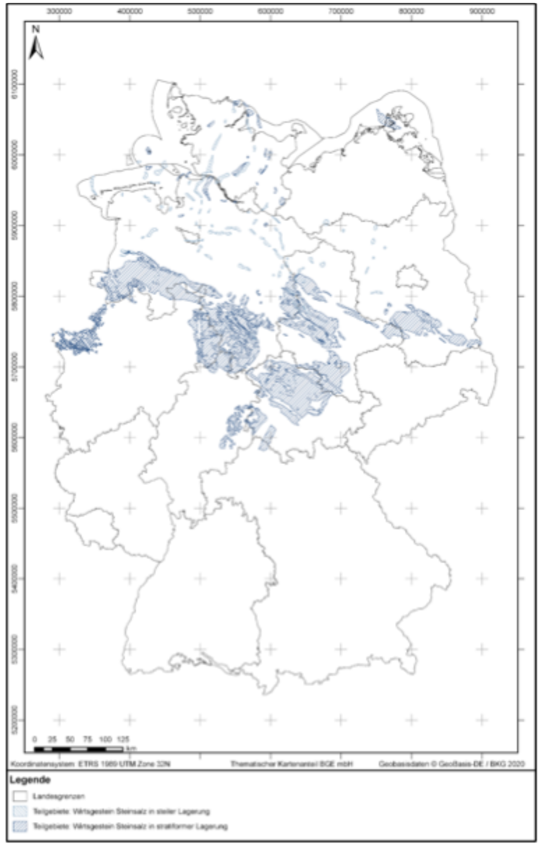




- Rock Salt







Python interface between Smart Data Hub and OGS

Smart Data Hub

send ‘main’ to the Forward Solver

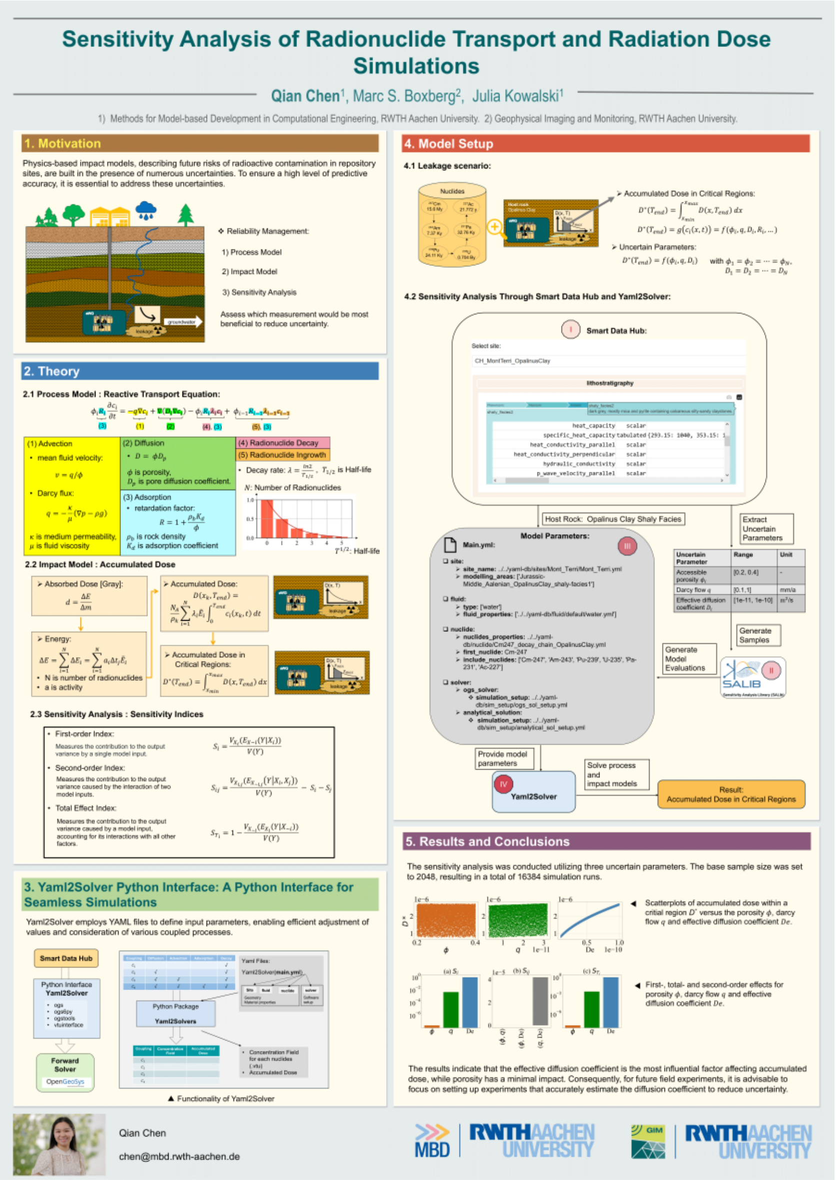
Python Interface
Yaml2Solver

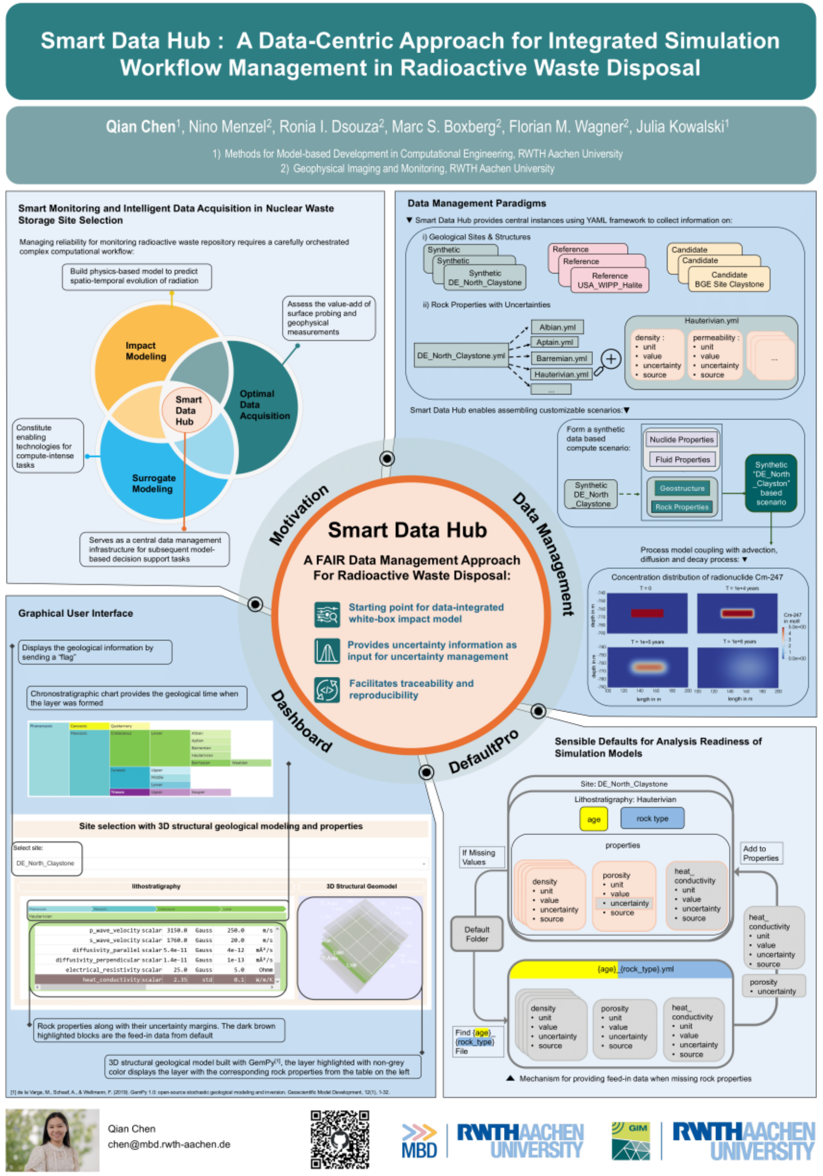
- ogs
- ogs6py
- ogstools
- vtuinterface

Forward Solver

OpenGeoSys
OPEN-SOURCE MULTI-PHYSICS

Poster Session





Surrogate Models in URS

Maria Fernanda Morales, Universität Stuttgart

My main points will be…

- URS needs surrogate models
    - Surrogates have their own uncertainty
- Python package for surrogate model building and UQ
- Surrogates are challenging for models
    - with many parameters
    - with many results

Surrogate Models: What and Why?

Model

Highly detailed geometry & processes uncertain parameters

(Limited) forward model runs

Surrogate

Surrogate model training

What: substitute original models

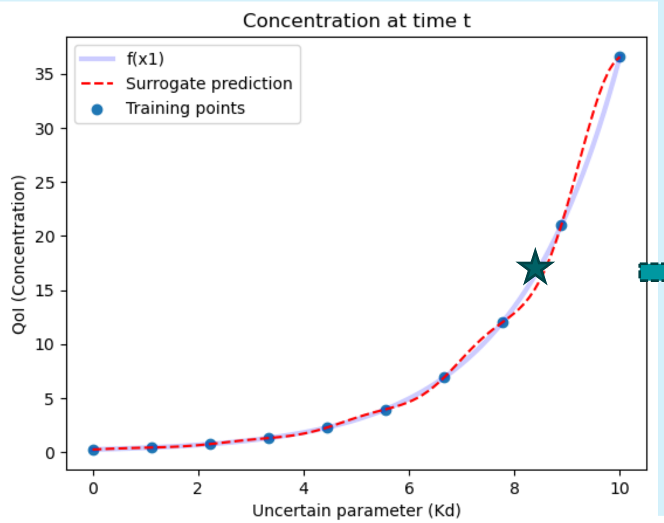
- Trained with few model runs

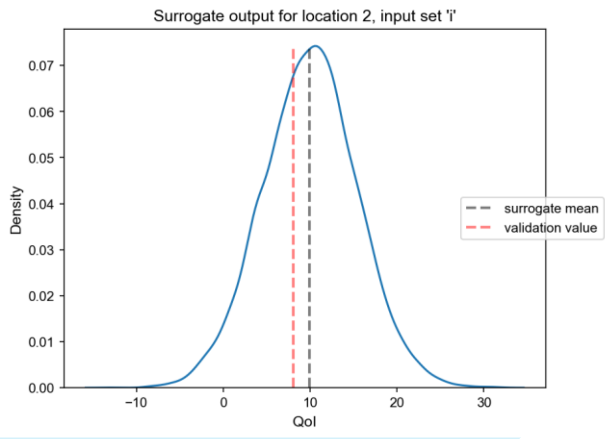
Why: run in a fraction of the time

But: are **approximations**

→ have errors

→ We use surrogates that quantify their uncertainty

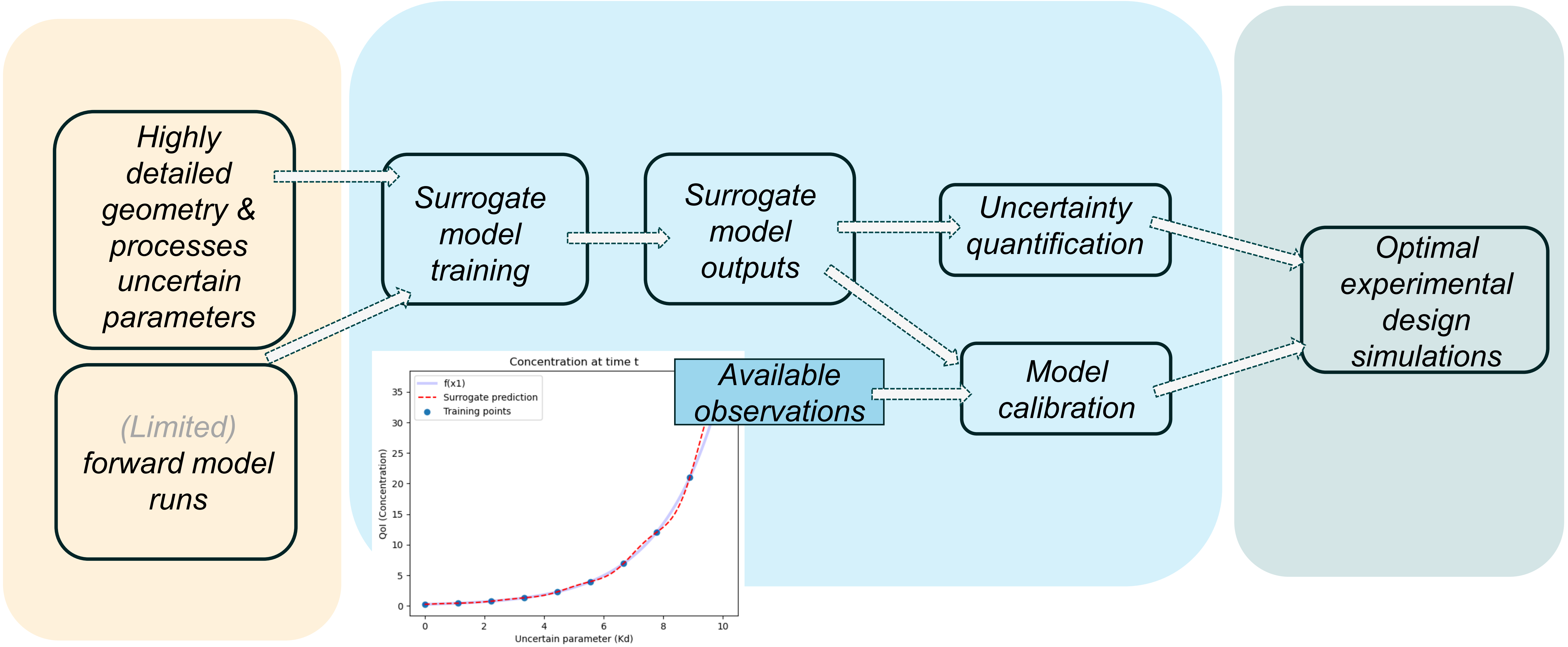




University of Stuttgart

RWTH AACHEN UNIVERSITY

Surrogate Models in Smart Monitoring & URS



Surrogate Model Python Package: BayesValidRox

Easy all-in-one **Bayesian** toolbox for:

- Surrogate model building
  - Optimizes model structure
  - Optimizes selection of original model runs (Active Learning)
- Uncertainty quantification and sensitivity analysis
- Bayesian model calibration
- Model comparison

Available in BayesValidRox 2.0

Website with documentation

In progress

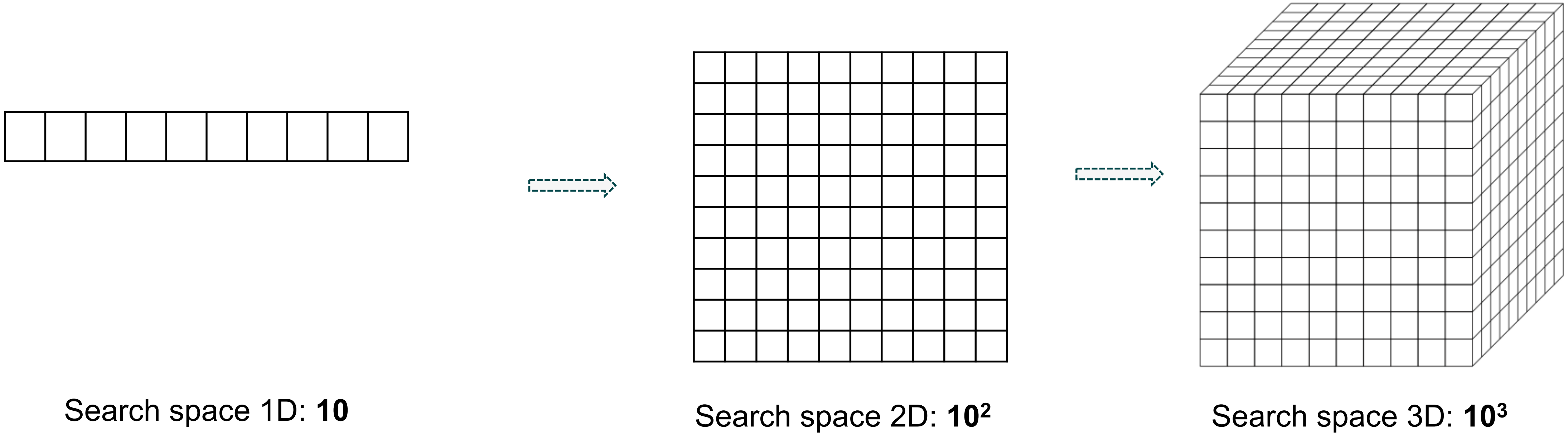
Paper in preparation

https://pypi.org/project/bayesvalidrox/

Phase I | Phase II | Phase III | Post-Closure

What's the problem with "high-dimensional" inputs?

- Surrogate training needs model runs with selected parameter combinations.



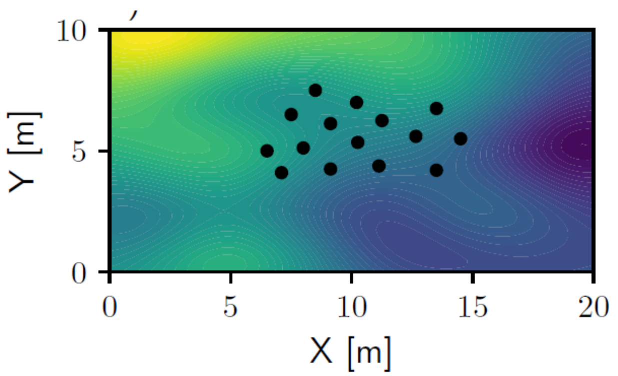
- **More parameters are "exponentially bad" (10, 100, 1000 parameters?)**
- **Our approach**: Input dimension reduction for surrogate training
  - There is an error associated to it, which we must consider during training and use of the surrogate

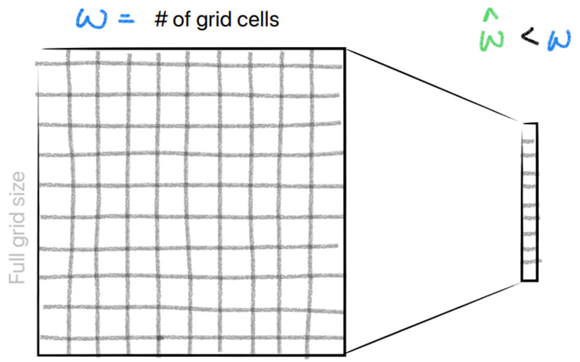
University of Stuttgart

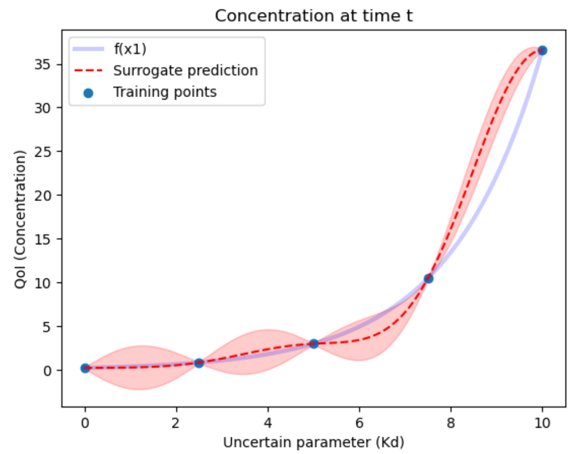
RWTH AACHEN UNIVERSITY

To the rescue: “input dimension reduction”

High dimensional problem → Input dimension reduction (IDR) approach → Surrogate training on reduced inputs





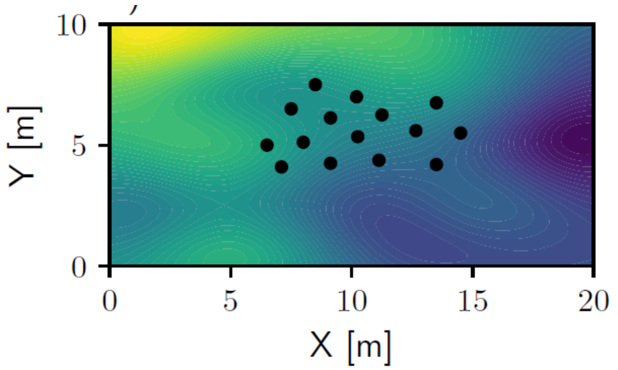


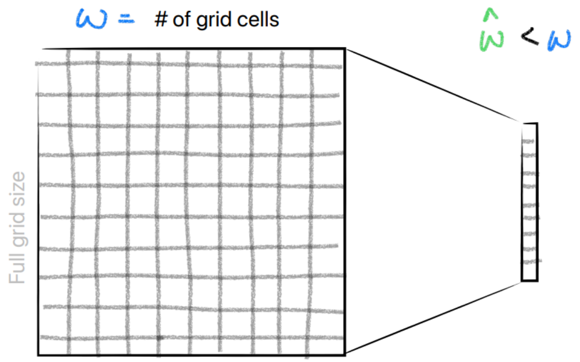
Goal: Find a **smaller** combination of parameters that represent most of the original parameter variability

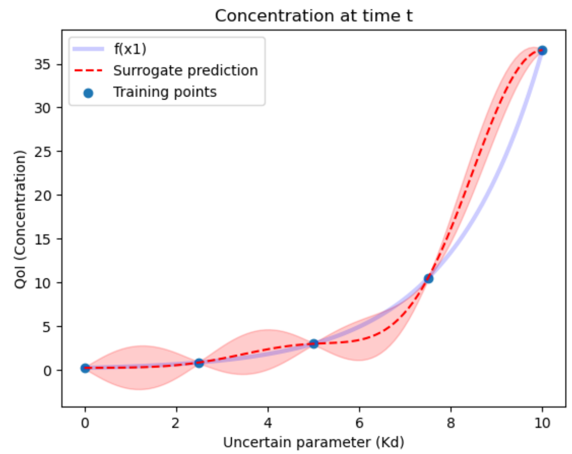
- Examples: PCA, SVD as in PEST, VAE
  - Makes surrogate building cheaper
- Quantify the uncertainty through input reduction together with all other uncertainties → **see our poster**

To the rescue: “input dimension reduction”

High dimensional problem → Input dimension reduction (IDR) approach → Surrogate training on reduced inputs







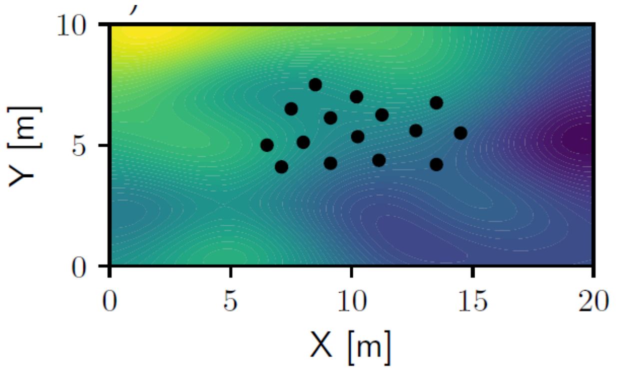
Goal: Find a **smaller** combination of parameters that represent most of the original parameter variability

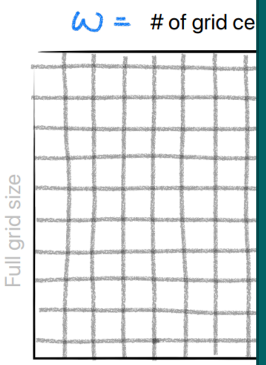
- Examples: PCA, SVD as in PEST, VAE
  - Makes surrogate building cheaper
- Quantify the uncertainty through input reduction together with all other uncertainties → **see our poster**

To the rescue: "input dimension reduction"

High dimensional problem → Input dimension reducti… app…

Surrogate training…





Methodologies presented in EGU2024[1] and GeoEnv2024[2] conferences and 1 paper publication[1]

Paper and technical note in preparation

[1] **Morales Oreamuno MF**, Oladyshkin S, Nowak W. Error-aware surrogate modelling with input dimension reduction for groundwater modelling in heterogeneous media. In: Geophys Res Abstr. Vienna: EGU General Assembly 2024; (Geophys. Res. Abstr.; Bde. 26, EGU24-12586).

[2] **Morales Oreamuno MF**, Oladyshkin S, Nowak W. Training surrogate models using input dimension reduction and Bayesian active learning techniques for inverse modelling in heterogeneous media applications. In: geoENV2024 Book of Abstracts. Chania, Crete, GR:; 2024. S. 191--192. (geoENV2024 Book of Abstracts).

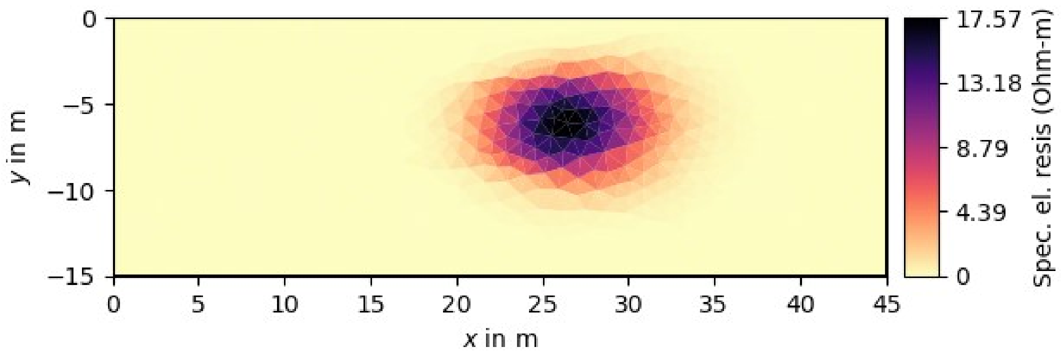
[3] Kröker I, Brünnette T, Wildt N, **Oreamuno MFM**, Kohlhaas R, Oladyshkin S, u. a. Bayesian3 Active Learning for Regularized Multi-Resolution Arbitrary Polynomial Chaos using Information Theory. International Journal for Uncertainty Quantification. September 2024;

Goal: Find a **smaller** combination of p…

- Examples: PCA, SVD as in PEST, VAE
  - Makes surrogate building cheaper
- Quantify the uncertainty through input reduction together with all other uncertainties → **see our poster**

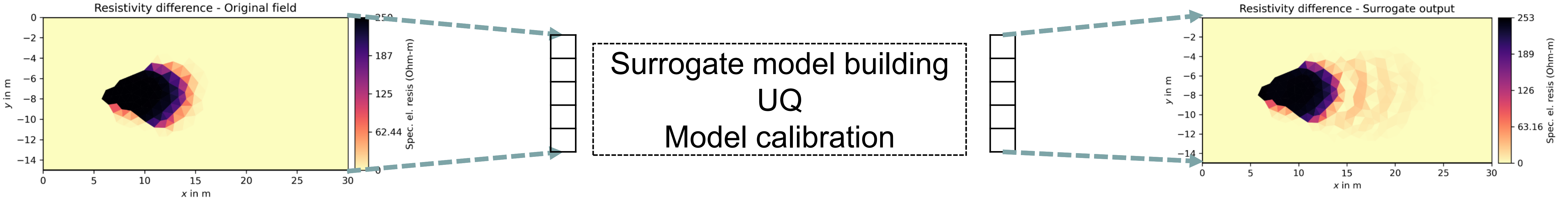
What's the problem with "high-dimensional" OUTPUTS?

- When we have outputs on a 2D/3D grid and over time → large number of outputs



- Potential problems:
  - Each output/cell = 1 individual surrogate
  - Computationally expensive to train and to evaluate
- For **optimal experimental design** you need the space/time resolution

To the rescue: "OUTPUT dimension reduction"



1. ZIP your problem

2. Do everything on ZIP level

3. Un-ZIP your problem

- **Goal**: smaller representation of our outputs (like a ZIP file):
  - Example: PCA, SVG, VAE
- Accelerate OED simulations
- **See our joint poster**

Work in progress

University of Stuttgart

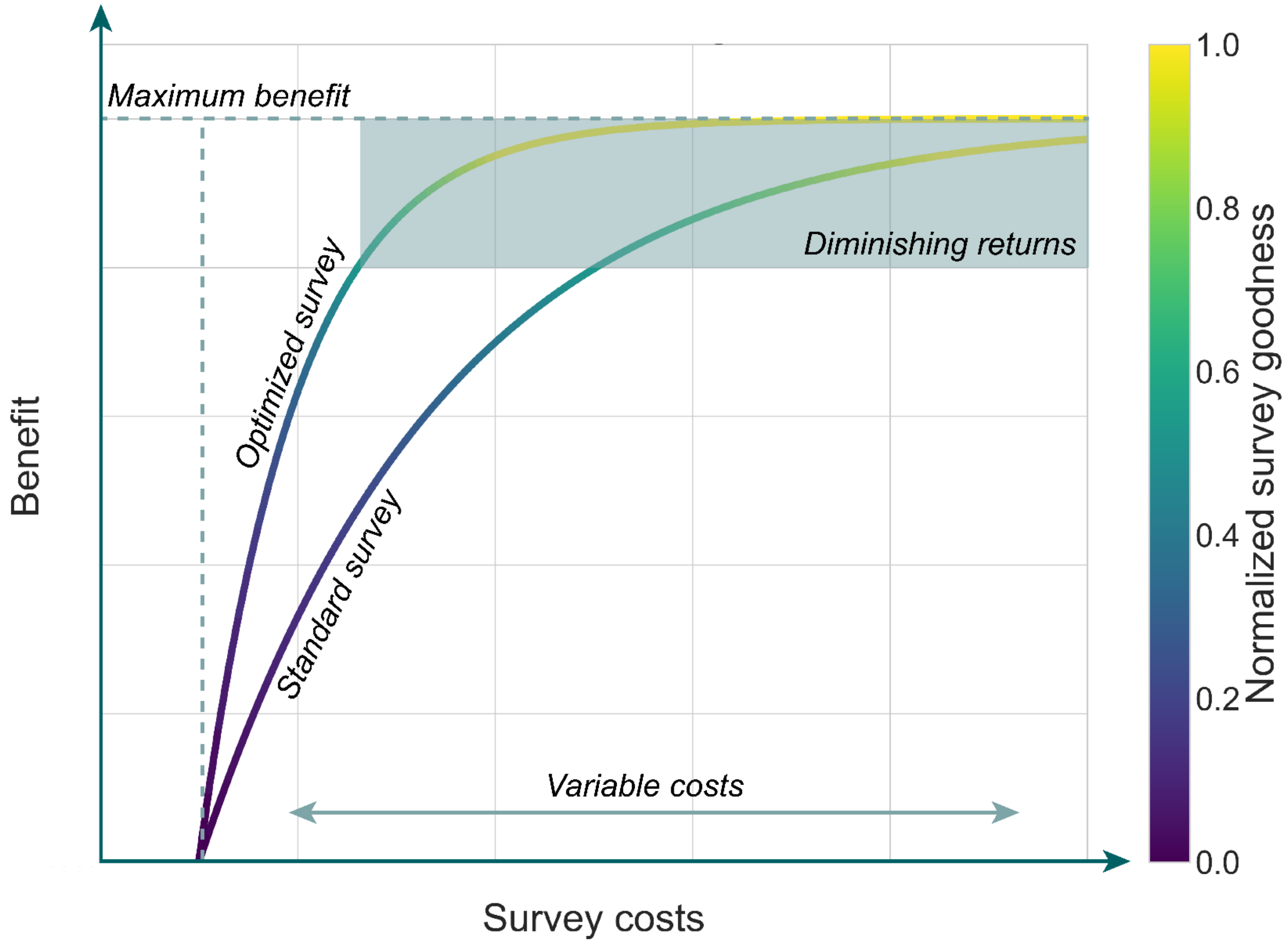
RWTH AACHEN UNIVERSITY

Nino Menzel, GIM, RWTH Aachen

My main points will be…

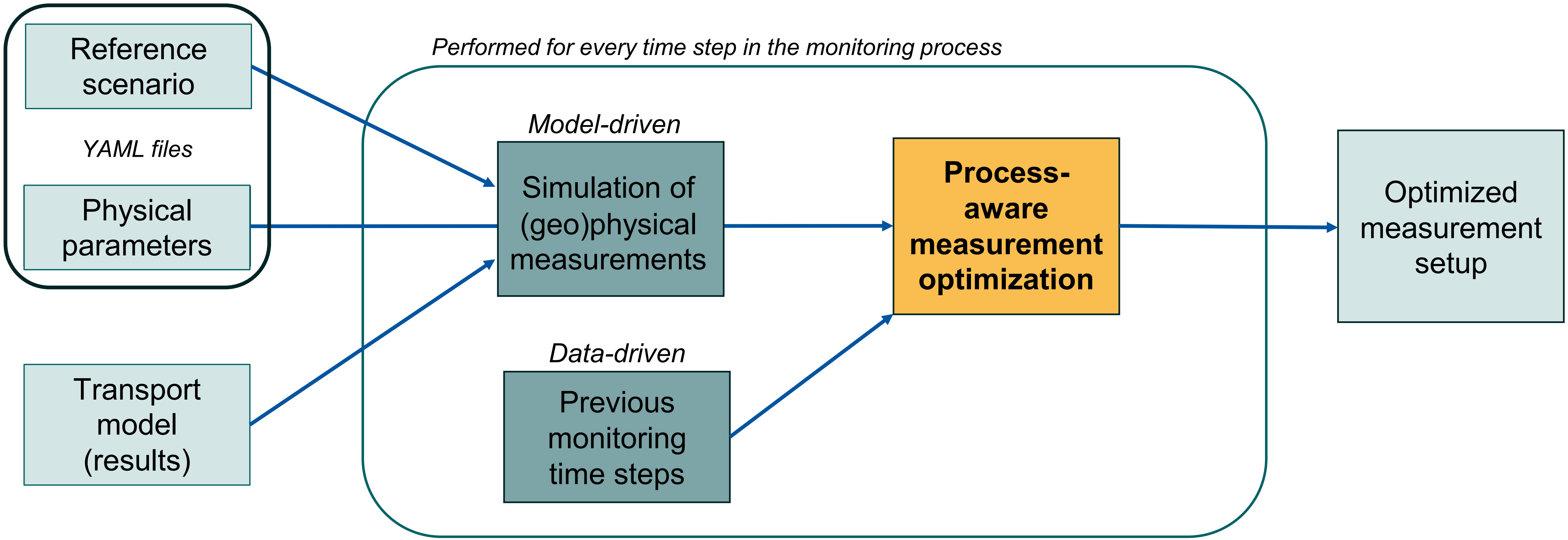
- URS needs smart monitoring strategies
- We use process-aware optimization strategies
- Survey optimization strategies are applicable to
  - surface exploration
  - subsurface exploration and process transport monitoring

Smart Data Acquisition – Overall Concept



- "**Smart**" data acquisition aims at reaching the point of **maximum benefit as fast as possible**
- **Benefit** of a survey: resulting **net increase in resolution** of model parameters of interest
- Overall goal: **limit** the amount of acquired data (and **variable survey cost**) without drastically reducing information content
- In our case: **effective monitoring of fluid transport processes** using (geo-)physical surveys

Smart Data Acquisition – Workflow



Input from ***DataHub*** | *Modelling in* ***pyGIMLi & OED software package*** *(in preparation at GIM)* | *Output to* ***DataHub***

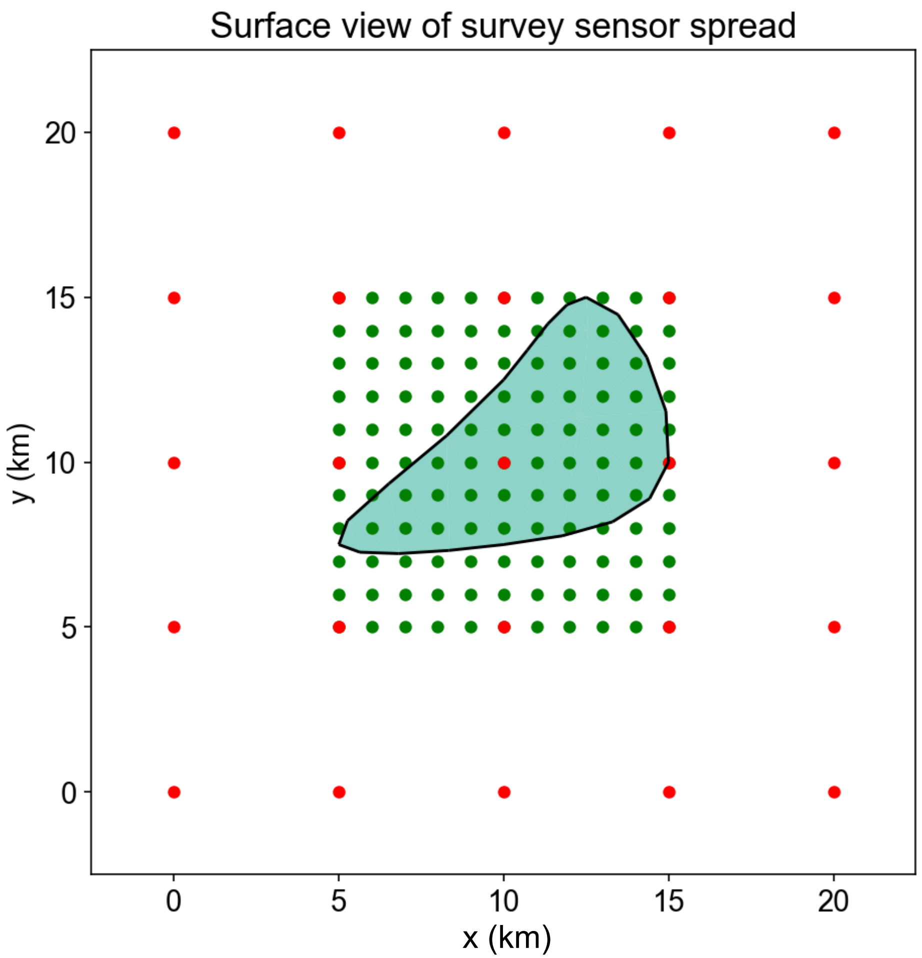
Smart Data Acquisition – Methodology

Inputs:

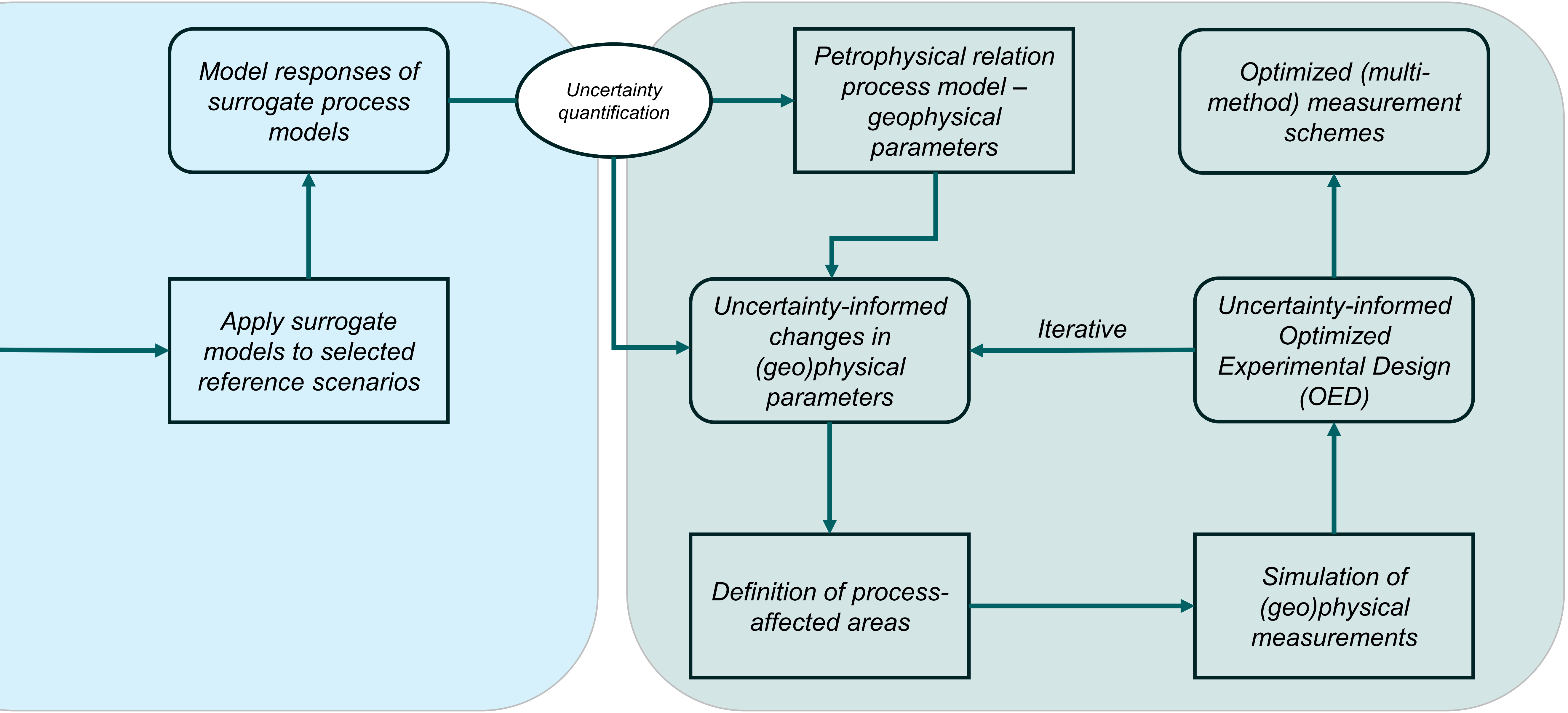
- **A-priori information** of the **target area** in the subsurface (transport process: *hydr. parameters*; geological structure: *geometrical parameters*)

- "Small" **base measurement** setup
  - Seismic tomography survey with *40* receivers and *5* shot points
  - Geoelectric survey using *20* electrodes

- **Densest** possible **measurement setup** (*comprehensive dataset*)
  - Seismic tomography survey with *n* receivers and *m* shot points
  - Geoelectric survey with *n* electrodes



Smart Data Acquisition – Optimization Strategy

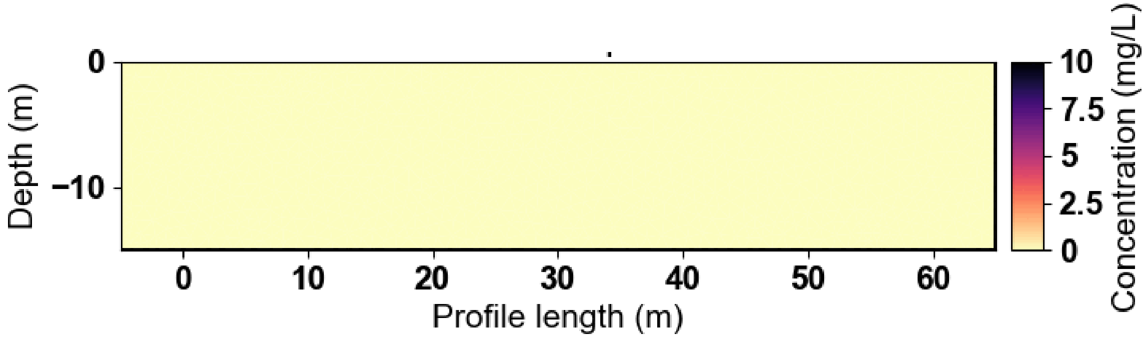


Phase I Phase II Phase III Post-Closure

University of Stuttgart RWTH AACHEN UNIVERSITY

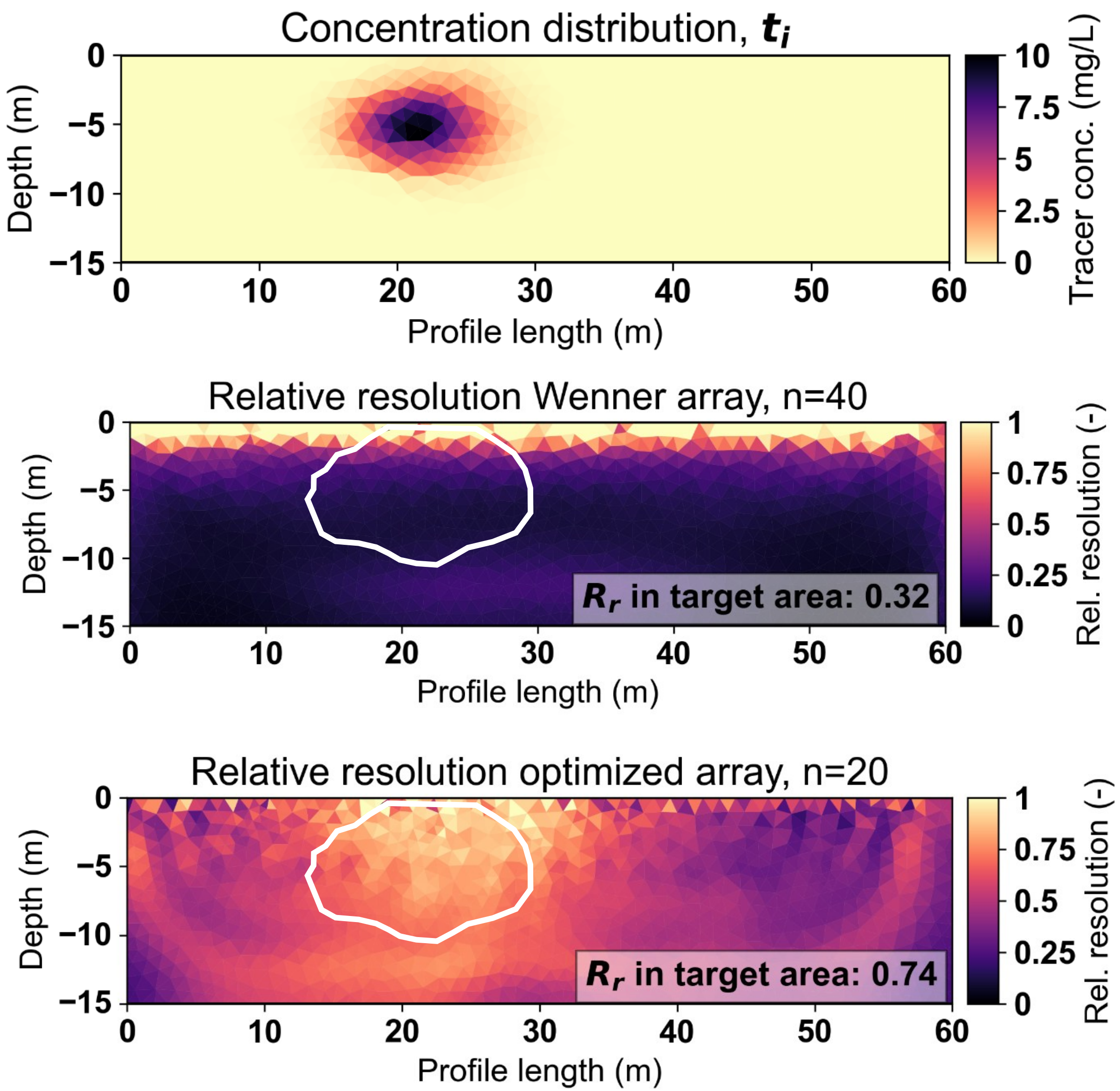
Smart Data Acquisition – Transport Process Monitoring Example

- Optimization algorithms **consider underlying transport process**
- **Focusing** of data acquisition on area that is affected by transport process
- Include **parameter uncertainties** in focusing
  - Multiple model runs with **different physical parameter sets**
  - Consider uncertainties **during optimization**
- Utilize both model predictions and inverse model of acquired data to evaluate simulation quality and **adjust underlying transport model** if necessary



Smart Data Acquisition – Transport Process Monitoring Example

- Optimization algorithms **consider underlying transport process**
- **Focusing** of data acquisition on area that is affected by transport process
- Include **parameter uncertainties** in focusing
  - Multiple model runs with **different physical parameter sets**
  - Consider uncertainties **during optimization**
- Utilize both model predictions and inverse model of acquired data to evaluate simulation quality and **adjust underlying transport model** if necessary
- Process information can come from **any transport simulator and/or real-time monitoring data**

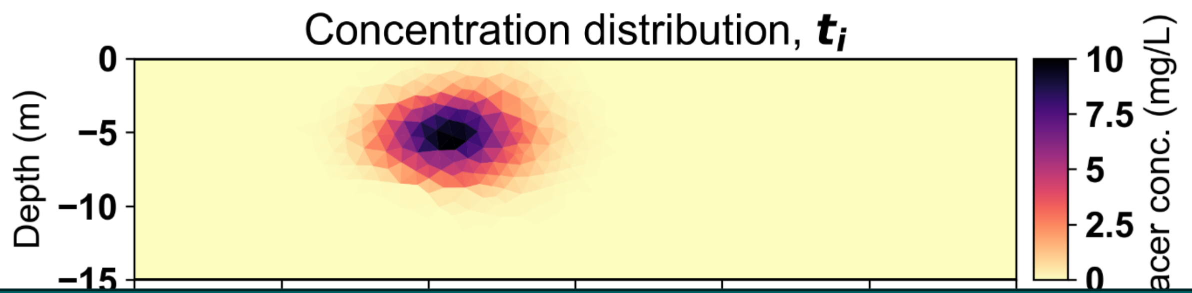


Phase I Phase II Phase III Post-Closure

University of Stuttgart RWTH AACHEN UNIVERSITY

Smart Data Acquisition – Transport Process Monitoring Example

- Optimization algorithms **consider underlying transport process**
- **Focusing** of data acquisition on area that is affected by transport process
- Include **parameter uncertainties** in focusing
  - Multiple model runs with **different physical parameter sets**
  - Consider uncertainties **during optimization**
- Utilize both model predictions and inverse model of acquired data to evaluate simulation quality and **adjust underlying transport model** if necessary
- Process information can come from **any transport simulator and/or real-time monitoring data**

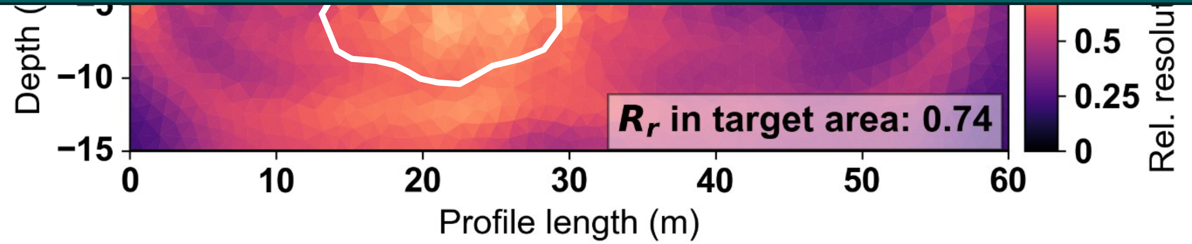


Menzel, N. and Uhlemann, S. and Wagner, F. M. (2024): Strategies for geoelectrical monitoring of subsurface fluid transport processes using Optimized Experimental Design.

EGU General Assembly, Vienna, 14-19 April 2024.

84. Jahrestagung der Deutschen Geophysikalischen Gesellschaft, 10.-14. März, Jena.

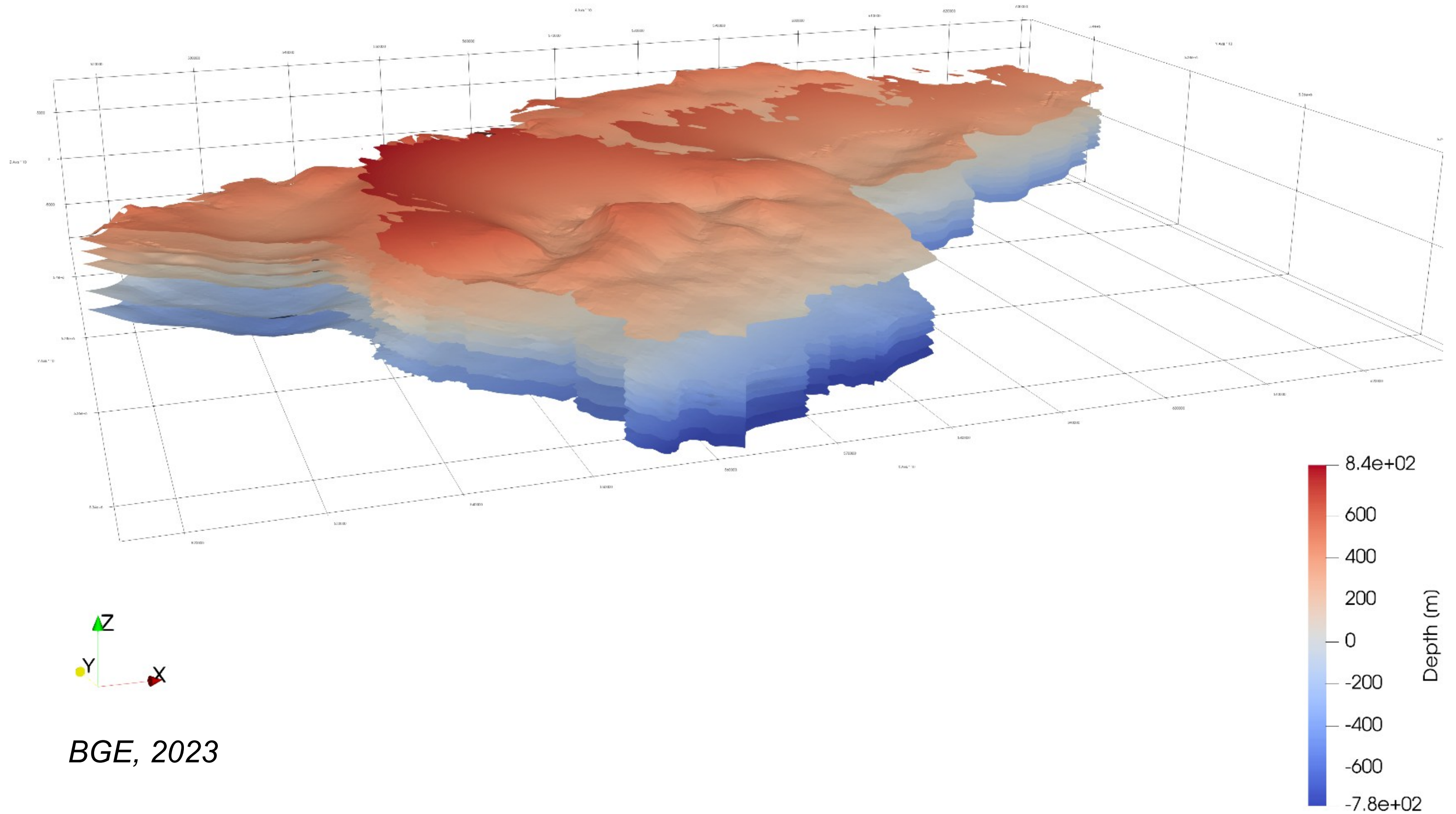
Paper in internal review



Smart Data Acquisition – OED for Site Exploration

OED can also be applied to focus surveys on **static targets**, e.g. **geological features**

- Optimize positions of sensors for surface or borehole exploration
- Optimize length and orientation of geophysical surface and borehole surveys
- Goal: **increase coverage** of measurements **in targeted area**



BGE, 2023

Key Takeaways

Key takeaways:

- **Smart Data Hub** provides **analysis-ready data with uncertainties** that **seamlessly integrate with simulation workflows**
- **Surrogate models enable uncertainty quantification** for computationally expensive models
- Method-agnostic and process-aware “Smart monitoring” strategies are **key for resource-efficient and reliable data acquisition.**
- Surrogates enable **uncertainty-aware OED methodologies**, which require a large number of model runs

University of Stuttgart RWTH AACHEN UNIVERSITY

Thank you for your attention!

Smart Data Hub
chen@mbd.rwth-aachen.de

BayesValidRox
maria.morales@iws.uni-stuttgart.de

SmartDOT
nino.menzel@gim.rwth-aachen.de