INVERSE PROBLEMS – A BAYESIAN PERSPECTIVE

Niklas Linde

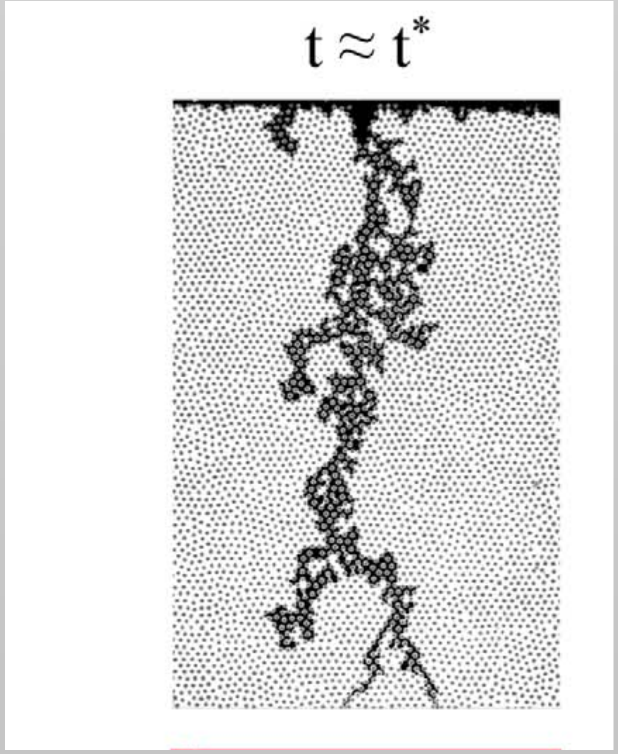
Institute of Earth Sciences, UNIL

Outline of lecture

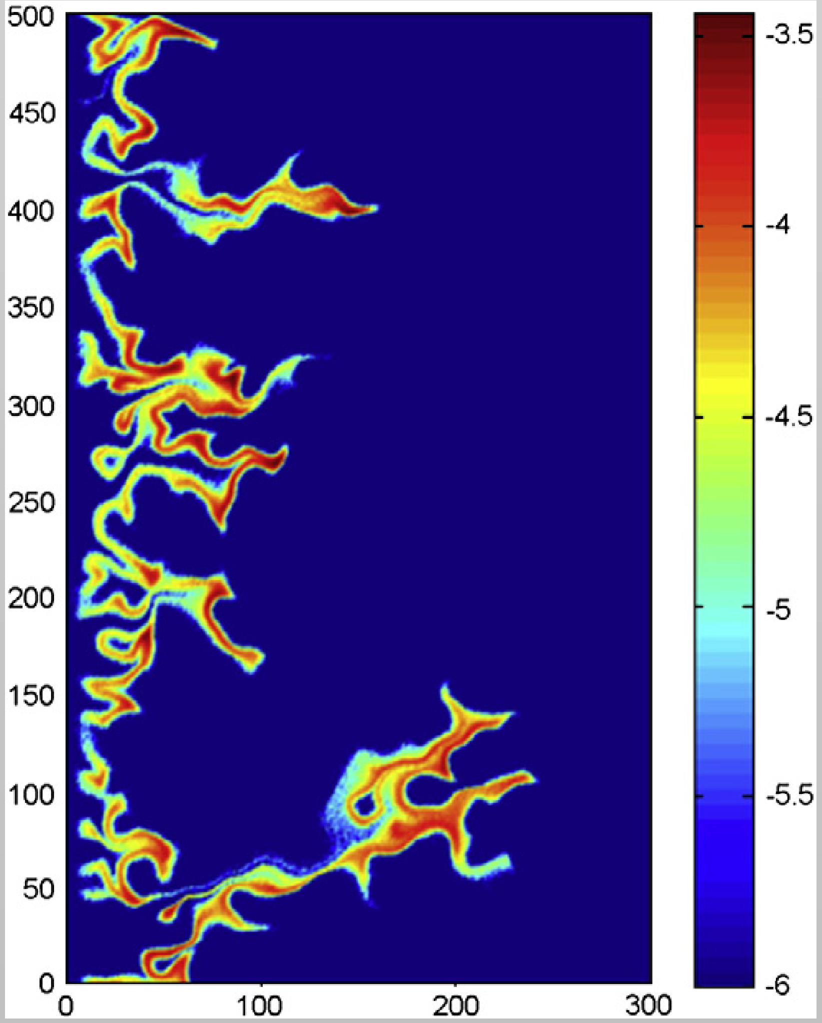
- Linear vs. non-linear finite-dimensional problems
- Bayes theorem (a “complete” solution to the inverse problem)
- Posterior inference for non-linear problems
- Complex geological priors
- Modeling and petrophysical errors (intractable likelihoods)
- Bayesian model selection

Mathematical treatment will be simplified and follow common usage in geophysics

Interesting problems are non-linear: connectivity matters



Drainage processes (Ferrari et al., 2015)



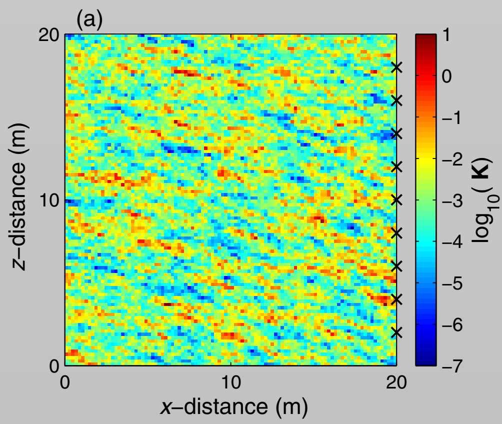
Location, timing and rate of chemical reactions (Dentz et al., 2011)

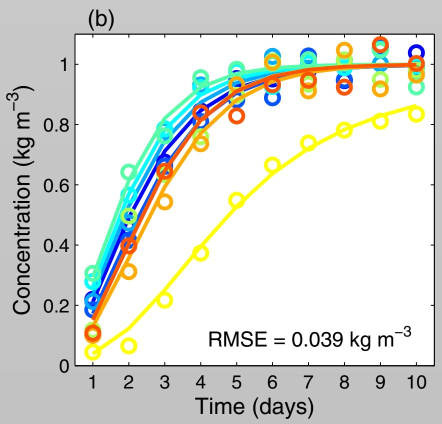
Models, data, and forward problems in finite dimensions

$$\mathbf{m} = \begin{bmatrix} m_1 \\ \cdots \\ m_j \\ \cdots \\ m_{M-1} \\ m_M \end{bmatrix}. \qquad \mathbf{d} = \begin{bmatrix} d_1 \\ \cdots \\ d_i \\ \cdots \\ d_{N-1} \\ d_N \end{bmatrix}.$$

Analytical or numerical (e.g., finite element, finite difference) of physics for a given Earth model, initial, and boundary conditions

$$g\left(\mathbf{m}^{\text{prop}}\right) = \mathbf{d}^{\text{sim}}$$

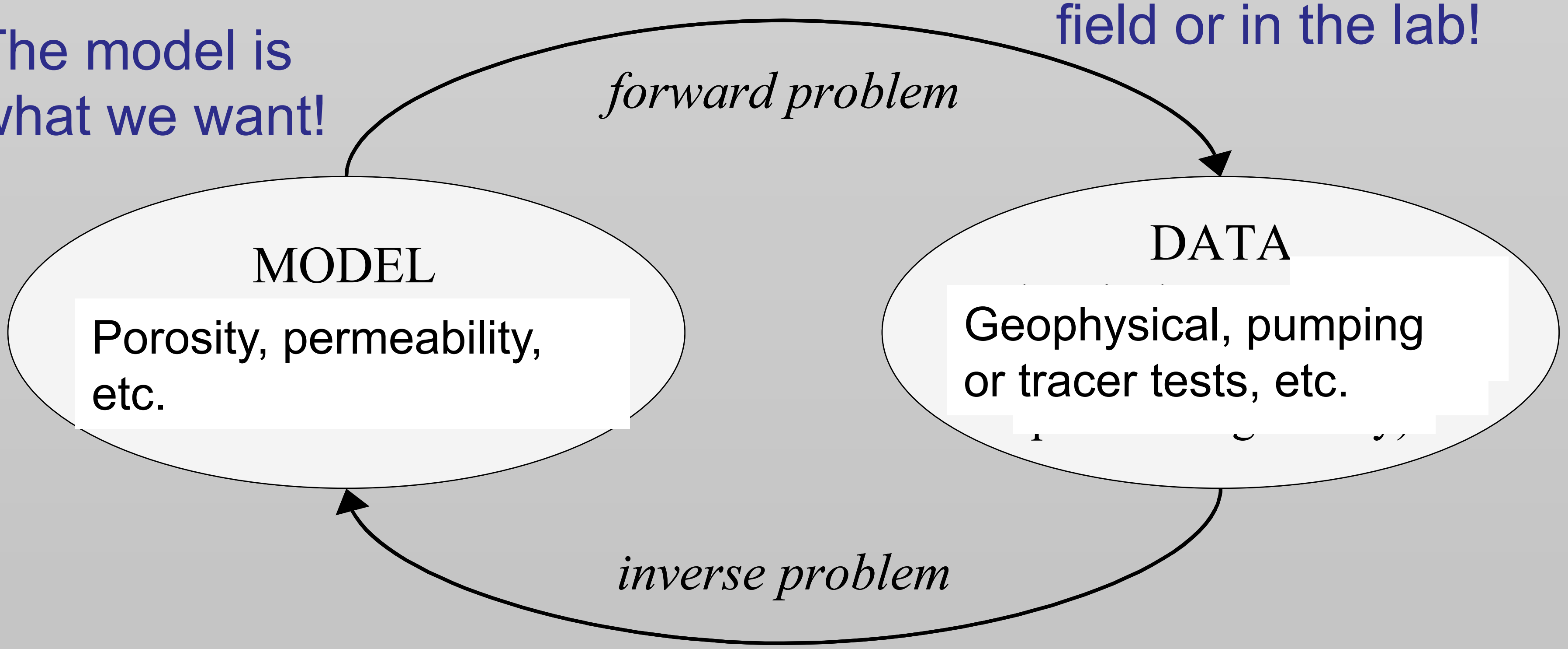




Forward and inverse problems

The data are what we can measure in the field or in the lab!

The model is what we want!



Adapted from Binley and Kemna (2005)

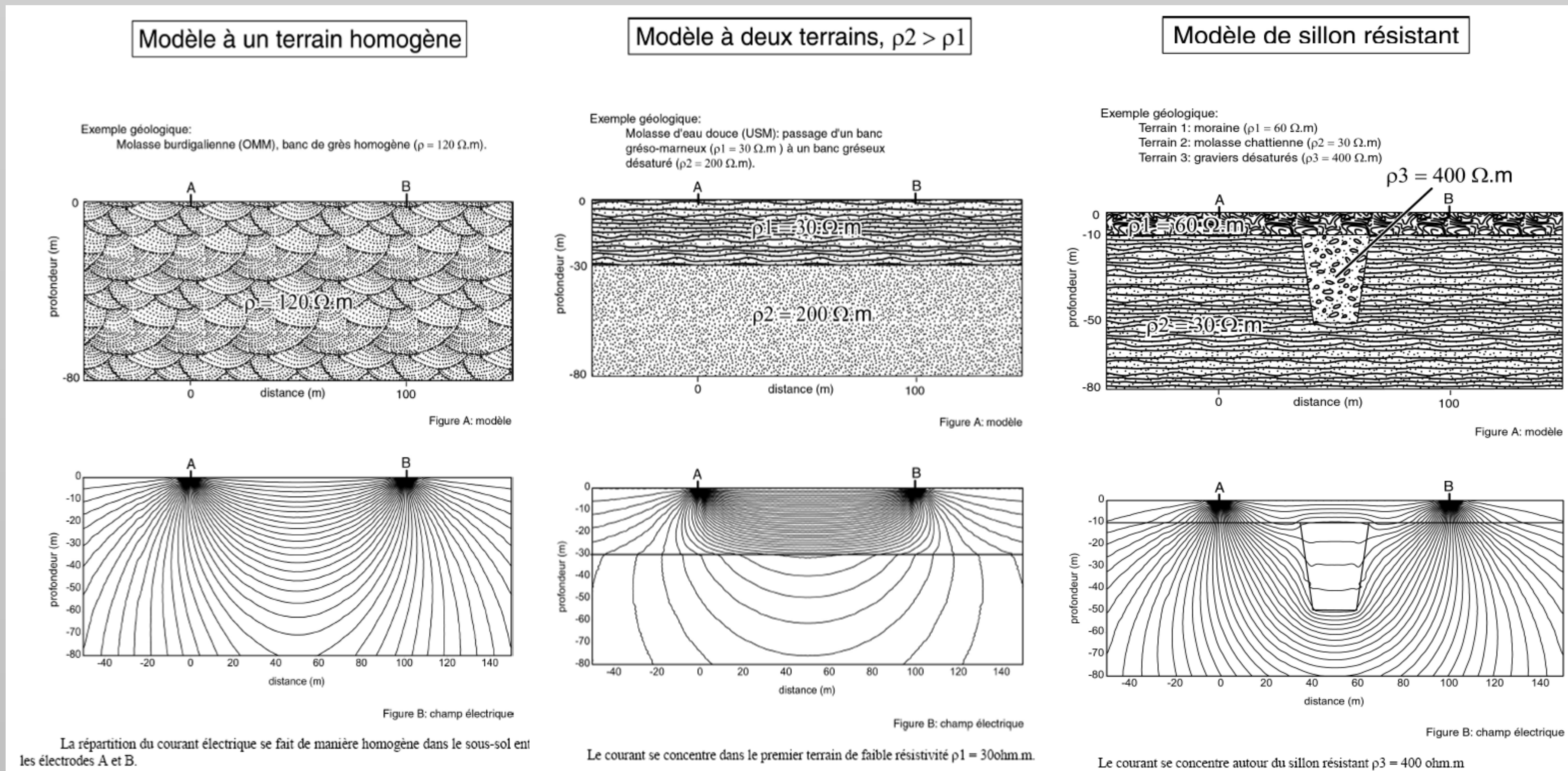
Linear and non-linear finite-dimensional forward problems

$$\mathbf{G}\mathbf{m}^{\mathrm{prop}} = \mathrm{d}^{\mathrm{sim}}$$ Linear case

The simulated data $\mathbf{d}^{\mathrm{sim}}$ $[N \times 1]$ can be calculated for a model $\mathbf{m}^{\mathrm{prop}}$ $[M \times 1]$ using the design matrix $\mathbf{G}$ $[N \times M]$. The design matrix describes the underlying physics and geometry of the experiment (e.g., using the finite-element method).

$$g(\mathbf{m}^{\mathrm{prop}}) = \mathbf{d}^{\mathrm{sim}}$$ Non-linear case

Ex. the electrical response is not proportional to electrical conductivity at a given location



Normal (Gaussian) distribution

- The most used probability density function is the normal (Gaussian) distribution

$$\rho(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2} \frac{(x - x_0)^2}{(\sigma)^2} \right),$$

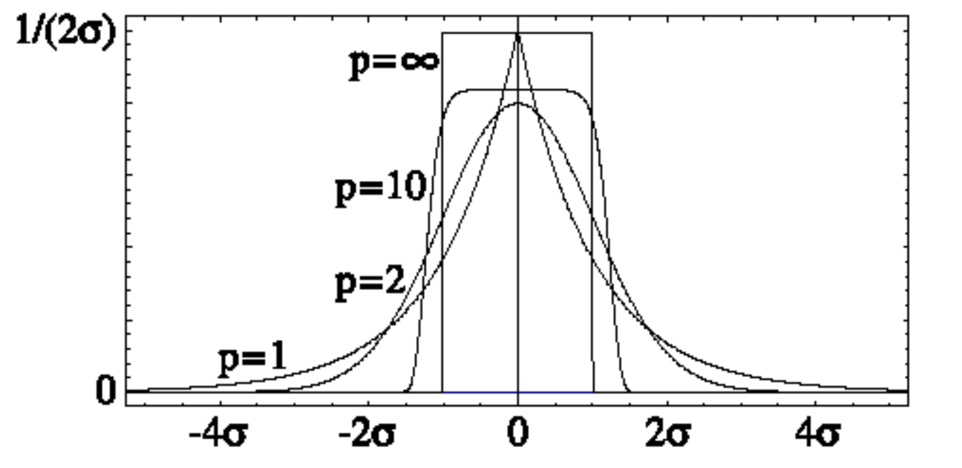
For one datum or model variable!

σ is the standard deviation;

x_0 is the mean value

- Probability density functions integrate to one.

More general



$$\rho_p(x) = \frac{p^{1-1/p}}{2\sigma_p \Gamma(1/p)} \exp\left(-\frac{1}{p} \frac{|x - x_0|^p}{(\sigma_p)^p} \right),$$

Generalized Gaussian

$$\rho_1(x) = \frac{1}{2\sigma_1} \exp\left(-\frac{|x - x_0|}{\sigma_1} \right),$$

Symmetric exponential

$$\rho_2(x) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{(x - x_0)^2}{(\sigma_2)^2} \right),$$

Gaussian function

$$\rho_\infty(x) = \begin{cases} 1/(2\sigma_\infty) & \text{for } x_0 - x_\infty \le \mathrm{x} \le x_o + \sigma_\infty \\ 0, & \text{otherwise} \end{cases}$$

Boxcar function

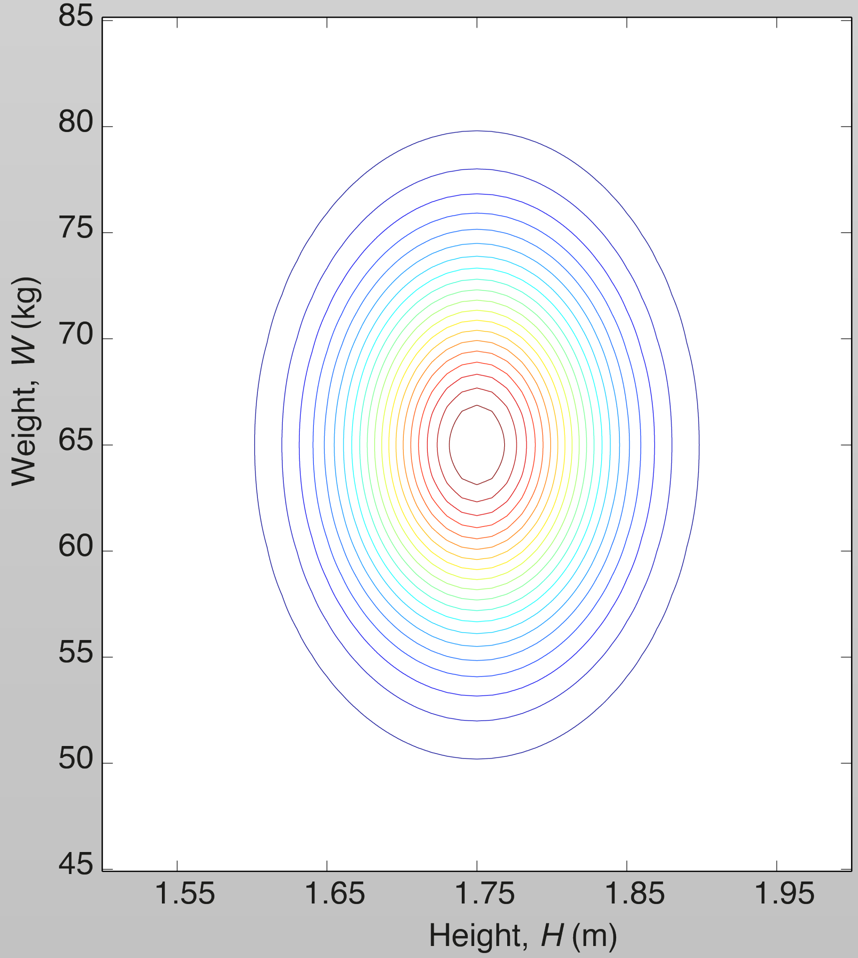
e.g., Tarantola, 2005

Joint probability density function (weight and height)

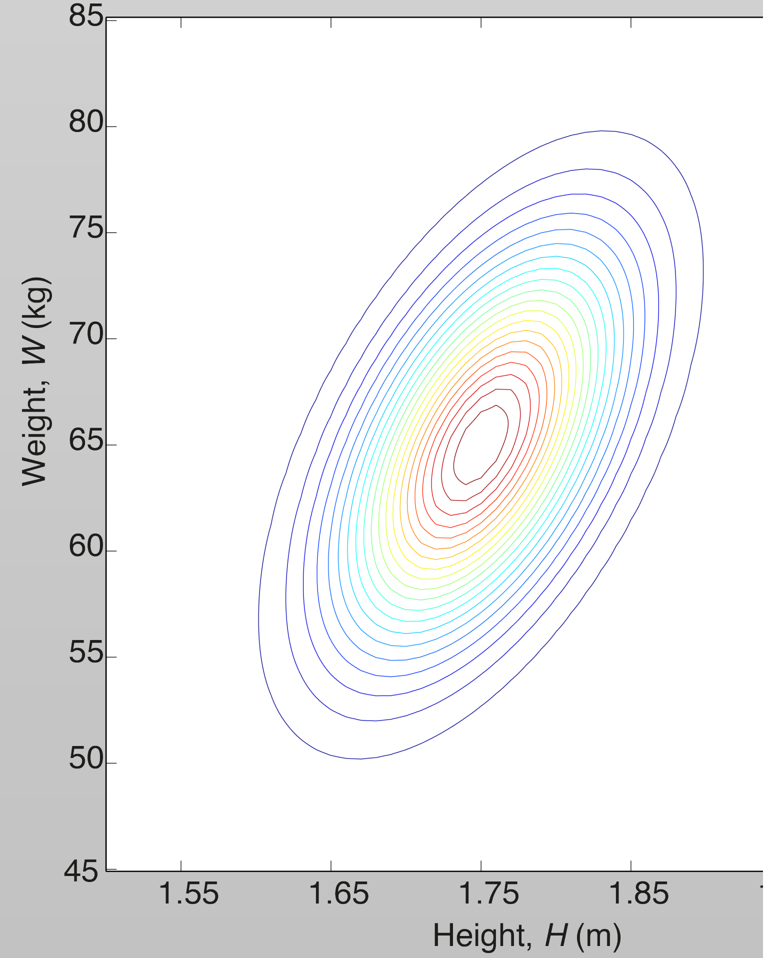
- Let $\rho(W,H)$ be the joint probability density function;
- For independent variables, we have that: $\rho(W,H) = \rho(W)\,\rho(H)$;
- This is unlikely for height and weight;
- It would imply that knowing something about height (e.g., a person is 2.50 m tall) does not carry any information about the persons supposed weight.

Joint probability density function

Uncorrelated variables



Correlated variables



Data errors, modeling errors and geological properties are correlated!

Conditional and marginalised pdfs

- Let $\rho(W \mid H)$ be the conditional pdf of W given H:

$$\rho(W, H) = \rho(H)\rho(W \mid H)$$

- The marginalised pdf of H is when all influence of W has been integrated away:

$$\rho(H) = \int \rho(H, W)\, dW$$

Shown as a figure a few slides back.

Bayes theorem

- From the rule of conditional probabilities, we have that

$$\rho(W,H) = \rho(W)\rho(H|W) = \rho(H)\rho(W|H)$$

- From this we can write:

$$\rho(W|H) = \frac{\rho(W)\rho(H|W)}{\rho(H)}$$

Tarantola and Valette (1982) describe a more general inversion framework that does not assume conditional probabilities (nor a distinction between m and d). Here, we follow a Bayesian formalism.

The complete solution to the inverse problem

$$\rho(\mathbf{m}|\mathbf{d}) = \frac{\rho(\mathbf{m})\rho(\mathbf{d}|\mathbf{m})}{\int \rho(\mathbf{d}|\mathbf{m})\rho(\mathbf{m})\,d\mathbf{m}}$$

$\rho(\mathbf{m}|\mathbf{d})$ Posterior probability density function

$\rho(\mathbf{m})$ Prior probability density function

$\rho(\mathbf{d}|\mathbf{m}) = L(\mathbf{m}|\mathbf{d})$ Likelihood function

$\int L(\mathbf{m}|\mathbf{d})\rho(\mathbf{m})\,d\mathbf{m}$ Evidence (marginal probability)

Often in practice

$$\rho(\mathbf{m}|\mathbf{d}) \propto \rho(\mathbf{m}) L(\mathbf{m}|\mathbf{d})$$

This proportionality is valid (i.e., the evidence is a constant) when the model parameterization is fixed during the inversion.

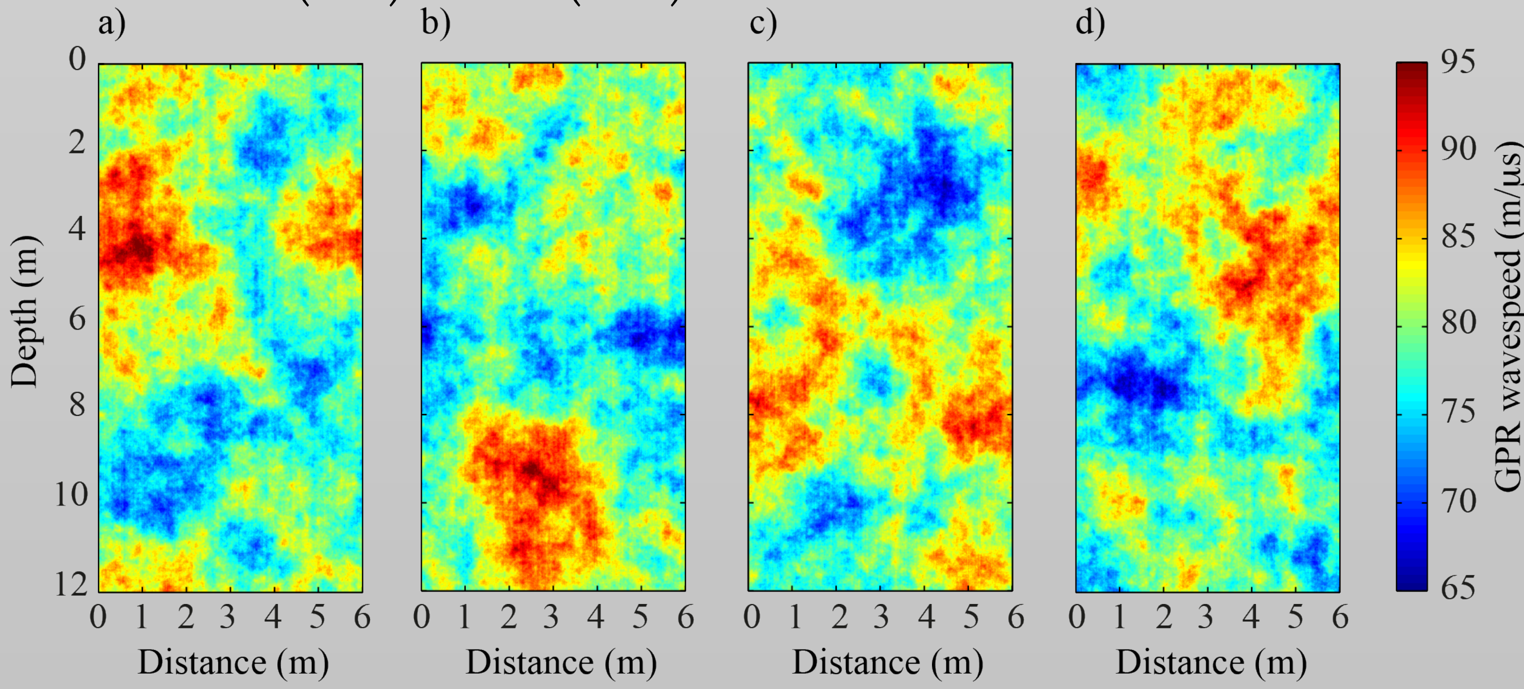
The normalizing constant underlies Bayesian model selection (see later)

Choosing a prior $\rho(\mathbf{m})$

- For positive physical constants, assuming a uniform prior of the logarithm of the property is the least “informative” (Jeffrey priors);
- The assumed distribution is the same regardless of if the property (e.g., electrical conductivity) or its reciprocal property (e.g., electrical resistivity) is used.
- It is very common to rely on two-point statistics in the form of multi-Gaussian prior models.

Prior model and likelihood function (Gaussian assumptions)

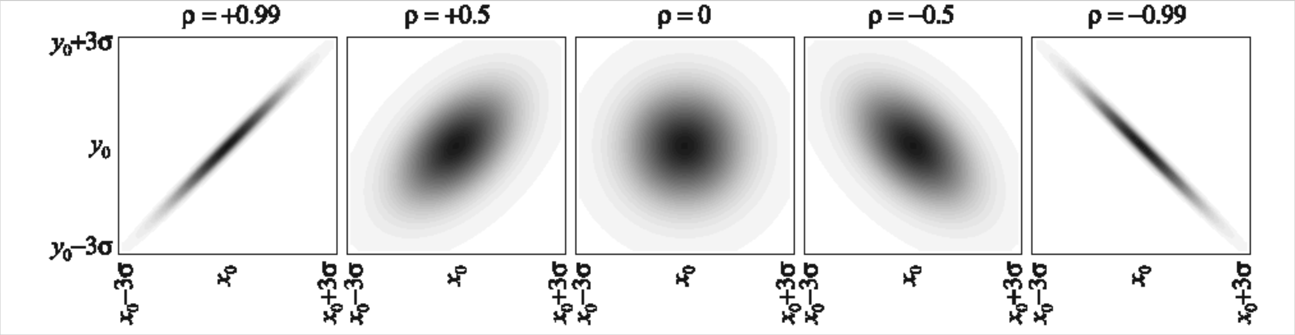
$$\rho(\mathbf{m}) = \frac{1}{(2\pi)^{M/2} \det(\mathbf{C}_{\mathrm{M}})^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{m}-\mathbf{m}_{\mathrm{prior}})^T \mathbf{C}_{\mathrm{M}}^{-1} (\mathbf{m}-\mathbf{m}_{\mathrm{prior}})\right)$$



$$L(\mathbf{m}|\mathbf{d}) = \frac{1}{(2\pi)^{N/2} \det(\mathbf{C}_{\mathrm{D}})^{1/2}} \exp\left(-\frac{1}{2}(g(\mathbf{m})-\mathbf{d})^T \mathbf{C}_{\mathrm{M}}^{-1} (g(\mathbf{m})-\mathbf{d})\right)$$

Pdfs for a joint Gaussian (normal) distribution: prior or likelihood

$$\rho(\mathbf{x}) = \left((2\pi)^N \det \mathbf{C}\right)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T \mathbf{C}^{-1}(\mathbf{x}-\mathbf{x}_0)\right)$$



The uncorrelated case: very common assumption

$$\rho(\mathbf{x}) = (2\pi)^{-N/2} \sigma_2^{-N} \exp\left(-\frac{1}{2}\left(\frac{\mathbf{x}-\mathbf{x}_0}{\boldsymbol{\sigma}_2}\right)^T \left(\frac{\mathbf{x}-\mathbf{x}_0}{\boldsymbol{\sigma}_2}\right)\right)$$

$$\rho(\mathbf{x}) = \left(\frac{1}{\sigma_1}\right)^{N/2} \exp\left(-\left|\frac{\mathbf{x}-\mathbf{x}_0}{\boldsymbol{\sigma}_1}\right|_1^1\right)$$

For the exponential (l_1) case

Posterior for a Gaussian/normal likelihood and prior (linear case)

In the linear Gaussian case, the posterior is fully described by a mean and a posterior covariance model. The analytical solution is:

$$\mathbf{d} = \mathbf{G}\mathbf{m}$$

$$\tilde{\mathbf{m}} = \left(\mathbf{G}^T \mathbf{C}_\mathrm{D}^{-1} \mathbf{G} + \mathbf{C}_\mathrm{M}^{-1}\right)^{-1} \left(\mathbf{G}^T \mathbf{C}_\mathrm{D}^{-1} \mathbf{d} + \mathbf{C}_\mathrm{M}^{-1} \mathbf{m}_\mathrm{prior}\right)$$

$$\tilde{\mathbf{C}}_\mathrm{M} = \left(\mathbf{G}^T \mathbf{C}_\mathrm{D}^{-1} \mathbf{G} + \mathbf{C}_\mathrm{M}^{-1}\right)^{-1}.$$

Tarantola (2005)

Rejection sampling: propose from prior, accept proportional to likelihood

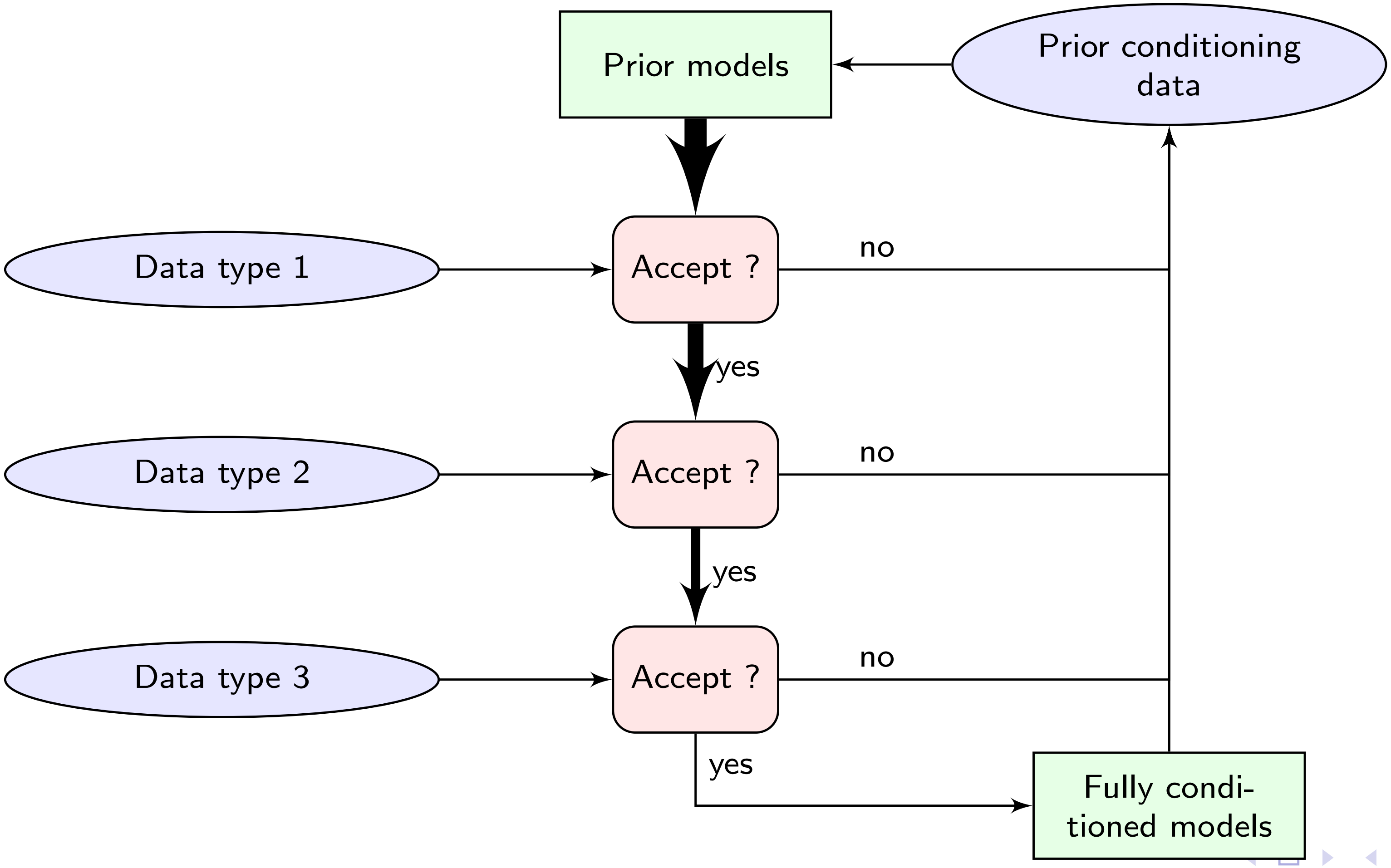
A. Draw $\mathbf{m}^{\text{prop}}$ from $\rho(\mathbf{m})$

B. Accept $\mathbf{m}^{\text{prop}}$ with probability $L(\mathbf{m}|\mathbf{d})/S_L$; return to A.

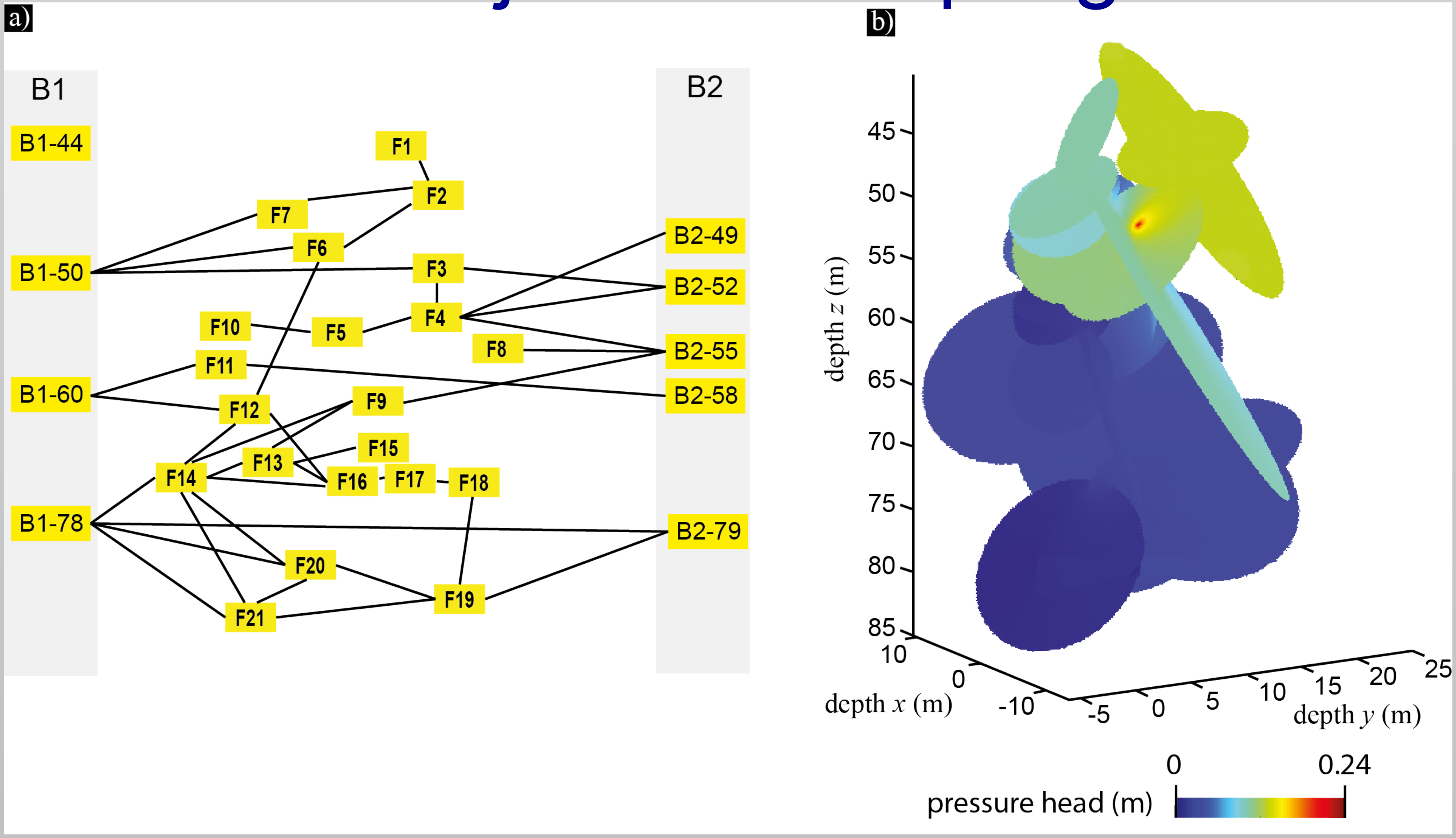
This is the only exact sampler of $\rho(\mathbf{m}|\mathbf{d})$ and it can deal with complex topologies. The supremum is the highest expected likelihood possible.

If supremum is chosen too low, then results are biased.
If chosen too high, then algorithm is inefficient.

- Independent sampling from posterior distribution
- Bayesian-style hierarchical conditioning to data
- Conditional independence between different data types



Integration of flow rates, tracers, borehole data, and geophysics with rejection sampling



Dorn et al. (2013) AWR

The curse of dimensionality

- Rejection sampling does not work well (acceptance rate is very low) in high parameter dimensions because the (hyper)volume of the prior space that contains significant likelihood becomes very small;
- For example, if 20% of the first dimension is significant (rejection sampling works), it would be less so in 100 dimensions, as the chance of hitting a high-likelihood area $0.2^{100}=10^{-70}$.

Metropolis-Hastings algorithm (Hastings, 1970)

- A. If at $\mathbf{m}_{curr}$, propose a move to $\mathbf{m}_{prop}$ according to a proposal distribution $q(\mathbf{m}_{curr} \rightarrow \mathbf{m}_{prop})$
- B. Calculate

Metropolis algorithm: if symmetric proposals

$$h = \min\left(1, \frac{L(\mathbf{m}_{prop}|\mathbf{d})\rho(\mathbf{m}_{prop})q(\mathbf{m}_{prop} \rightarrow \mathbf{m}_{curr})}{L(\mathbf{m}_{curr}|\mathbf{d})\rho(\mathbf{m}_{curr})q(\mathbf{m}_{curr} \rightarrow \mathbf{m}_{prop})}\right) = \min\left(1, \frac{L(\mathbf{m}_{prop}|\mathbf{d})\rho(\mathbf{m}_{prop})}{L(\mathbf{m}_{curr}|\mathbf{d})\rho(\mathbf{m}_{curr})}\right)$$

- C. Move to $\mathbf{m}_{prop}$ with probability h, else remain at $\mathbf{m}_{curr}$; go to A.

The resulting chain will asymptotically sample from a stationary distribution that is proportional to $\rho(\mathbf{m}|\mathbf{d})$.

Model proposals (states) with higher posterior probability are always accepted, but those with lower probability might also be accepted. After burn-in and convergence, the sampled states describe the posterior distribution.

Log-likelihoods

- The Metropolis ratio is not easy to calculate due to the limited numerical accuracy (e.g., smallest double-precision number is on the order of 10^{-323}):

$$h = \min\left(1, \frac{L(\mathbf{m}_{\text{prop}}|\mathbf{d})\,\rho(\mathbf{m}_{\text{prop}})}{L(\mathbf{m}_{\text{curr}}|\mathbf{d})\,\rho(\mathbf{m}_{\text{curr}})}\right)$$

- We are only interested in obtaining the ratio. Can be obtained as (numerical trick)

$$\frac{L(\mathbf{m}_{\text{prop}}|\mathbf{d})\,\rho(\mathbf{m}_{\text{prop}})}{L(\mathbf{m}_{\text{curr}}|\mathbf{d})\,\rho(\mathbf{m}_{\text{curr}})} = \exp\Big(\log\big(L(\mathbf{m}_{\text{prop}}|\mathbf{d})\big) + \log\big(\rho(\mathbf{m}_{\text{prop}})\big) - \log\big(L(\mathbf{m}_{\text{curr}}|\mathbf{d})\big) - \log\big(\rho(\mathbf{m}_{\text{curr}})\big)\Big)$$

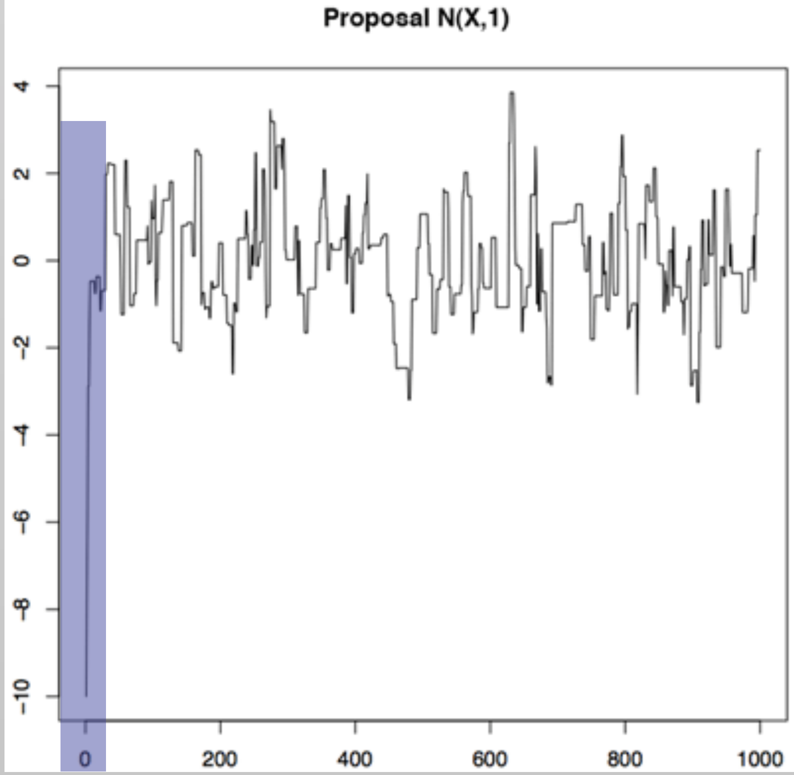
Extension to Metropolis-Hastings is straightforward.

Acceptance rate (AR)

- One should monitor the acceptance rate (the proportion of actual updates in the MCMC chain).
- The proposal distribution is often chosen to get an AR of 25% (10-40% is typically good).
- As long as the AR is above 0 or below 1, the chain will converge at some point, but perhaps after too many steps to be possible/ convenient (say 10^{10} steps).

Burn-in

- It takes time before the chain is independent of the starting point and starts to sample proportionally to the posterior.
- This is called the burn-in time and the preceding steps should be removed.
- Often approximated as the time when the sampled posterior probabilities start to fluctuate around a constant value.

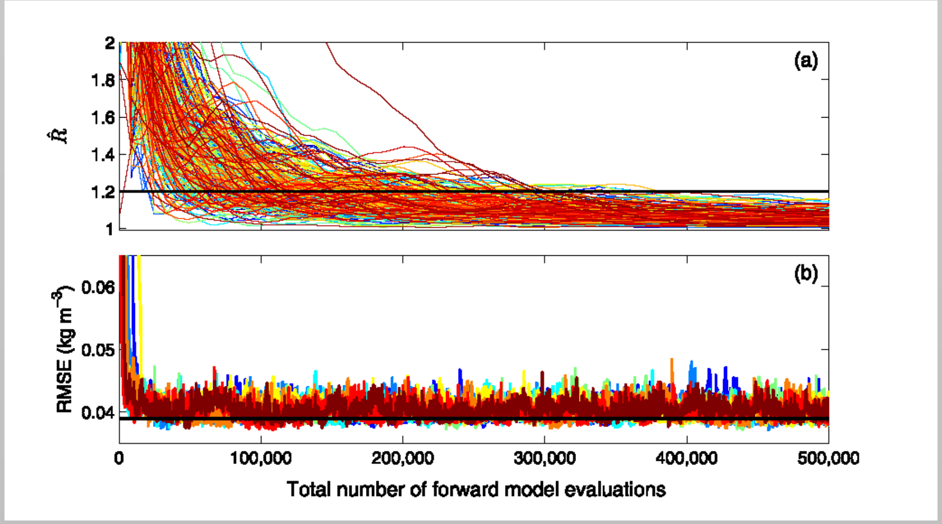


The shaded area indicates the burn-in time. It can be a few samples or many (hundreds of) thousands of samples.

Assessing MCMC convergence

- A common measure to assess convergence is the potential scale reduction factor by Gelman and Rubin (1992).
- It compares the average within-chain variance with the across-chain variance of the (within chain) means for the second half of the chain.

Laloy et al. (2015)



Convergence when $\hat{R} < 1.2$ for all parameters; Example to the left suggests that burn-in is only 5% of the time needed to have a proper sampling of the posterior.

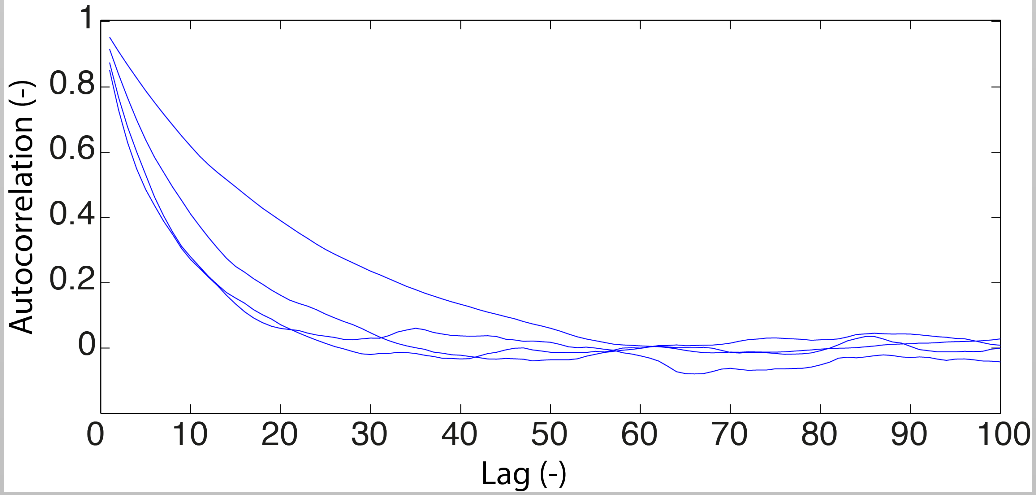
Reaching the posterior vs. exploring the posterior

- The first stage in MCMC corresponds to locating the posterior (burn-in period) and the next step corresponds to sampling the posterior distribution;
- Sometimes, one might only aim at finding the global minimum (not the full posterior) or a set of realizations that explain the data without a formal assessment of uncertainty;
- Global optimization methods are then suitable.
- **If convergence is not achieved, then one is left with a few samples from the posterior with no ability to make a full probabilistic assessment!**

Auto-correlation

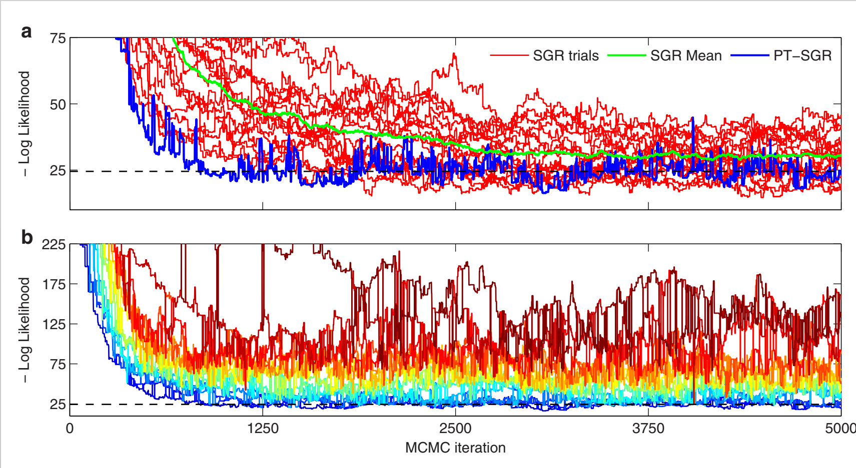
- Neighboring samples/states in an MCMC chain are highly correlated. **That is, the number of independent draws from the posterior are much fewer than the number of samples**;
- The auto-correlation function describes how the correlation decrease as a function of the lag τ:

$$R(\tau) = \frac{E\left[\left(X_t - \mu\right)\left(X_{t+\tau} - \mu\right)\right]}{\sigma^2}$$

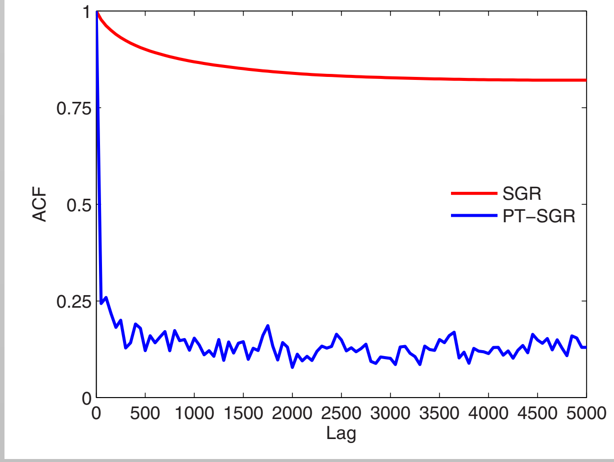


Better exploration with parallel tempering

- Tempering raises the posterior distribution or only the likelihood with the inverse of a temperature *T;*
- Several chains are developed in parallel using different temperatures;
- Within-chain and between chain proposals are used to exchange information.



Laloy et al. (2016)



$$p(\mathbf{m}|\mathbf{d},T) \propto p(\mathbf{m})L(\mathbf{m}|\mathbf{d})^{1/T}$$

Constraint 1 on MCMC: Ergodicity

- Two conditions are needed for an MCMC chain to converge asymptotically: Ergodicity and detailed balance.
- **Ergodicity:** the chain is irreducible (it can get from any state to any other state after a number of steps), aperiodic (the chain does not repeat itself), it is positive recurrent (it will return to a given state after a finite number of steps).

Constraint 2 on MCMC: Detailed balance

- Detailed balance states that the transition kernel to move from one state to another (essentially a combination of a model proposal step and an acceptance step) ensures that:

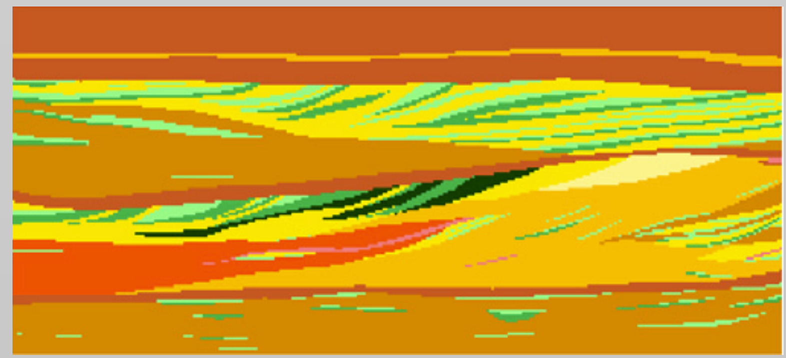
$$T\left(\mathbf{m}^*\middle|\mathbf{m}\right)\rho\left(\mathbf{m}\middle|\mathbf{d}\right)=T\left(\mathbf{m}\middle|\mathbf{m}^*\right)\rho\left(\mathbf{m}^*\middle|\mathbf{d}\right)$$

- Example, if $\rho\left(\mathbf{m}^*\middle|\mathbf{d}\right)$ is twice as likely as $\rho\left(\mathbf{m}\middle|\mathbf{d}\right)$ then detailed balance states that $T\left(\mathbf{m}^*\middle|\mathbf{m}\right)/T\left(\mathbf{m}\middle|\mathbf{m}^*\right)$ must be 2.

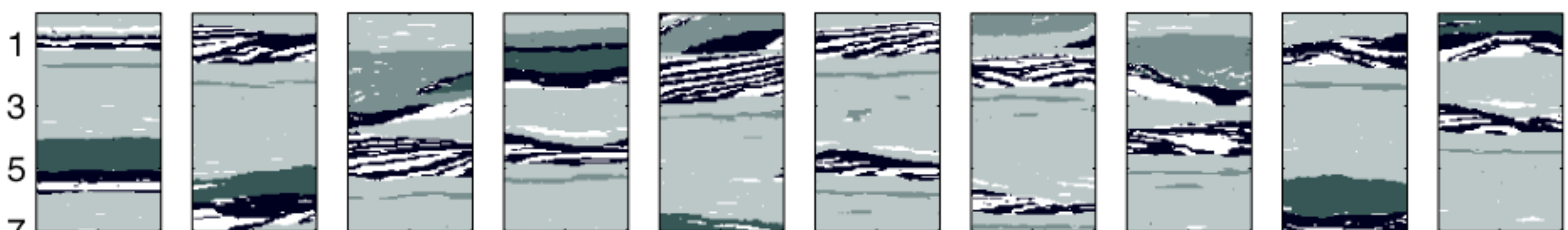
Choosing a prior $\rho(\mathbf{m})$

- The prior doesn't need to be defined mathematically; it may be represented by geostatistical simulations.

Bayer *et al*. (2011)



Comunian *et al*. (2011)

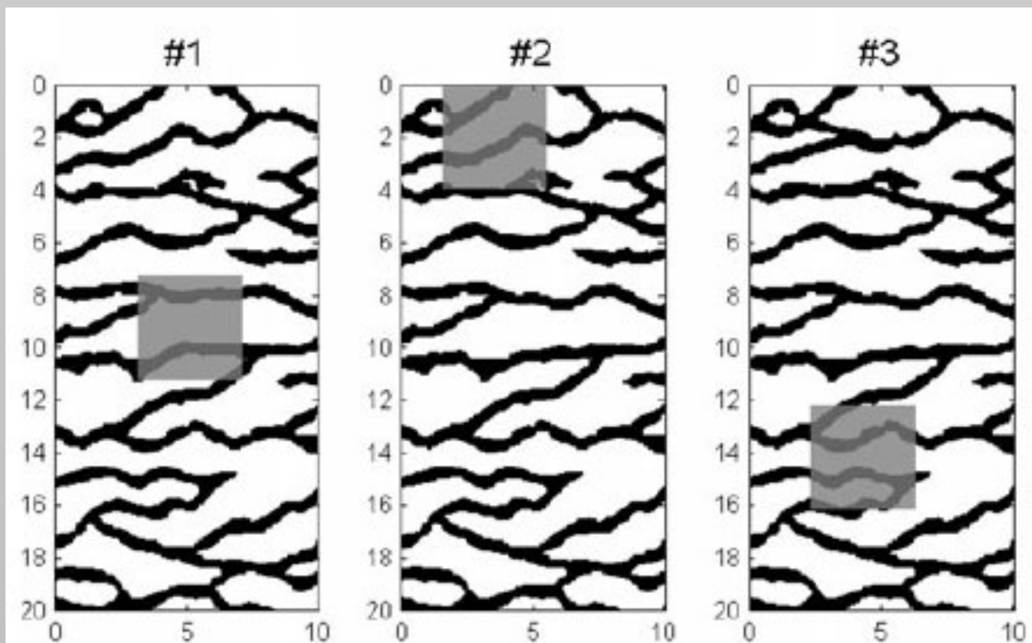


The extended Metropolis algorithm (Mosegaard and Tarantola, 1995)

- If at $\mathbf{m}_{\mathrm{curr}}$, conditional resimulation of a portion of the model according to the prior by using a geostatistical algorithmresimulation;
- Calculate

$$h = \min\left(1, \frac{L\left(\mathbf{m}_{\mathrm{prop}} \middle| \mathbf{d}\right)}{L\left(\mathbf{m}_{\mathrm{curr}} \middle| \mathbf{d}\right)}\right)$$

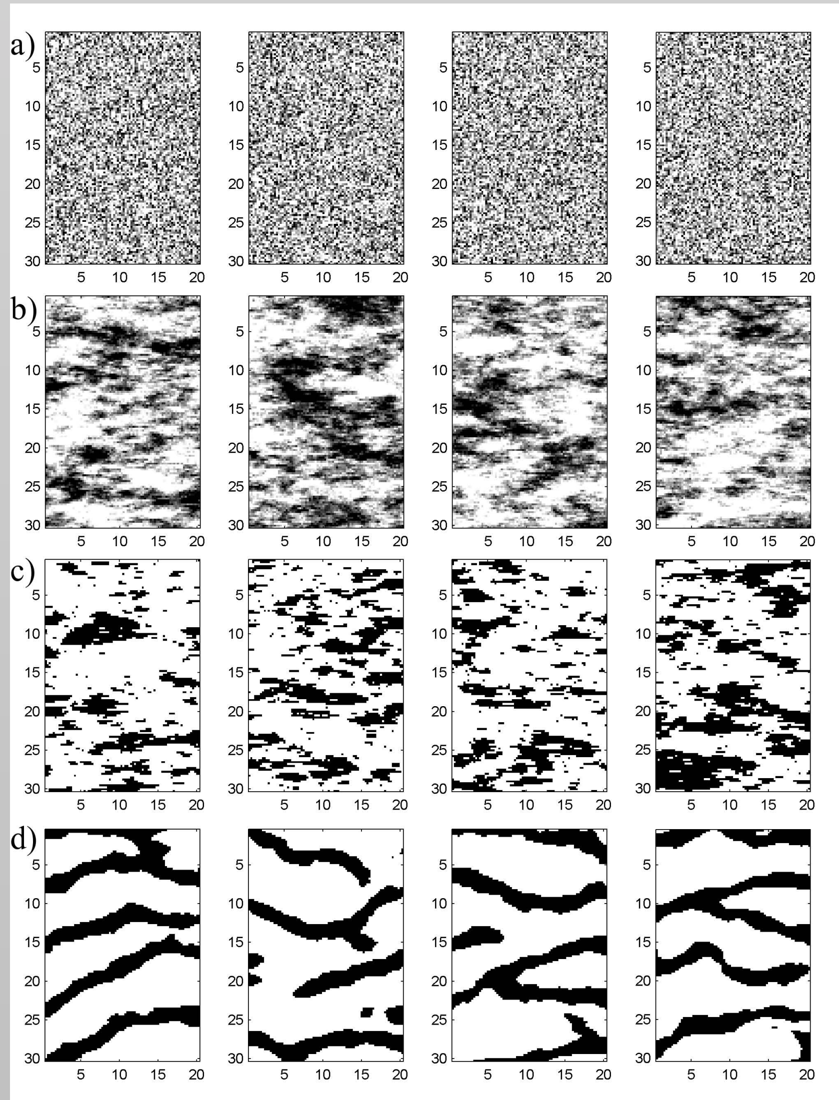
- Move to $\mathbf{m}_{\mathrm{prop}}$ with probability h, else remain at $\mathbf{m}_{\mathrm{curr}}$; go to A.



General approach, but can be inefficient because of (i) low acceptance rate and (ii) costly geostatistical resimulation.

Hansen et al. (2012) Comput. Geosci.

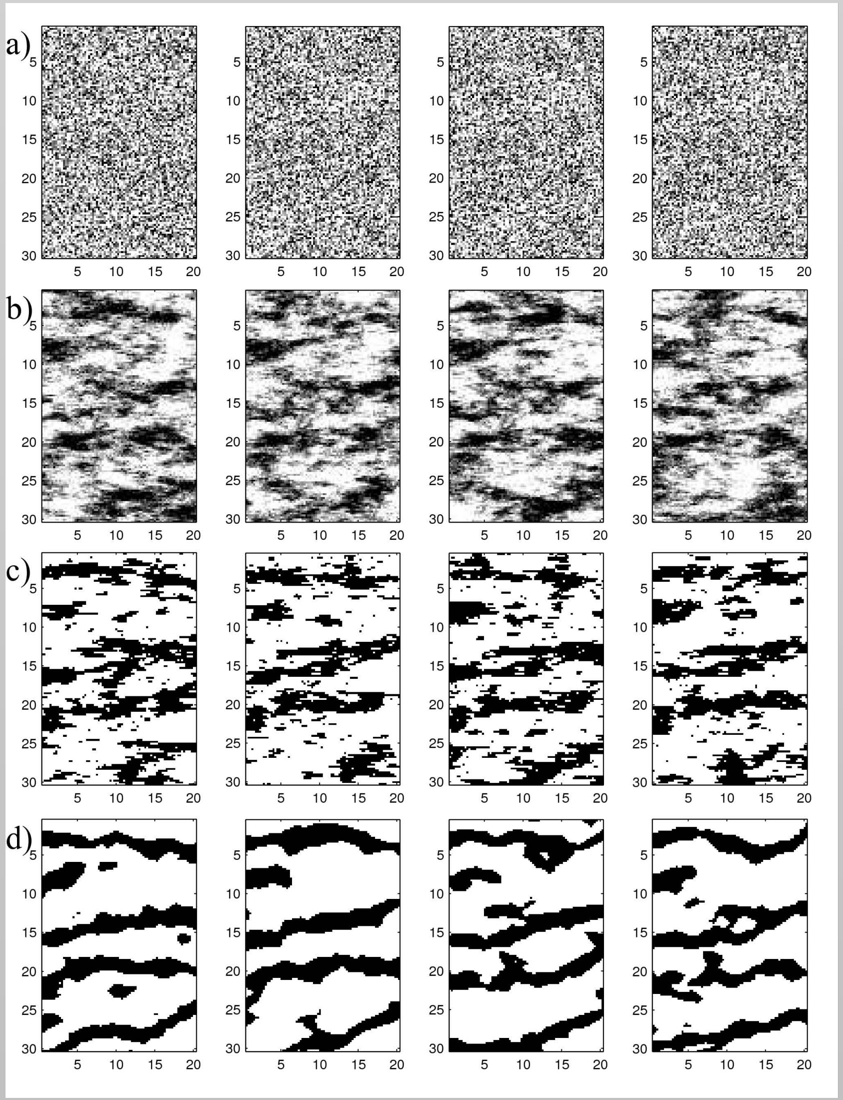
Some prior states for different prior models (different lines)



Hansen et al. (2012) Comput. Geosci.

Some posterior states for different prior models (different lines)

Crosshole GPR data are used in these examples.



Hansen et al. (2012) Comput. Geosci.

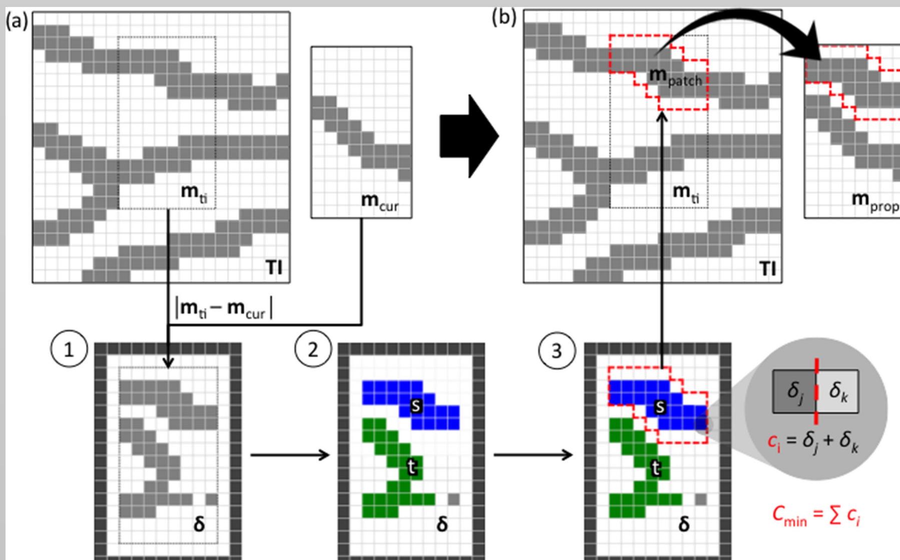
The spatial statistics of the posterior are largely determined by the prior;

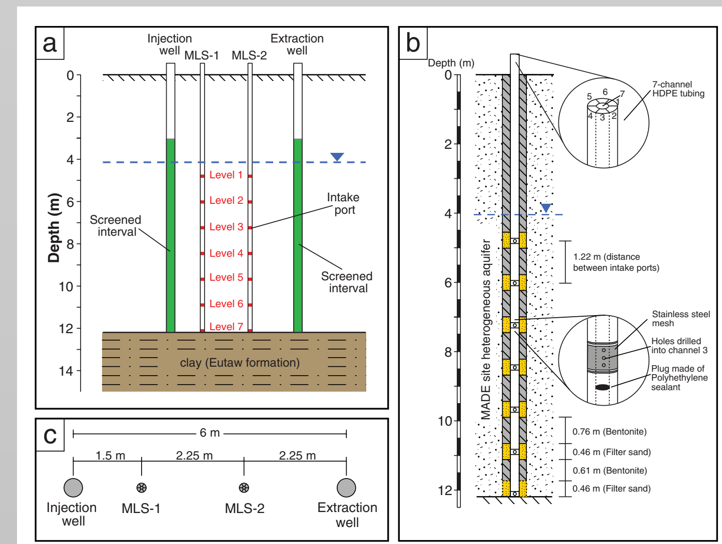
Any uncertainty quantification (UQ) is highly dependent on the prior;

Insufficient research on "geological" priors.

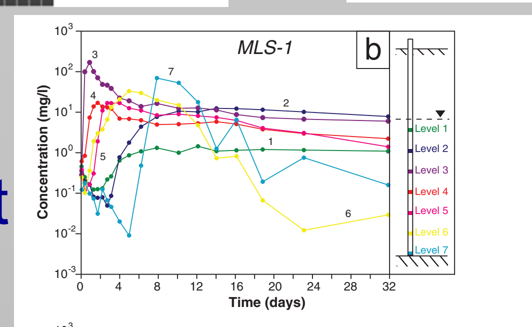
Application to the MADE site using graph cuts

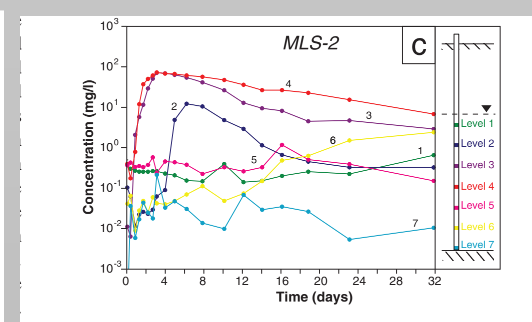
- Model proposals with graph cuts (Zahner et al., 2016) bring down the model proposal time to <1% than classical multiple-point statistics resimulation

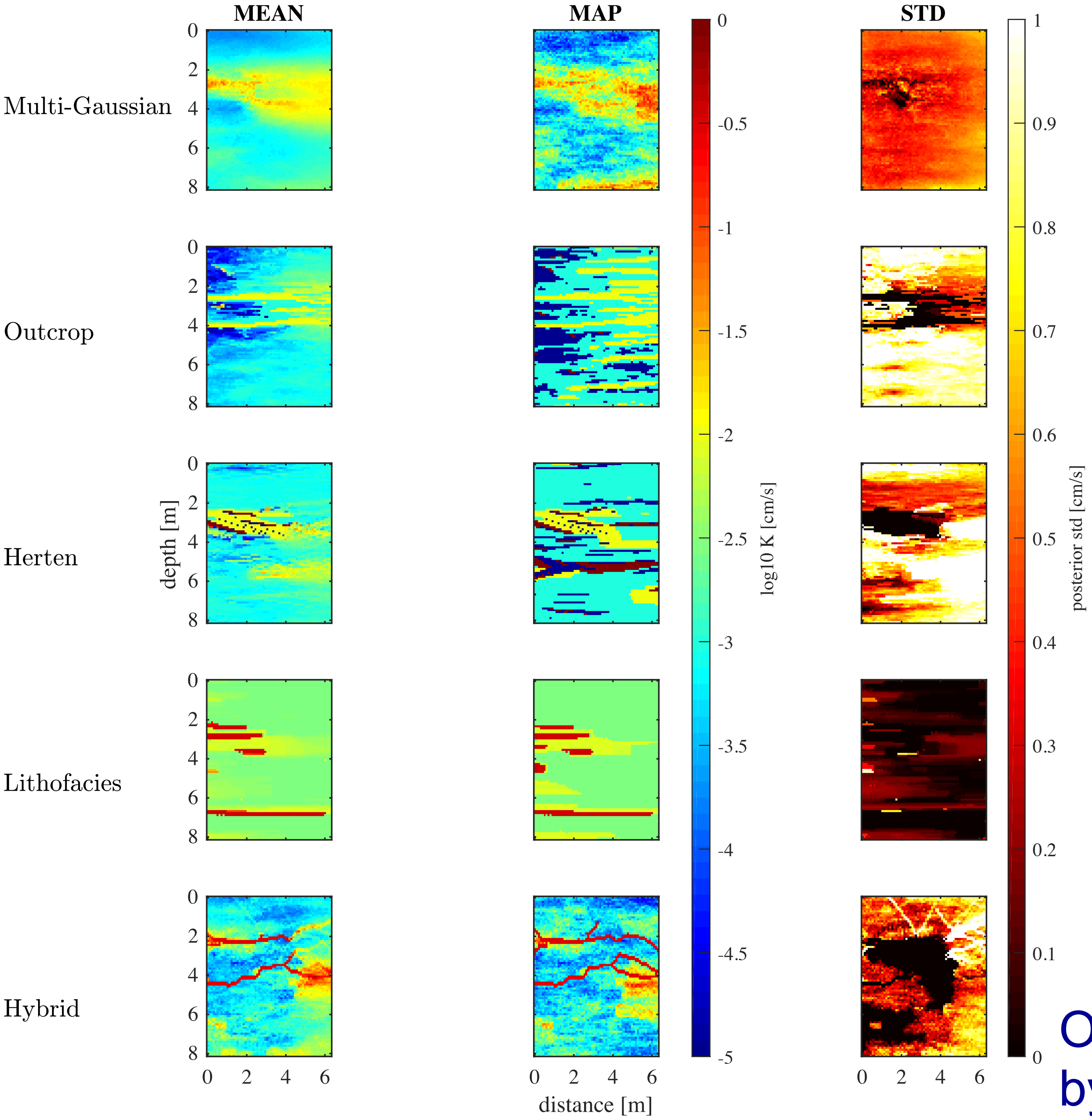




Data from Bianchi et al. (2011): MADE-5 experiment







High likelihood

Low likelihood

High likelihood

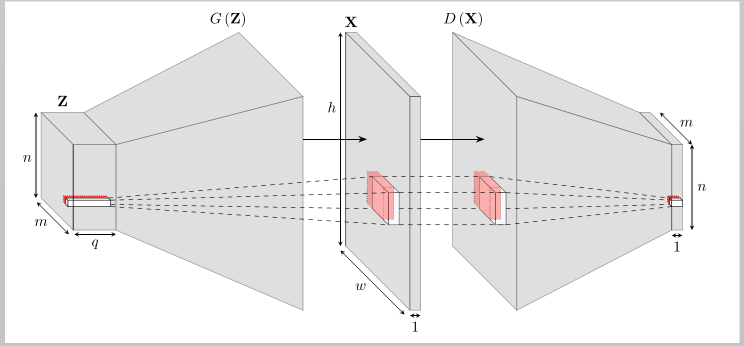
Low likelihood

High likelihood

Ongoing PhD work by Carlotta Brunetti

Low-dimensional representation of priors using deep learning

- If complex priors can be captured by uncorrelated (e.g., standard normal, bounded uniform) coefficients, then any state-of the-art MCMC method can be used.
- Wavelets, discrete cosine transforms, etc. do not offer this, but deep learning may.



Spatial generative adversarial neural networks (SGANs) can be trained on a training image (TI) to produce a low-dimensional representation of uncorrelated standard normals (Laloy et al., 2018).

See also:
Mosser et al. on arxiv.org

Examples (Laloy et al., 2018)

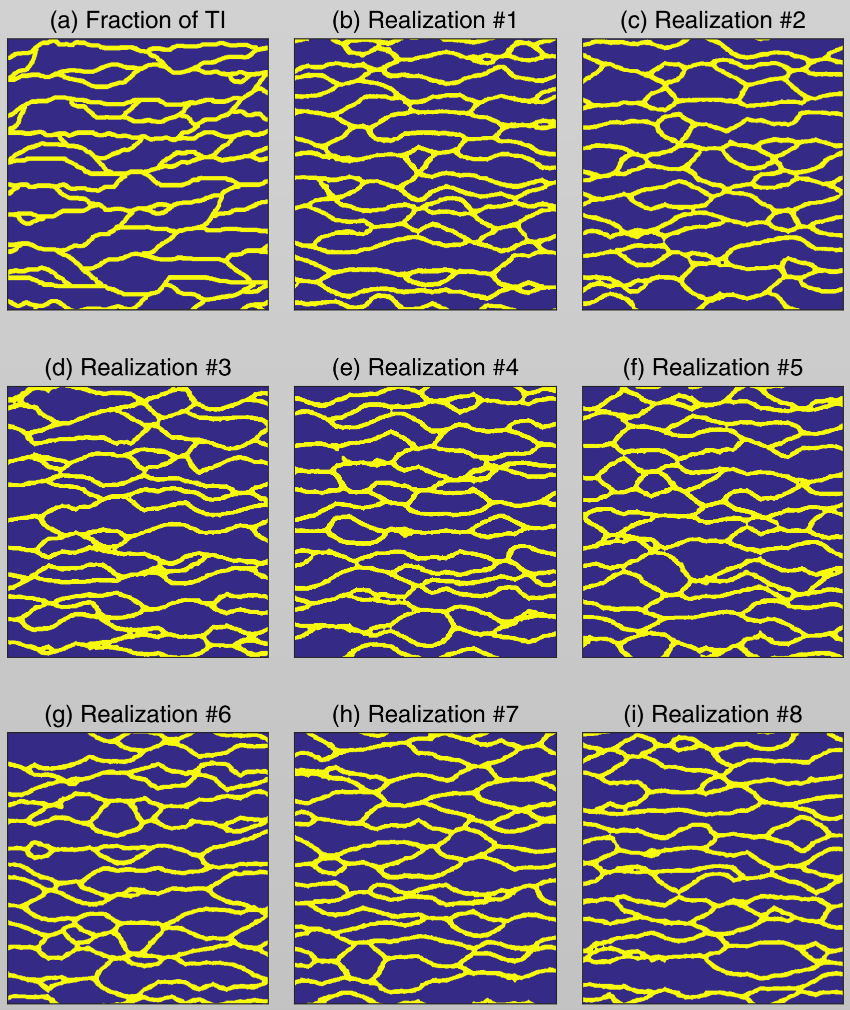


Figure 3. (a) Fraction of size 609 × 609 of the TI shown in Figure 2a and (b–i) randomly chosen 609 × 609 realizations derived by our SGAN. Each realization is generated by sampling 400 random numbers from a uniform distribution, $U(-1, 1)$.

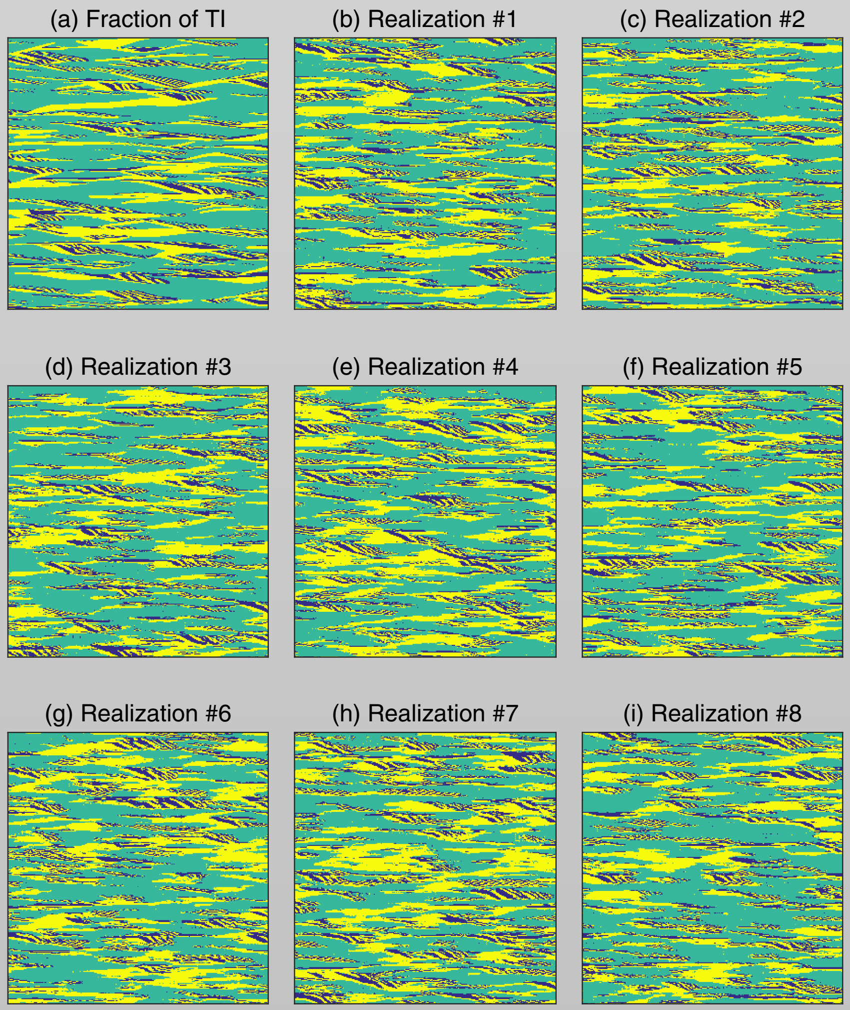


Figure 6. (a) Fraction of size 289 × 289 of the TI shown in Figure 2b and (b–i) randomly chosen 289 × 289 realizations derived by our SGAN. Each realization is generated by sampling 300 random numbers from a uniform distribution, $U(-1, 1)$.

Lithological tomography (Bosch, 1999)

$\mathbf{m}_{\text{structure}}$ is lithology (e.g., porosity, permeability)
$\mathbf{m}_{\text{geophysics}}$ is geophysical properties (e.g., electrical conductivity)

Joint posterior probability density function

Advanced geostatistics enters here

$$\rho\left(\mathbf{m}_{\text{structure}},\mathbf{m}_{\text{geophysics}}\middle|\mathbf{d}_{\text{hydrology}},\mathbf{d}_{\text{geophysics}}\right) \propto \rho\left(\mathbf{m}_{\text{structure}}\right)\times$$

$$\rho\left(\mathbf{m}_{\text{geophysics}}\middle|\mathbf{m}_{\text{structure}}\right) L\left(\mathbf{m}_{\text{structure}}\middle|\mathbf{d}_{\text{hydrology}}\right) L\left(\mathbf{m}_{\text{geophysics}}\middle|\mathbf{d}_{\text{geophysics}}\right)$$

Petrophysics

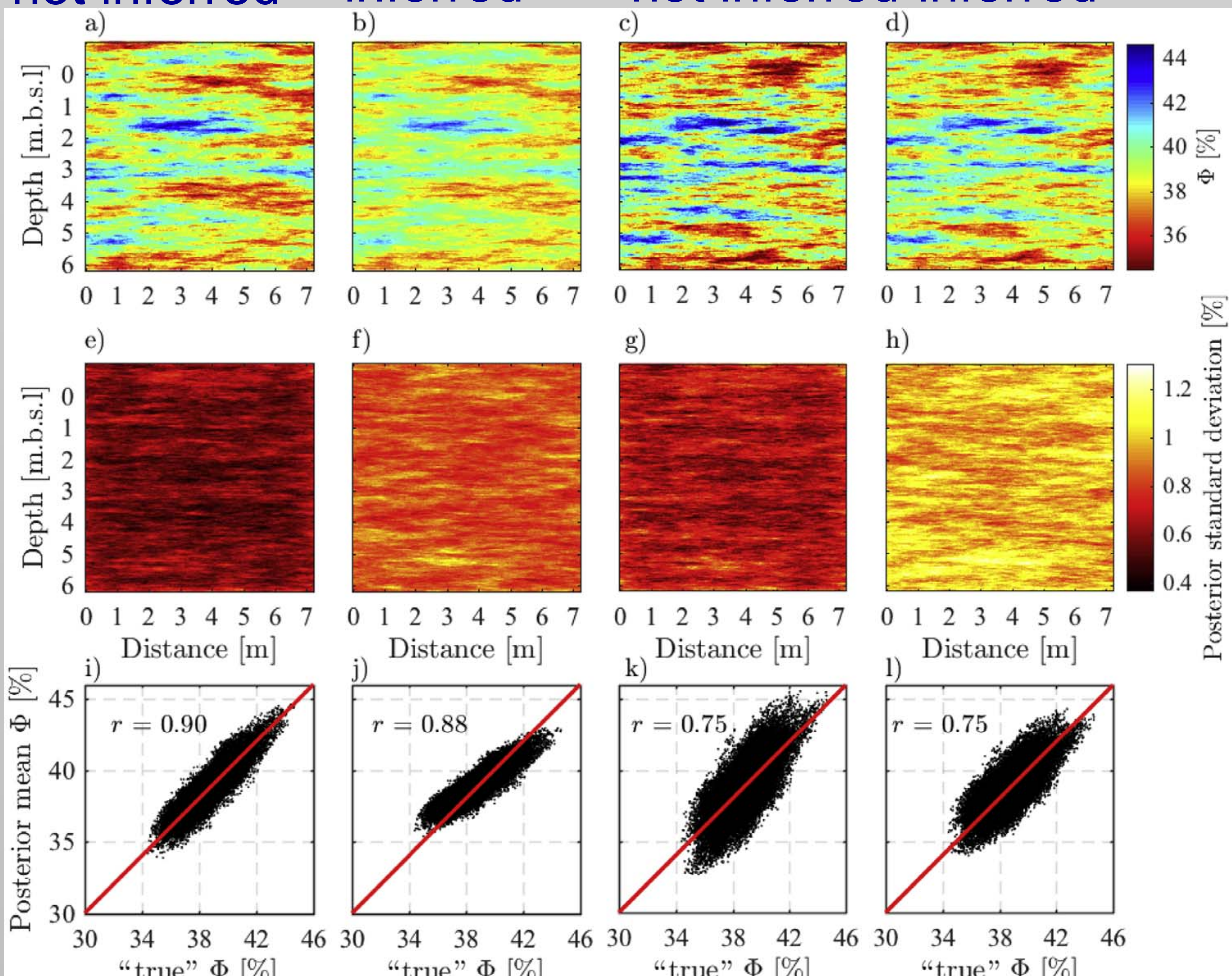
Comparison with hydrological data

Comparison with geophysical data

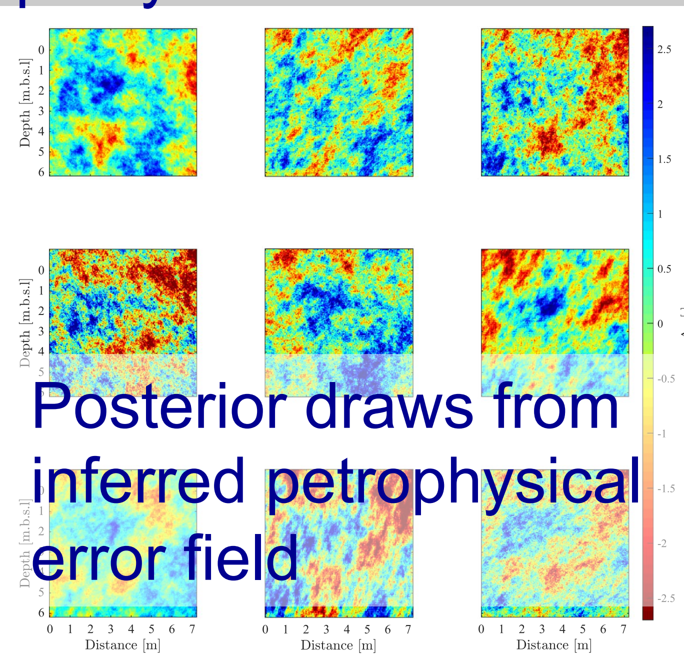
$\rho\left(\mathbf{m}_{\text{structure}}\middle|\mathbf{d}_{\text{hydrology}},\mathbf{d}_{\text{geophysics}}\right)$ obtained by marginalization

Crosshole GPR and petrophysical error (Brunetti and Linde, 2018)

not present not inferred	not present inferred	present not inferred	present inferred



Bias and overoptimistic errors if petrophysical errors are ignored;
Statistics can only be partly resolved.



Posterior draws from inferred petrophysical error field

Challenge 1: Model errors and intractable likelihoods

- Why not directly infer:

$$\rho\left(\mathbf{m}_{\text{structure}} \middle| \mathbf{d}_{\text{geophysics}}\right) \propto \rho\left(\mathbf{m}_{\text{structure}}\right) L\left(\mathbf{m}_{\text{structure}} \middle| \mathbf{d}_{\text{geophysics}}\right).$$

If petrophysical errors are important, then $L\left(\mathbf{m}_{\text{structure}} \middle| \mathbf{d}_{\text{geophysics}}\right)$ is much less peaky than $L\left(\mathbf{m}_{\text{geophysics}} \middle| \mathbf{d}_{\text{geophysics}}\right)$, so inference should be much easier. Right?

- Straightforward for linear Gaussian problems, but likelihood is otherwise intractable.
- Solutions? Linearization around present model (approximate), pseudo-marginal MCMC (Beaumont, 2003), approximate Bayesian computation (Marjoram, 2003). Hopefully, more on this in the next Summer School.

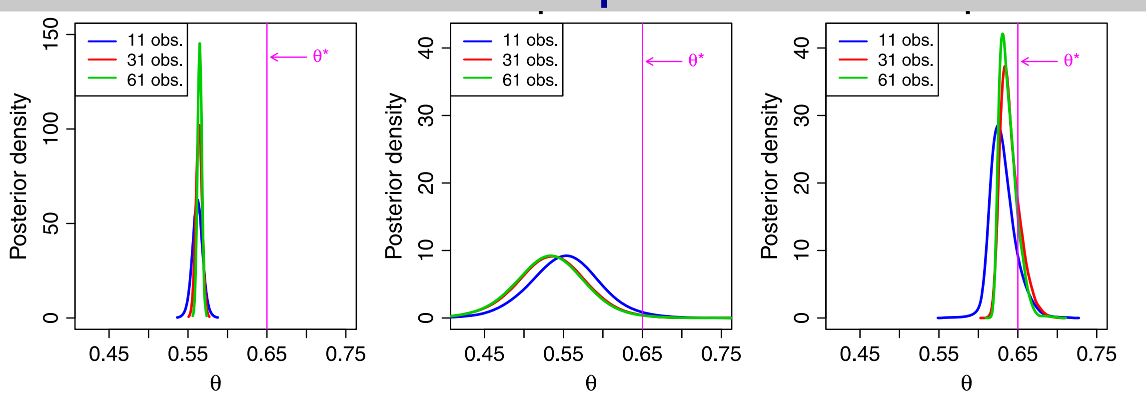
Challenge 2: Model errors

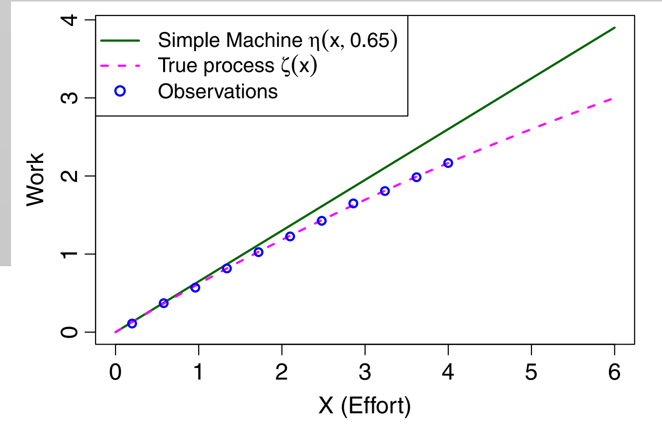
- Approaches to handle model errors (petrophysical errors, simplified physics, discretization, numerical) exist (Kaipio and Somersalo, 2007; Calvetti et al., 2014) that work well for idealized situations.

No model error treatment

Gaussian Process description

Informed Gaussian Process





Allowing for model errors inflate;
Informed priors needed
to remove bias.

Brynjarsdottir and O'Hagan, 2014

Towards model selection

- Comparison of alternative conceptual models is ideally based on the Bayes factor (ratio of evidences/marginal likelihoods:

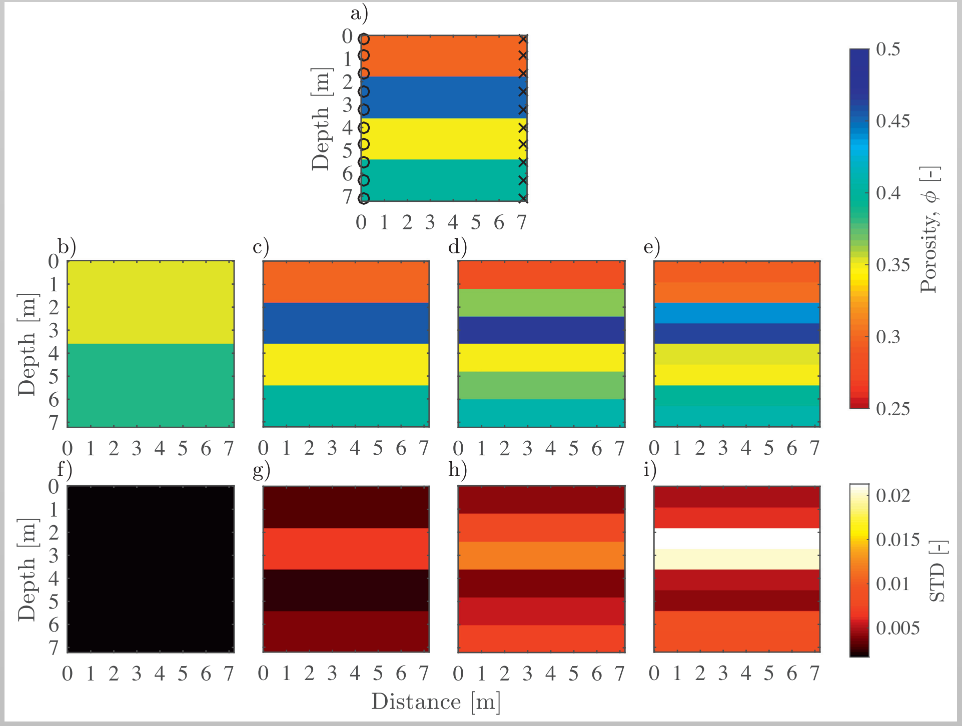
$$B_{ij} = \frac{\rho(\mathbf{d}|H_i)}{\rho(\mathbf{d}|H_j)} = \frac{\int L(\mathbf{m}_i|\mathbf{d},H_i)\rho(\mathbf{m}_i|H_i)d\mathbf{m}_i}{\int L(\mathbf{m}_j|\mathbf{d},H_j)\rho(\mathbf{m}_j|H_j)d\mathbf{m}_j} \qquad \frac{\rho(H_i|\mathbf{d})}{\rho(H_j|\mathbf{d})} = \frac{\rho(\mathbf{d}|H_i)}{\rho(\mathbf{d}|H_j)}\frac{\rho(H_i)}{\rho(H_j)}$$

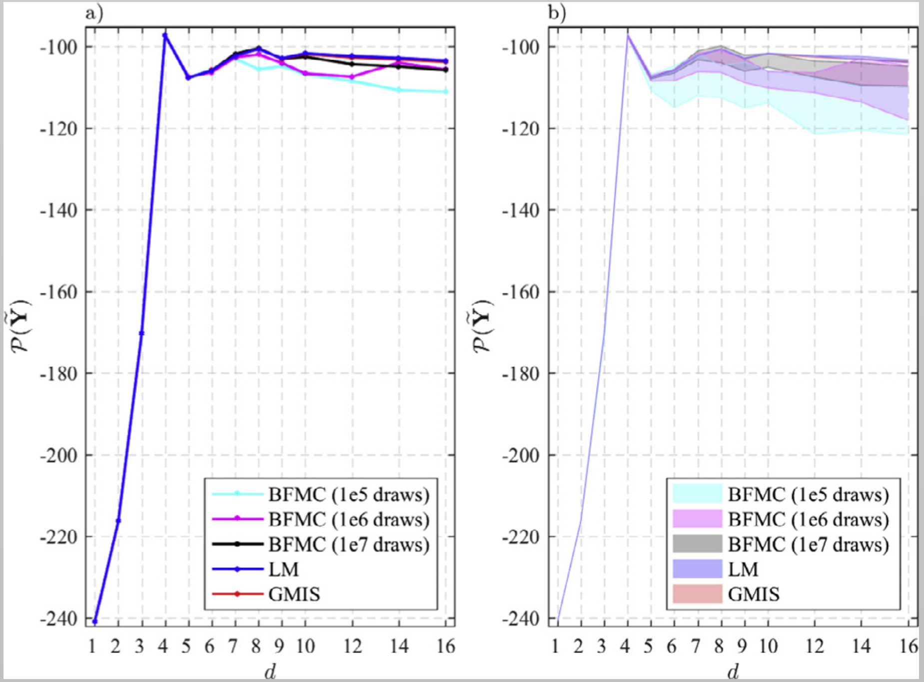
- Integrals can be solved by Monte Carlo, Nested sampling (Skilling, 2006), Laplace-Metropolis (quadratic expansion around MAP), importance sampling from posterior (Volpi et al., 2017), thermodynamic integration (Lartillot and Philippe, 2008), stepping-stone sampling (Xie et al., 2011).

Excellent introduction is given by Schöniger et al. (2014), WRR

A GPR toy example (Brunetti et al., 2017)

- True number of layers resolved. See paper for field-application with multi-Gaussian priors. Carlotta is currently addressing model selection using the tracer test at MADE with complex geological priors.





Summary

- Posterior distributions are only meaningful if priors and likelihoods are well chosen (Garbage In, Garbage Out);
- Geological realism is needed for meaningful uncertainty quantification (deep-learning: low-dimensional representations; efficient MCMC);
- Unaccounted model errors (discretization, numerical, petrophysical) give overly optimistic uncertainty estimates and estimates are biased;
- Bayesian model selection can help to answer questions about the most appropriate process description and conceptual model.

Suggested reading

Reader's Digest

More heavy

- Linde, N., D. Ginsbourger, J. Irving, F. Nobile, A. Doucet, 2017. On uncertainty quantification in hydrogeology and hydrogeophysics. *AWR 110*, 161-181.
- Linde, N., P. Renard, T. Mukerji, J. Caers, 2015. Geological realism in hydrogeological and geophysical inverse modeling: a review. *AWR 86*, 86-101.
- Sambridge, M., K. Mosegaard, 2002. Monte Carlo methods in geophysical inverse problems. RG, 40, 1009.
- Tarantola, A. 2005. Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM.
- Tarantola, A., B. Valette. 1982. Inverse problems=Quest for information. JG 50, 159-170.
- Robert, C. and G. Casella, 2004, Monte Carlo Statistical Methods. Springer.
- Stuart, A. M., 2010. Inverse problems: a Bayesian perspective. Acta Numerica 19, 451-559.