Stochastic inverse methods for near-surface geophysical problems

James Irving
University of Lausanne

UNIL | Université de Lausanne
Faculté des géosciences
et de l'environnement

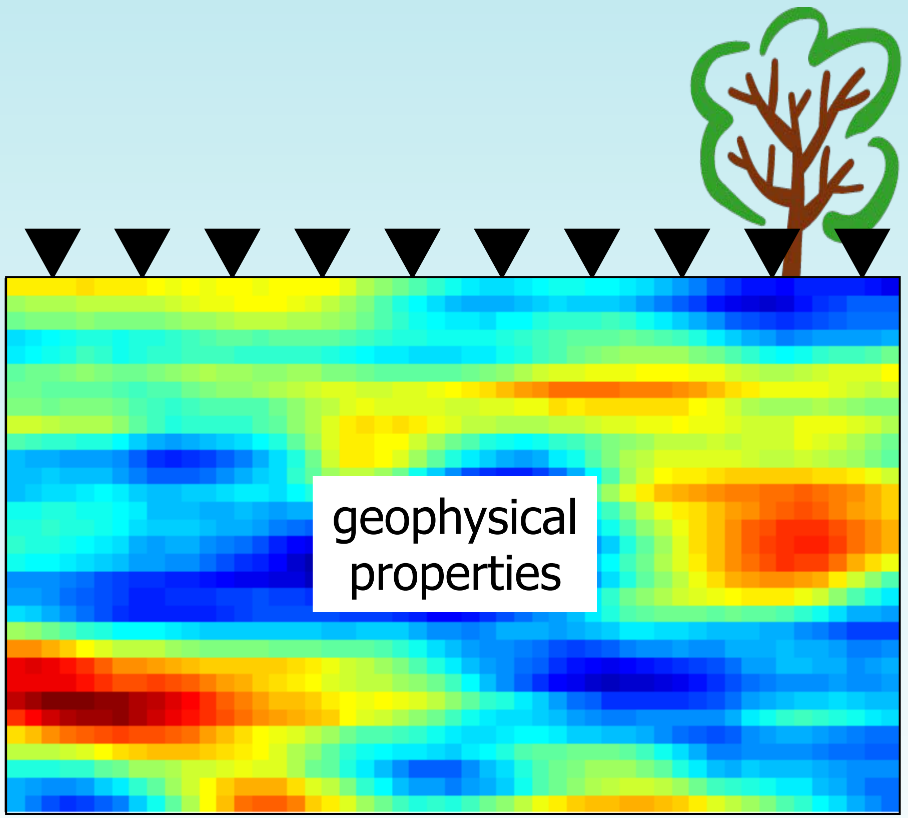
Outline

- Introduction to the domain of near-surface applied and environmental geophysics
- Motivation for uncertainty quantification
- Bayesian-MCMC inversion approach
  - extended Metropolis algorithm
- Two example applications and corresponding challenges
  - Use of crosshole ground-penetrating radar data to estimate unsaturated hydraulic parameters
  - Joint use of crosshole geoelectrical and tracer-test measurements to estimate the spatial configuration of saturated hydraulic conductivity
- Future needs and research directions

My hopes with this presentation

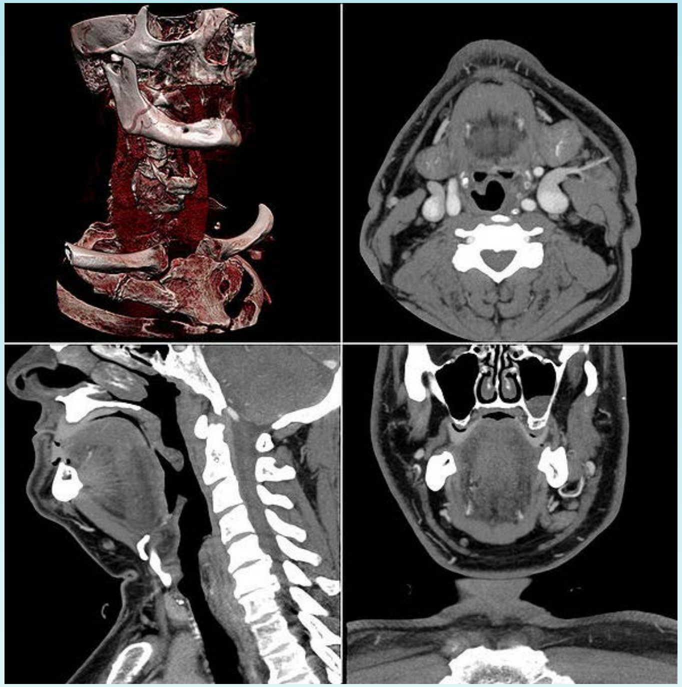
- Introduce a new research domain (for many/most of you) where there exists a strong need for robust uncertainty quantification and related research.
- Highlight through the two examples some key aspects of geophysical inverse problems that make UQ particularly challenging, which may have parallel challenges/solutions in your various fields of research.
- Summarize what I believe to be some of the critical issues that must be addressed in the future with regard to UQ in near-surface environmental geophysical work.
- Generate some discussion, ideas, and potential collaborative opportunities.

What is geophysics?

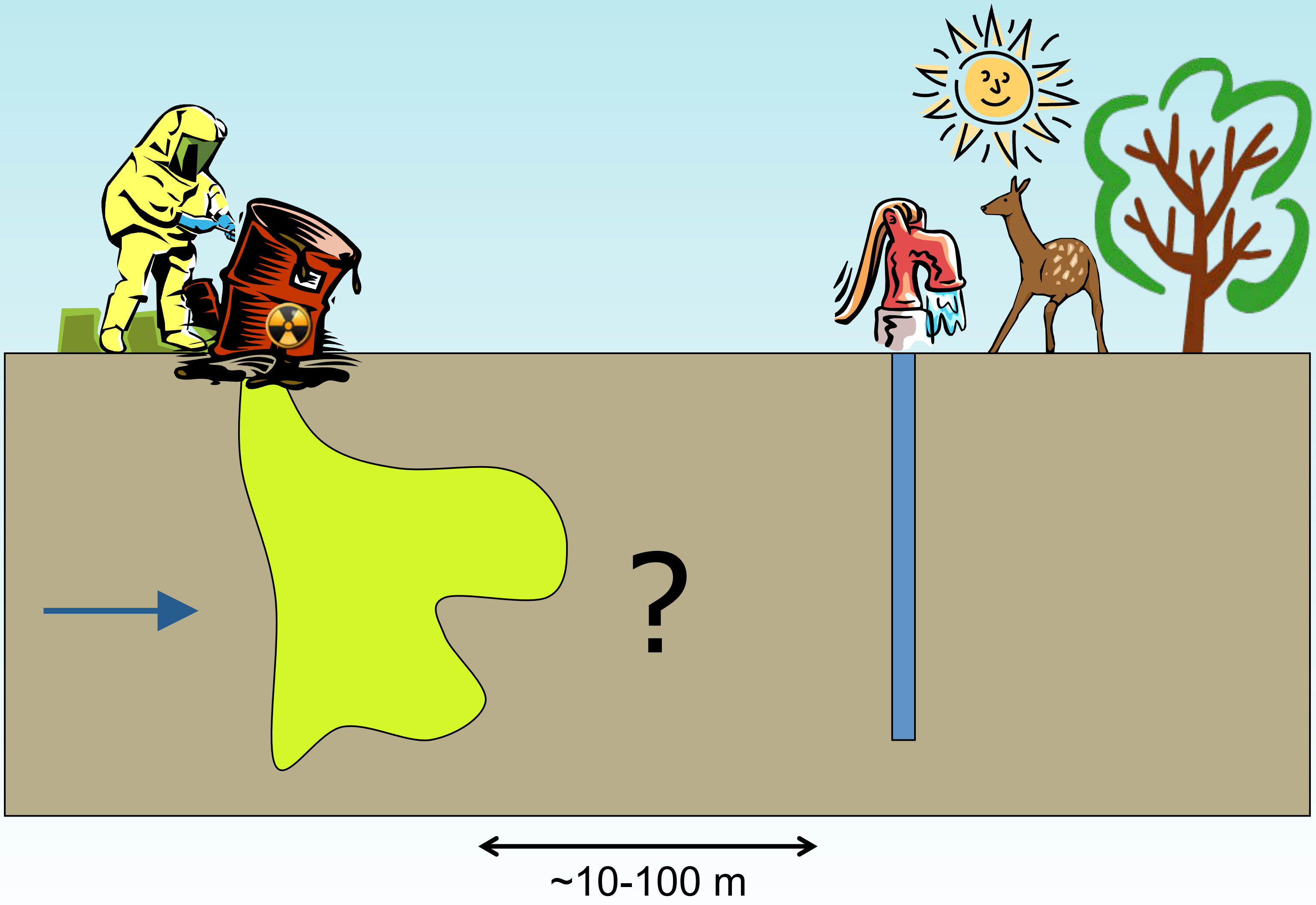


Geophysical Survey	**Property Inferred**
seismic	acoustic and/or elastic wave velocity and attenuation
ground-penetrating radar	electromagnetic wave velocity and attenuation
gravity	density
geoelectrical	electrical resistivity
magnetic	magnetic susceptibility
electromagnetic induction	electrical conductivity
induced polarization	“chargeability”

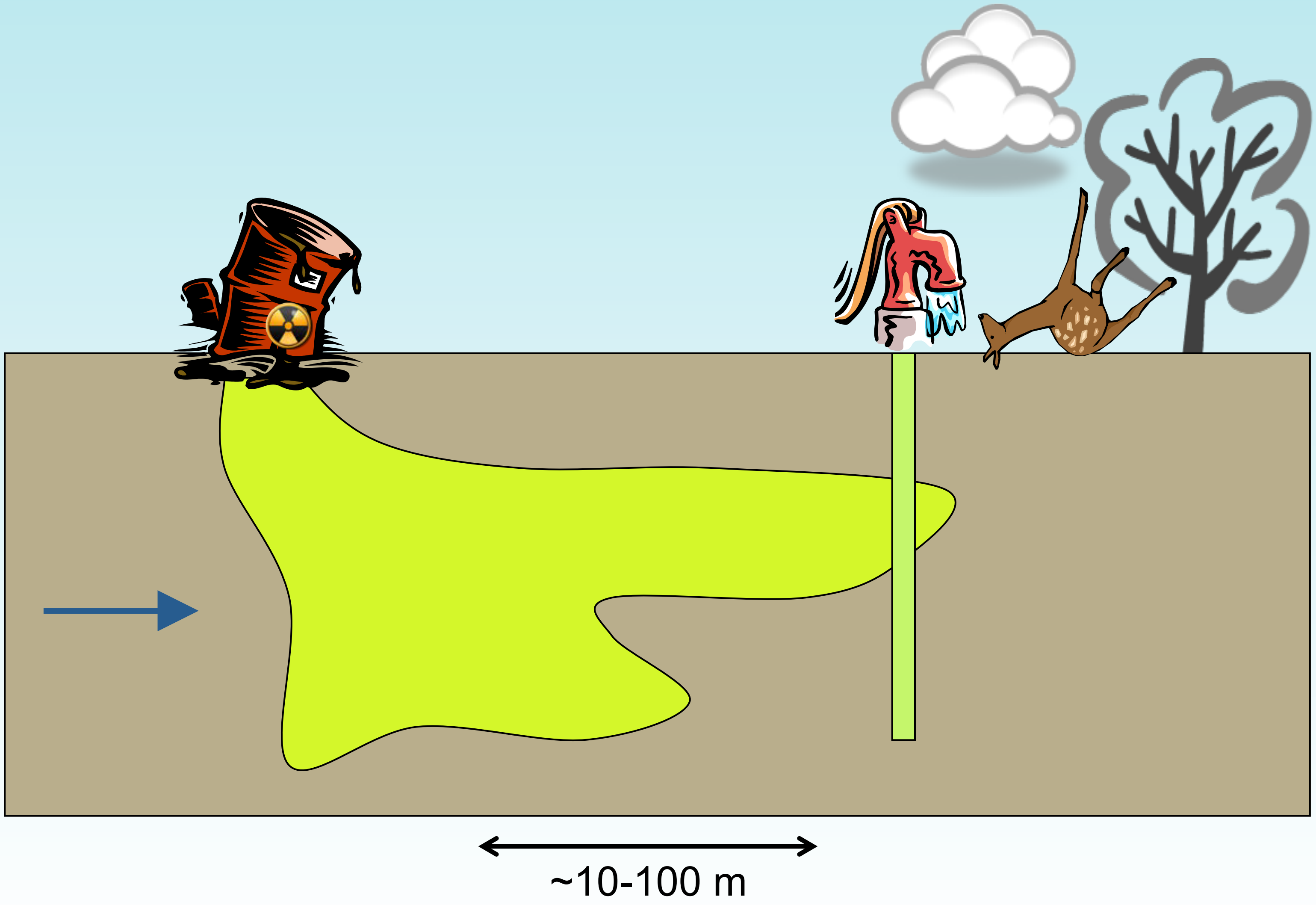
Analogous to medical imaging



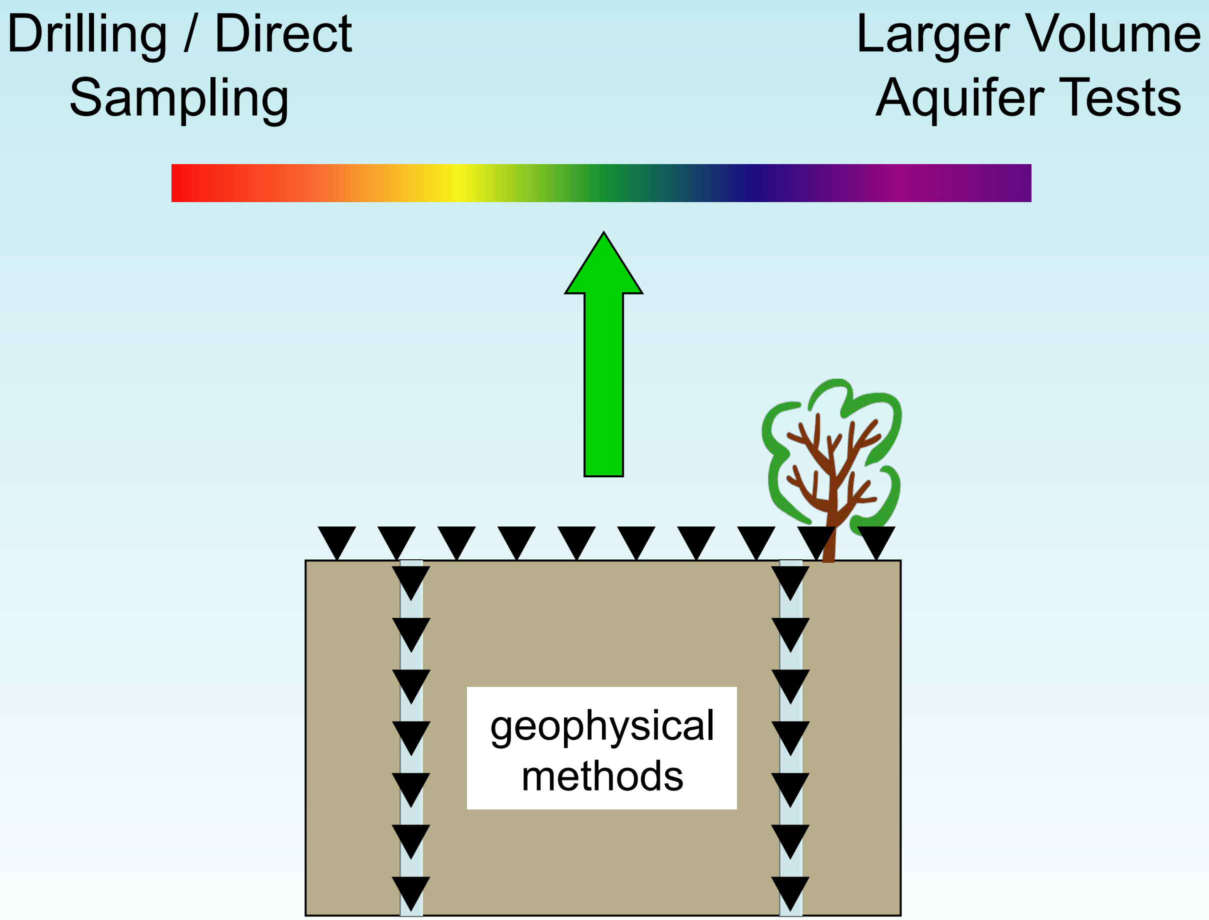
Some motivation for near-surface applied environmental geophysics



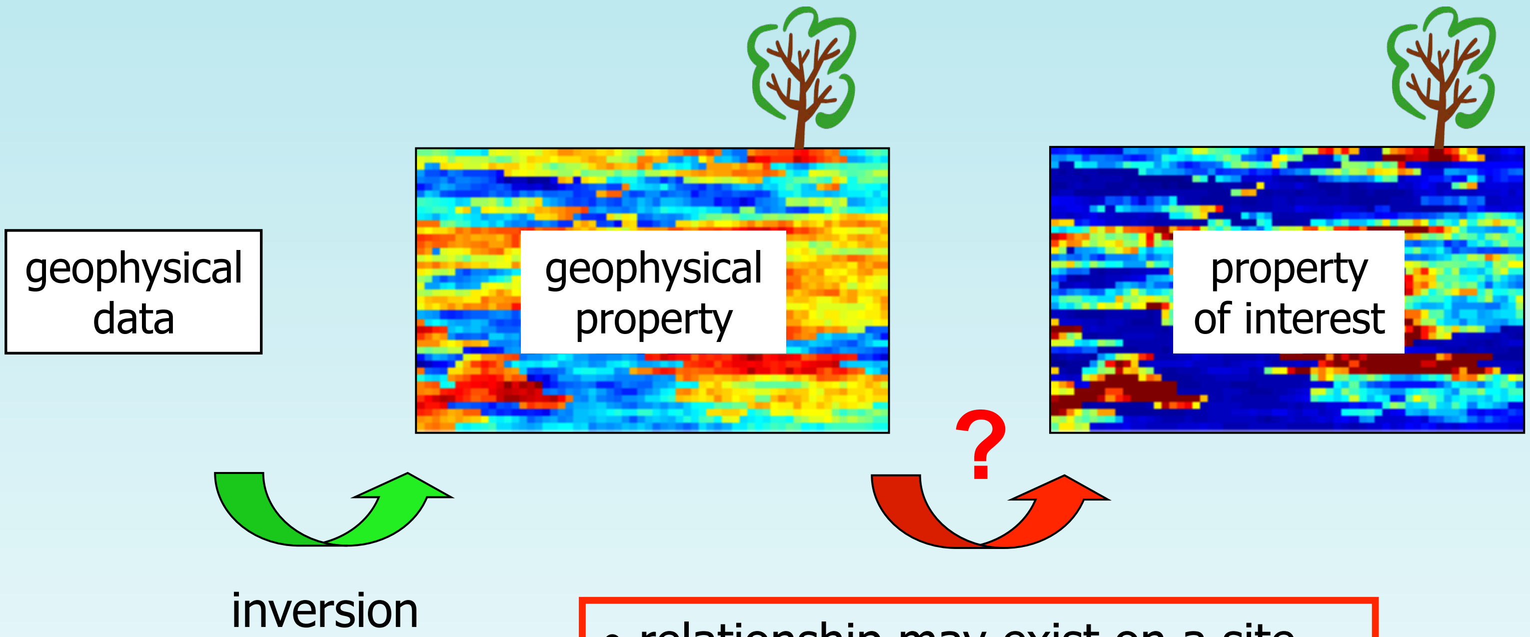
Some motivation for near-surface applied environmental geophysics



Geophysics fills a niche in terms of resolution and subsurface coverage…



What's the catch?



- relationship may exist on a site, scale, and/or facies specific basis

HOWEVER...

- often complex and non-unique
- can be difficult to establish

Why do I care about uncertainty quantification in my work?

- essential for making meaningful hydrological and environmental predictions
  - risk assessment
  - decision making
  - groundwater management and remediation
- allows us to evaluate worth of data
  - how much can geophysics help?
  - how should we best carry out our measurements?
- provides a framework for data assimilation
  - many diverse sources of information available
  - how can we bring them together and build a consistent and useful representation of the subsurface?

Challenges we face in UQ for near-surface geophysical inverse problems

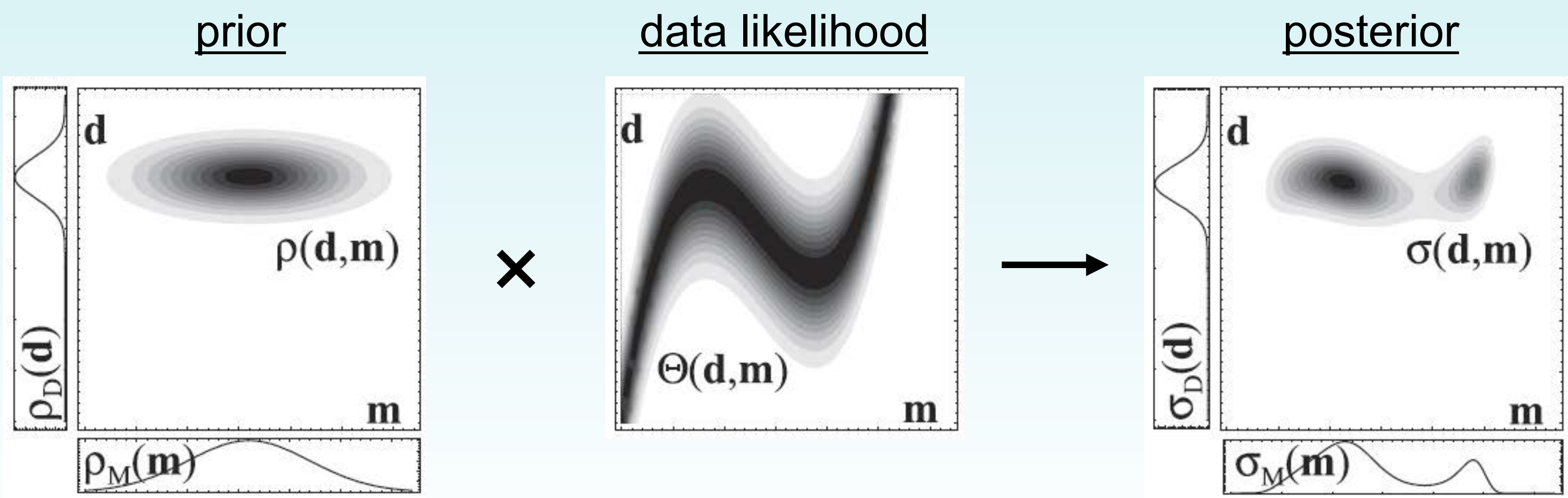
- high-dimensionality
  - unknowns are usually spatially distributed subsurface property fields (1D, 2D, 3D)
- long forward simulator computation times
- highly non-linear and non-unique relationships between measured data and model parameters
  - common in many geophysical problems
  - coupling geophysical simulators to other process-based models (e.g., flow and transport) only makes it worse
- must respect complex geological prior knowledge
- wide variety of data to be considered / integrated
  - e.g., borehole measurements, multiple geophysical and/or hydrological and other data sets, outcrop studies, etc.

Bayesian approach

- most common means of UQ in geophysical studies

$$\sigma(\mathbf{m}) = k\, L(\mathbf{m})\rho(\mathbf{m})$$

- intuitive framework for stochastic parameter estimation
- naturally suited to the assimilation of various types of data
- <u>**Markov chain Monte Carlo sampling**</u>



(from Tarantola, 2005)

MCMC sampling from σ(m)

Metropolis-Hastings

$$P_{acc} = \min\left[1, \frac{\sigma(\mathbf{m}')\,Q(\mathbf{m}_i|\mathbf{m}')}{\sigma(\mathbf{m}_i)\,Q(\mathbf{m}'|\mathbf{m}_i)}\right]$$

Metropolis

$$Q(\mathbf{m}_i|\mathbf{m}') = Q(\mathbf{m}'|\mathbf{m}_i)$$

$$P_{acc} = \min\left[1, \frac{\sigma(\mathbf{m}')}{\sigma(\mathbf{m}_i)}\right]$$

Extended Metropolis

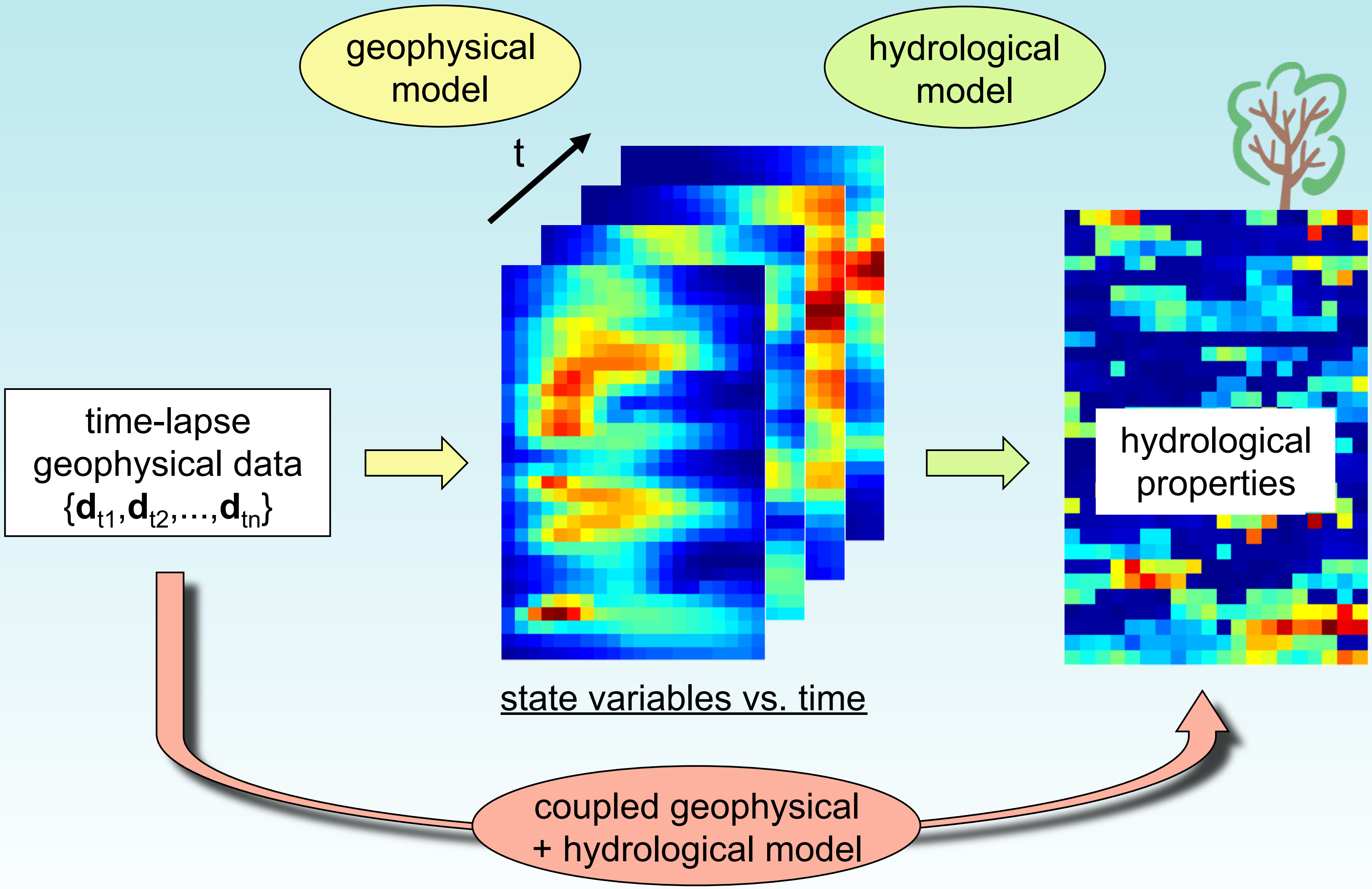
(Mosegaard and Tarantola, 1995)

$$\frac{Q(\mathbf{m}_i|\mathbf{m}')}{Q(\mathbf{m}'|\mathbf{m}_i)} = \frac{\rho(\mathbf{m}_i)}{\rho(\mathbf{m}')}$$

$$P_{acc} = \min\left[1, \frac{L(\mathbf{m}')}{L(\mathbf{m}_i)}\right]$$

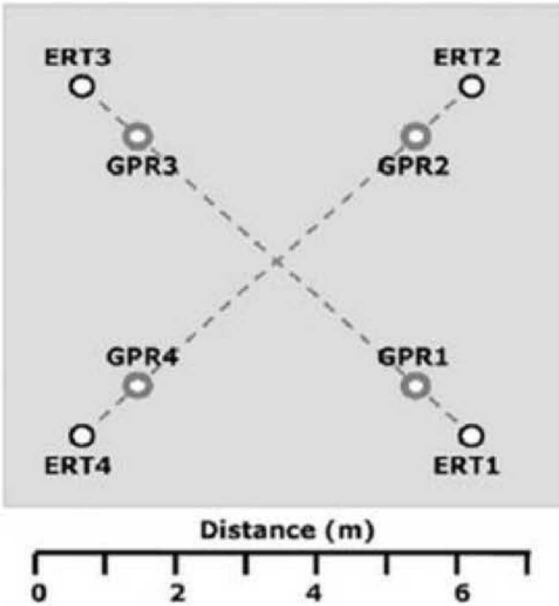
Examples

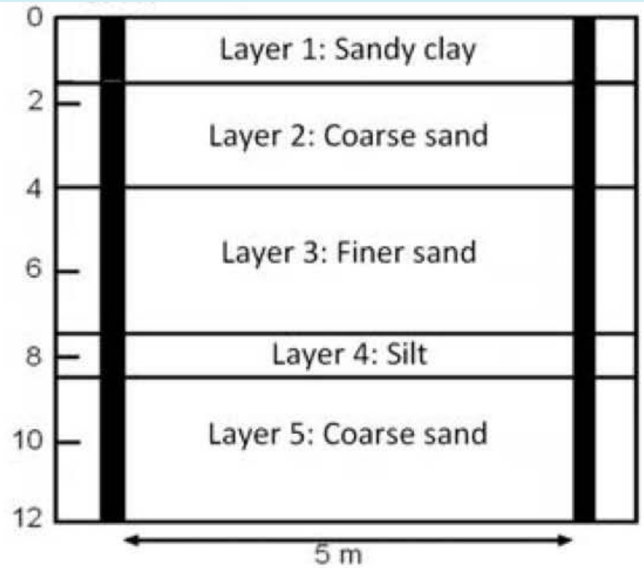
Dynamic geophysical measurements



Problem 1: Arrenaes test site

- developed to study flow and transport processes in the unsaturated zone using geophysical and hydrological methods
- water table at ~30 m depth
- 8 shallow boreholes installed and equipped for hydrological and geophysical measurements

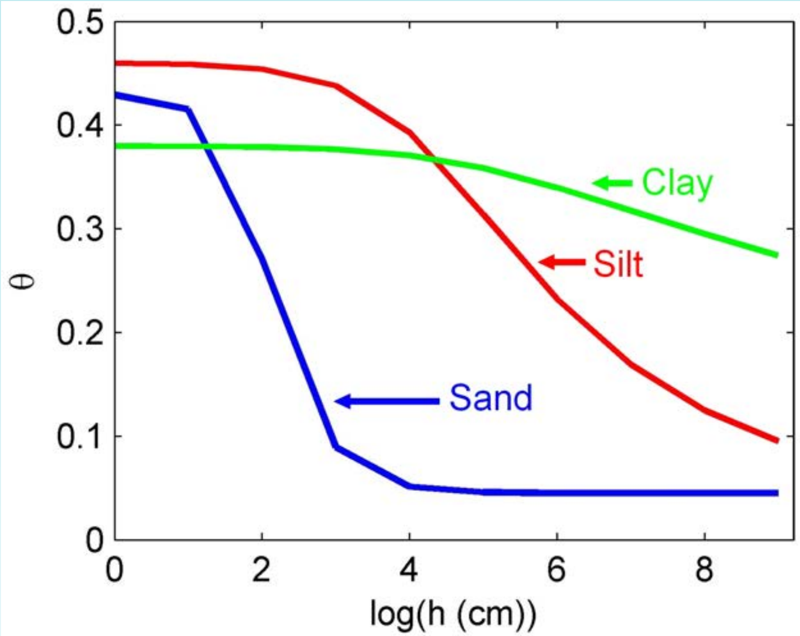




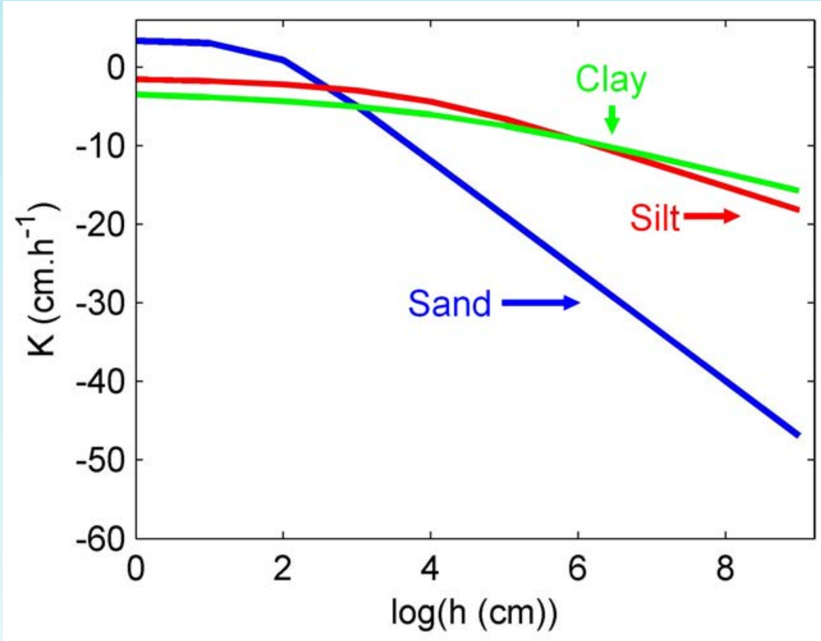
Scholer, M., Irving, J., Looms, M., Nielsen, L., and Holliger, K. Bayesian Markov-chain-Monte-Carlo inversion of time-lapse crosshole ground-penetrating radar data to characterize the vadose zone at the Arrenaes field site, Denmark. Vadose Zone Journal, 11(4), 2012.

Unsaturated hydraulic properties

Water retention curve



Unsaturated hydraulic conductivity



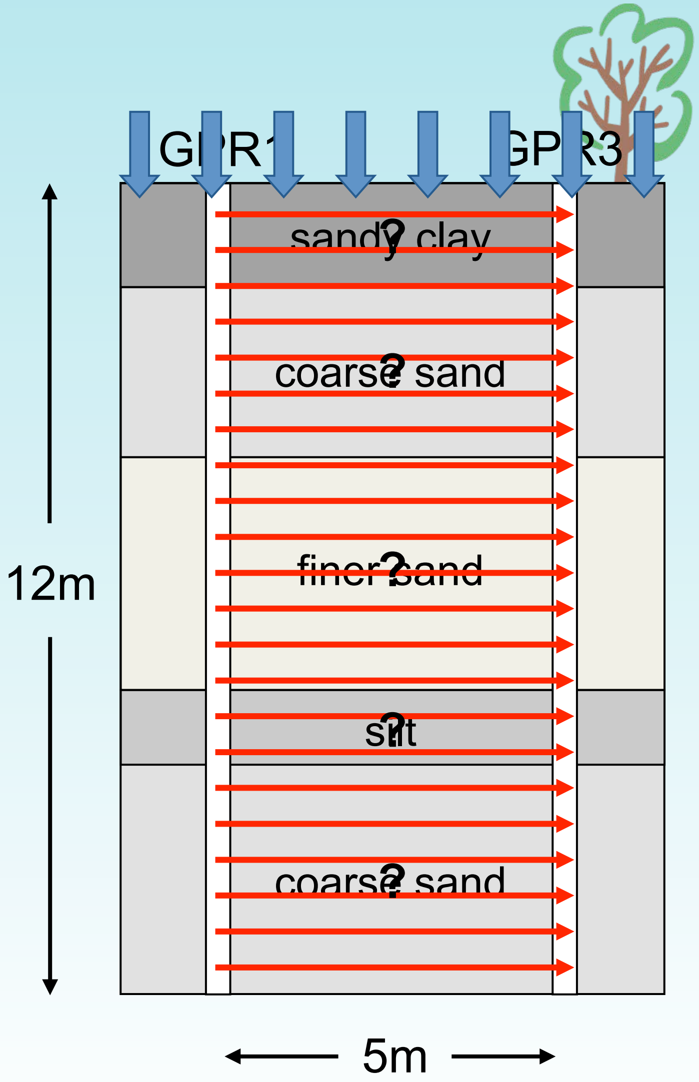
VGM model:

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{(1 + |\alpha h|^n)^m}$$

$$K(h) = K_s S_e^{\frac{1}{2}} \left[1 - (1 - S_e^{\frac{1}{m}})^m\right]^2$$

$$\text{where} \quad S_e = \frac{\theta(h) - \theta_r}{\theta_s - \theta_r} \quad \text{and} \quad m = 1 - \frac{1}{n}$$

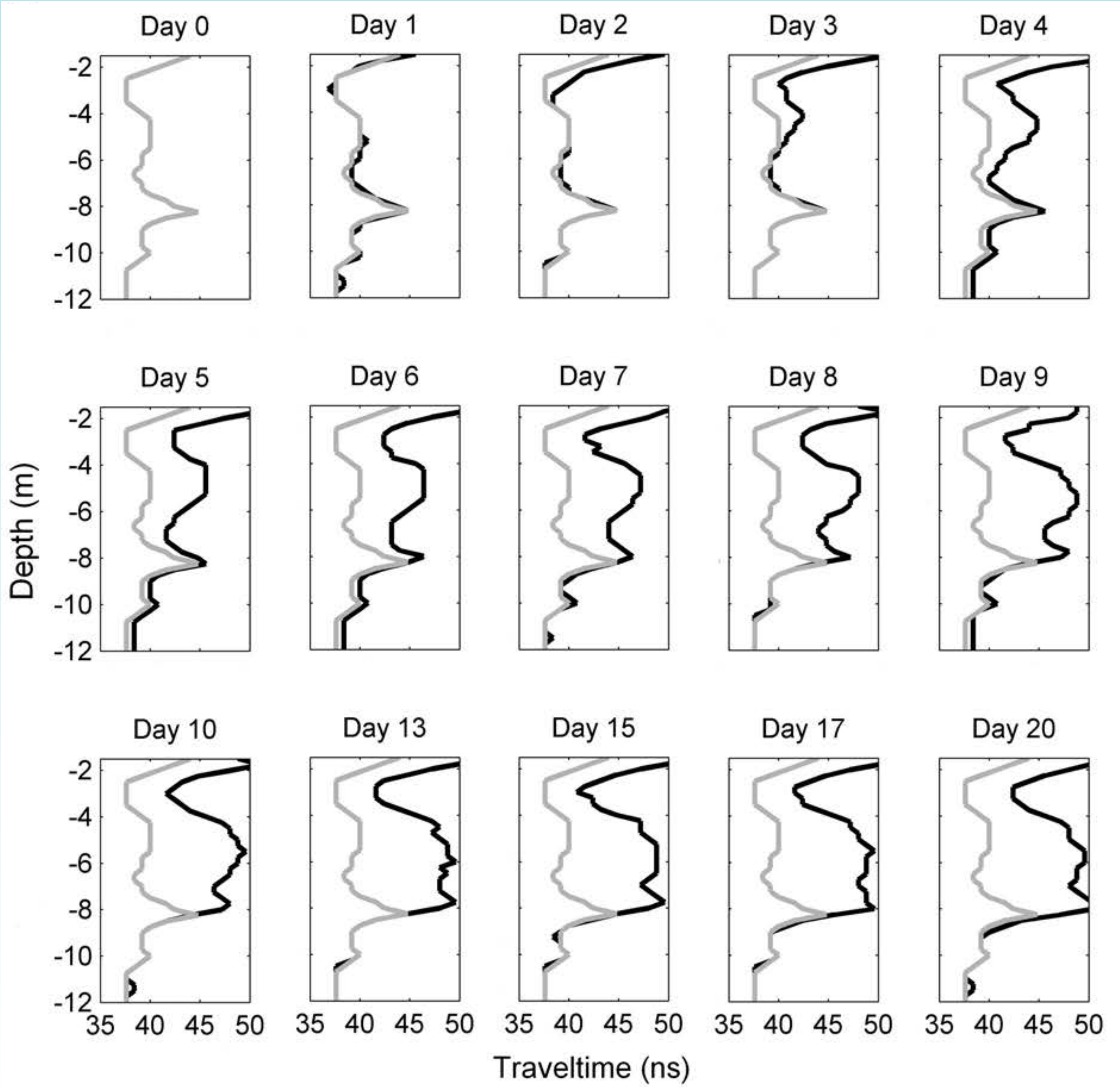
Field experiment



- forced infiltration experiment conducted by by irrigating 95,000 L of clean water over a 7 x 7 m area over 20 days (Looms et al., 2008)
- collect zero-offset-profile (ZOP) crosshole GPR traveltime data once per day until Day 10, and then on Days 13, 15, 17, and 20
- estimate VGM parameters of each subsurface layer (K_s, θ_r, α, n) from the time-lapse GPR traveltime data using MCMC posterior sampling
- i.e., characterize *in situ* field-scale soil hydraulic properties and their corresponding uncertainties

Field experiment

ZOP GPR traveltime data



Linking model parameters to data

- assume 1D vertical infiltration into the subsurface
- couple hydrological and geophysical numerical models to link soil VGM parameters to dynamic GPR traveltime measurements

1D Richards' equation: $$\frac{\partial}{\partial z}\left[K(h)\frac{\partial h}{\partial z} + K(h)\right] = \frac{\partial \theta(h)}{\partial t}$$

Topp equation: $$\epsilon_r = 3.03 + 9.3\theta + 146\theta^2 - 76.7\theta^3$$

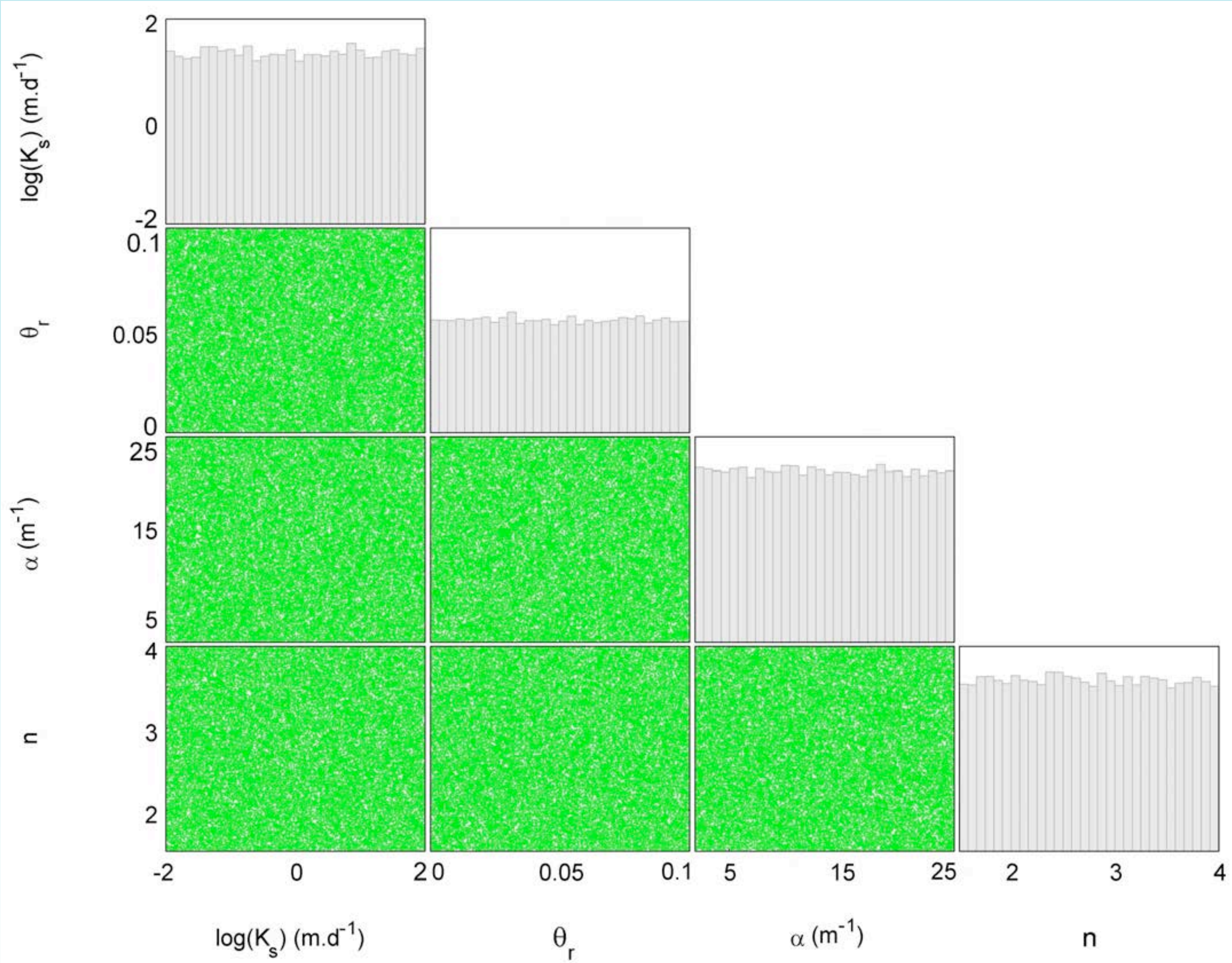
low-loss approximation: $$v = \frac{c}{\sqrt{\epsilon_r}}$$

eikonal equation: $$|\nabla T(\mathbf{r})|^2 = \frac{1}{v(\mathbf{r})^2}$$

MCMC details

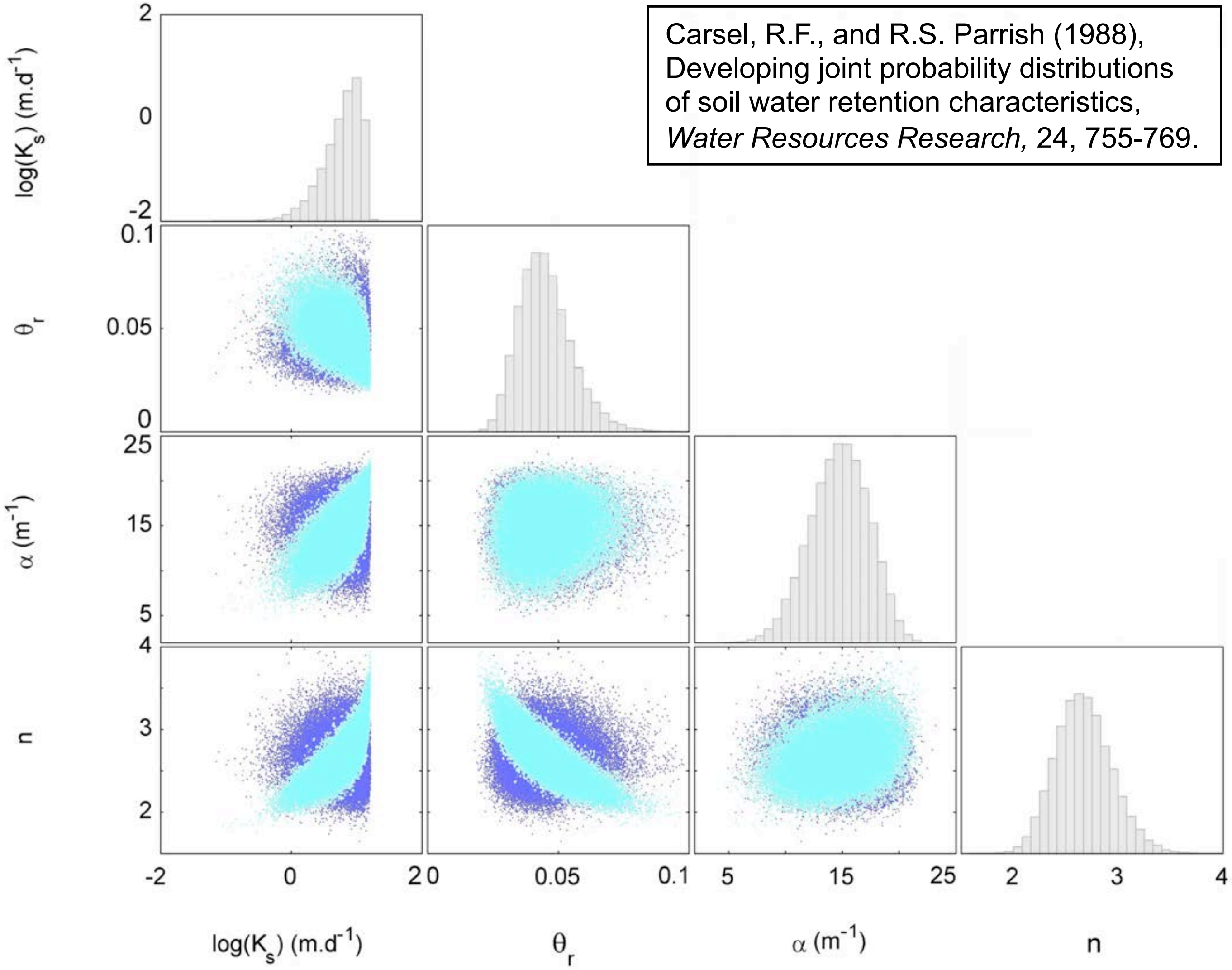
- 20-parameter problem
  - 5 layers x 4 VGM parameters in each layer
- layer interfaces assumed known from borehole data
- 3 different Bayesian prior distributions considered
  - flat priors (same in each layer)
  - refined priors based on observed soil type
    - correlated
    - uncorrelated
- assume zero-mean, identically normally distributed, uncorrelated residuals
  - corresponds to estimated errors in measured GPR traveltimes
  - results in simple (and standard) Gaussian likelihood function
- multiple parallel chains run to assess burn-in and speed the generation of independent posterior samples

Flat prior distributions



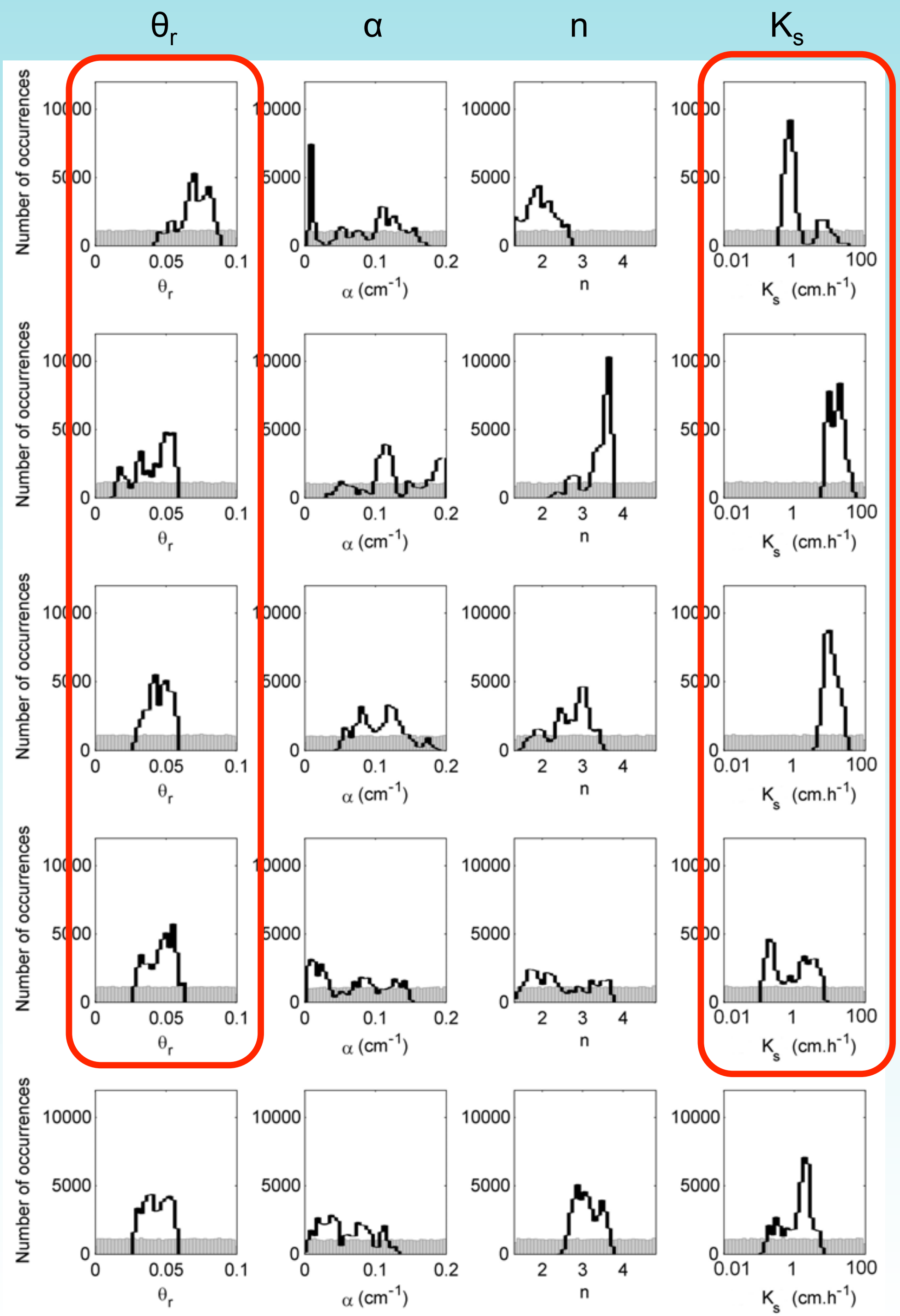
Refined prior distributions

Carsel, R.F., and R.S. Parrish (1988), Developing joint probability distributions of soil water retention characteristics, *Water Resources Research,* 24, 755-769.



Results: VGM parameters

(flat priors)



Layer 1 (sandy clay)

Layer 2 (coarse sand)

Layer 3 (fine sand)

Layer 4 (silt)

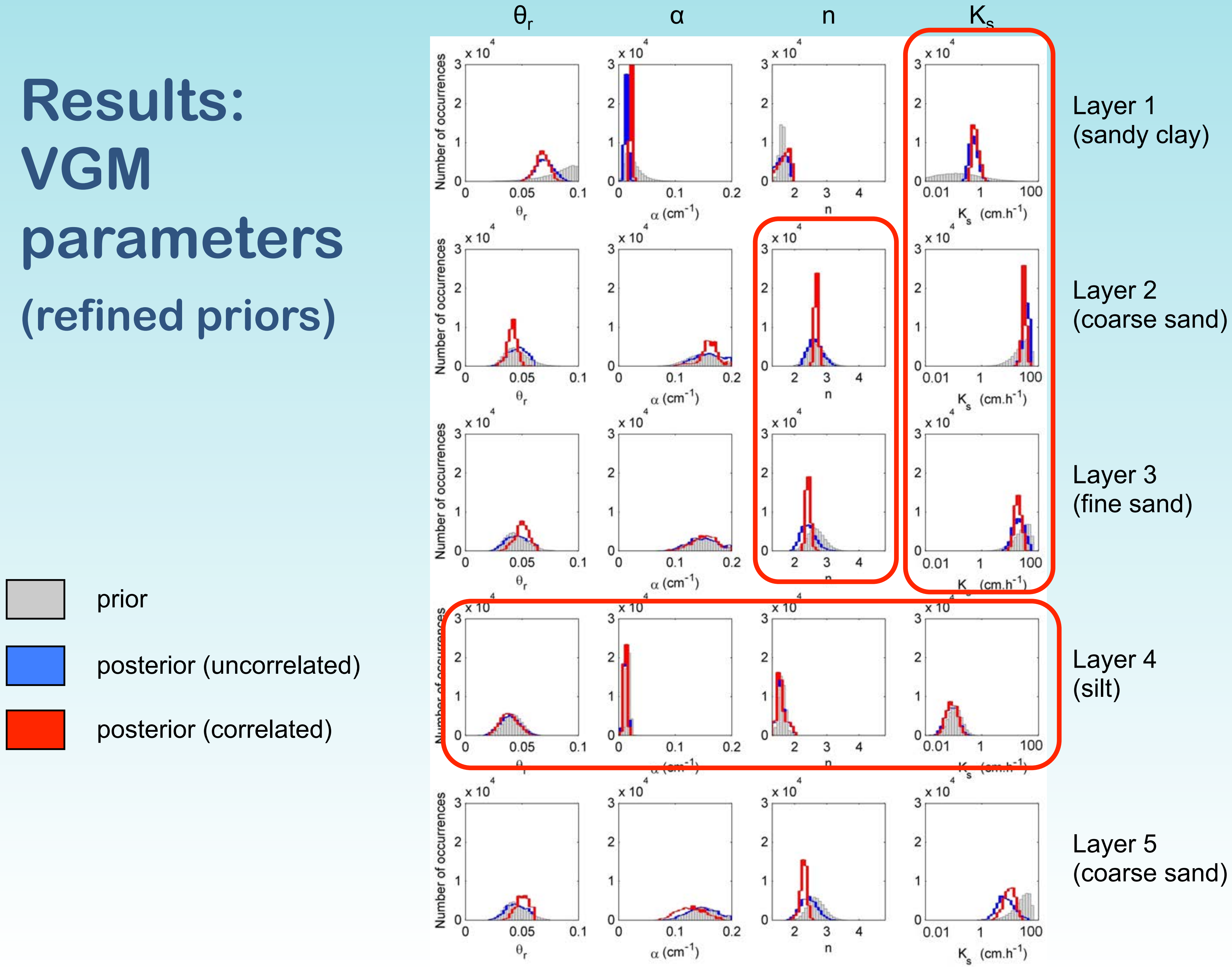
Layer 5 (coarse sand)

prior

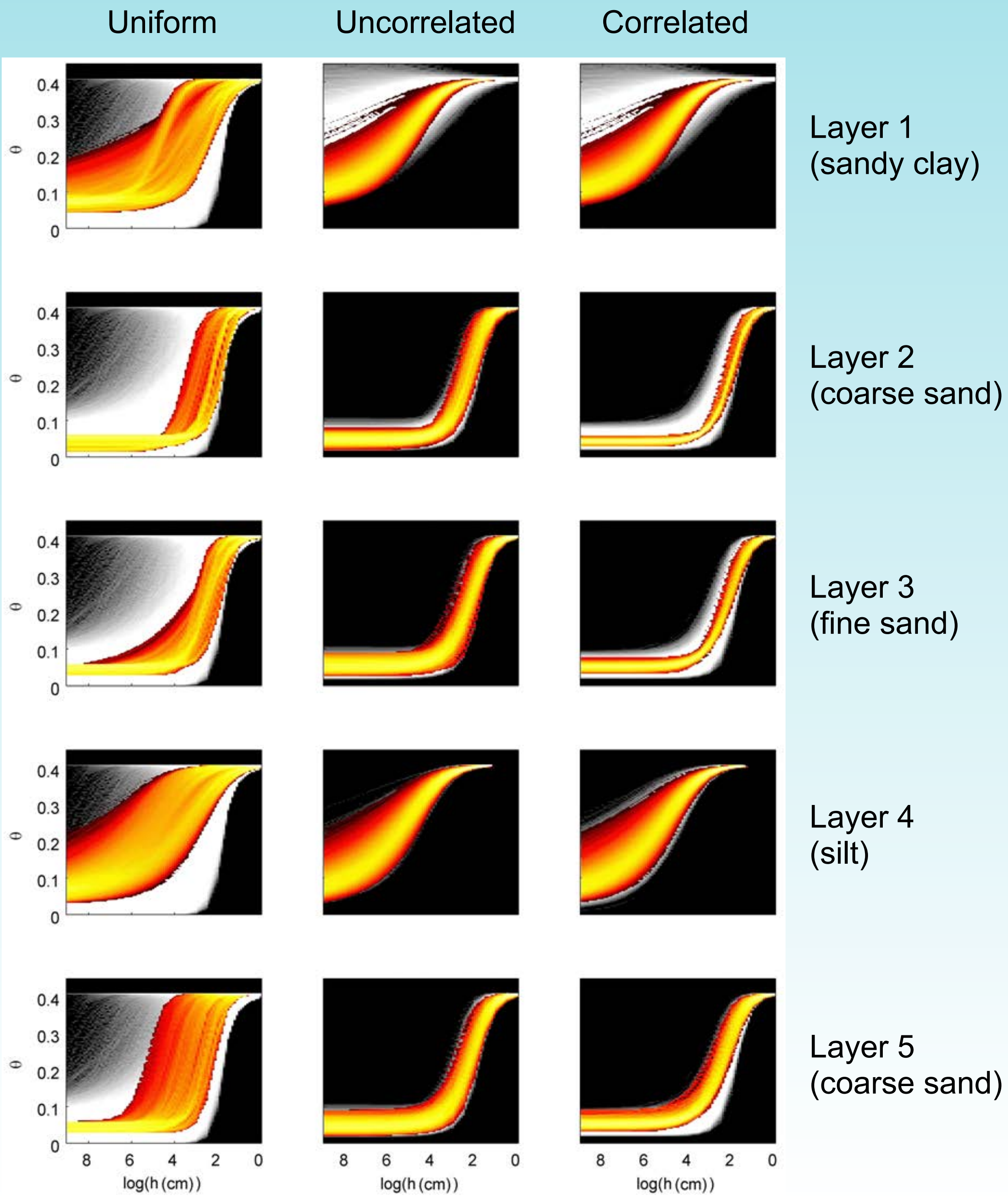
posterior

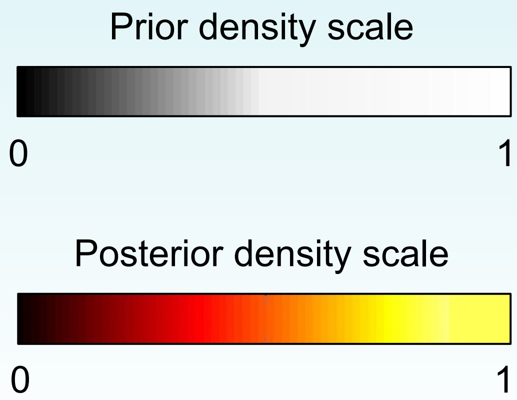
Results: VGM parameters

(refined priors)

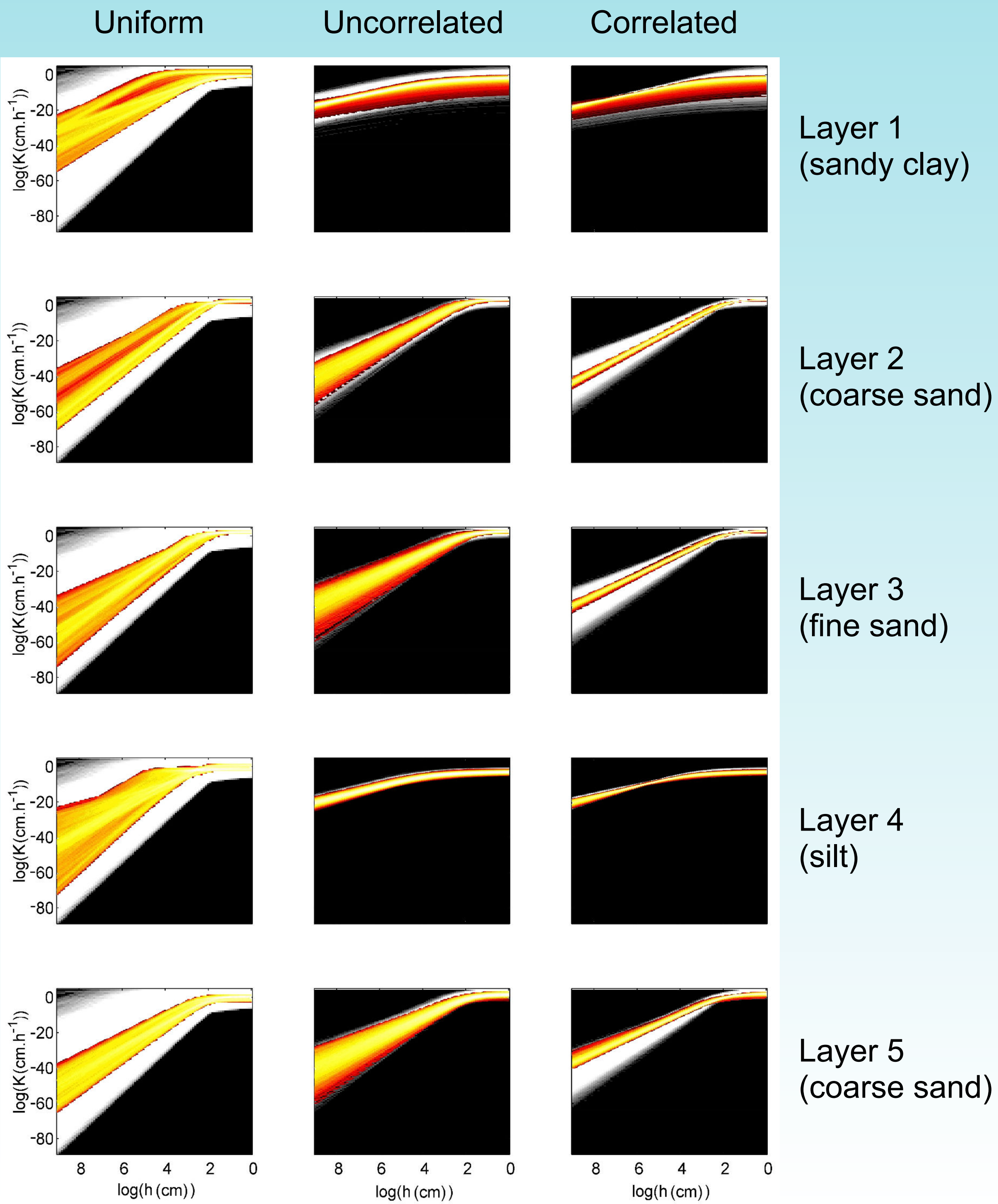


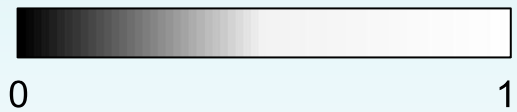
Results: Water retention functions



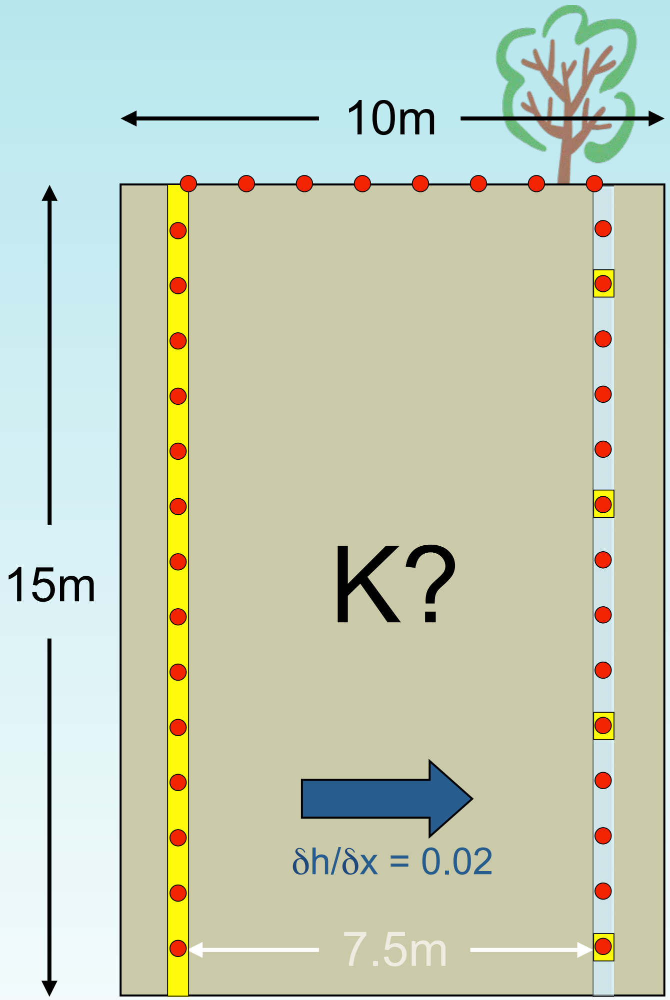


Results: Unsaturated hydraulic conductivity functions





Problem 2: Saturated zone

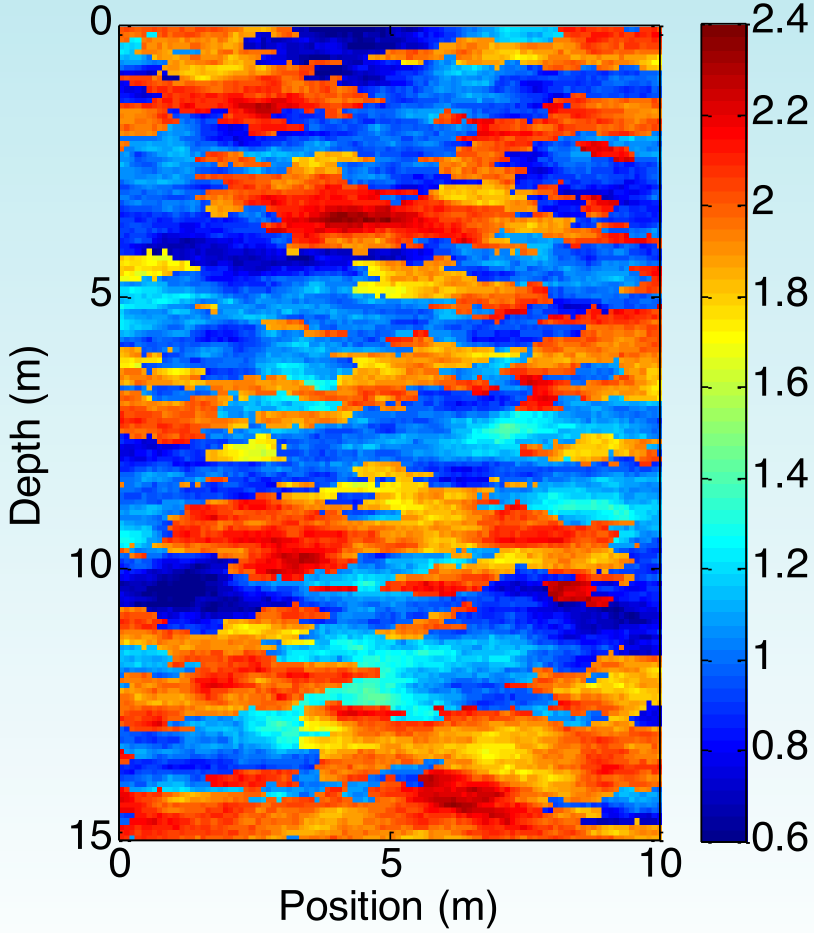


- natural-gradient NaCl groundwater tracer experiment conducted over a period of 25 days
- tracer concentration measured at 4 locations every 3 hours in right (observation) well
- geoelectrical measurements taken once per day using 36 borehole and surface electrodes
- estimate K(x,z) and quantify its uncertainty using
  - tracer concentration data
  - electrical resistivity data
  - concentration + resistivity data

Irving, J. and Singha, K. Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities. Water Resources Research, 46(11), 2010.

True model and MCMC assumptions

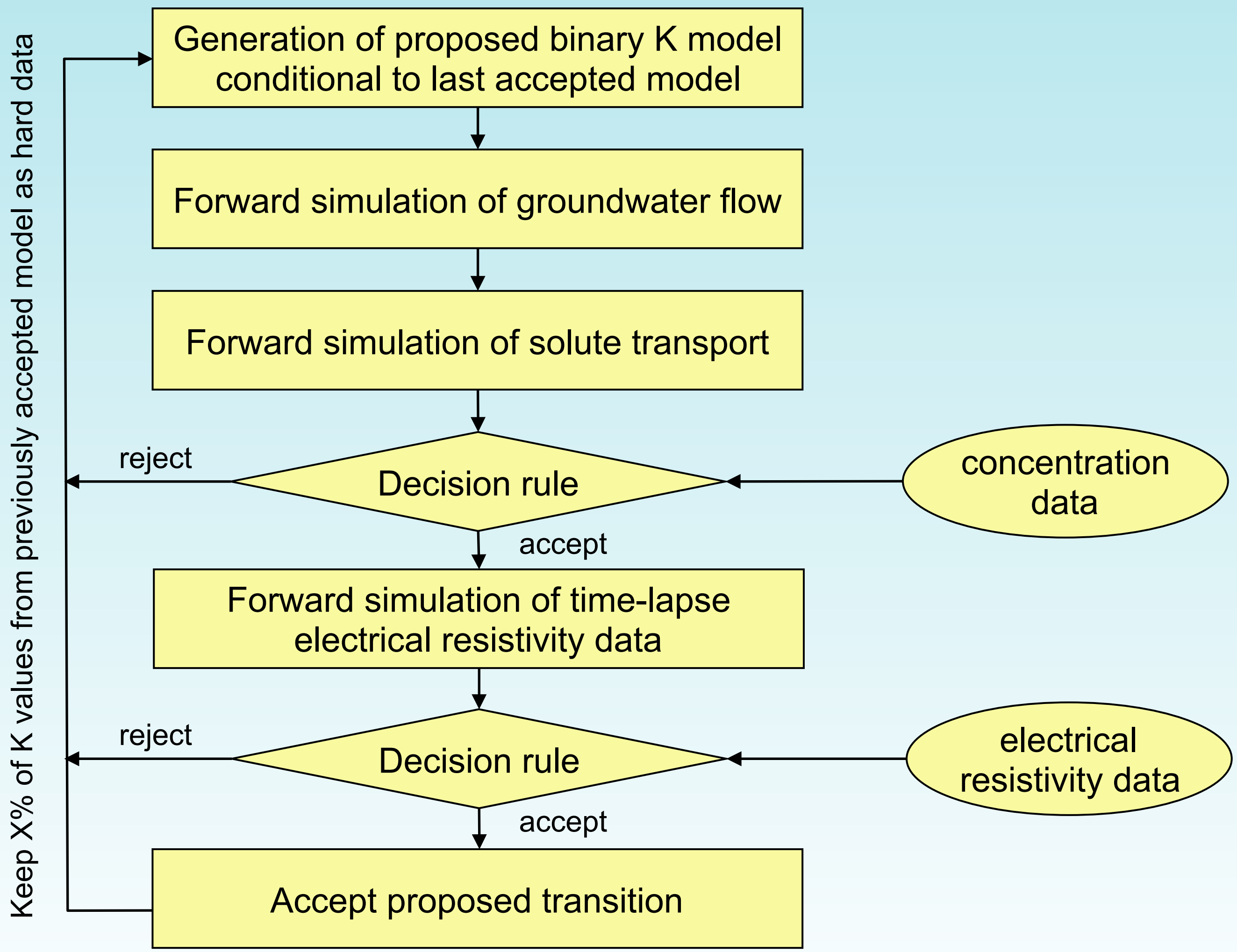
True $\log_{10}$(K) model (m/d)



MCMC assumptions

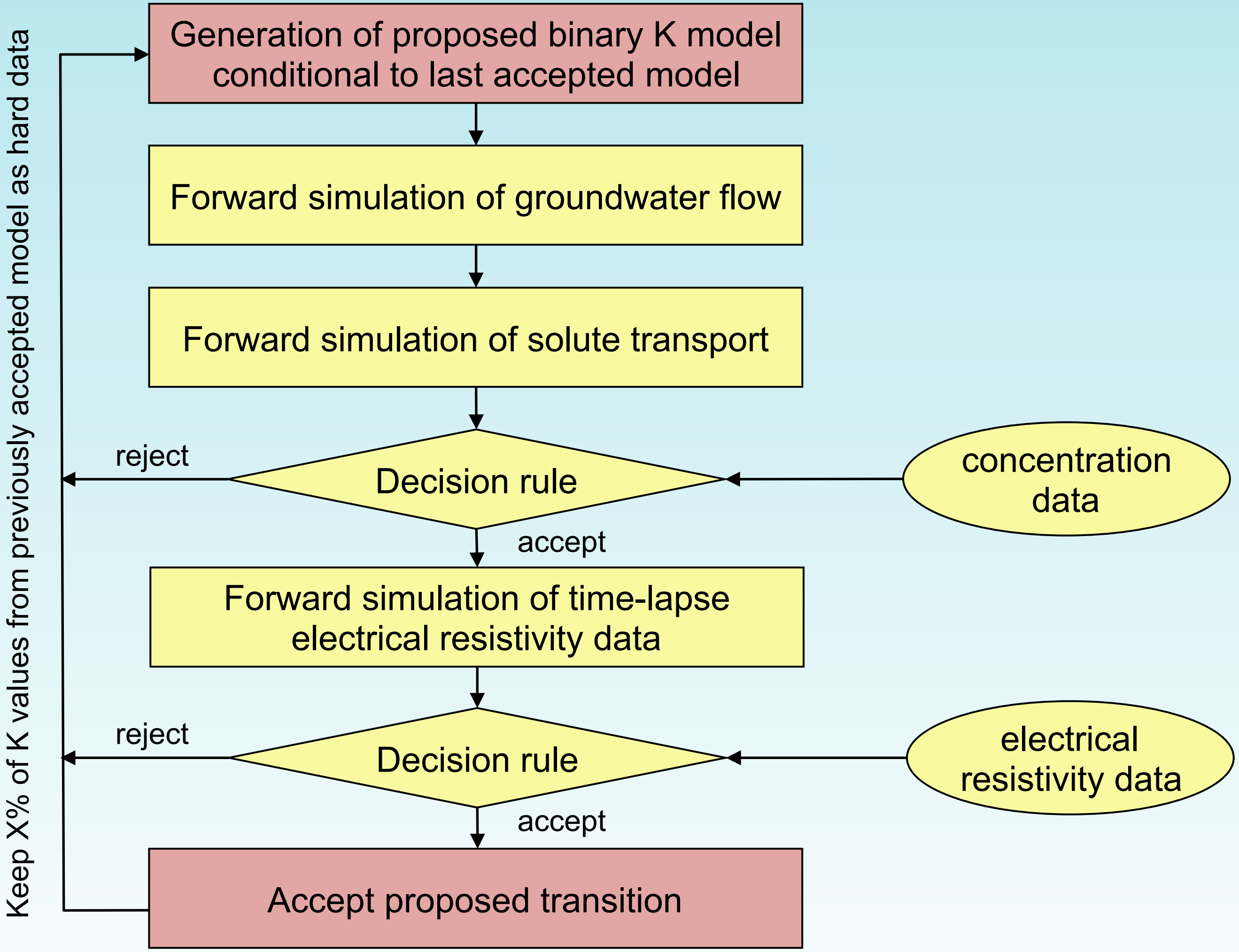
- coarser 20 x 30 mesh (0.5 m cells)
- binary distribution of facies having constant hydrological properties
  - facies 1: K = 10 m/d, θ = 0.2
  - facies 2: K = 100 m/d, θ = 0.3
- facies indicators are spatially correlated, but correlation lengths are unknown
- allow for uncertainty in relationship between solute concentration and soil electrical resistivity
- zero-mean, Gaussian, IID residuals again assumed

Simplified inversion flowchart



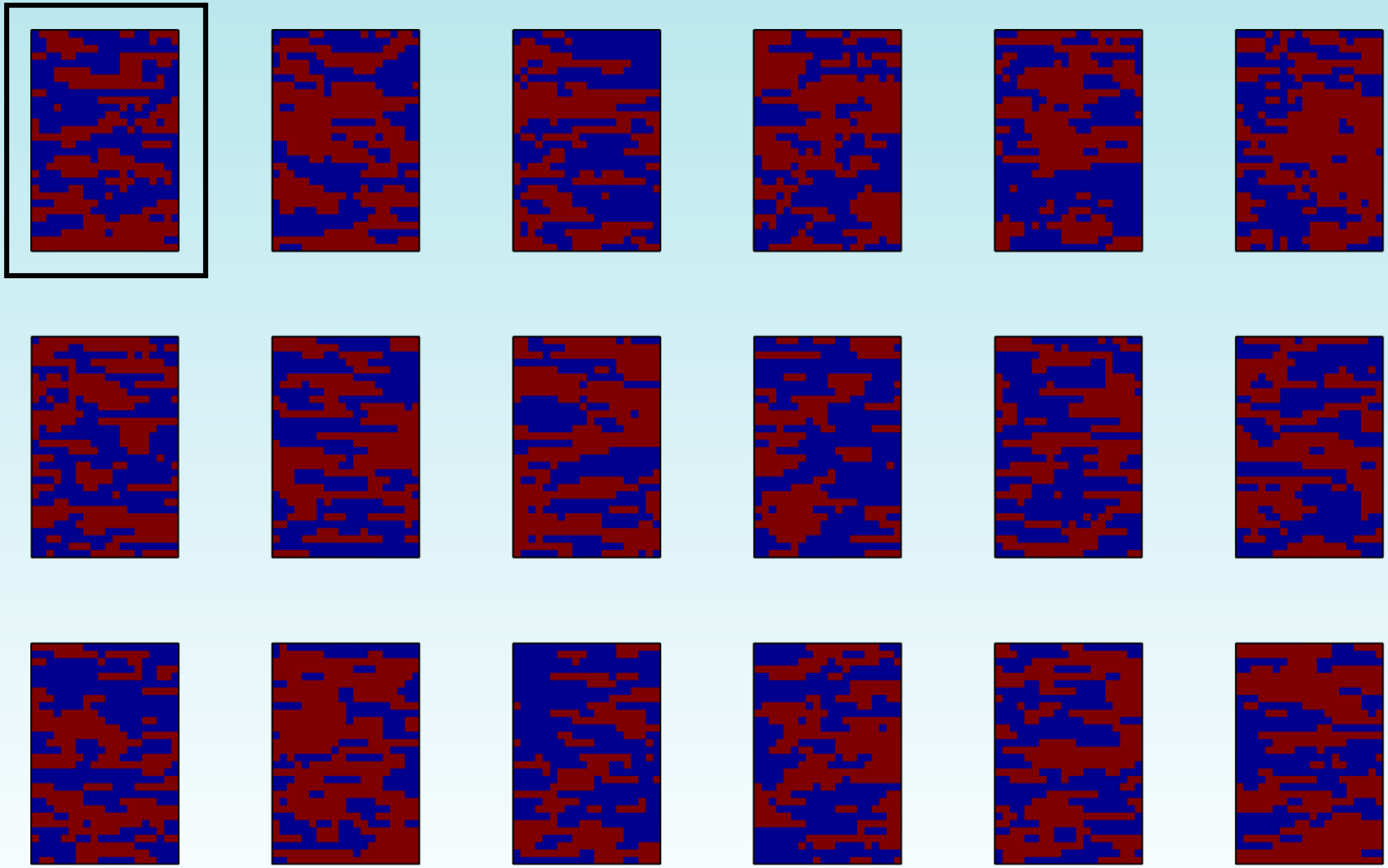
Based on cascaded Metropolis algorithm
proposed by Mosegaard and Tarantola (1995)

Prior distribution (no data)

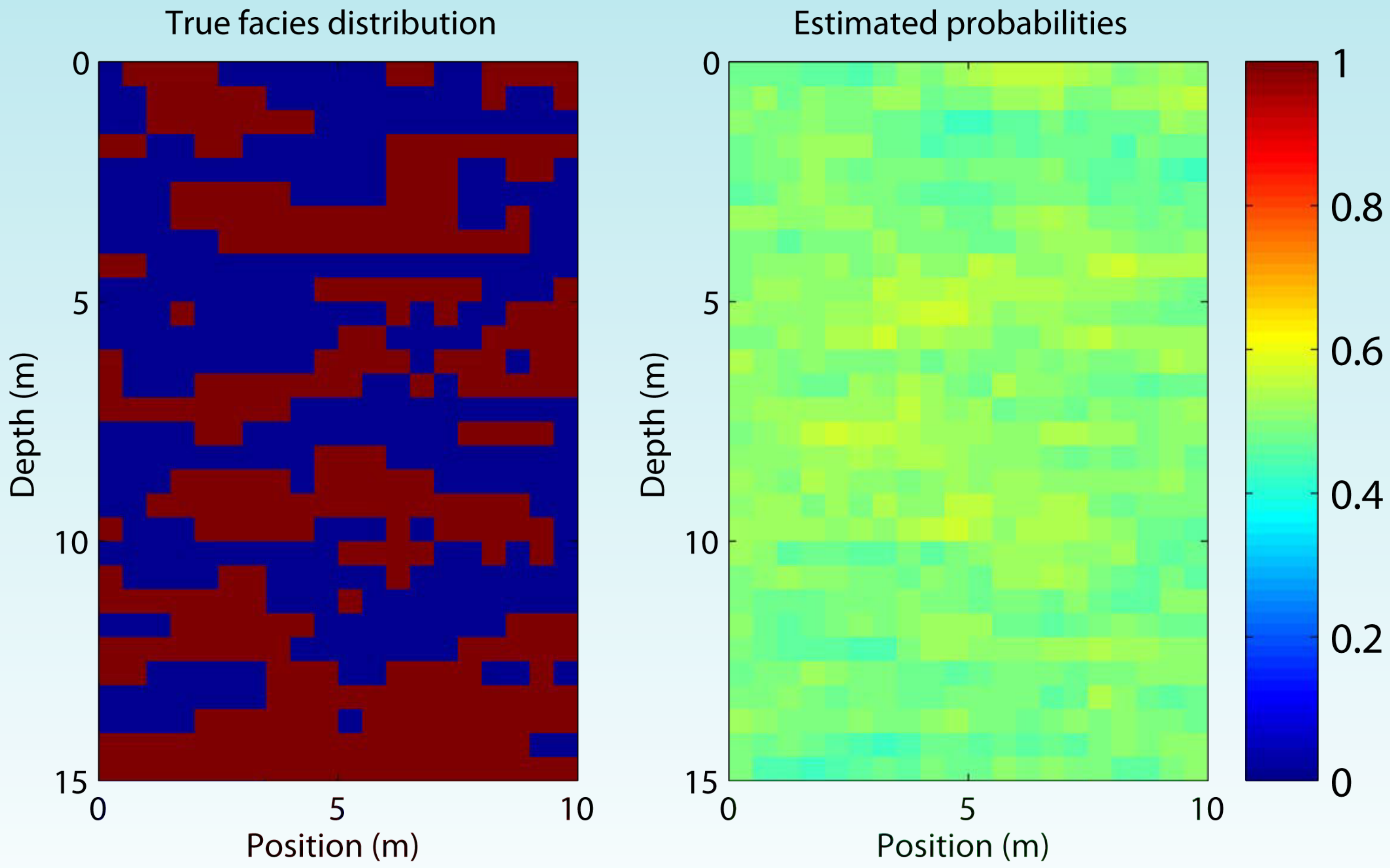


Based on cascaded Metropolis algorithm proposed by Mosegaard and Tarantola (1995)

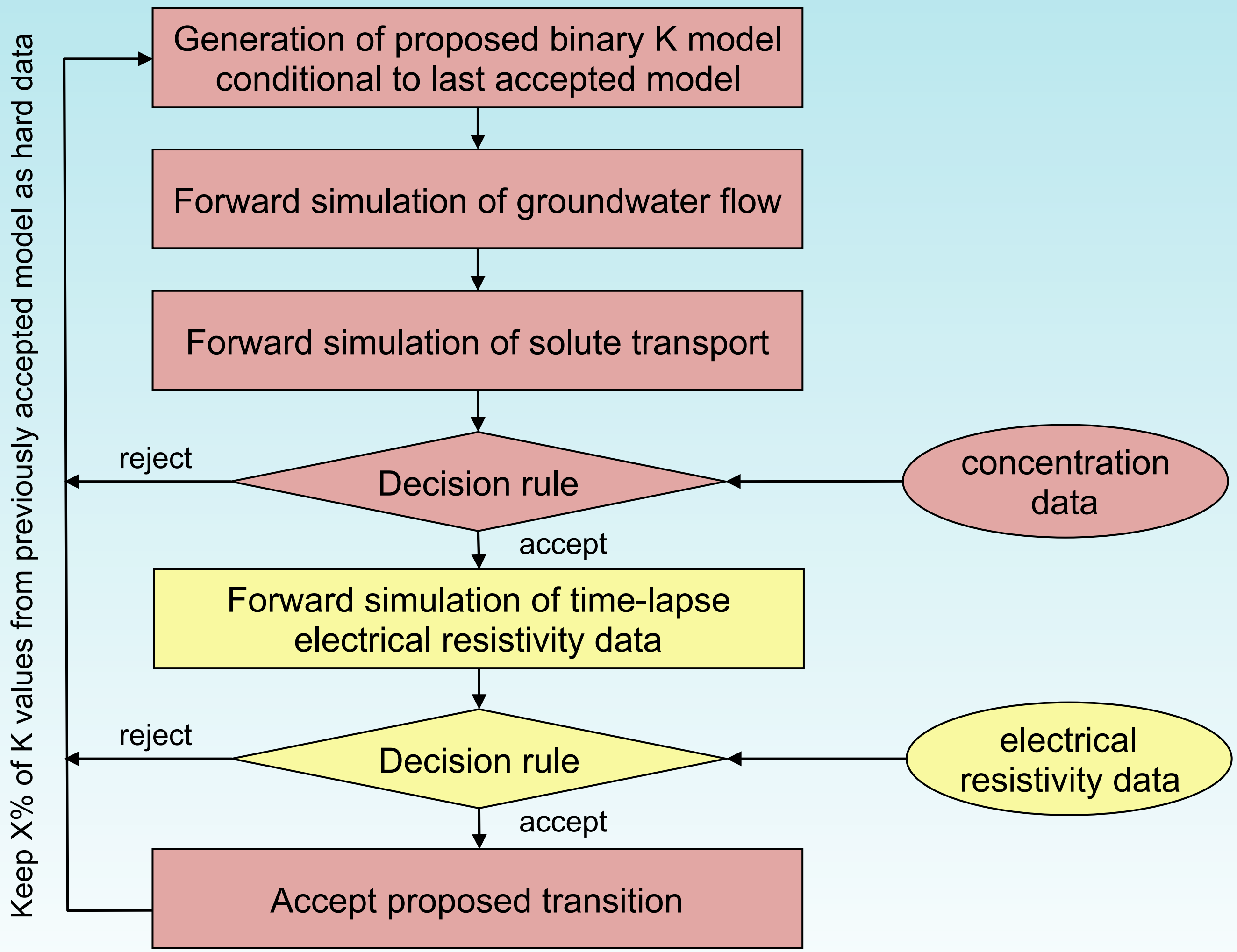
Prior distribution (no data)



Prior distribution (no data)

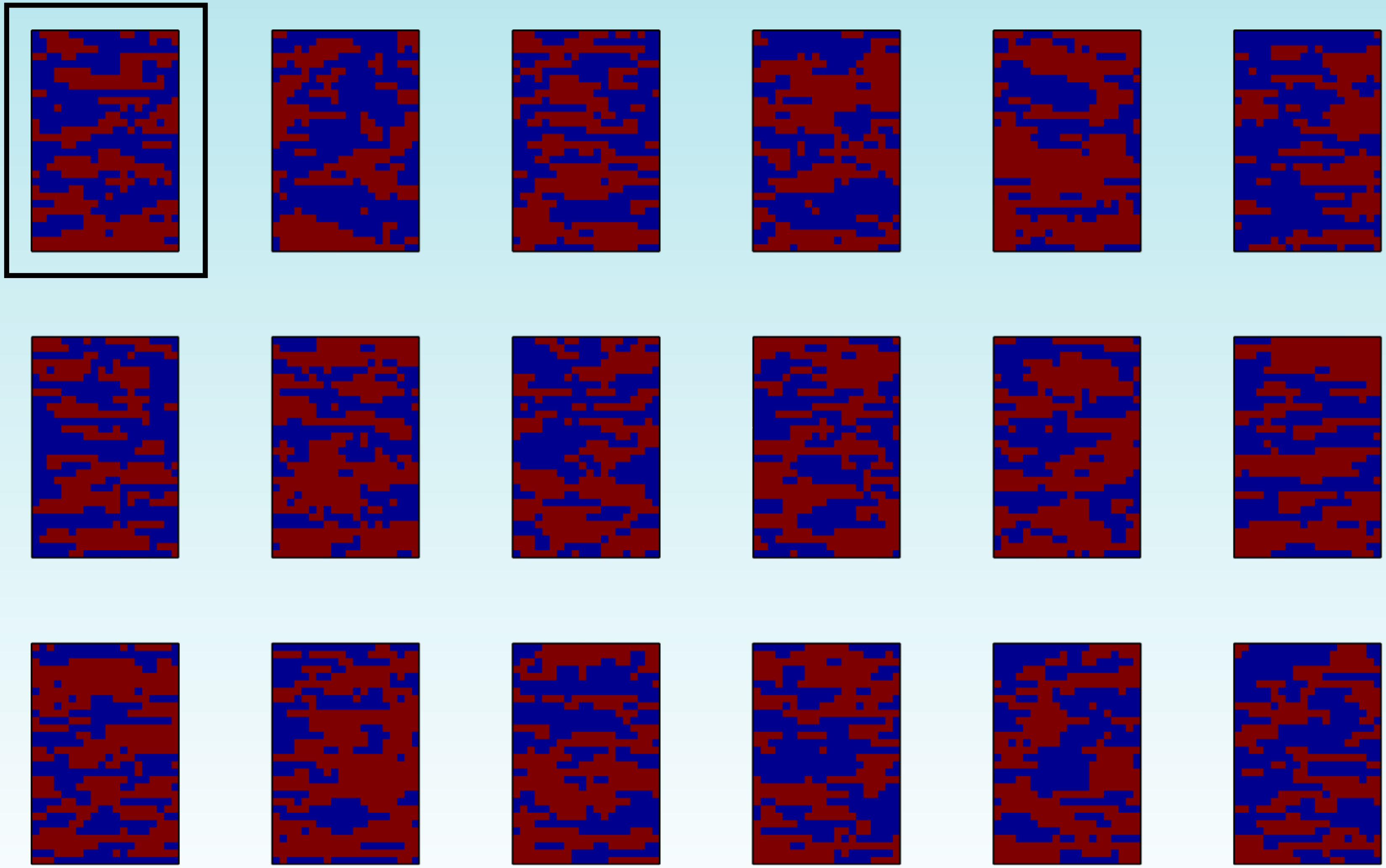


Tracer concentration data only

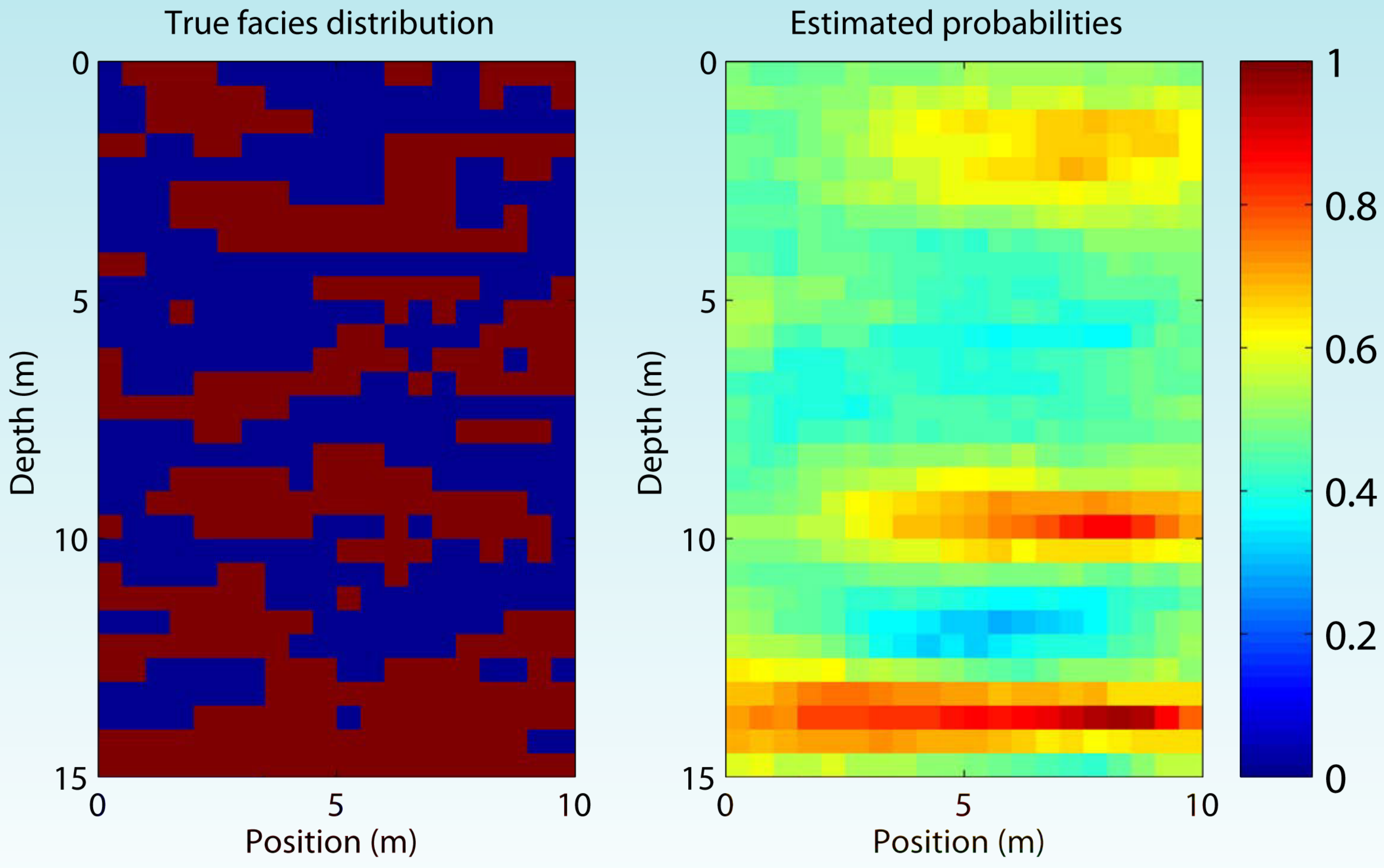


Based on cascaded Metropolis algorithm proposed by Mosegaard and Tarantola (1995)

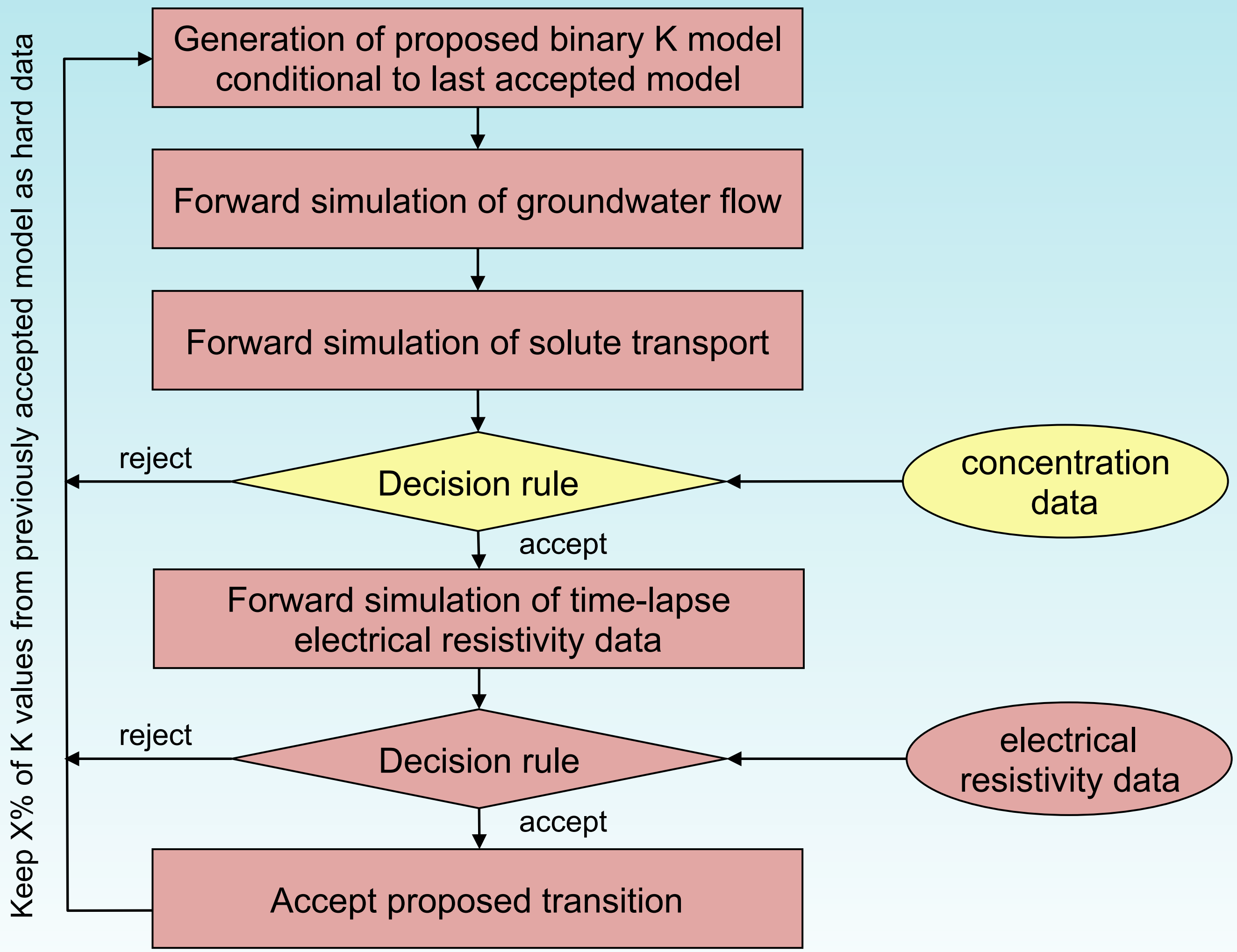
Tracer concentration data only



Tracer concentration data only

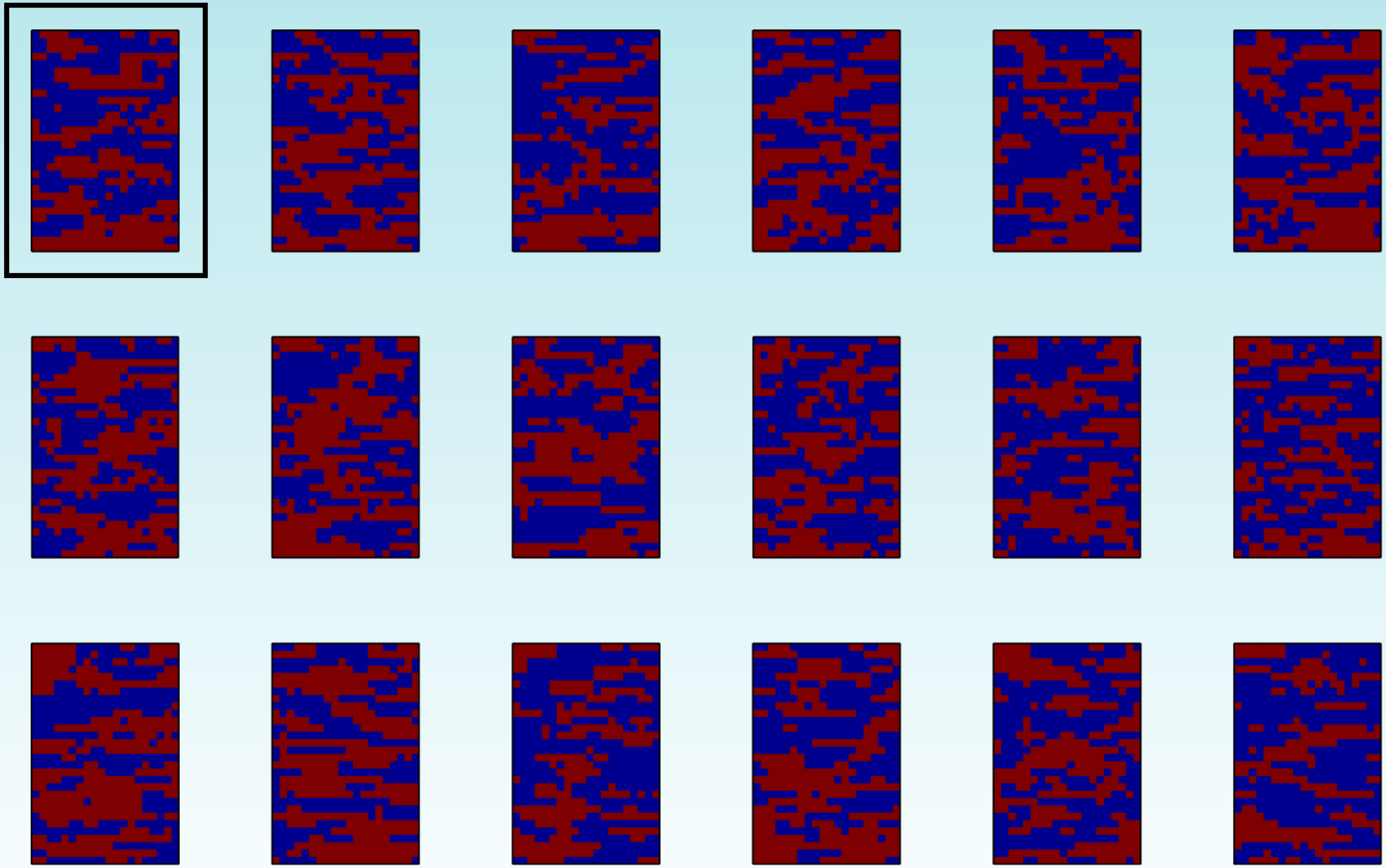


Electrical resistivity data only

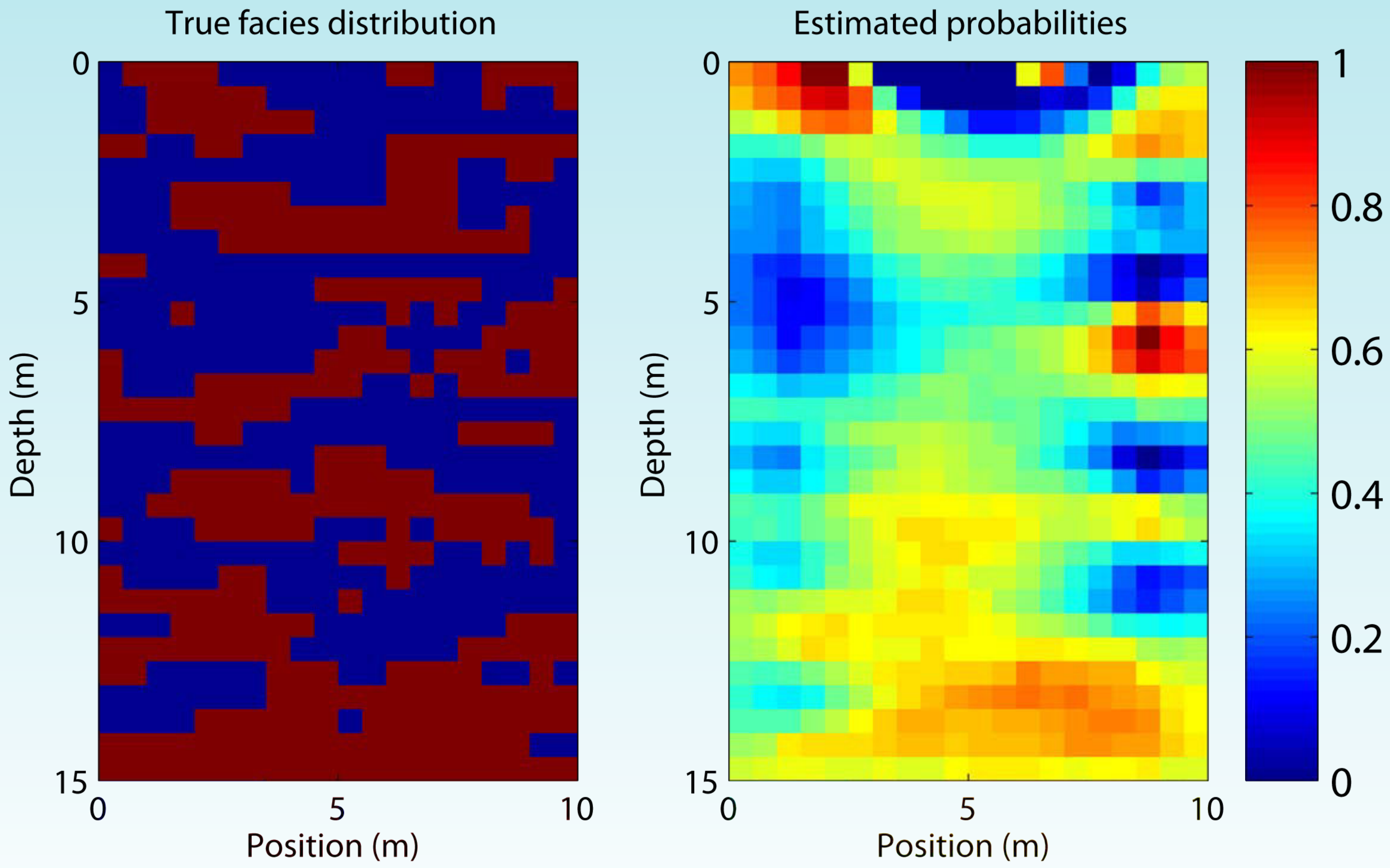


Based on cascaded Metropolis algorithm
proposed by Mosegaard and Tarantola (1995)

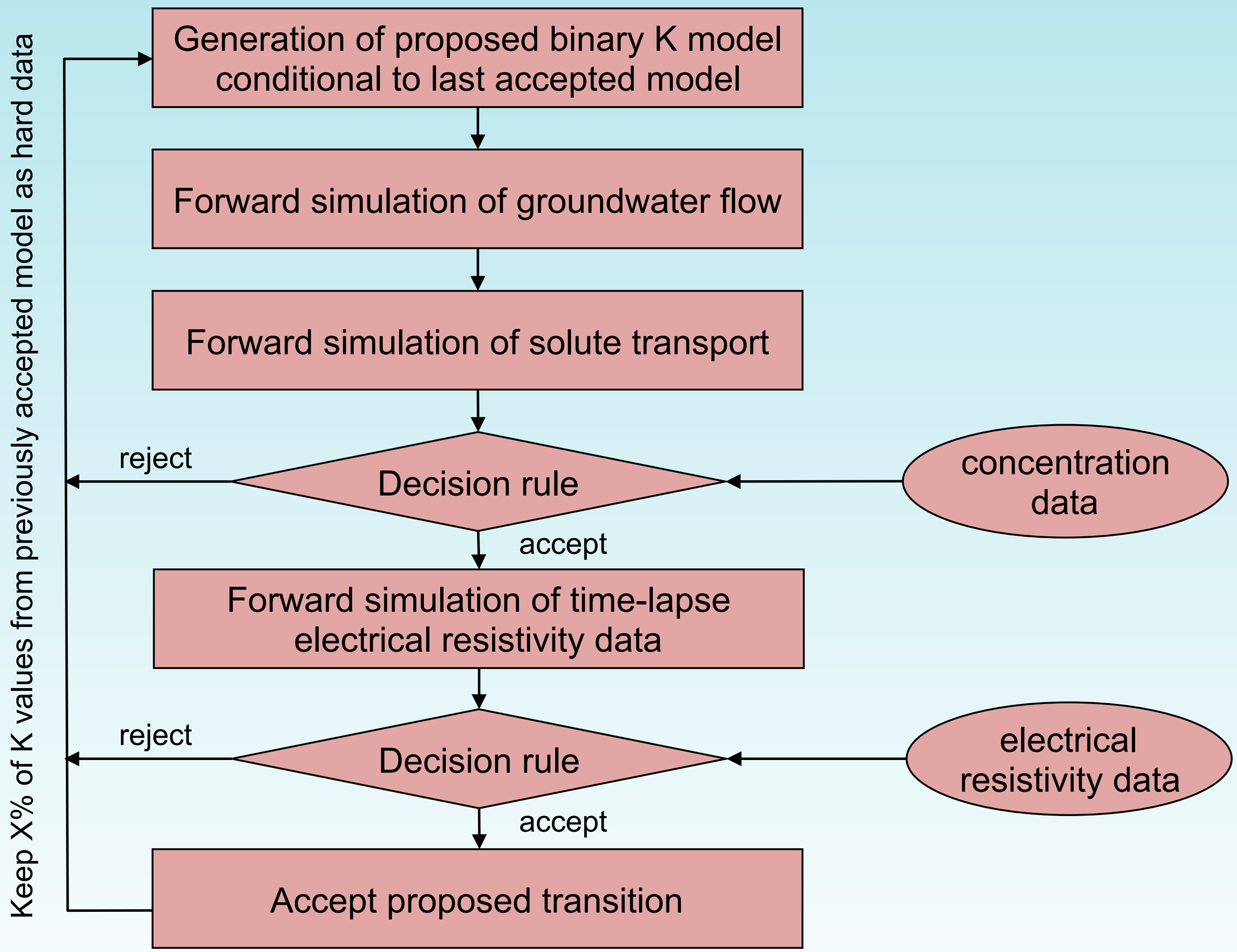
Electrical resistivity data only



Electrical resistivity data only

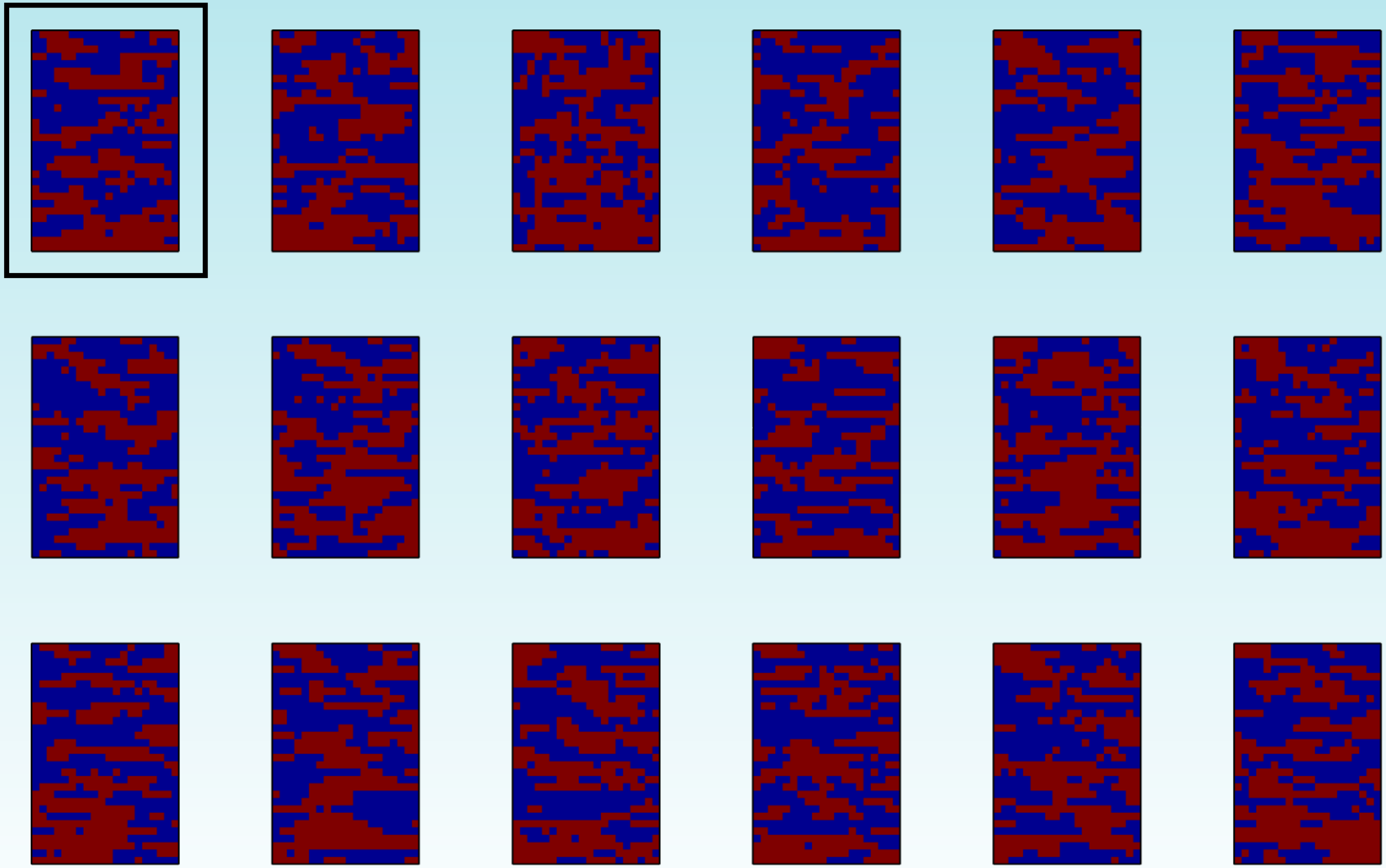


Concentration + resistivity data

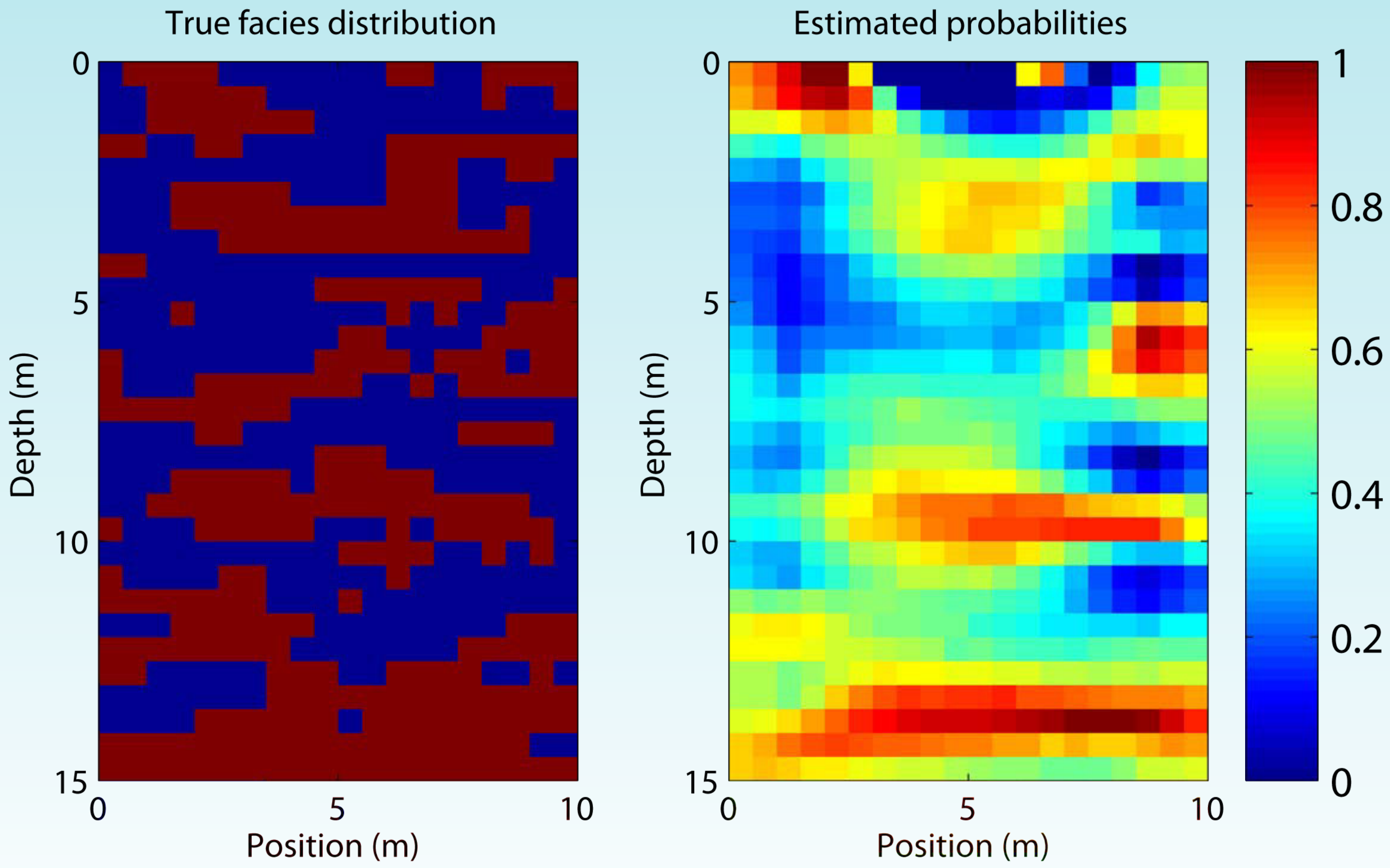


Based on cascaded Metropolis algorithm
proposed by Mosegaard and Tarantola (1995)

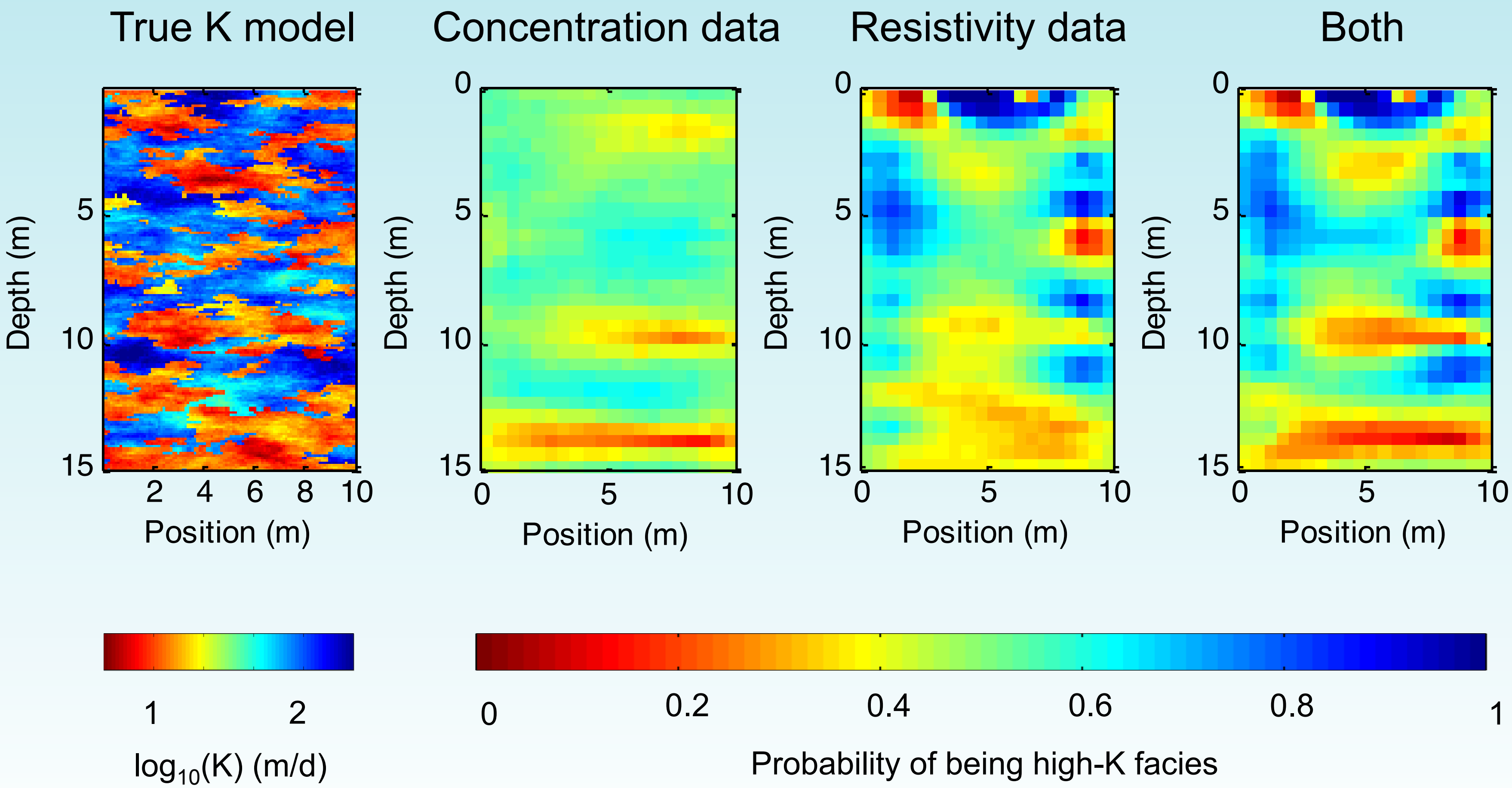
Concentration + resistivity data



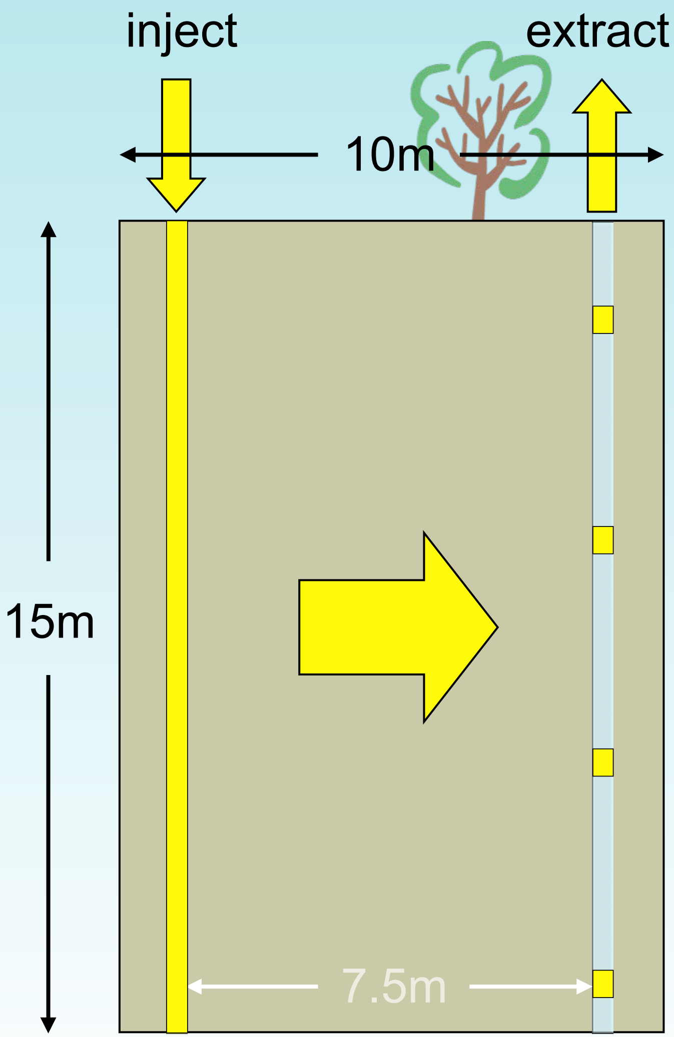
Concentration + resistivity data



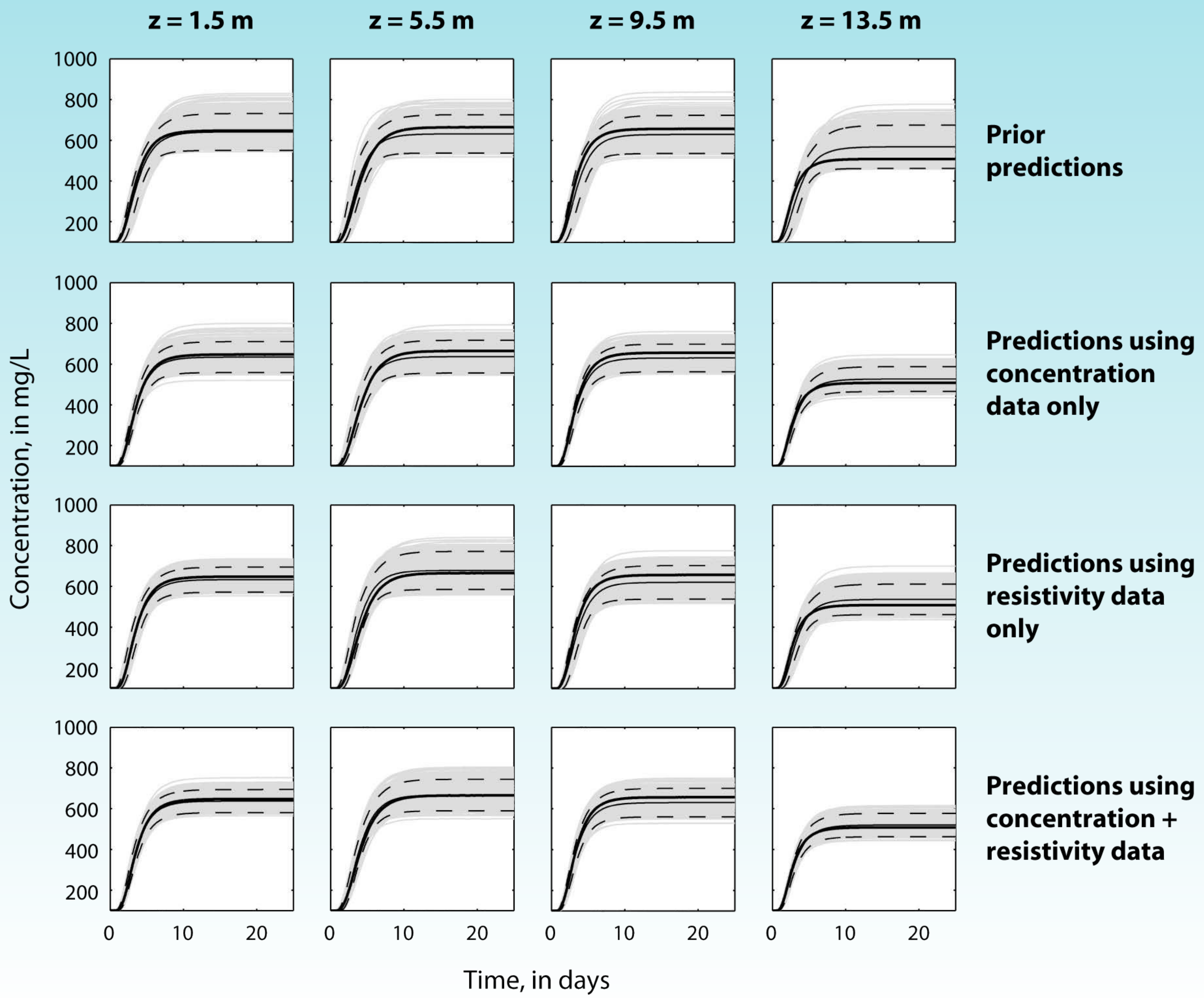
Results: Posterior facies probabilities



Model validation



- investigate how realizations predict a new and different hydrological experiment
- linear injection/extraction test simulated over a period of 25 days
- solute having a concentration of 1000 mg/L is continuously injected at 4 L/min into source well and pumped out at sampling well
- concentration and ERT data were simulated as before
- natural hydraulic gradient perturbed by injection/extraction



z = 1.5 m
z = 5.5 m
z = 9.5 m
z = 13.5 m
Prior predictions
Predictions using concentration data only
Predictions using resistivity data only
Predictions using concentration + resistivity data
Concentration, in mg/L
1000
800
600
400
200
0
10
20
Time, in days

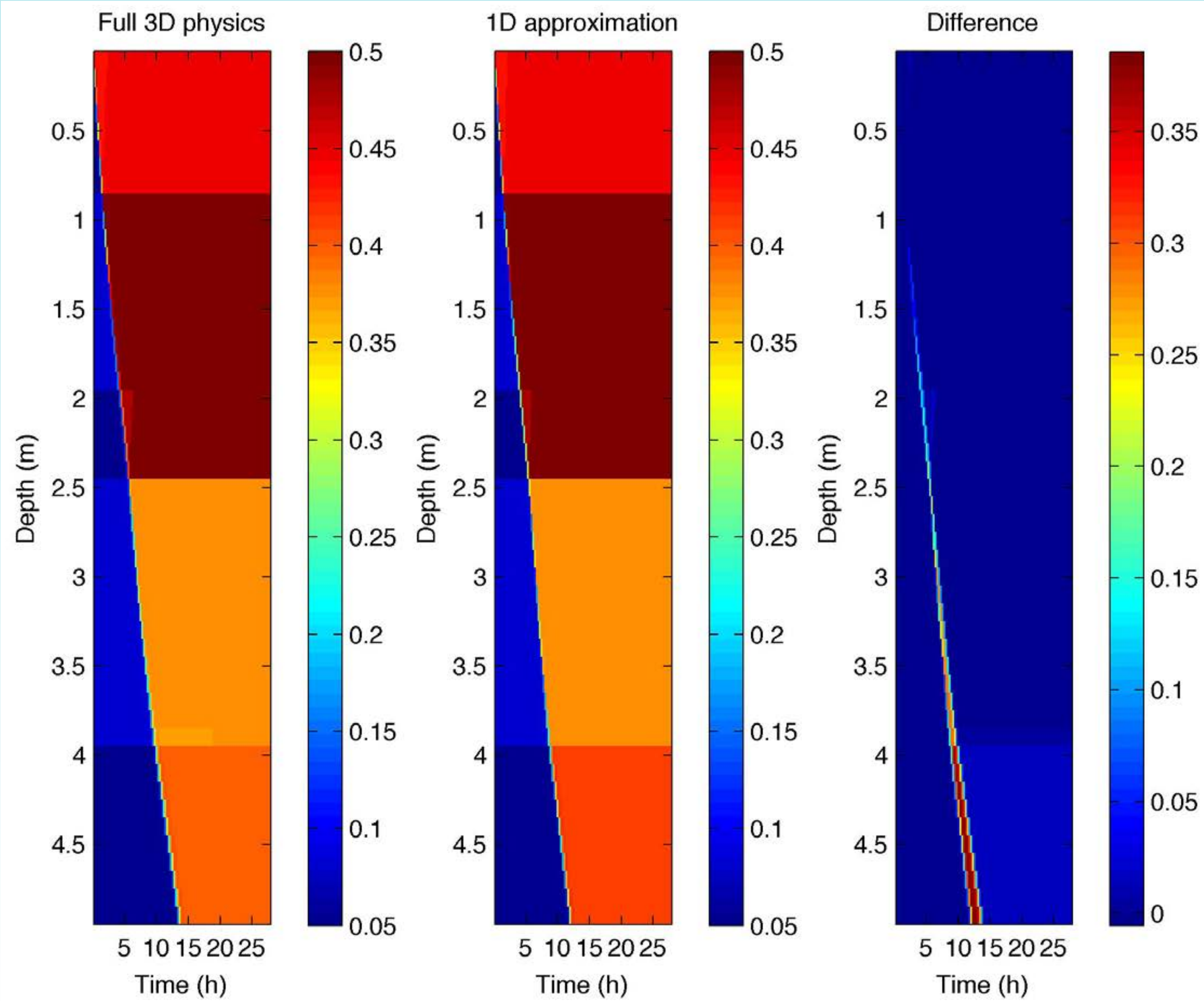
Conclusions

- in the context of both near-surface geophysical problems examined, Bayesian-MCMC appears to be a useful tool
  - reasonable estimates of subsurface hydrological parameters and their corresponding uncertainties
  - successful integration of different data
- HOWEVER...
  - approximations were made for computational tractability!
    - simplified forward model (e.g., 1D unsaturated flow)
    - reduced parameterization (e.g., larger cell size; binary distribution)
  - corresponding modelization errors are not accounted for through the simple Gaussian likelihood functions assumed
    - true residuals will be correlated, heteroscedastic, and non-Gaussian
  - we are also pushing the limits of our modest computational resources with these admittedly "simple" geophysical problems
- this naturally leads us to question...
  - can we trust our results? (decision making, risk assessment, etc.)
  - can we expect MCMC to be practical for more complex problems?

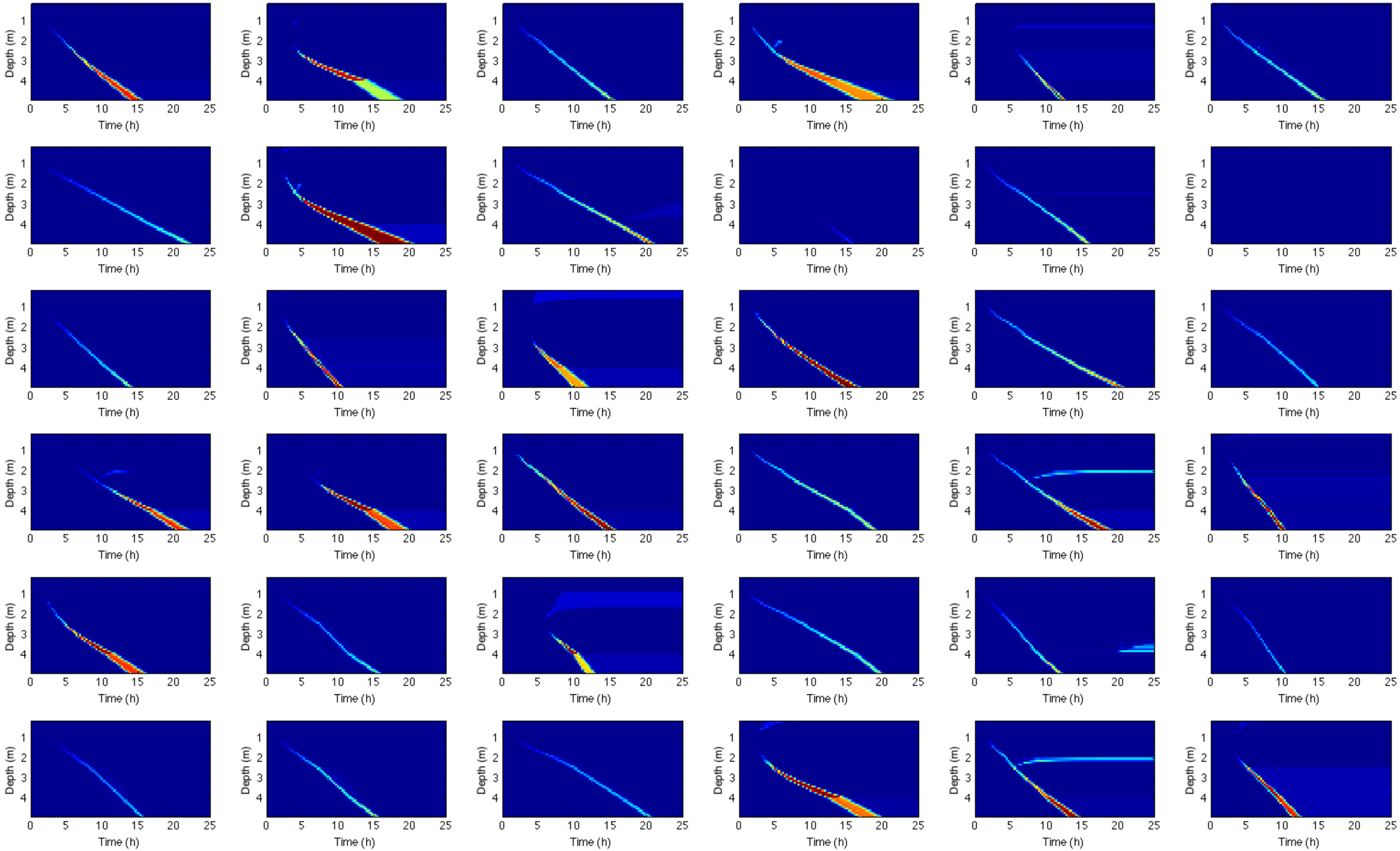
Current research directions

- modelization errors
  - can we use synthetic modeling combined with knowledge of our model approximations/simplifications to build a statistical model for the modelization errors?
  - use to build a more appropriate Bayesian likelihood function, and thus be more “honest” about the nature of the residuals
  - develop bias corrections for simplified models
- dealing with dimensionality
  - how can we best deal with the high-dimension of spatially distributed geophysical inverse problems with MCMC methods?
  - model space compression (e.g., PCA, DCT, …)
    - how to avoid bias in the prior specification?
    - how to condition to complex geological scenarios?
  - sequential geostatistical simulation
    - flexibility to condition to a wide range of prior geostatistical constraints through the extended Metropolis algorithm
    - still has efficiency issues for high-dimensional problems

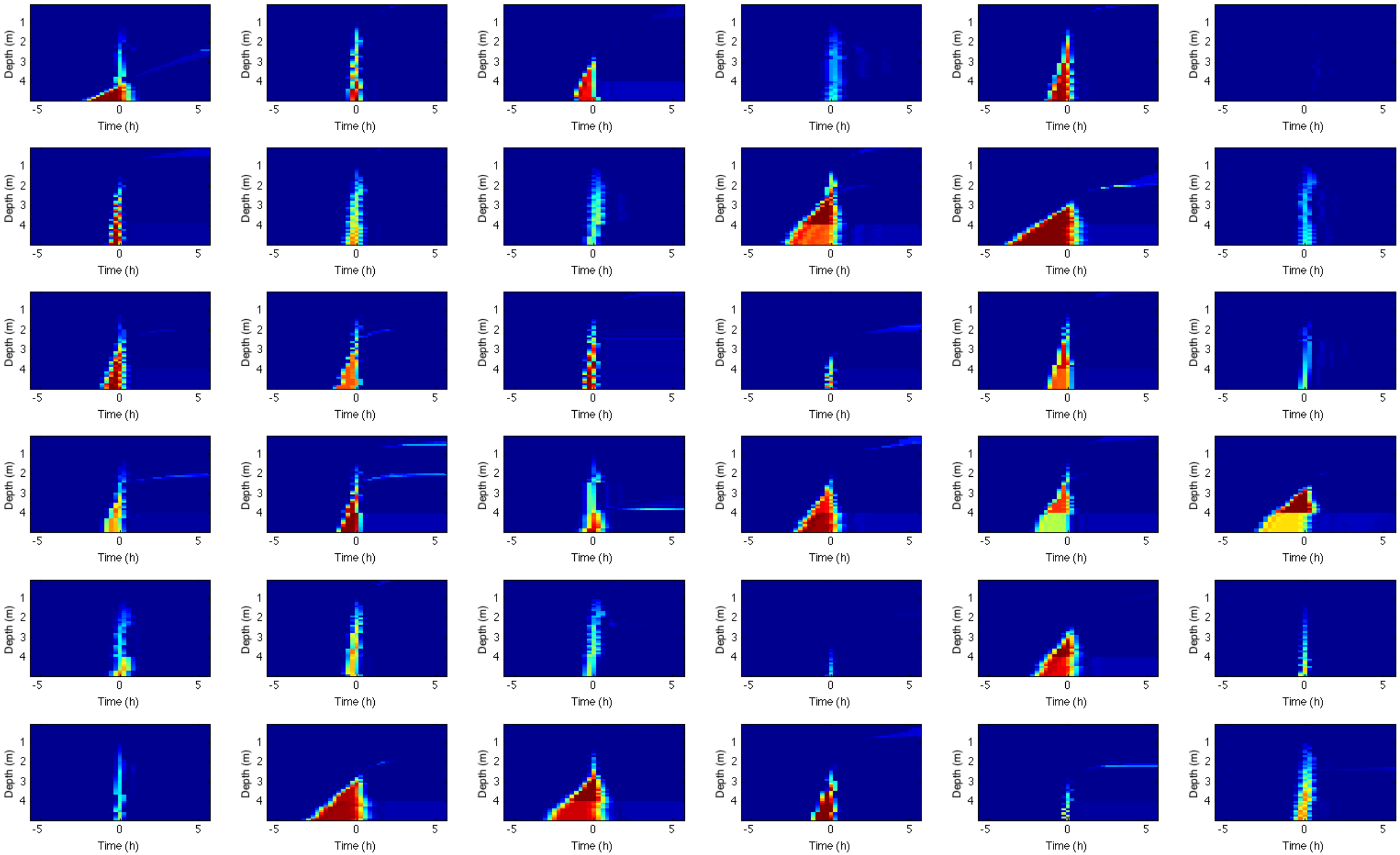
Water content errors resulting from the 1D unsaturated flow assumption



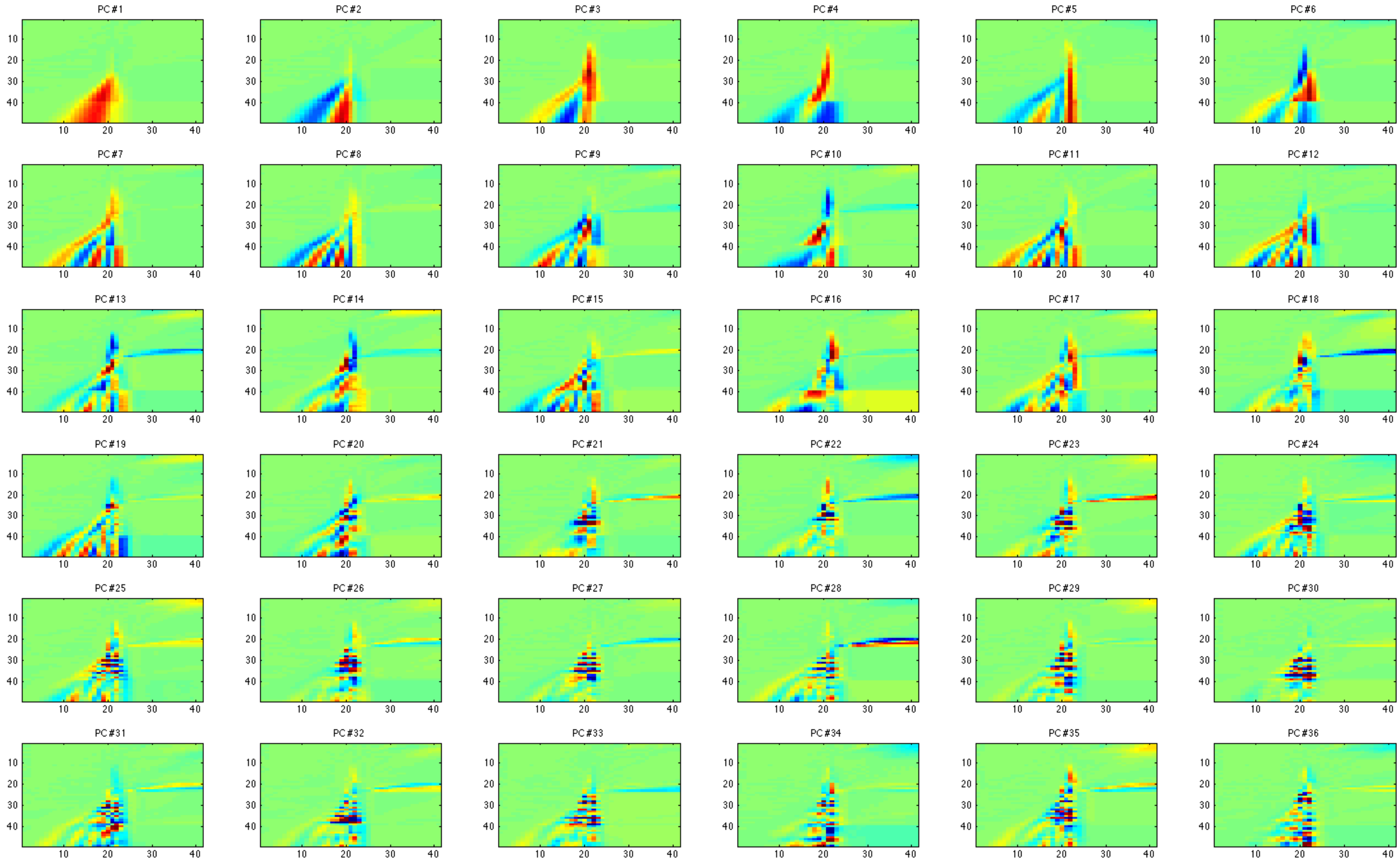
Errors resulting from a 1D flow assumption for different VGM parameter configurations



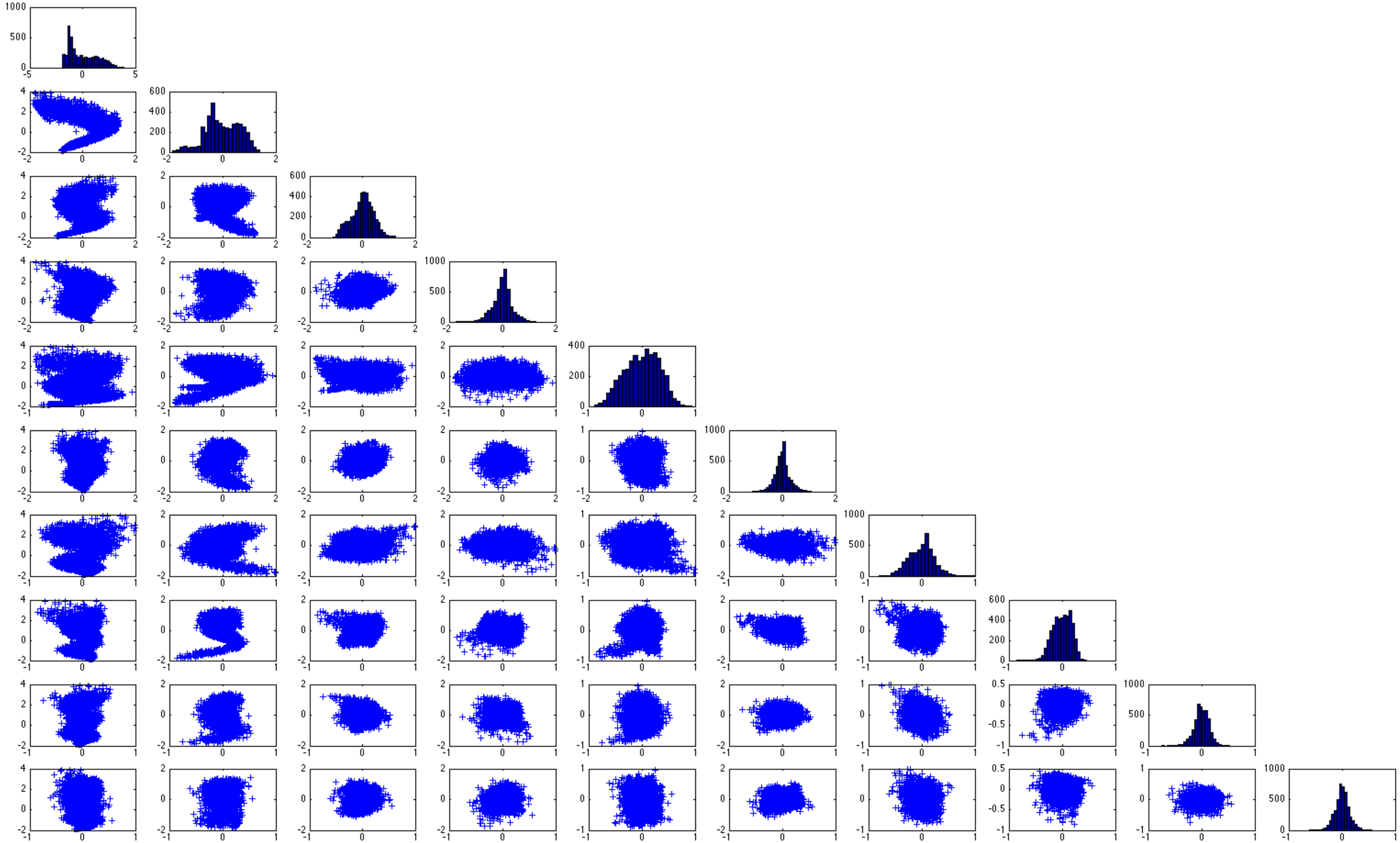
Errors plotted relative to the arrival time of the water front at each depth



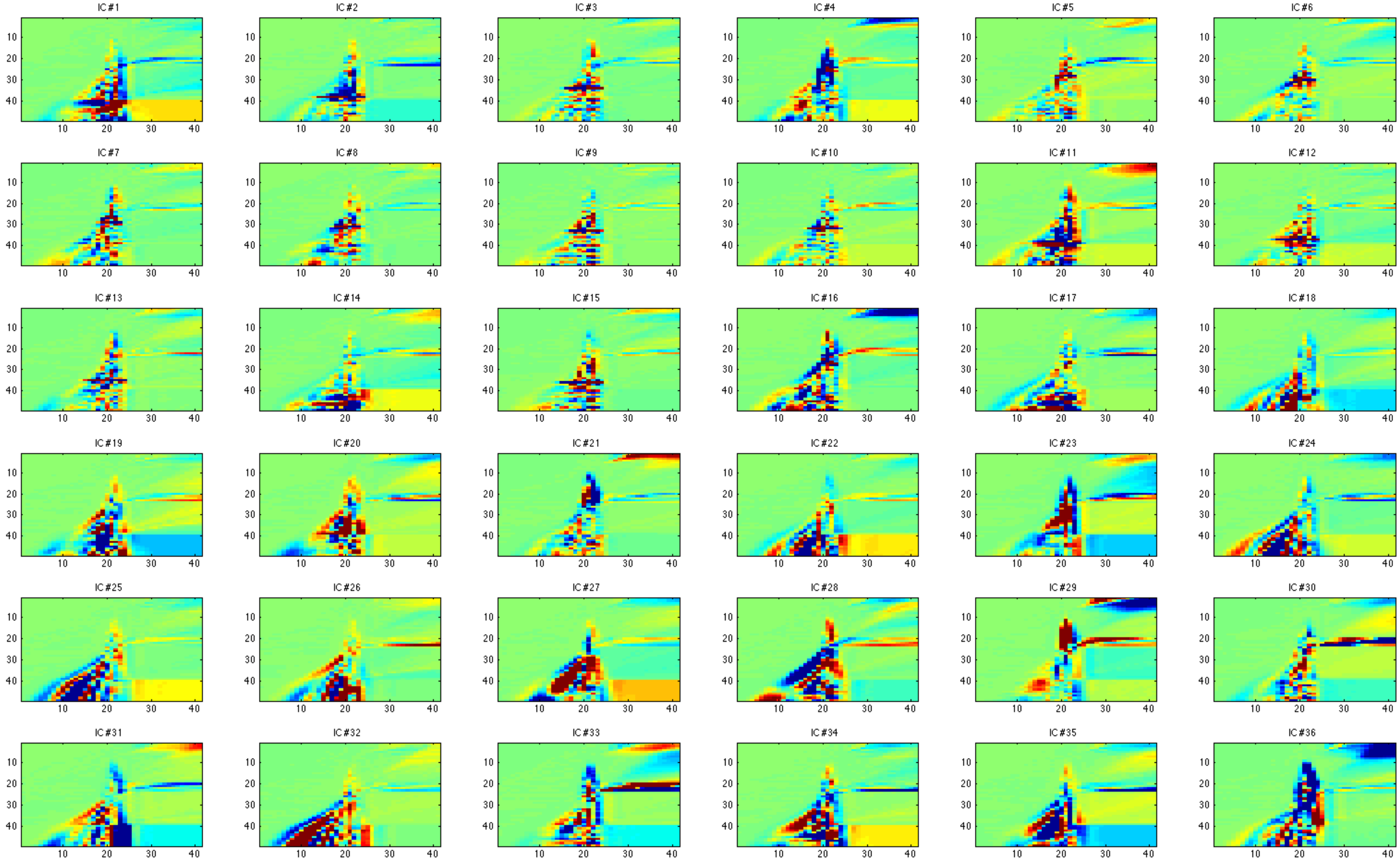
Leading 36 principal components accounting for 98% of the error variance



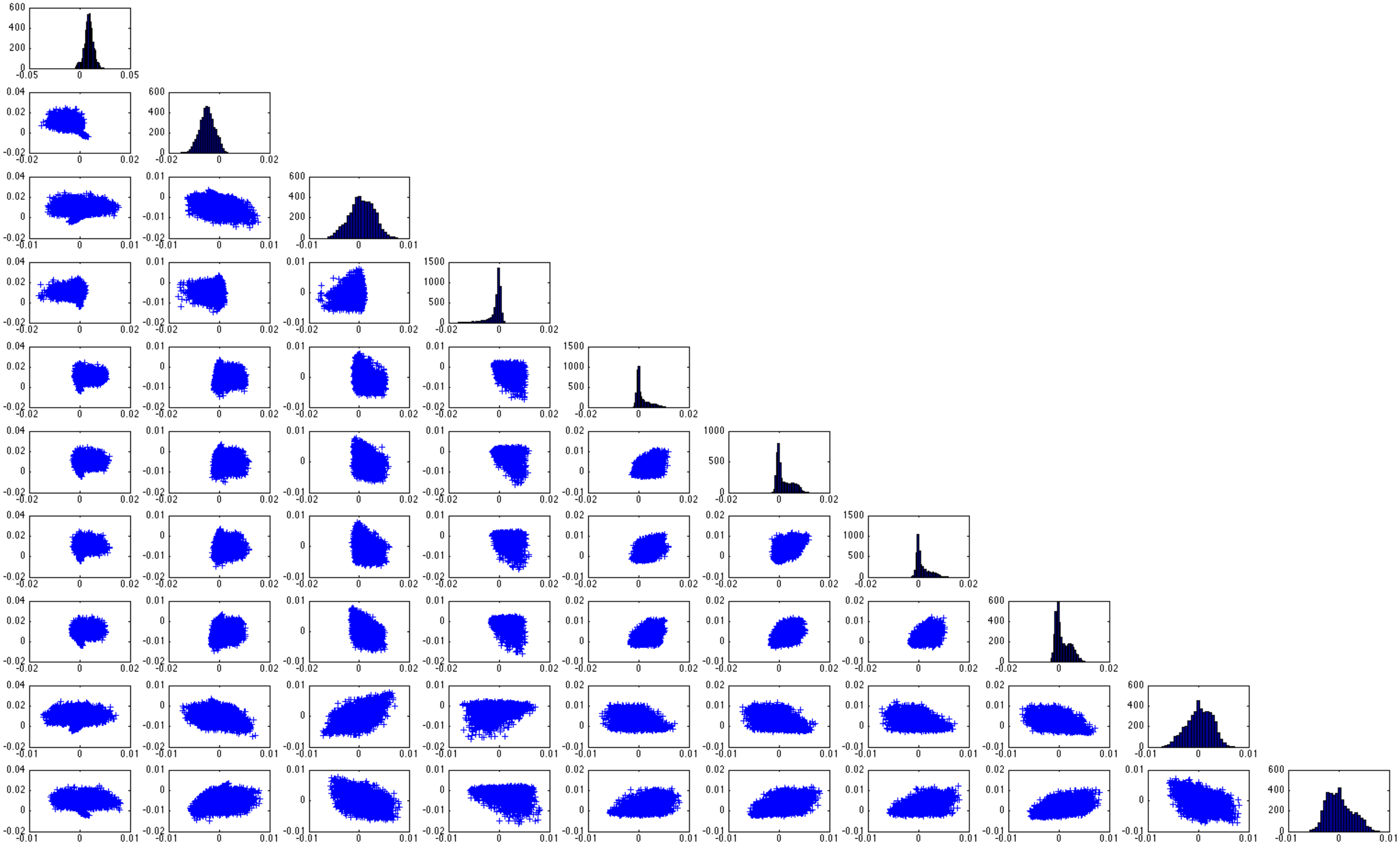
Basis weight statistics for the top 10 PCs (uncorrelated but not independent)



First try, independent component analysis with 98% of the error variance captured



Basis weight statistics for the 36 ICs
Can we use ICA to build an error model?



Acknowledgments

PhD students:
Baptiste Dafflon, Marie Scholer,
Paolo Ruggeri, Corinna Koepke

Collaborators:
Kamini Singha, Colorado School of Mines
Majken Looms, University of Copenhagen
Lars Nielsen, University of Copenhagen
Andrew Binley, Lancaster University

Funding agencies:
Swiss National Science Foundation
UNIL Fondation Herbette