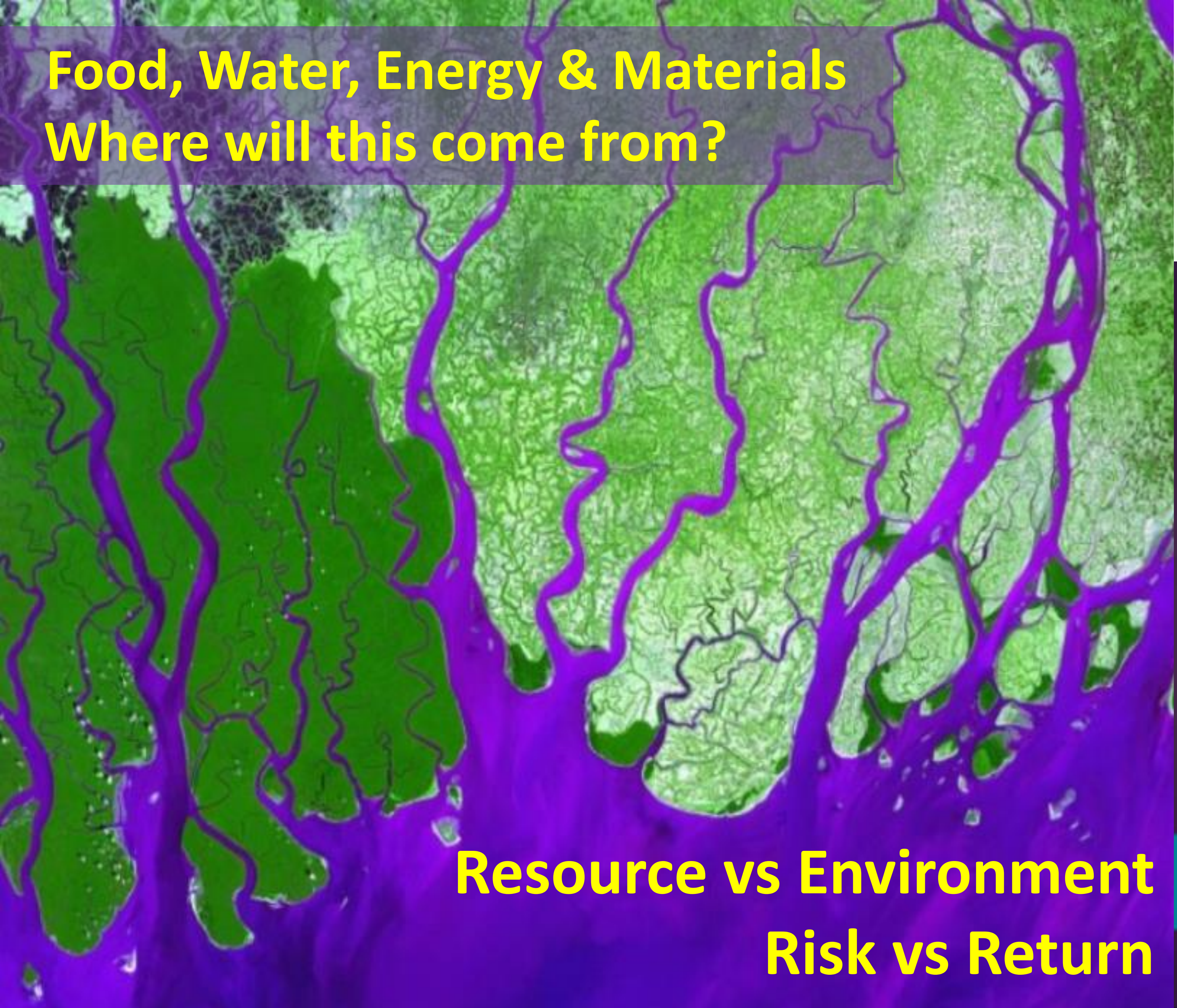
Bayesian Evidential Learning
A scientific protocol for uncertainty quantification

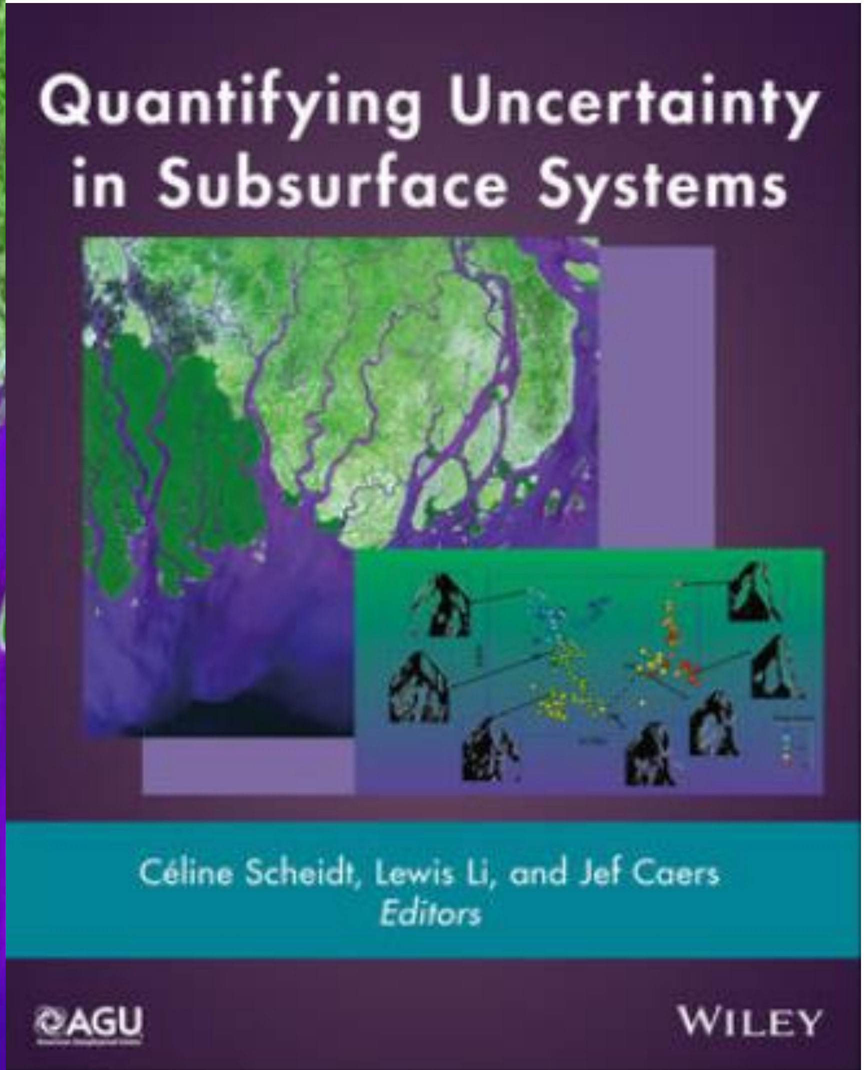
Jef Caers
Geological Sciences
Stanford University, California, USA



EOS Earth & Space Science News

Quantifying Uncertainty About Earth's Resources

A new book explores how uncertainty quantification can enable optimal decisions in the exploration, appraisal, and development of subsurface resources.



What is Bayesian Evidential Learning (BEL)?

- **Bayesianism**:
  - relies on the fundamental notion of prior knowledge
  - an inductive-deductive form of reasoning about science
- **Evidential**: evidence is provided from both field observations and modeling
- **Learning**: it relies on machine learning from Monte Carlo

Six stages of decision making, UQ with BEL

Formulating the decision question and statement of prediction variables

Statement of model complexity and prior uncertainty

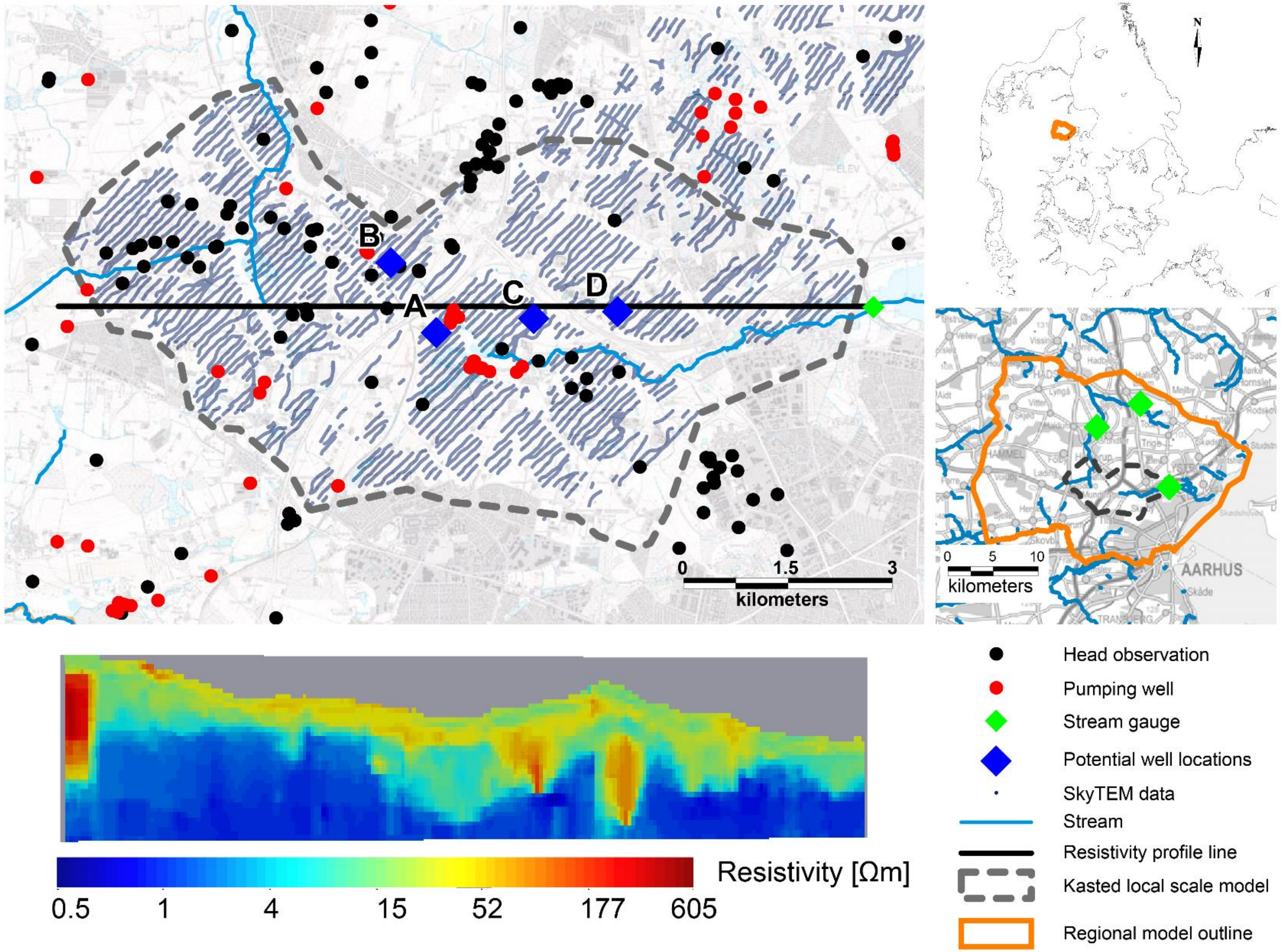
Monte Carlo & falsification of prior uncertainty using data

Sensitivity analysis on both data and prediction variables

Design of uncertainty reduction on prediction variables based on data

Posterior falsification & sensitivity, decision making

Formulating the decision question: groundwater management in Denmark



Re-allocate well field to A,B,C or D?

Formulating the decision question and statement of prediction variables

What problem do I want to address

and

what would it take to solve it?

Decision objectives: “narratives”

minimize *drawdown extraction*: preferably, the new location should be able to bear the burden of the 20% extraction due to re-allocation and anything more is an additional plus. A large drawdown is a risk, and hence needs to be minimized.

maximize *streamflow reduction potential*: depends on the flow in the stream given the existing abstraction, and the flow on the stream if we move 20% of the groundwater abstraction from the existing wells to the new well at any of the four locations.

maximize *increased groundwater outflow of water to wetlands*: due to re-allocation, the aim is to restore the water table, thereby increasing the outflow of groundwater to the wetlands proximate to the existing well field.

minimize *risk of contamination of drinking water*: the abstracted groundwater from the new well originates from within a so-called well catchment zone. This catchment zone intersects land-use, such as “nature”, “city”, “farmland” and industry

Prediction variables

***minimize** drawdown extraction*: preferably, the new location should be able to bear the burden of the 20% extraction due to re-allocation and anything more is an additional plus. A large drawdown is a risk, and hence needs to be minimized. **WDD**

***maximize** streamflow reduction potential*: depends on the flow in the stream given the existing abstraction, and the flow on the stream if we move 20% of the groundwater abstraction from the existing wells to the new well at any of the four locations. **SFRP**

***maximize** increased groundwater outflow of water to wetlands*: due to re-allocation, the aim is to restore the water table, thereby increasing the outflow of groundwater to the wetlands proximate to the existing well field. **WLR**

***minimize** risk of contamination of drinking water*: the abstracted groundwater from the new well originates from within a so-called well catchment zone. This catchment zone intersects land-use, such as “nature”, “city”, “farmland” and industry. **INDUS** & **FARM**

Objectives vs Alternatives

alternatives

Objectives

	Loc A	Loc B	Loc C	Loc D
Farming pollution Farm (units)	?	?	?	?
Industry pollution Indus (units)	?	?	?	?
Streamflow restoration SFRP (units)	?	?	?	?
Wetland restoration WLR (units)	?	?	?	?
Drawdown WDD (units)	?	?	?	?

Statement of model complexity and prior uncertainty

How do I conceptualize “Earth” in all its aspects

and

what does the community already know about it?

Statement of model complexity and prior uncertainty

- Define **model complexity** and **prior model uncertainty**

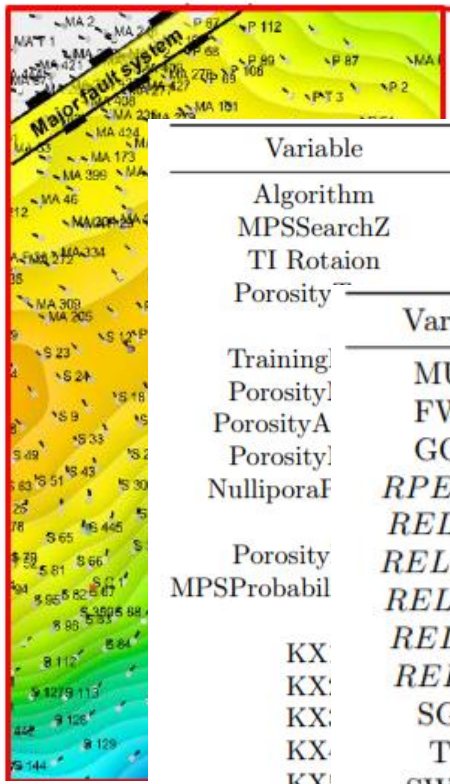
$$\mathbf{m} \qquad\qquad\qquad\qquad f(\mathbf{m})$$

- Need to define this for each field of science involved **jointly**
  - Sedimentology, structural geology, rock physics, geomechanics, hydrogeophysics, multi-phase flow, reactive transport, engineering etc…

- This is just an **initial hypothesis**, no need for accuracy, yet: likely iterate between prior uncertainty stage and falsification stage

Statement of model complexity and prior uncertainty

Matzen: largest oil field in Europe



Geology & Geophysics

Variable	Description	Information	Prior Distribution
Algorithm	Geostatistic Algorithm	Logs, Expert	U[MPS, MPSC]
MPSSearchZ	Z range in K layers for scanning hard data	Expert	U[1, 4]
TI Rotaion	Rotation angle: training image vs hard data	Expert	U[1, 229]
Porosity			
Training			
Porosity			
PorosityA			
Porosity			
NulliporaF			
Porosity			
MPSProbabil			
KX			
KX			
KX			
KX			
KX			
KX			
KX			
KZM			
KZM			

Flow & Transport

Variable	Description	Information	Prior Distribution
MUE	Oil Viscosity (cP)	Lab	U[13, 20]
FWL	Free Water Level (m)	Logs, Lab, Expert	U[1098, 1106]
GOC	Gas Oil Contact (m)	Logs, Expert	U[1078, 1080]
$RPERM_i$	Base RelPer	Analogy	U[5, 9]
REL_{SOR}	Change of be	Analogy	U[-0.06, -0.02]
REL_{KRW}	Change of be	Analogy	U[0.05, 0.15]
REL_{KRO}	Change of be	Analogy	U[-0.15, -0.05]
REL_{NW}	Change of be	Analogy	U[-1.0, -0.2]
REL_{NO}	Change of be	Analogy	U[0.2, 1.0]
SGR	Residual Gas	pert	U[0.05, 0.1]
TZ	Scaling of tra	pert	U[1, 2.3]
SWR1	Minimum wa	pert	U[0.5, 0.65]
SWR2	Minimum wa	pert	U[0.4, 0.6]
SWR3	Minimum water saturation of SATNUM 3	Lab, Expert	U[0.35, 0.5]
SWR4	Minimum water saturation of SATNUM 4	Lab, Expert	U[0.25, 0.4]
SWR5	Minimum water saturation of SATNUM 5	Lab, Expert	U[0.1, 0.25]
$RELPERM^{(i)}$	Relative Permeability = f($RPERM_i$, REL_{SOR}, REL_{KRW}, REL_{KRO}, REL_{NW}, SWR_i)		

All sources of uncertainty need to be considered jointly

Statement of model parameterization and prior uncertainty

Parameter Name	Parameter code	Amount of parameters	Type of uncertainty	Established from	Fields of science
Lithological Model	ma	N/A	Scenario	Geophysics wells	geophysics, glaciology, geostatistics
Regional and local permeability	Kh	22	Log-normal pdfs	Head data Well data	hydrogeology
pressure boundary conditions	ch	5	Uniform	Experience	hydrology engineering
River flows and conductance	riv	8	Conductance: log-normal DEM: uniform	Experience from previous studies	River science
Drain conductance	drn	8	Conductance: log-normal DEM: uniform	Experience from previous studies	hydrology
Aquifer Recharge	rch	1	Trapezoidal	Base-flow estimates	hydrology, meteorology, climate science

Monte Carlo & falsification of prior uncertainty using data

$\mathbf{d}_{obs}$

$\mathbf{d} \longleftrightarrow \mathbf{d}^{(\ell)}, \ell = 1, \dots L$

$g_d(.)$

$\boxed{f(\mathbf{m}) \quad \mathbf{m}}$ $\xleftrightarrow{\text{Monte Carlo}}$ $\mathbf{m}^{(\ell)}, \ell = 1, \dots L$

$g_h(.)$

$\mathbf{h} \longleftrightarrow \mathbf{h}^{(\ell)}, \ell = 1, \dots L$

$g_d(.)$: data forward model

$g_h(.)$: prediction forward model

$\mathbf{d}_{obs}$: actual measurements

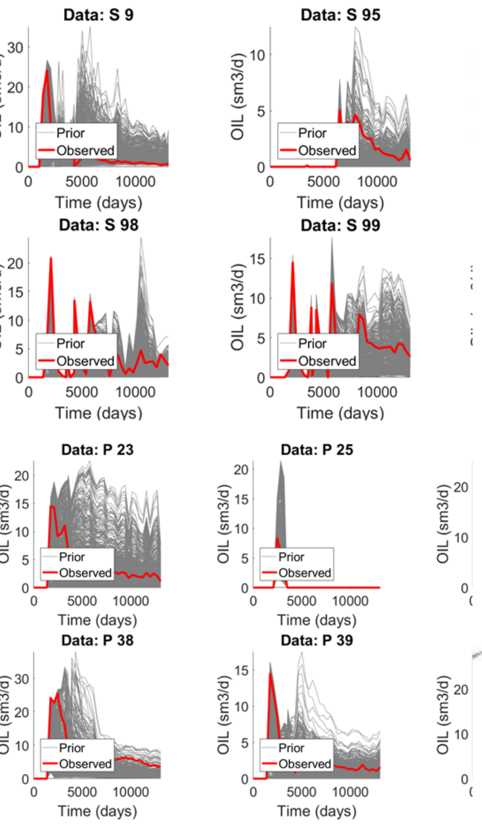
$\mathbf{d}$: **data variables**

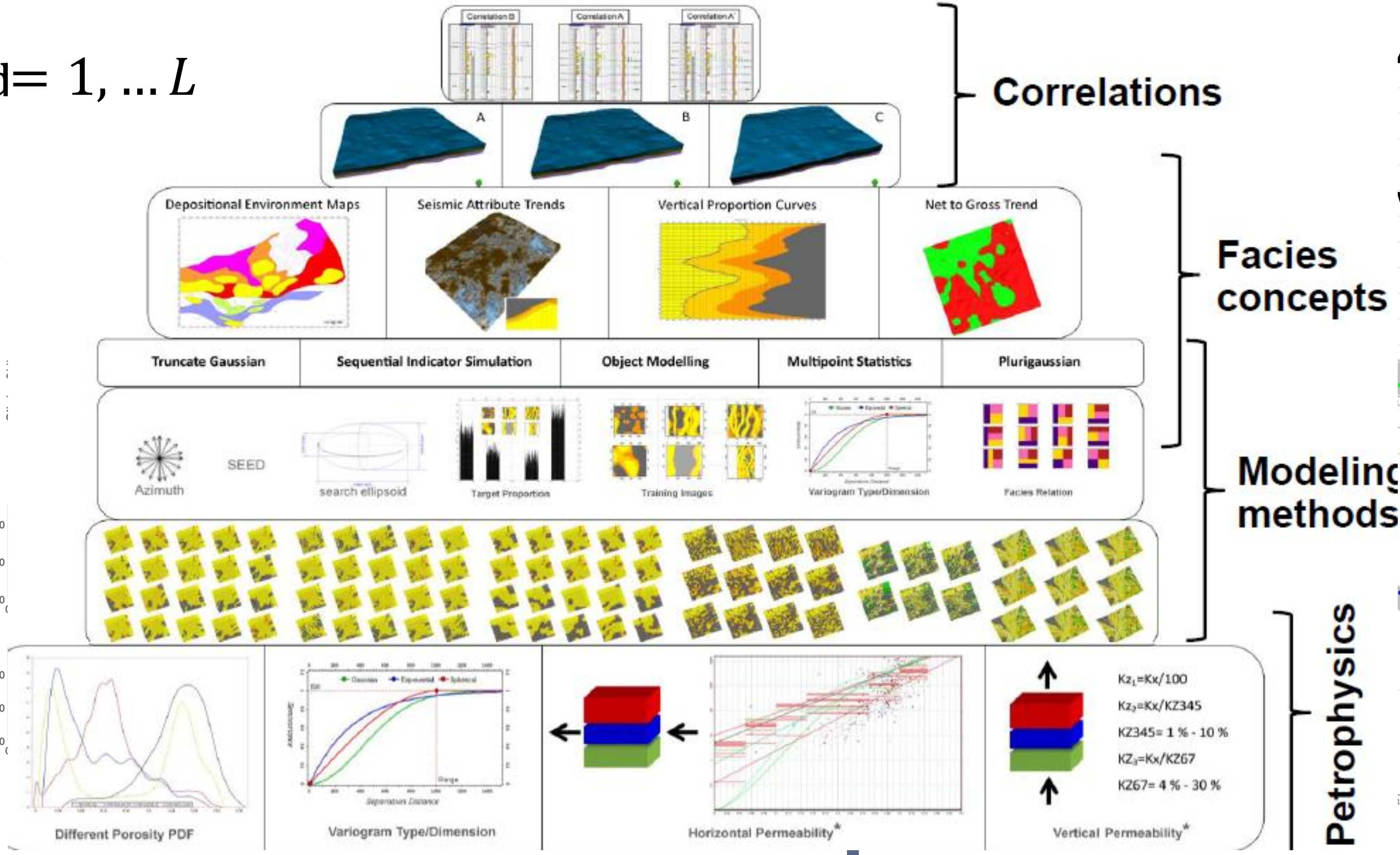
$\mathbf{h}$: **prediction variables**

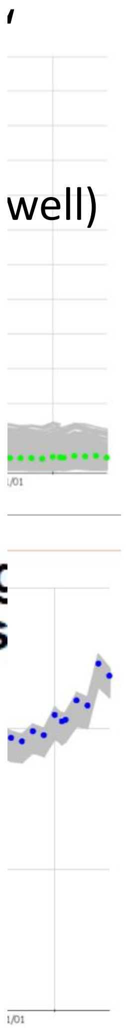
In complex, non-linear systems, Monte Carlo is required for UQ

Monte Carlo: a lot of information is generated

$\mathbf{m}^{(\ell)}, \ell\mathrm{d}= 1, \ldots L$





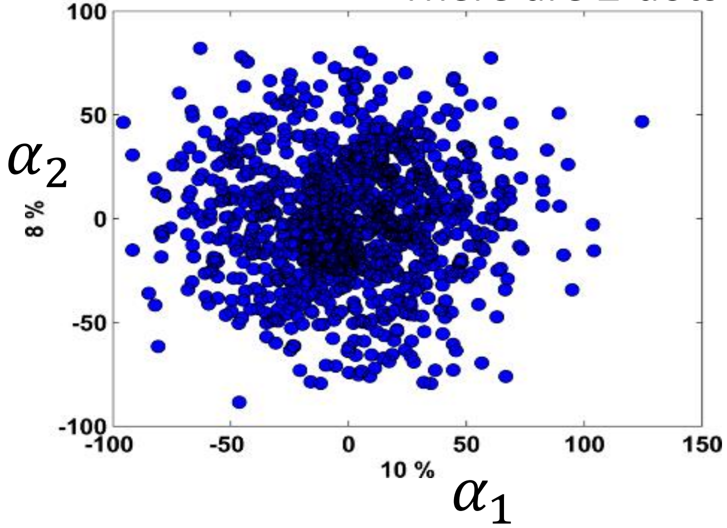


Monte Carlo: dimension reduction

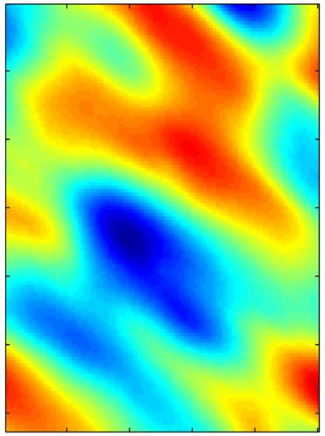
$$\mathbf{x} = \sum_{k=1}^{K} \underbrace{\alpha_k}_{\text{variance}} \overbrace{\boldsymbol{\varphi}_k}^{\text{fixed shapes}}$$

$\mathbf{x} = \mathbf{d}, \mathbf{h}, \mathbf{m}, \ldots$
$K = \min(L, N)$
$\boldsymbol{\varphi}_k^T \boldsymbol{\varphi}_{k'} = 0$

There are L dots

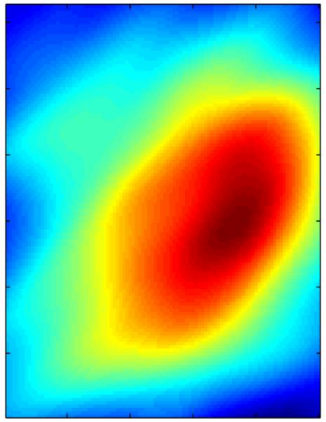


$\mathbf{x}^{(\ell)}$



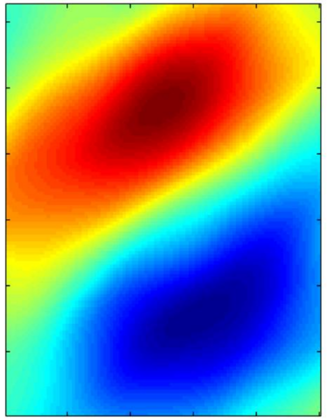
$= \alpha_1^{(\ell)}$

$\boldsymbol{\varphi}_1$



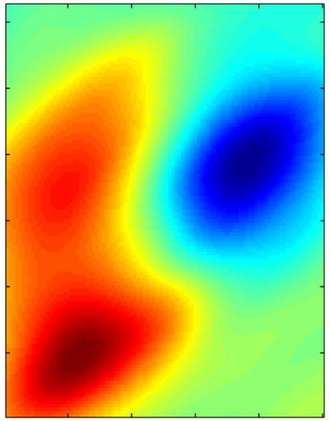
$+ \alpha_2^{(\ell)}$

$\boldsymbol{\varphi}_2$



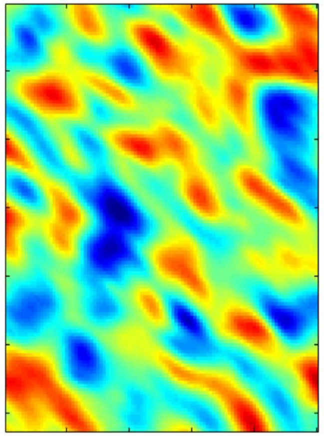
$+ \alpha_3^{(\ell)}$

$\boldsymbol{\varphi}_3$



$\ldots + \alpha_K^{(\ell)}$

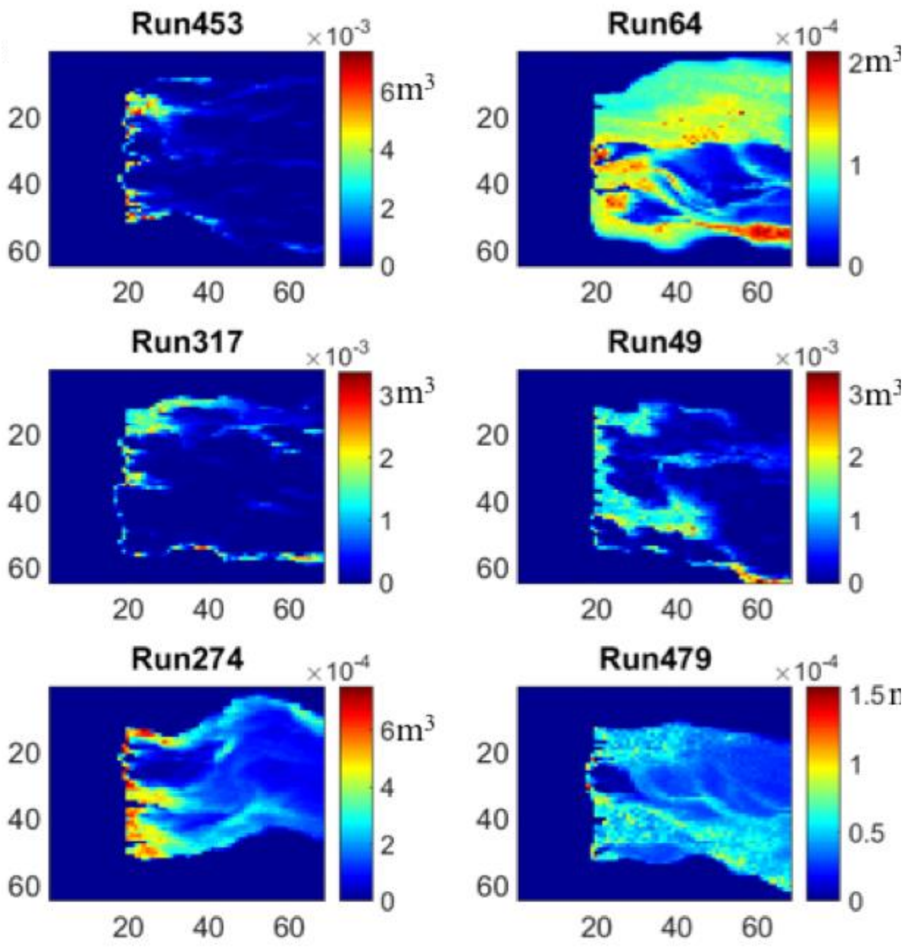
$\boldsymbol{\varphi}_K$



Monte Carlo: reactive transport model example

	Parameter Name (code)	Parameter Type	Parameter Distribution
Geological	Mean permeability (Meanlogk)	Continuous	Logk~U(-11,-10) m^2
	Variance of permeability(VarioVar)	Continuous	Logk~U(0.26,0.69) m^2
	Variogram type of permeability(VarioType)	Discrete	Discrete [1 2 3]
	Variogram correlation length(VarioCorrL)	Continuous	U(3.3, 6.6) m
	Variogram azimuth(VarioAzimuth)	Continuous	U(50,90)^o
	Mean porosity(MeanPoro)	Continuous	U(0.12,0.17)
	Porosity correlation with permeability(PoroCorr)	Continuous	U(0.5,0.8)
	Method to simulate porosity(MethodSimPoro)	Categorical	Discrete [1 2 3 4]
	Spatial Uncertainty(SpatialLogk)	Spatial	500 realizations
Geochemical	Mean Fe(III) mineral content(MeanFerric)	Continuous	U(2.5,10) µmol/g
	Method to simulate Fe(III) mineral content(MethodSimFerric)	Categorical	Discrete [1 2 3 4]
	Fe(III) correlation coefficient with permeability(FerricCorr)	Continuous	U(-0.8, -0.5)
	Mineral surface area(SurfaceArea)	Continuous	U(0.1,2) *base value
	Reaction rate(FerricRate,FerrousRate, UraniumRate, SRBRate)	Continuous	U(0,2) off base reaction rate (varies for different reactions)
	Initial concentrations of different species (ICSulfate,ICFerrous, ICUranium)	Spatial	500 conditioned spatial realizations

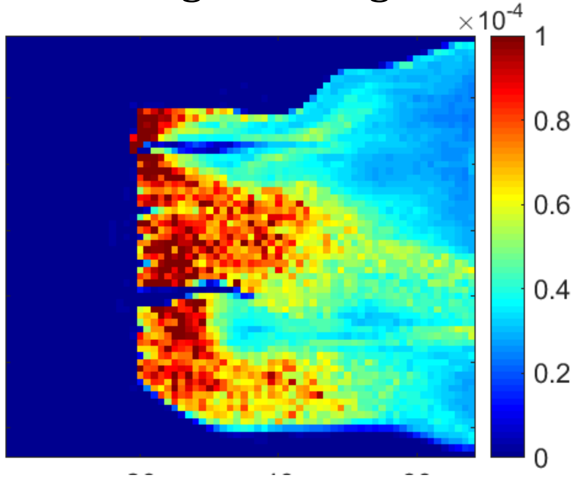
CRUNCHFLOW



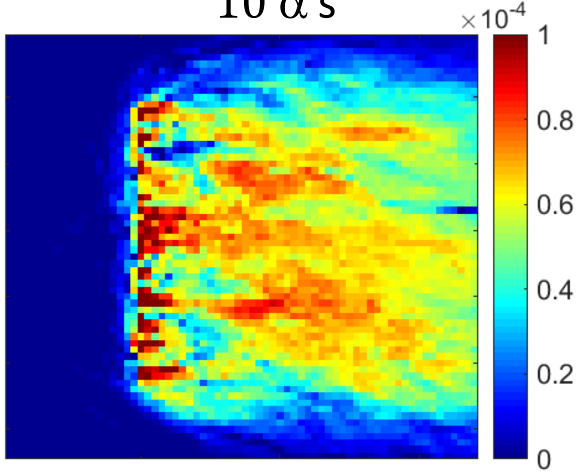
Immobilized uranium

Monte Carlo: reactive transport model example

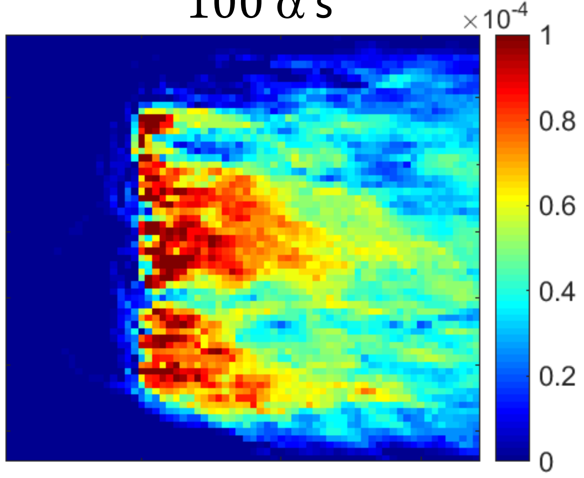
Original image



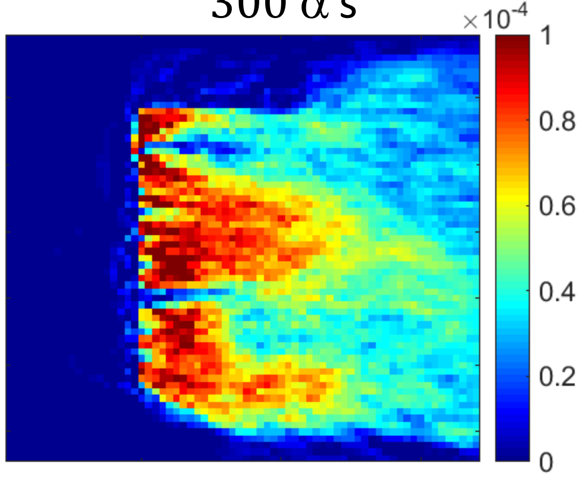
Reconstruction using
10 α's



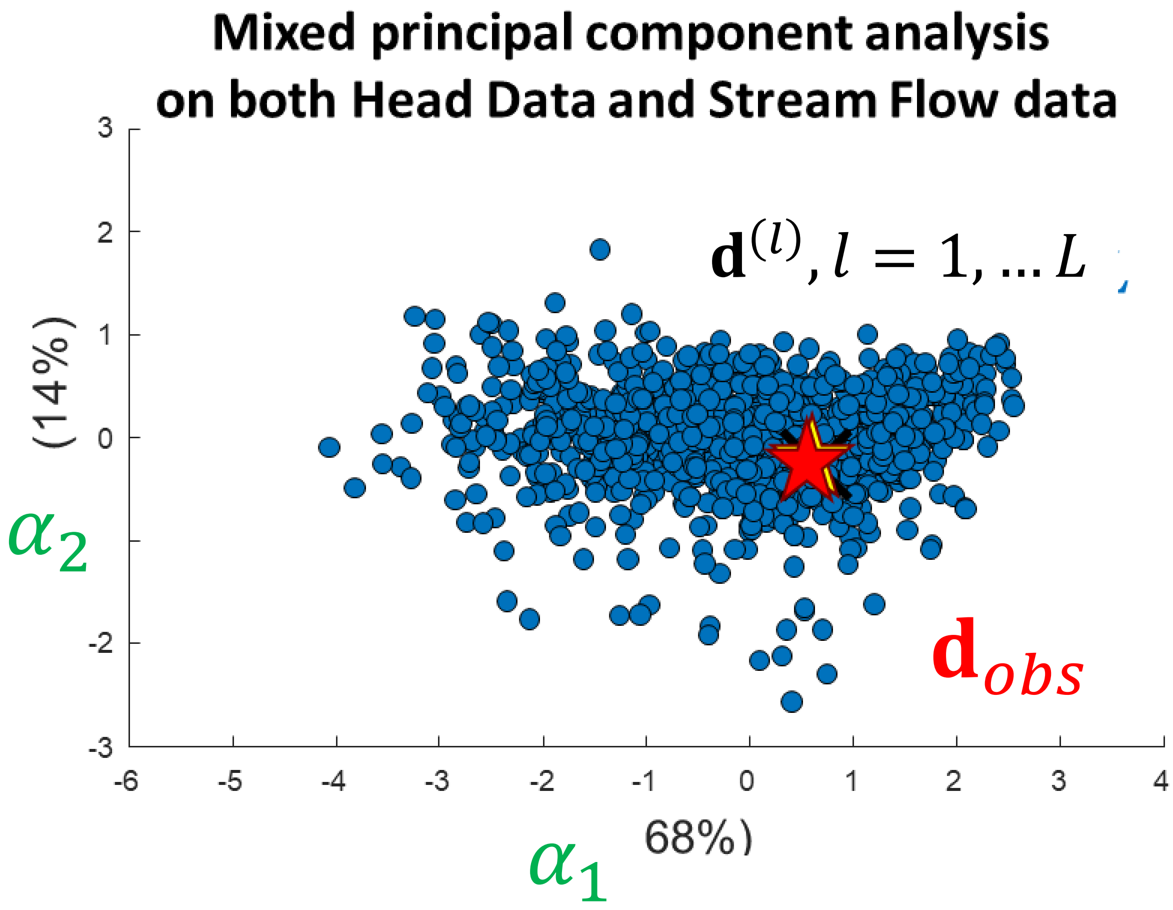
Reconstruction using
100 α's

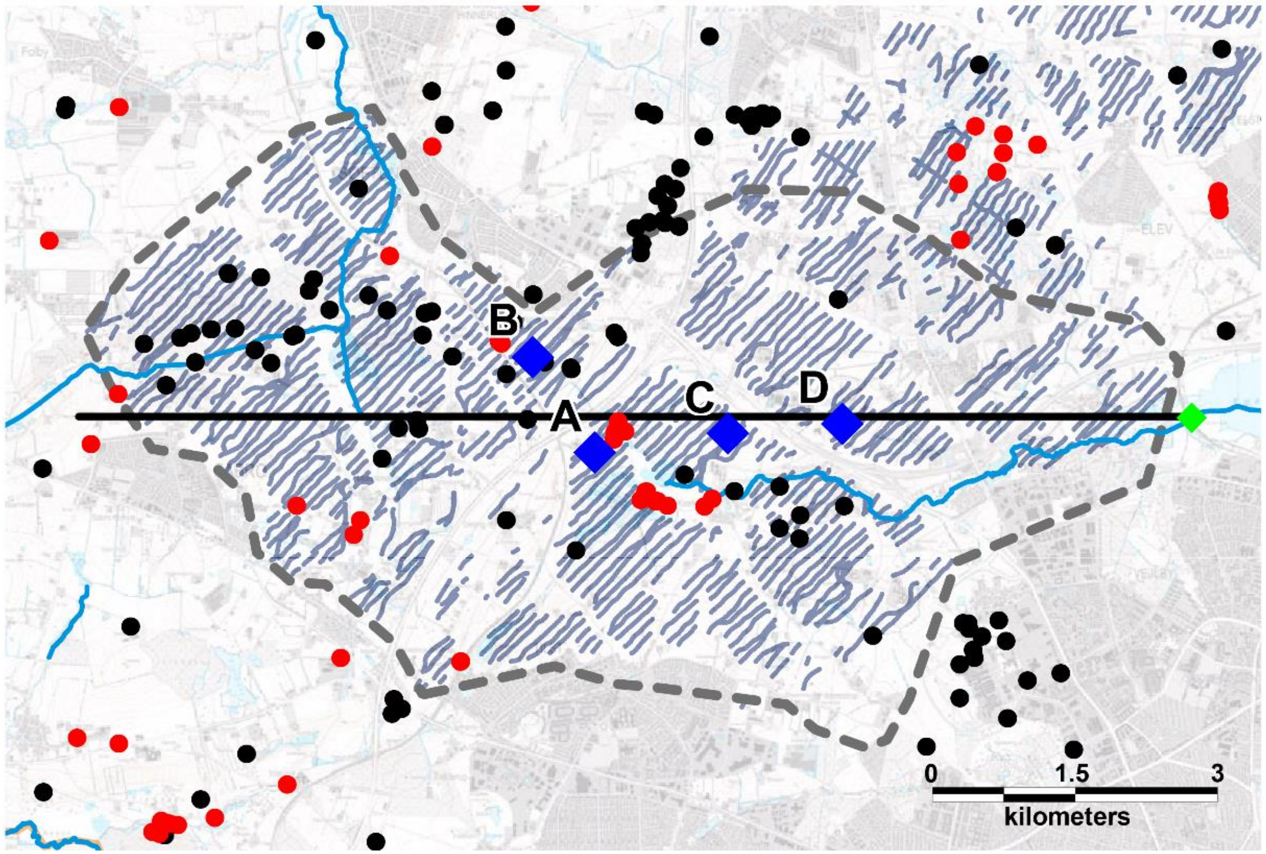


Reconstruction using
300 α's



Monte Carlo & falsification of prior uncertainty using data





Monte Carlo & falsification of prior uncertainty using data

- What does falsification **not** mean: the prior is “correct”, “validated”, “verified”, “calibrated”, “inverted”, “updated”….

- **Falsification makes a hypothesis stronger if not falsified**

- **What should happen when prior uncertainty is falsified?**
  - You need to increase model complexity
  - You need to increase model uncertainty
  - or both….and you don’t know which of the three

Sensitivity analysis on both data and prediction variables

- Sensitivity for prediction variables: *learn* what in the model variables impacts most each prediction variable, allows focus on important model parameters

- Sensitivity for data variables: *learn* what model variables can be reduced in uncertainty with the observations

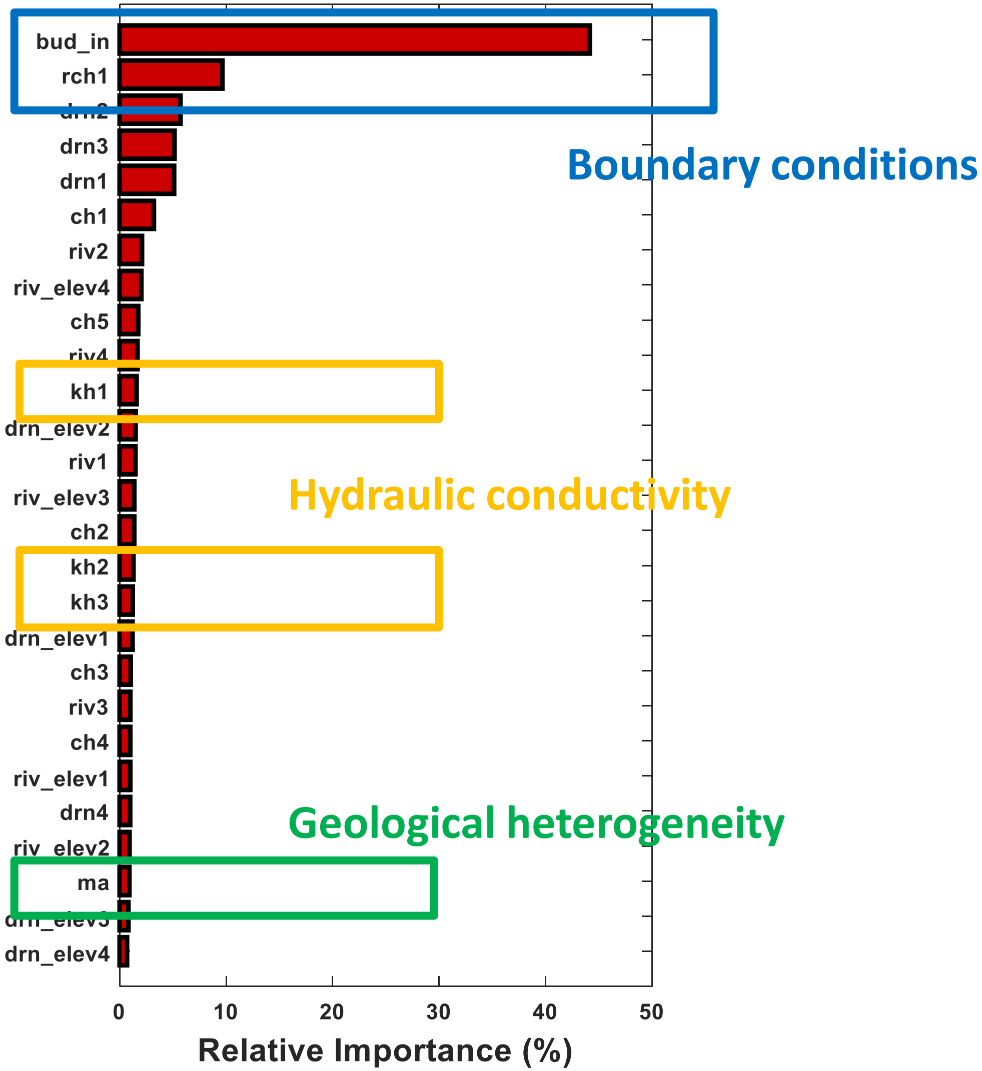
- Joint analysis of both sensitivity analyses: *learn* what method should be used for uncertainty reduction on predictions

Monte Carlo-based or Global Sensitivity Analysis

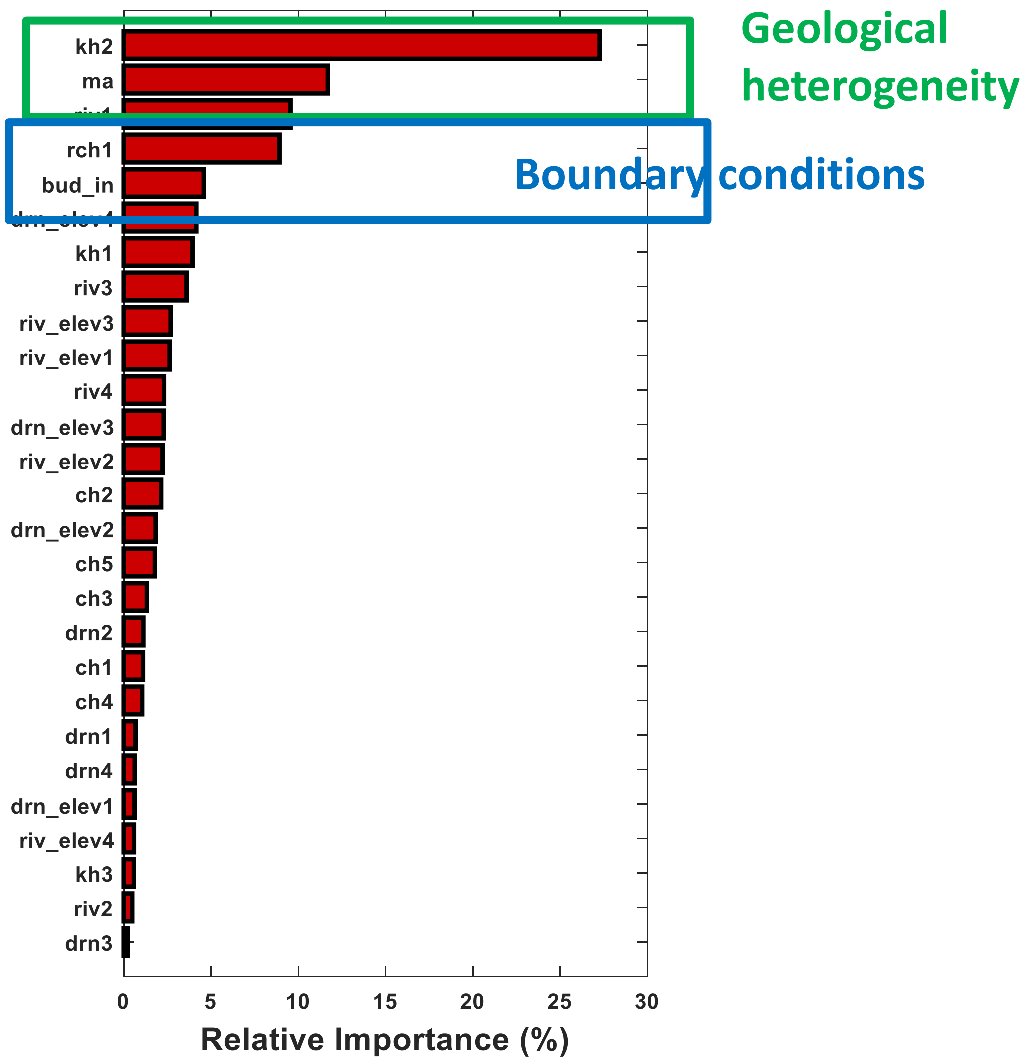
no local sensitivity analysis

Sensitivity analysis on both data and prediction variables

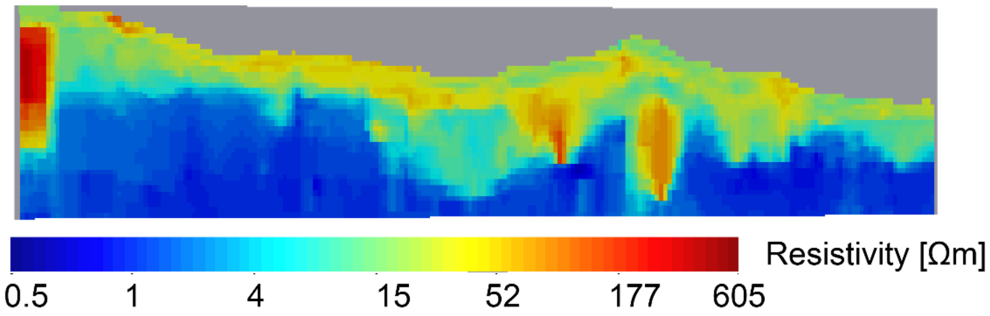
Head & streamflow data

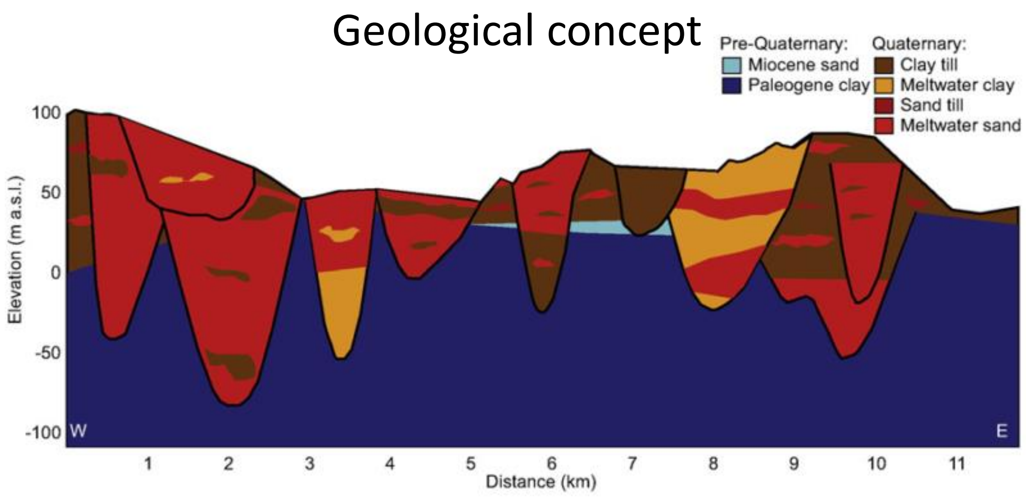


Industrial pollution prediction



Design of uncertainty reduction on prediction variables based on data





+ Head and streamflow data

Many uncertainty sources

Boundary conditions
Geological interpretations
SkyTEM data uncertainty
Spatial distribution lithologies
Streamflow models
Rock physics models
Reactive transport models
Biogeochemical models

Inversion modeling?

Design of uncertainty reduction on prediction variables based on data

Strategy

Transform a non-linear non-Gaussian problem to a linear-Gaussian problem

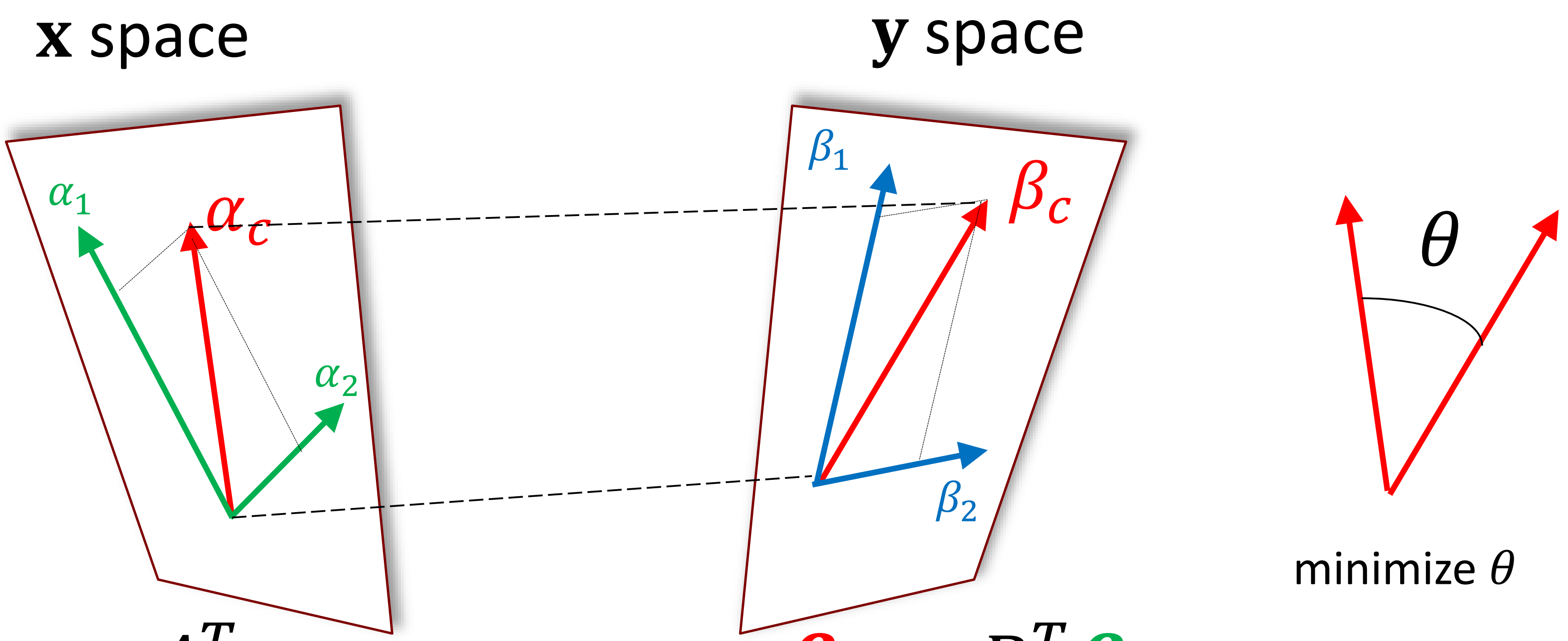
Predict with the linear-Gaussian problem

Back-transform the results into the original problem

No traditional inverse modeling, only statistical calculations from Monte Carlo

Design of uncertainty reduction on prediction variables based on data

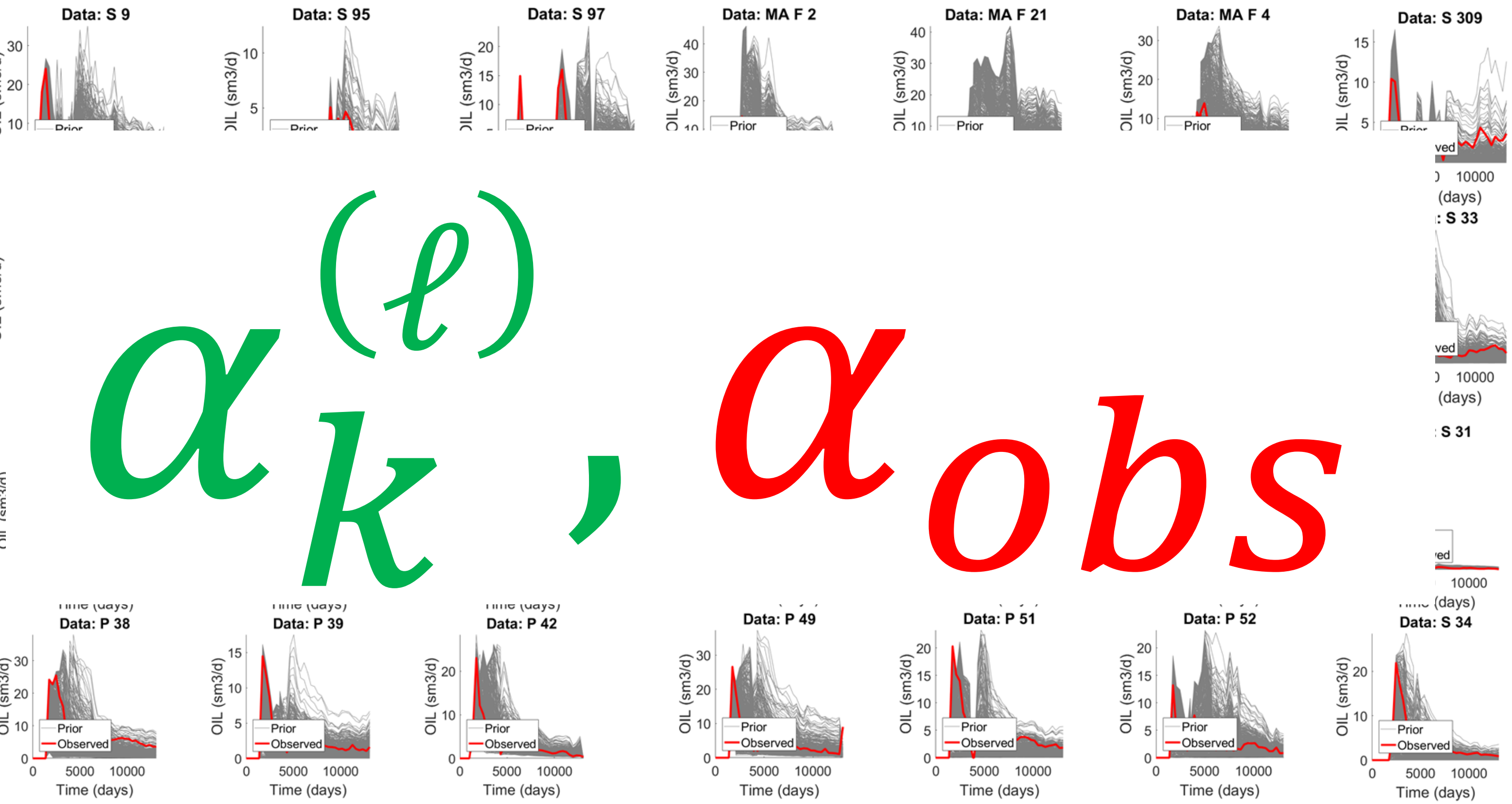
$$\mathbf{x} = \sum_{k=1}^{K} \alpha_k \boldsymbol{\varphi}_k \qquad \mathbf{x}, \mathbf{y} = \mathbf{d}, \mathbf{h} \text{ or } \mathbf{m} \qquad \mathbf{y} = \sum_{k=1}^{K'} \beta_k \boldsymbol{\psi}_k$$



$$\boldsymbol{\alpha}_c = A^T \boldsymbol{\alpha} \qquad \boldsymbol{\beta}_c = B^T \boldsymbol{\beta}$$

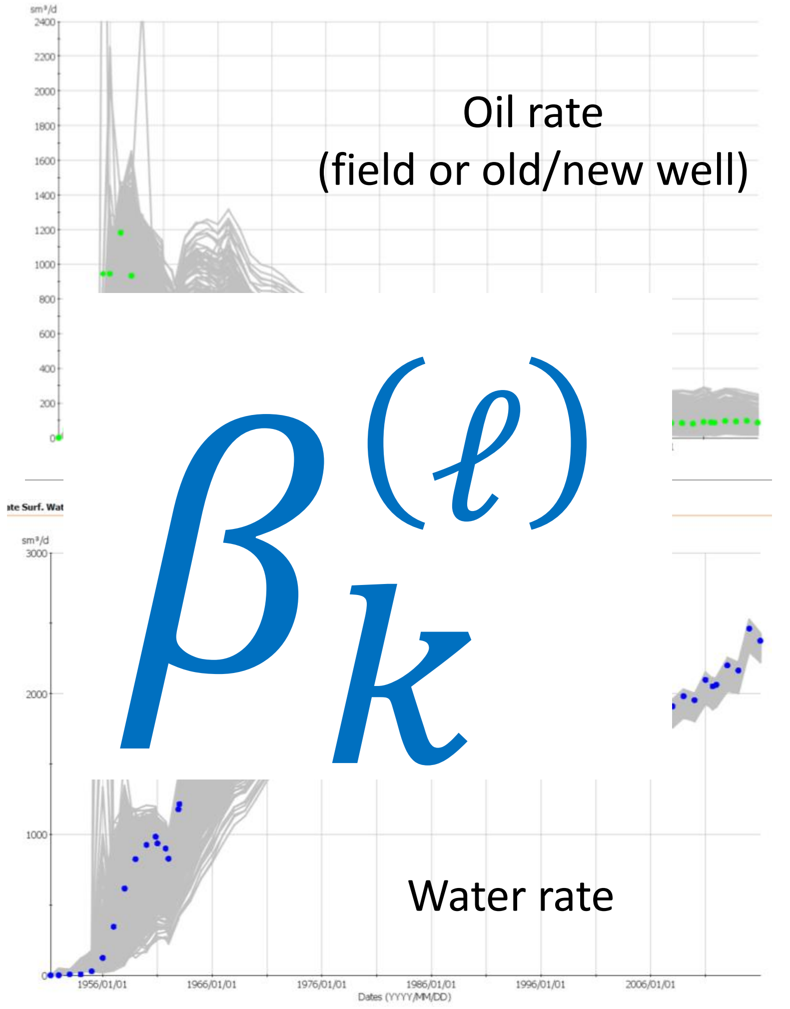
Design of uncertainty reduction on prediction variables based on data

$\mathbf{d}^{(\ell)}, \ell = 1, \dots L$ $\quad$ $\mathbf{d}_{obs}$



$$\alpha_k^{(\ell)}, \alpha_{obs}$$

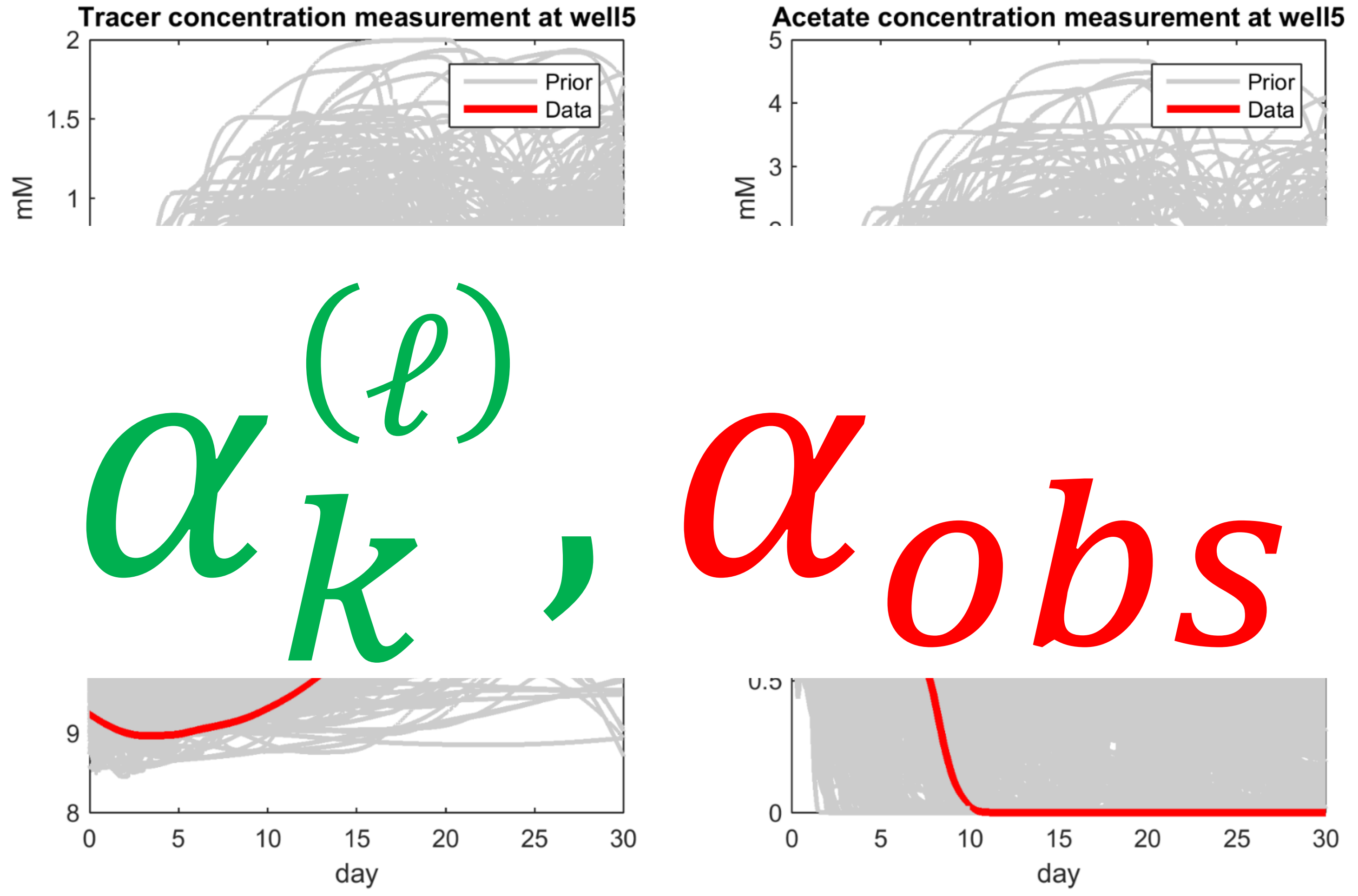
$\mathbf{h}^{(\ell)}, \ell = 1, \dots L$



$$\beta_k^{(\ell)}$$

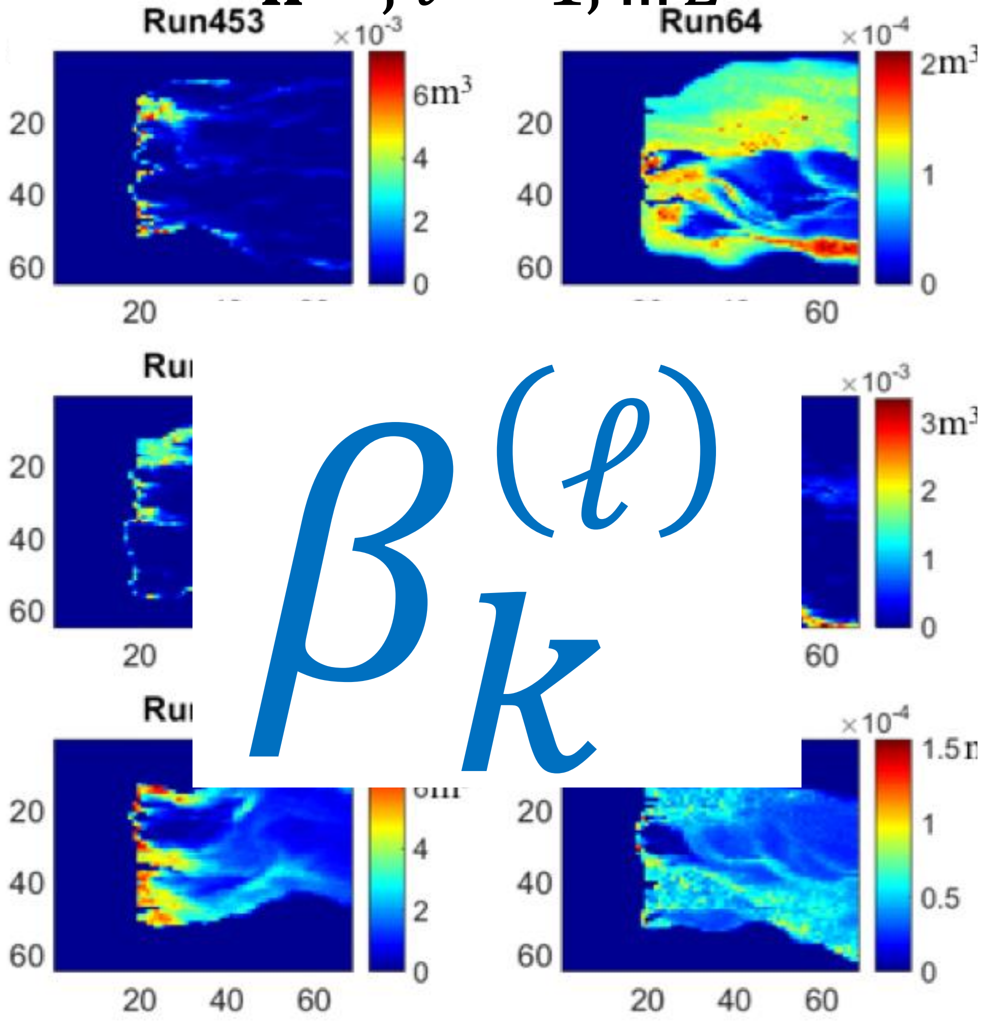
Design of uncertainty reduction on prediction variables based on data

$\mathbf{d}^{(\ell)}, \ell = 1, \dots L$ $\quad$ $\mathbf{d}_{obs}$



$\alpha_k^{(\ell)}, \alpha_{obs}$

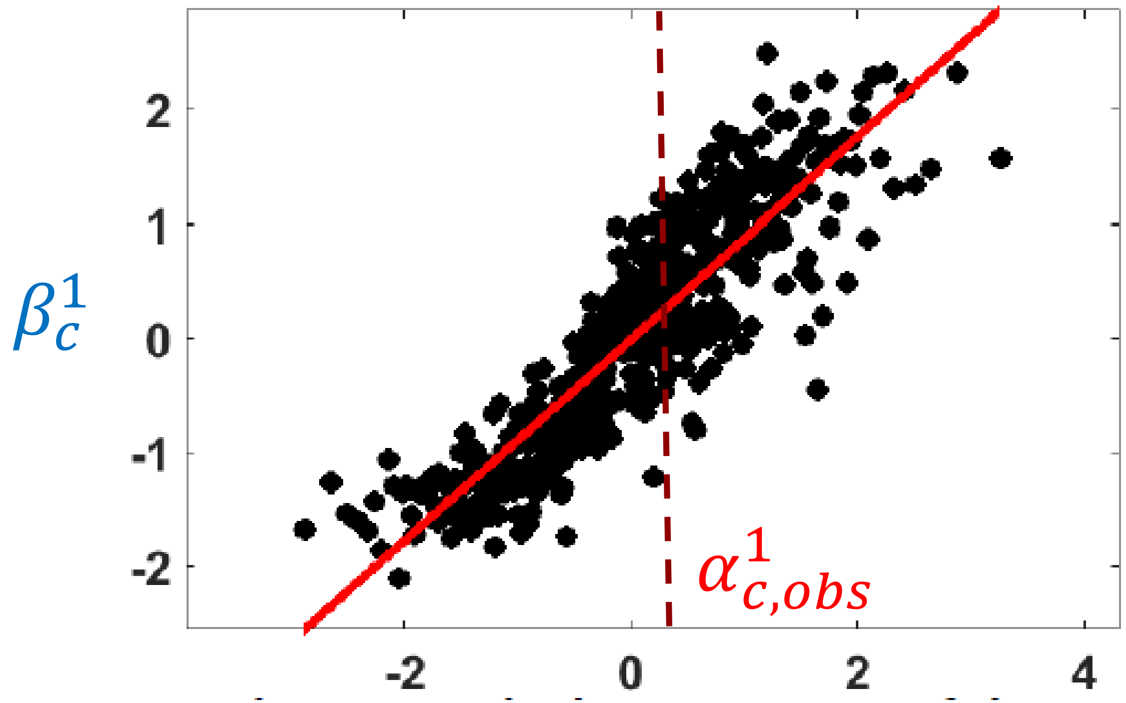
$\mathbf{h}^{(\ell)}, \ell = 1, \dots L$



$\beta_k^{(\ell)}$

Design of uncertainty reduction on prediction variables based on data

Linear-Gauss problem



α_c^1

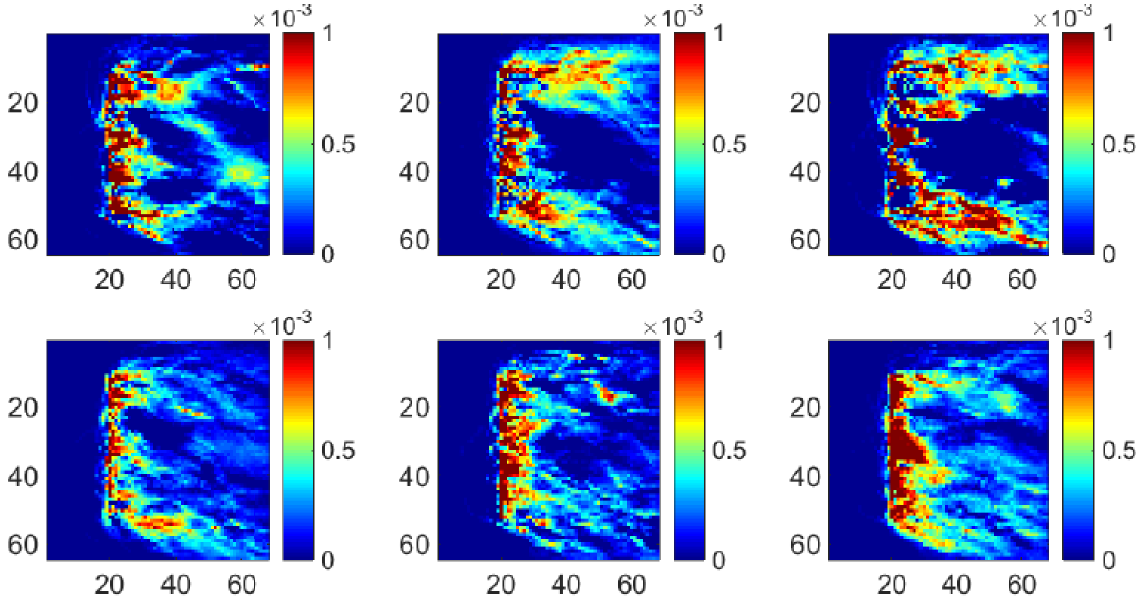
The tracer data of all species in all wells

The spatial distribution of uranium oxide

$$B^T \boldsymbol{\beta}_c = \boldsymbol{\beta}$$

Back-transformation

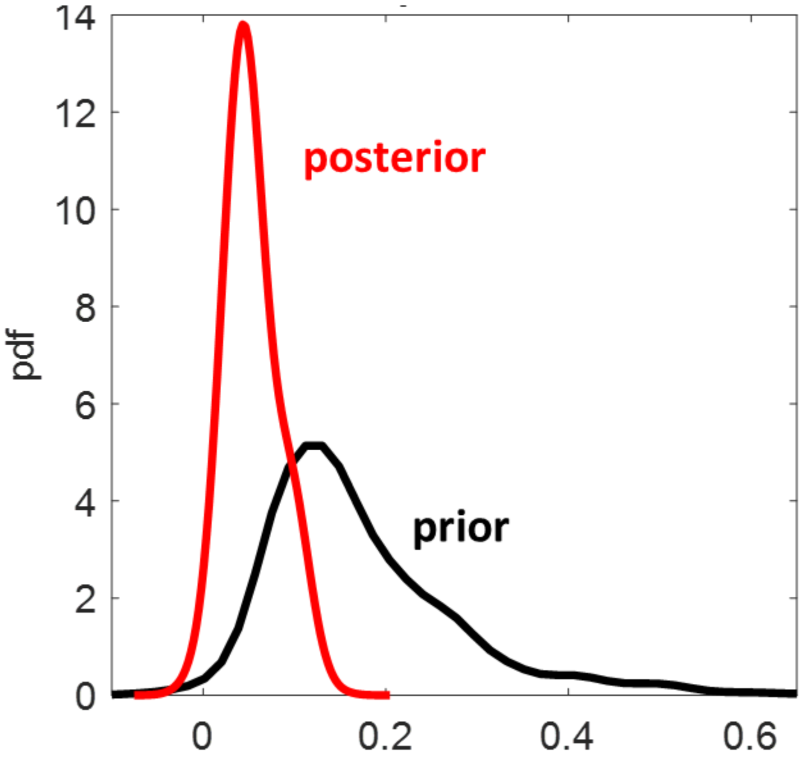
$$\mathbf{h} = \sum_{k=1}^{K'} \beta_k \boldsymbol{\psi}_k$$



Immobilized uranium uncertainty

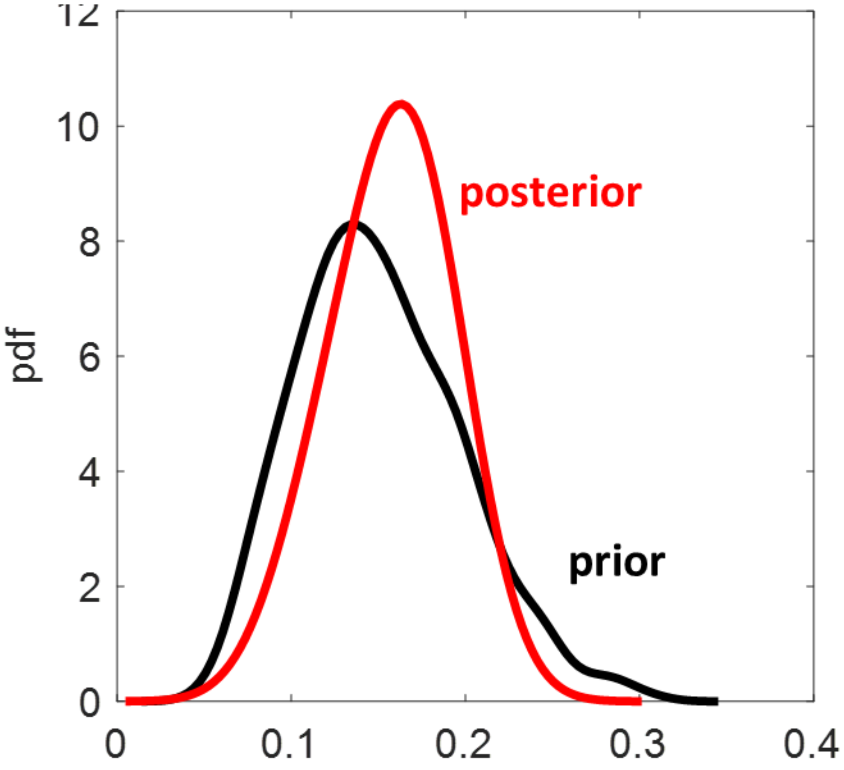
Design of uncertainty reduction on prediction variables based on data

Location A; SFRP



update

Location C; INDU



No update

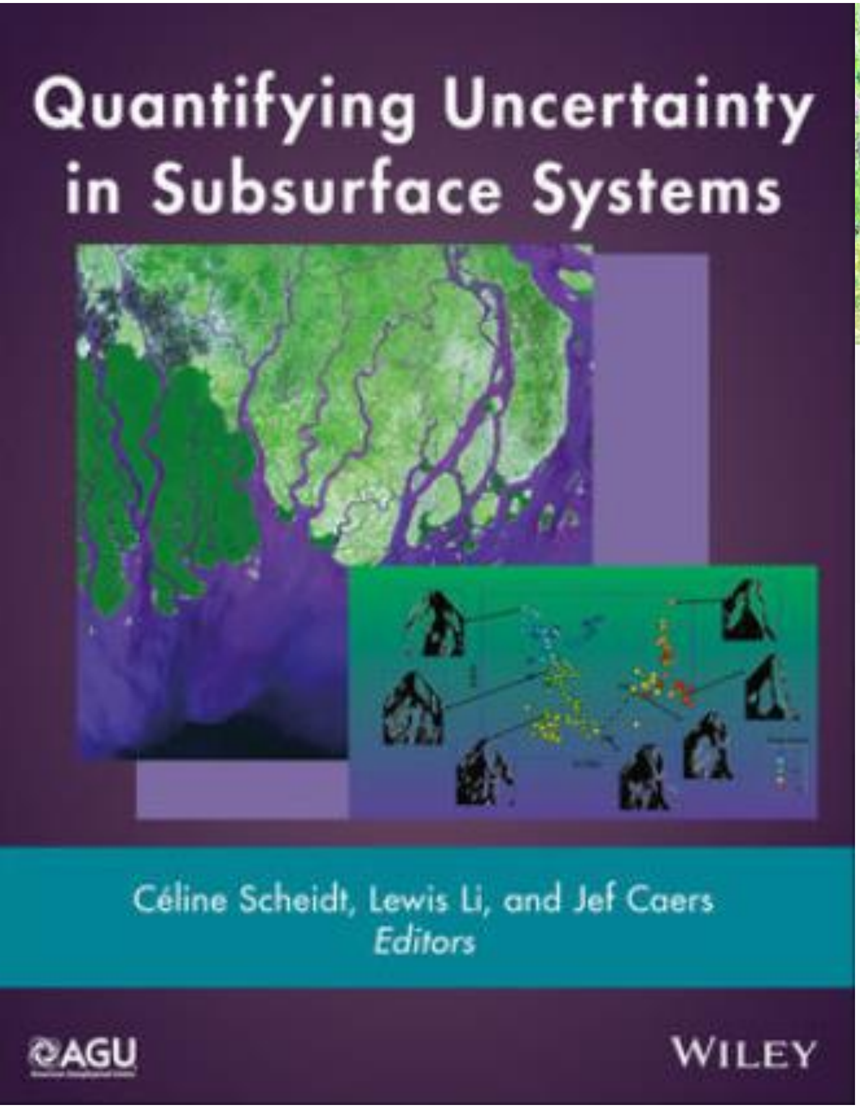
Decision making; Posterior falsification & sensitivity

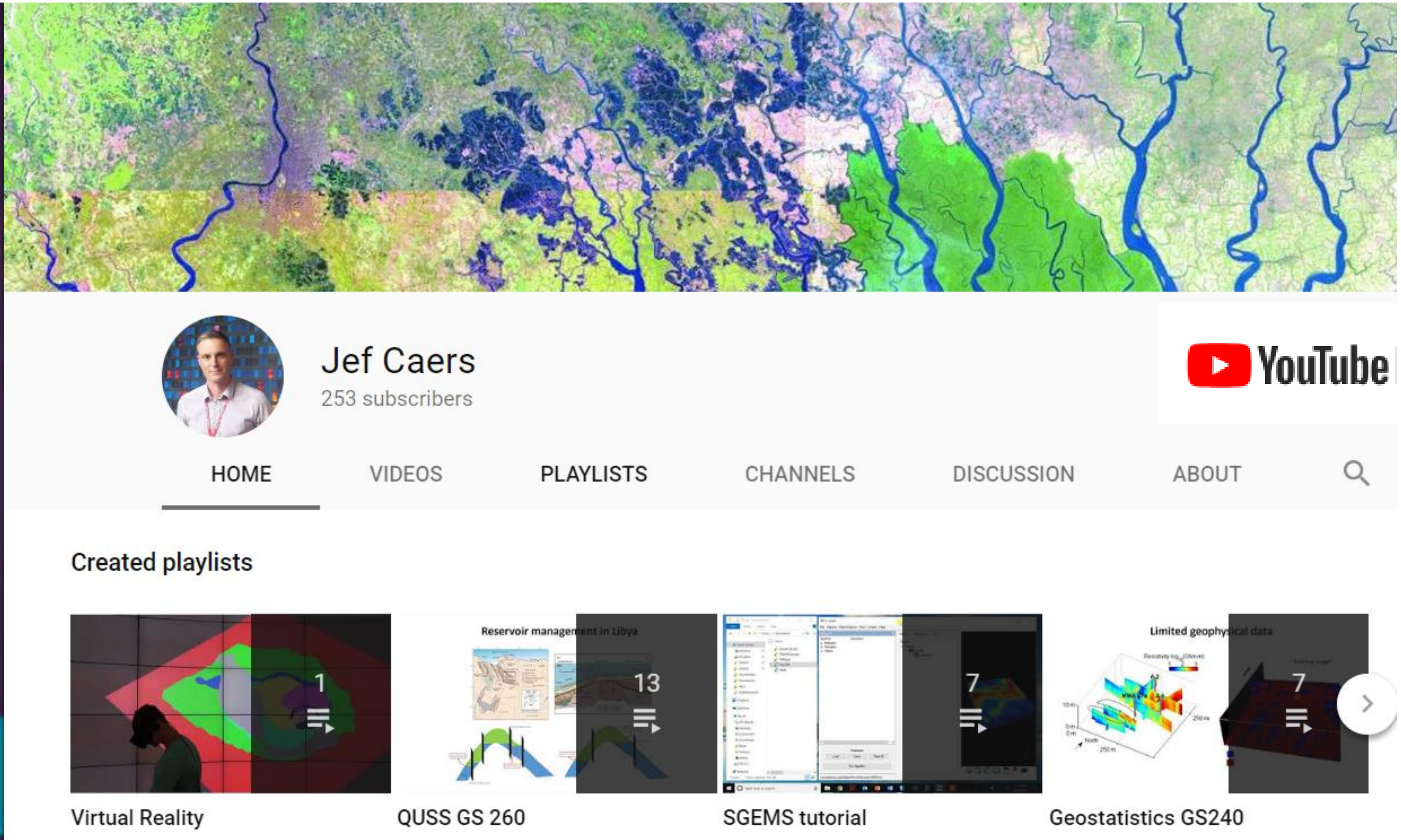
RISK AVERSE DECISION MAKER

Objectives	rank	weight	Loc A	Loc B	Loc C	Loc D	type
Farming pollution Farm (units)	5	0.067	0.00	3	48	100	Risk / $ cost
Industry pollution Indus (units)	2	0.267	84	100	35	0	Risk / $ cost
Streamflow restoration SFRP (units)	1	0.333	62	0	22	100	Return / $ benefit
Wetland restoration WLR (units)	3	0.200	0	70	71	100	Return / $ benefit
Drawdown WDD (units)	4	0.133	100	0	87	57	Risk / $ cost
		Total:	56.7	40.8	42.6	61.0	
		Return-benefit score	20.74	14.08	21.66	53.33	
		Risk-cost score	35.92	26.85	24.16	14.30	

Table accounts for objective (swing) weighting, preferences through value functions

Reference material





Software

scerf.stanford.edu

Stanford
SCHOOL OF EARTH, ENERGY & ENVIRONMENTAL SCIENCES

Stanford Center for Earth Resources Forecasting

Home People Events Research Resources Education Affiliates Contact Us

https://github.com/SCRFpublic/QUSS

Quantifying Uncertainty in Subsystem Systems

This repository contains the companion code repository for Quantifying Uncertainty in Subsurface Systems by Céline Scheidt, Lewis Li, and Jef Caers (John Wiley & Sons).

About This Repository

This repository implements the various UQ strategies discussed in the book. The source code for the algorithms can be found under the src folder. The codebase was developed in MATLAB

For illustrative purposes, a set of Jupyter tutorials have been prepared. They are as follows:

1. Dimension Reduction: Showcase of various dimension reduction techniques discussed in Chapter 3.
2. DGSA: Implementation of Distance Based Sensitivty Analysis from Chapter 4.
3. Bayesian Evidential Learning. Methodology discussed in Chapter 7, implemented for the Libyan Oil Reservoir case.
4. SIR: The Sequential Importance Resampling methodology from Chapter 7 applied to the same Libyan Oil Reservoir.

These tutorials can be viewed directly in the browser, or download and re-run.

Six stages of Decision Making, UQ with BEL

Formulating the decision question and statement of prediction variables

Statement of model parameterization and prior uncertainty

Monte Carlo & falsification of prior uncertainty using data

Sensitivity analysis on both data and prediction variables

Design of uncertainty reduction on prediction variables based on data

posterior falsification & sensitivity, decision making