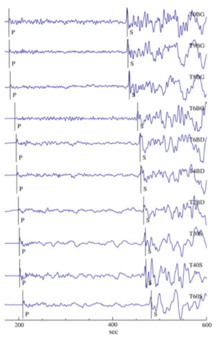
UQ IN HIGHLY NON-LINEAR, MULTI-PARAMETER INVERSE PROBLEMS

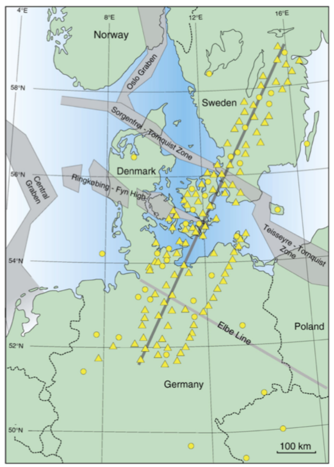
Klaus Mosegaard
Niels Bohr Institute
University of Copenhagen
Denmark

The Probabilistic Inverse Problem

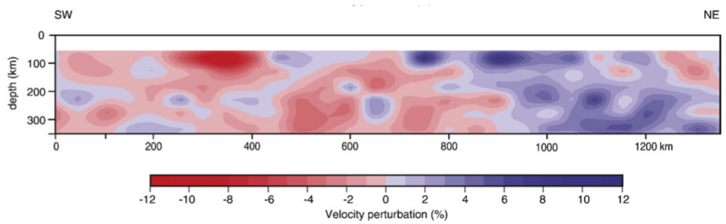
Example: Seismic Tomography as Classical Least-Squares

- Seismic signals from distant earthquakes, recorded along a profile
- Time delays compared to a reference Earth model



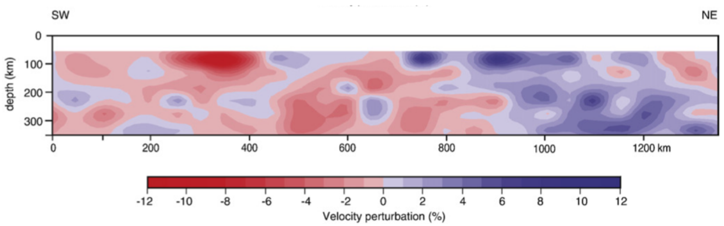


Example: Seismic Tomography as Classical Least-Squares



- Deviation from a reference Earth model
- Smooth models (least-squares/Gaussian)

Example: Seismic Tomography as Classical Least-Squares



- Deviation from a reference Earth model
- Smooth models (least-squares/Gaussian)

> The smoothness of Least squares models is not physically and geologically acceptable!

Probabilistic Inversion

Bayes Theorem

From the definition of conditional probability density

$$f(\mathbf{x}|\mathbf{y}) \equiv \frac{f(\mathbf{x},\mathbf{y})}{f(\mathbf{y})} \quad \text{or} \quad f(\mathbf{y}|\mathbf{x}) \equiv \frac{f(\mathbf{x},\mathbf{y})}{f(\mathbf{x})}$$

we get:

$$f(\mathbf{m}|\mathbf{d}) \equiv \frac{f(\mathbf{d}|\mathbf{m})f(\mathbf{m})}{f(\mathbf{d})}$$

Bayes Theorem

From the definition of conditional probability density

$$f(\mathbf{x}|\mathbf{y}) \equiv \frac{f(\mathbf{x},\mathbf{y})}{f(\mathbf{y})} \quad \text{or} \quad f(\mathbf{y}|\mathbf{x}) \equiv \frac{f(\mathbf{x},\mathbf{y})}{f(\mathbf{x})}$$

we get:

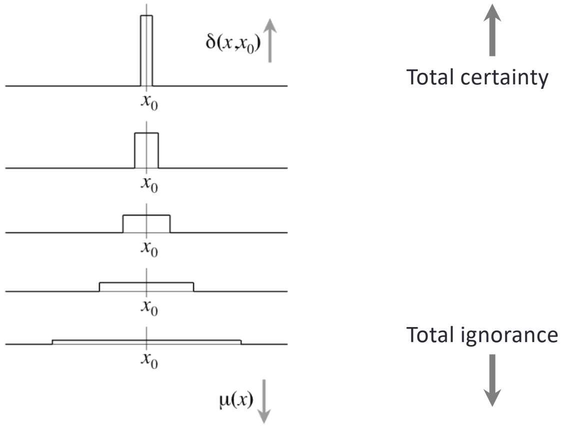
$$f(\mathbf{m}|\mathbf{d}) \equiv \frac{f(\mathbf{d}|\mathbf{m})f(\mathbf{m})}{f(\mathbf{d})}$$

prior

posterior

likelihood

Information Theory Formulation: Tarantola-Valette



Information Theory Formulation: Tarantola-Valette

From the definition of conjunction of information:

$$(\rho \wedge \theta)(\mathbf{x}) \equiv \frac{\rho(\mathbf{x})\theta(\mathbf{x})}{\mu(\mathbf{x})}$$

we get, in the joint space $\mathcal{D}\times\mathcal{M}$:

$$\sigma(\mathbf{d},\mathbf{m}) \equiv (\rho \wedge \theta)(\mathbf{d},\mathbf{m}) = \frac{\rho(\mathbf{d},\mathbf{m})\theta(\mathbf{d},\mathbf{m})}{\mu(\mathbf{d},\mathbf{m})}$$

from which

$$\sigma_m(\mathbf{m}) = \int_{\mathcal{D}} \sigma(\mathbf{d},\mathbf{m})\, d\mathbf{d}$$

Tarantola-Valette Formulation

From the definition of conjunction of information:

$$(\rho \wedge \theta)(\mathbf{x}) \equiv \frac{\rho(\mathbf{x})\theta(\mathbf{x})}{\mu(\mathbf{x})}$$

Null Information Density

we get, in the joint space $\mathcal{D}\times\mathcal{M}$:

posterior

prior

$$\sigma(\mathbf{d},\mathbf{m}) \equiv (\rho \wedge \theta)(\mathbf{d},\mathbf{m}) = \frac{\rho(\mathbf{d},\mathbf{m})\theta(\mathbf{d},\mathbf{m})}{\mu(\mathbf{d},\mathbf{m})}$$

Forward density

from which

$$\sigma_m(\mathbf{m}) = \int_{\mathcal{D}} \sigma(\mathbf{d},\mathbf{m})\, d\mathbf{d}$$

Two Formulations Compared

	Classical Bayes	Tarantola-Valette
Spaces	Separate $\mathcal{D}$ and $\mathcal{M}$	Joint $\mathcal{D}\times\mathcal{M}$
Information	(Realizations of) $\mathbf{d}$ and $\mathbf{m}$	Distributions over $\mathcal{D}\times\mathcal{M}$
Forward Relation	$\mathbf{d} = g(\mathbf{m})$	$\theta(\mathbf{d},\mathbf{m})$
Conditionals	YES	NO

TV Formulation under simplifying conditions

If the data are a priori independent of the model:

$$\rho(\mathbf{d},\mathbf{m}) = \rho_d(\mathbf{d})\rho_m(\mathbf{m})$$

and if the data-model relation is exact:

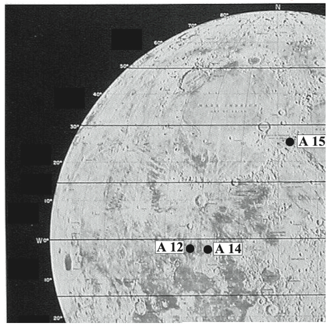
$$\theta(\mathbf{d},\mathbf{m}) = \theta(\mathbf{d}|\mathbf{m})\theta(\mathbf{m}) = \theta(\mathbf{d}|\mathbf{m})\mu(\mathbf{m}) = \delta\big(\mathbf{d}-g(\mathbf{m})\big)\mu(\mathbf{m})$$

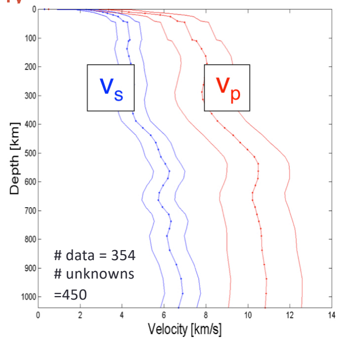
and if $\mu(\mathbf{d},\mathbf{m})$ is constant, we get the Bayes-like:

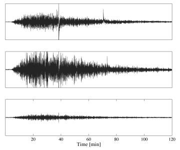
$$\begin{aligned}\sigma_m(\mathbf{m}) &= \int_{\mathcal{D}} \sigma(\mathbf{d},\mathbf{m})\, d\mathbf{d} \\ &= \int_{\mathcal{D}} \frac{\rho_d(\mathbf{d})\rho_m(\mathbf{m})\delta\big(\mathbf{d}-g(\mathbf{m})\big)}{\mu(\mathbf{d},\mathbf{m})}\, d\mathbf{d} \\ &= \rho_d\big(g(\mathbf{m})\big)\rho_m(\mathbf{m}) \equiv L_d(\mathbf{m})\rho_m(\mathbf{m})\end{aligned}$$

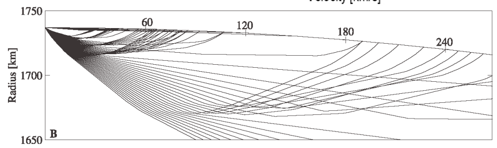
likelihood

Example: Lunar Tomography









From: Khan and Mosegaard (Journ. Geoph. Res. 2002)

Model Parameters, observable parameters and their relation

- Data:

$$\mathbf{d} = (d_1, d_2, \ldots, d_M)$$

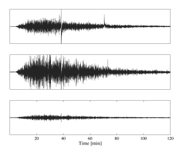
- Physical relation:

$$\mathbf{d} = g(\mathbf{m})$$

- Model parameters:

$$\mathbf{m} = (m_1, m_2, \ldots, m_M)$$

Model Parameters, observable parameters and their relation



- Data:

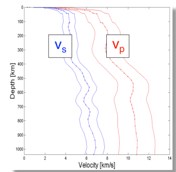
$$\mathbf{d} = (d_1, d_2, \ldots, d_M)$$

- Physical relation:

$$\mathbf{d} = g(\mathbf{m})$$

- Model parameters:

$$\mathbf{m} = (m_1, m_2, \ldots, m_M)$$



TV Formulation under simplifying conditions

The **prior** probability density (data independent of model:

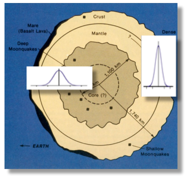
$$\rho(\mathbf{d},\mathbf{m}) = \rho_d(\mathbf{d})\rho_m(\mathbf{m})$$

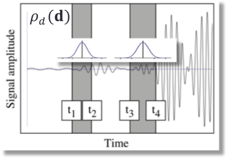
The **Likelihood Function** (exact theory):

$$L_d(\mathbf{m}) = \rho_d(g(\mathbf{m}))$$

The **posterior** probability density:

$$\sigma_m(\mathbf{m}) = L_d(\mathbf{m})\rho_m(\mathbf{m})$$



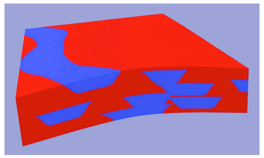


Tarantola and Valette, 1982

PARAMETERIZATION OF A COMPLEX SYSTEM

Choosing the Parameterization

The complete inversion process



- Parameterize the Earth structure m: $\mathbf{m} = f(m)$ to obtain a finite set of *model parameters* $\mathbf{m}$.
- Solve an inverse problem $\mathbf{d} = g(\mathbf{m})$ to infer information about $\mathbf{m}$ from data $\mathbf{d}$.
- Go backwards from the parameters $\mathbf{m}$ to arrive at statements about the Earth structure: $\mathbf{m} \rightarrow m$.

Model Parameters, observable parameters and their relation

- Data:

$$\mathbf{d} = (d_1, d_2, \dots, d_M)$$

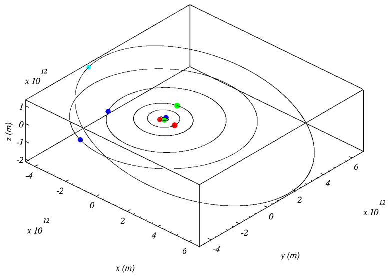
- Physical relation:

$$\mathbf{d} = g(\mathbf{m})$$

- Model parameters:

$$\mathbf{m} = (m_1, m_2, \dots, m_M)$$

Conditions satisfied by physical laws



- They must have a unique solution (for given initial/boundary conditions)
- Their predictions must be independent of the reference frame

The parameterization process

- An infinite set of orthonormal basis functions $\varphi_1(x,y,z), \varphi_2(x,y,z), \varphi_3(x,y,z), \dots$
- Parameters $m_1, m_2, \dots$

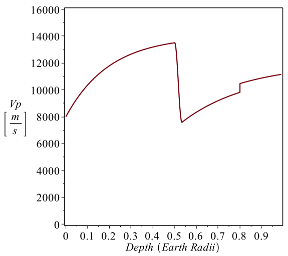
$$m(x,y,z) = \sum_{n=1}^{\infty} m_n\, \varphi_n(x,y,z)$$

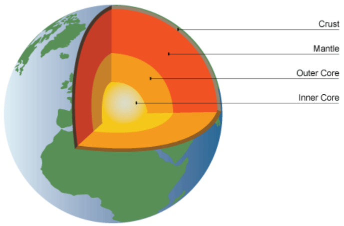
The parameterization process

- Truncate the expansion if necessary
- Keep many parameters to ensure an accurate representation

$$m(x,y,z) = \sum_{n=1}^{M} m_n\, \varphi_n(x,y,z)$$

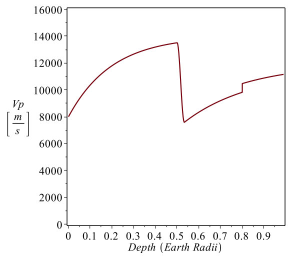
Example: A seismic model of the Earth



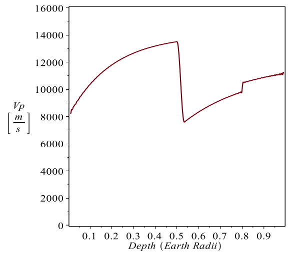


A model not unlike the P-wave velocity in the Earth's interior

Example: A seismic model of the Earth



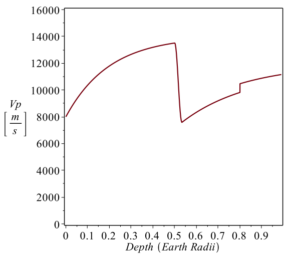
True model



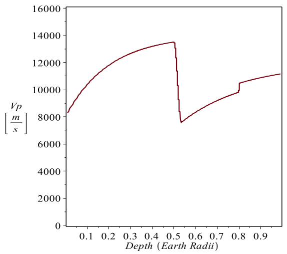
Approximated model

Representation through 128 **Fourier** (sin/cos)-basis functions

Example: A seismic model of the Earth



True model



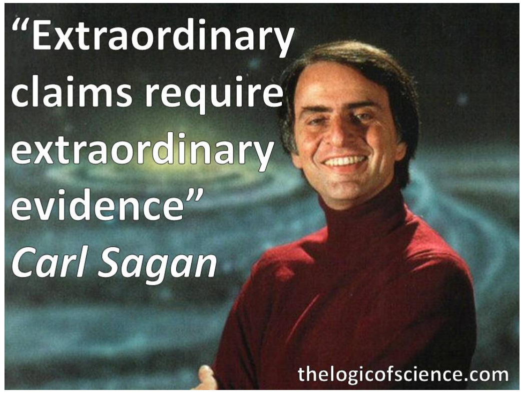
Approximated model

Representation through 256 **Haar**-basis functions

Invariant results from different bases

- Even when two analysts choose different set of base functions, they will obtain (almost) the same model.
- The result is **invariant** under a change of base functions.
- The method is **consistent**: There is agreement between the results from different analysts.

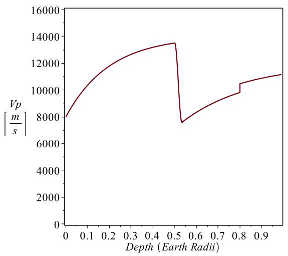
Sparse Models



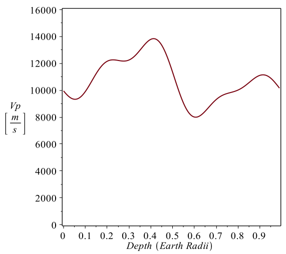
Reasons for sparsity

- To avoid unnecessary detail (Occam’s Razor)
- To minimize the number of model parameters
- To build-in prior knowledge about structure

Example: Different sparse models of the Earth with the same misfit: The Fourier basis



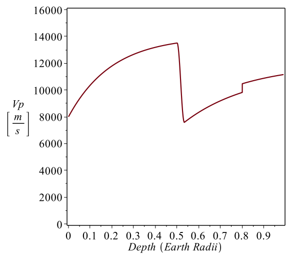
True model



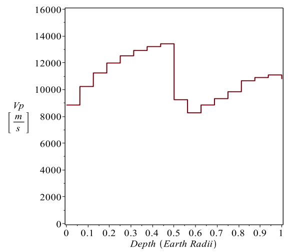
Approximated model

Representation through 4 **Fourier** (sin/cos)-basis functions

Example: Different sparse models of the Earth with the same misfit: The Haar basis



True model



Approximated model

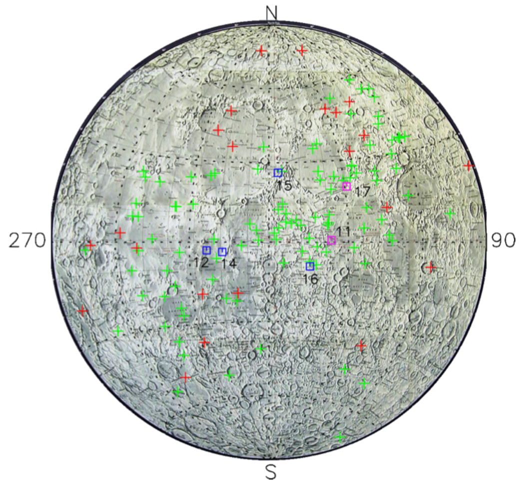
Representation through 16 **Haar**-basis functions

Sparse Models: Conclusions

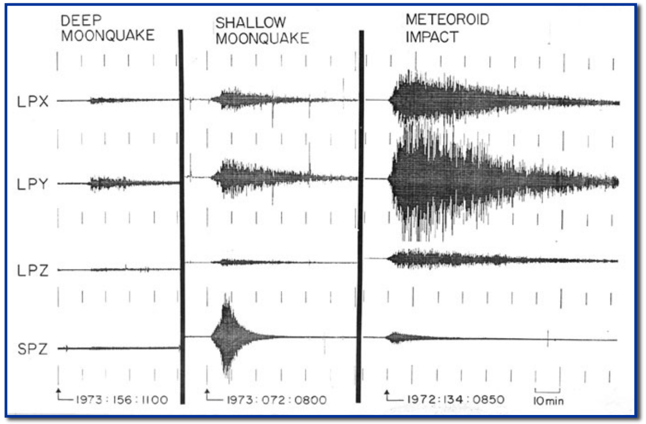
- Sparse models are inconsistent (the result depends on the analyst's choice of basis functions).
- Simplicity, measured as the number of model parameters, is not an objective concept (depends on basis functions).
- If it can be established from external information that the model can be represented sparsely by a certain basis, sparse methods may be useful.

Front View of the Moon with Apollo Stations and Epicenters

Red: Shallow MQ
Green: Deep MQ

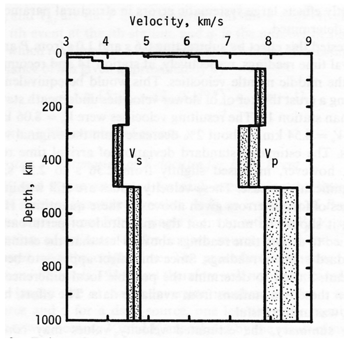


Sample Compressed Seismograms



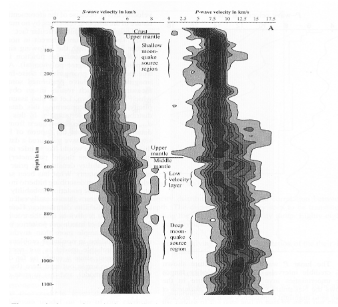
[from Nakamura et al., *Proc. LPSC* **5th**, 2883-2890, 1974]

Seismic Velocity Profile to 1000 km depth



[From Nakamura, *JGR* **88**, 677-686, 1983]

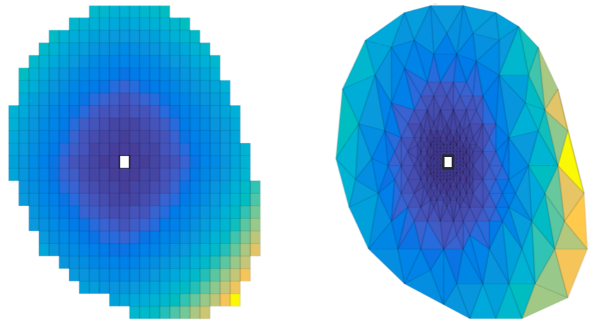
Seismic Velocity Profile by MCMC Inversion



[From Khan and Mosegaard, *JGR* **107**, 10.1029/2001JE001658, 2002]

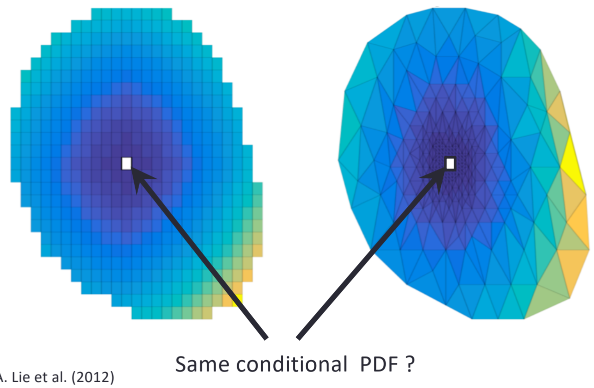
The Borel Paradox

Computing a conditional probability density



K.-A. Lie et al. (2012)

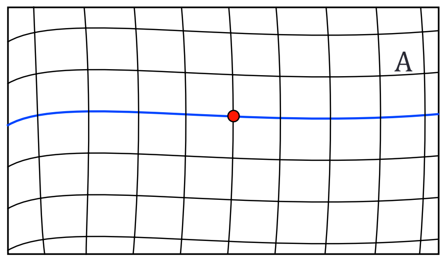
Computing a conditional probability density



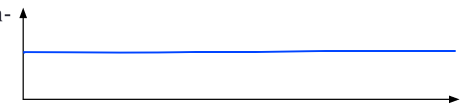
K.-A. Lie et al. (2012)

The Borel Paradox

- Near-Cartesian reference frame
- Equal volumes have equal probabilities

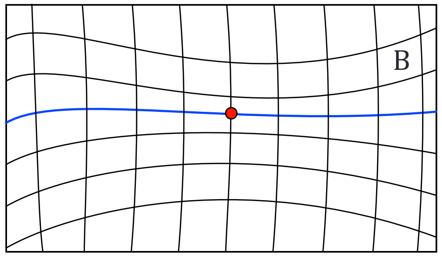


- Conditional probability density is constant

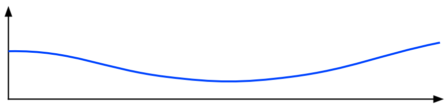


The Borel Paradox

- Non-Cartesian reference frame
- Equal volumes have equal probabilities



- Conditional probability density is **not** constant



Conclusion on conditional probability densities

- Conditional probability densities are **inconsistent**, because different analysts may arrive at different (conflicting) results.

Conclusion on conditional probability densities

- Conditional probability densities are **inconsistent**, because different analysts may arrive at different (conflicting) results.

Borel's paradox disappears if $f(\mathbf{x})$ is replaced with

$$g(\mathbf{x}) = f(\mathbf{x})/\mu(\mathbf{x})$$

where $\mu(\mathbf{x})$ is the *homogeneous* pdf (a constant *volume density*)

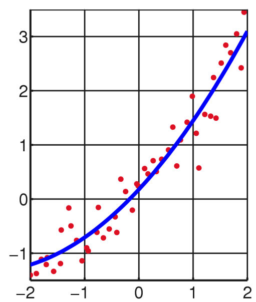
(Mosegaard and Tarantola, 2002)

END OF PART 1

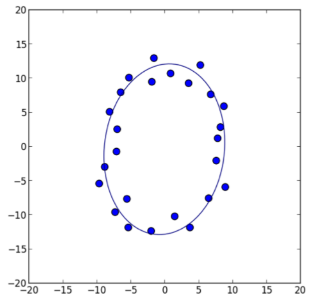
PROBABILISTIC INFERENCE ABOUT COMPLEX STRUCTURE

The Least-Squares Method

Fitting a curve with the smallest sum-of-deviations



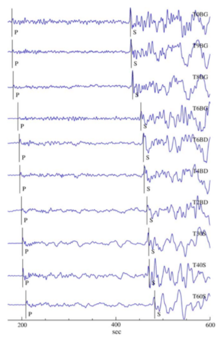
Fitting thermal conductivity and heat production to an observed temperature profile

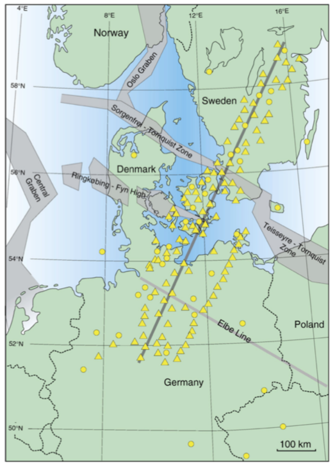


Fitting orbital parameters to an observed orbit of a double star component

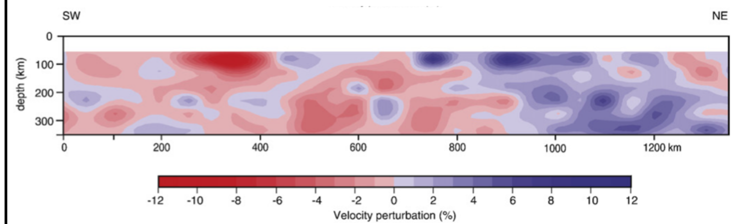
A classical example: Seismic tomography

- Seismic signals from distant earthquakes, recorded along a profile
- Time delays compared to a reference Earth model





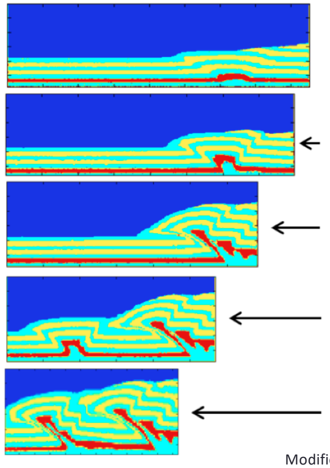
Seismic tomography: A least-squares result



- Deviation from a reference Earth model
- Smooth models (least-squares/Gaussian)

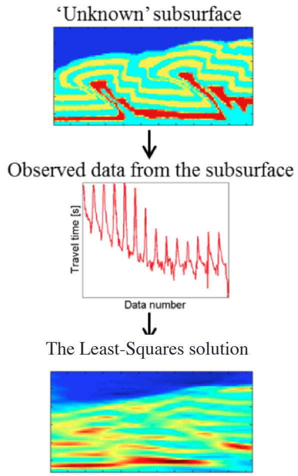
The smoothness of Least squares models is not physically and geologically acceptable!

Complexity: Horizontally compressed sediments



Modified from Egholm et al., 2007

Complexity erased by Least Squares



Cordua & Mosegaard, 2014

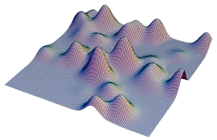
Solutions with Geostatistical Constraints

Example: The Braided River Model

Classical Least-Squares Inversion

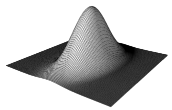
Likelihood Function:

$$L(\mathbf{m}) = C \cdot \exp\left(-\frac{\|\mathbf{d}_{obs} - f(\mathbf{m})\|^2}{2\sigma_d^2}\right)$$



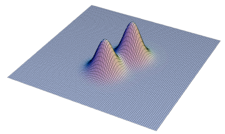
A priori distribution:

Gaussian distribution

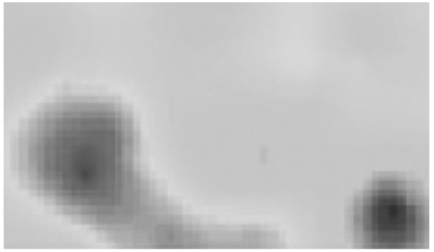


A posteriori distribution

$$\sigma(\mathbf{m}) = C \cdot L(\mathbf{m})\rho(\mathbf{m})$$



A Classical Least-Squares Solution

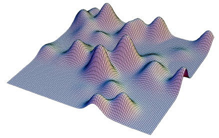


Least-squares solutions are usually too smooth to be geologically acceptable.

Probabilistic Inversion

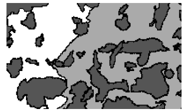
Likelihood Function:

$$L(\mathbf{m}) = C \cdot \exp\left(-\frac{\left\| \mathbf{d}_{obs} - f(\mathbf{m}) \right\|^2}{2\sigma_d^2} \right)$$



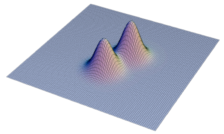
A priori distribution:

Prior information defined by geological prototype examples

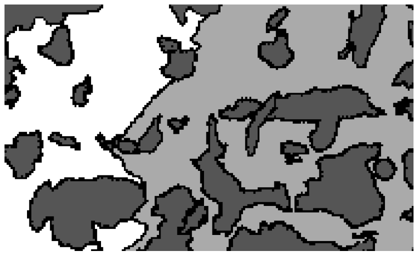


A posteriori distribution

$$\sigma(\mathbf{m}) = C \cdot L(\mathbf{m}) \rho(\mathbf{m})$$

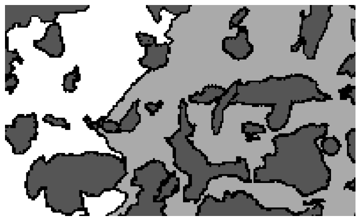


Prior information defined by a geological prototype example



“The subsurface model **m** is statistically similar to a training image“

The Geostatistical Prior



Training Image

A statistically similar image

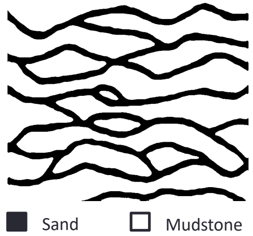
“The subsurface model **m** is statistically similar to a training image“

Examples of geo-information: Braided rivers

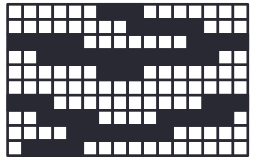
Rakaia River, New Zealand. (Google Earth)

Congo River at the border of Congo and Democratic Republic of Congo. (Google Earth)

Examples of geo-information: Braided rivers

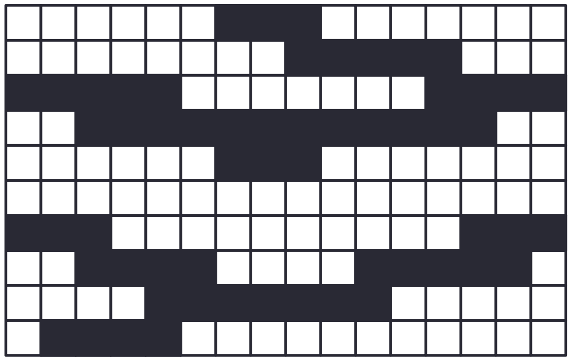


A simple model of a braided river (Strebelle, 2002)

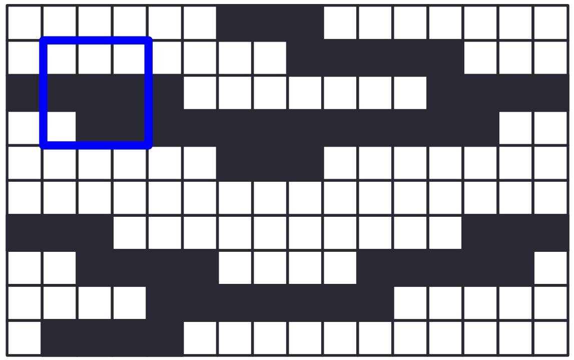


A close-up of part of the pixeled model

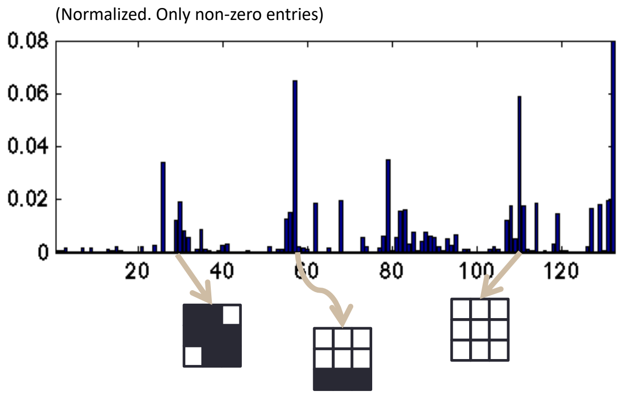
Pattern statistics from a geological model



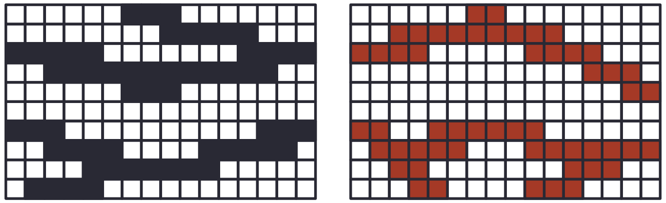
Pattern statistics from a geological model



The frequency distribution

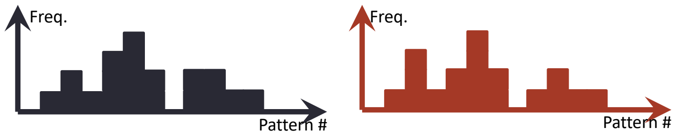


Computing the probability of an image



Reference Model

What is the probability that this model is a realization of the same random process?



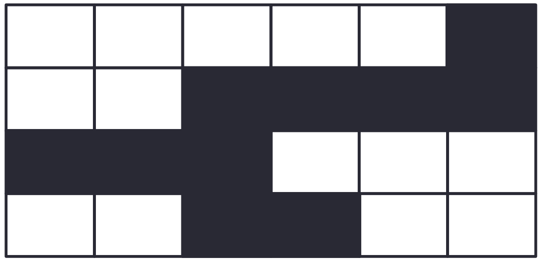
Combining data and the braided river model into a solution to the inverse problem

$$\sigma(\mathbf{m}) = L(\mathbf{m}|\mathbf{d}^{obs})\rho_{training}(\mathbf{m})$$

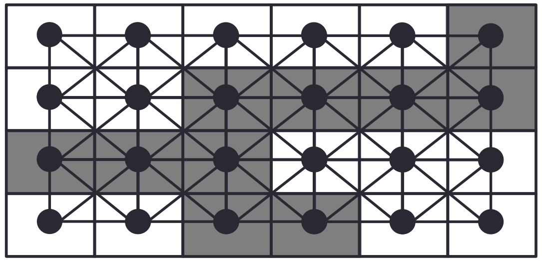
where $\mathbf{d}^{\mathrm{obs}}$ is the observed data, $\mathbf{m}$ is the Earth model, L is a likelihood function, and $\rho_{training}(\mathbf{m})$ is the probability of $\mathbf{m}$, given the reference model.

Joint probability from MRF theory

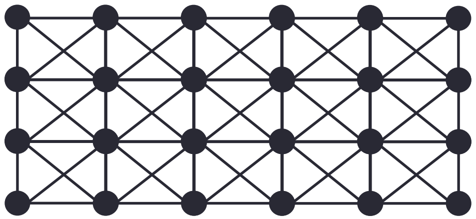
Joint probability from MRF theory



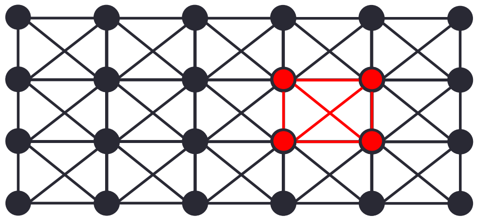
Joint probability from MRF theory



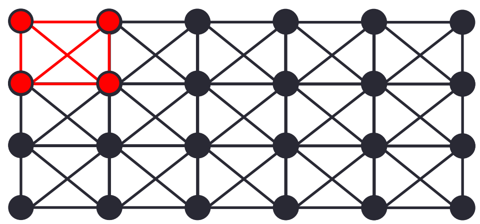
Joint probability from MRF theory



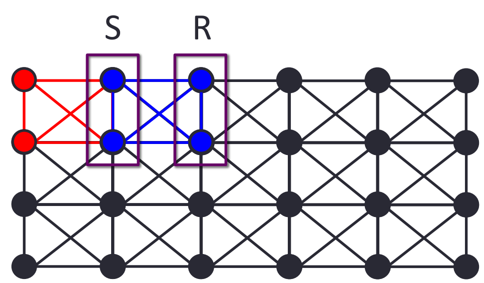
Joint probability from MRF theory



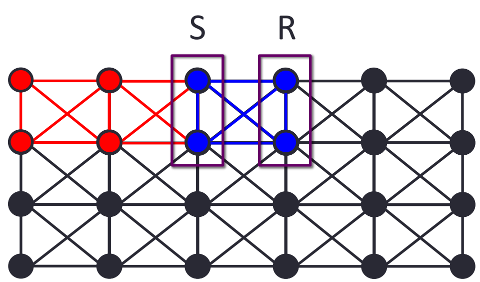
Joint probability from MRF theory



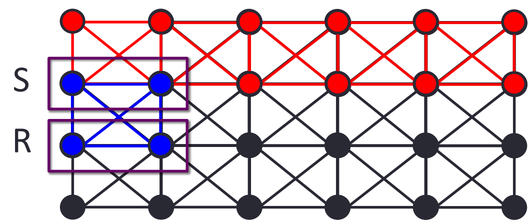
Joint probability from MRF theory



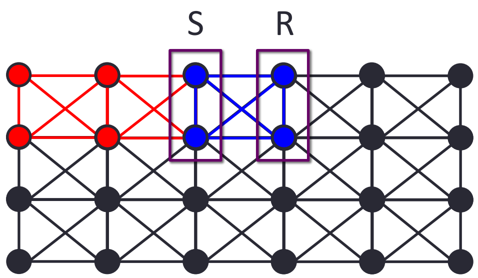
Joint probability from MRF theory



Joint probability from MRF theory



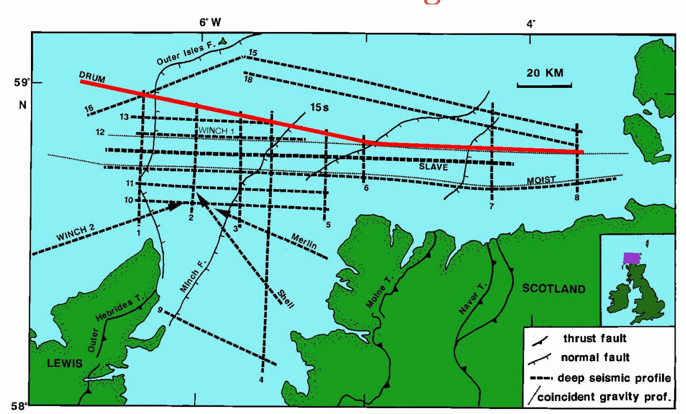
Joint probability from MRF theory



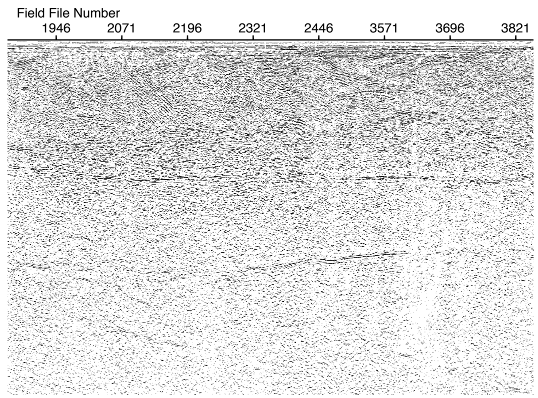
$$p(\mathbf{x}) = \prod_{k=1}^{K} p(R_k | S_k)$$

A PRACTICAL EXAMPLE

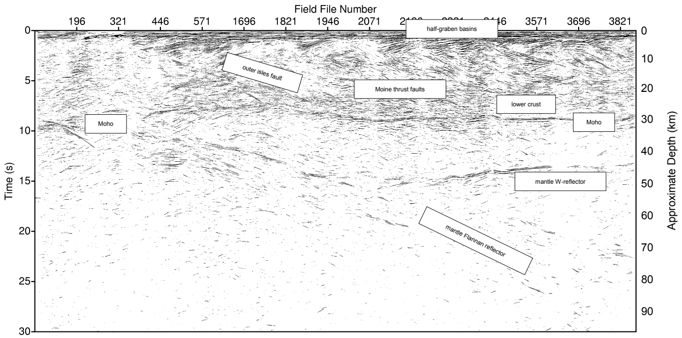
Example: Inversion of seismic reflection data: The Inverse Scattering Problem



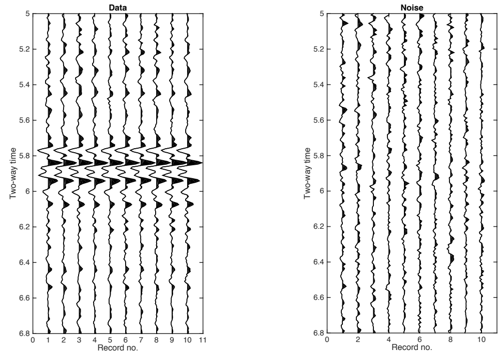
Example: Inversion of seismic reflection data



Example: Inversion of seismic reflection data



Example: Inversion of seismic reflection data



Data

Noise from stationary assumption

Example: Inversion of seismic reflection data

$$\boldsymbol{d} = \boldsymbol{Gm} = \boldsymbol{WDm}$$

where
$\boldsymbol{m}$ is the unknown acoustic impedance in the Earth
$\boldsymbol{D}$ is a differentiation operator
$\boldsymbol{W}$ is convolution with the seismic pulse (wavelet)

Bayesian inversion

If

- Noise and prior information is Gaussian
- The noise covariance matrix is $\boldsymbol{C}_n$
- $\boldsymbol{m}_0$ is a reference model ("best guess")
- $\boldsymbol{C}_m$ is the prior covariance matrix (a priori tolerance)

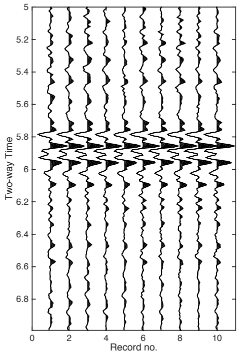
we can compute the maximum posterior model:

$$\boldsymbol{m}_{post} = \boldsymbol{m}_0 + (\boldsymbol{G}^T \boldsymbol{C}_n^{-1} \boldsymbol{G} + \boldsymbol{C}_m^{-1})^{-1} \boldsymbol{G}^T \boldsymbol{C}_n^{-1} (\boldsymbol{d} - \boldsymbol{G}\boldsymbol{m}_0)$$

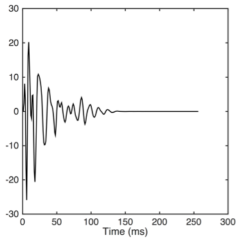
and the posterior covariance (the model parameter uncertainty)

$$\boldsymbol{C}_{post} = (\boldsymbol{G}^T \boldsymbol{C}_n^{-1} \boldsymbol{G} + \boldsymbol{C}_m^{-1})^{-1}$$

Example: Inversion of seismic reflection data

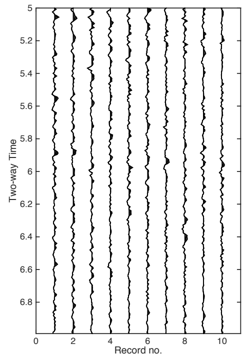


Data

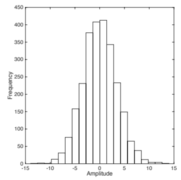


Seismic pulse (wavelet)

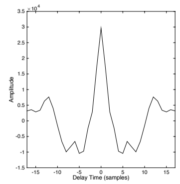
Example: Inversion of seismic reflection data



Estimated noise

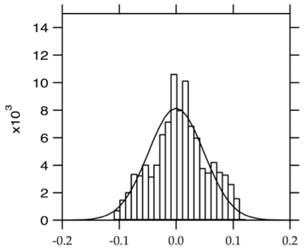


Noise distribution



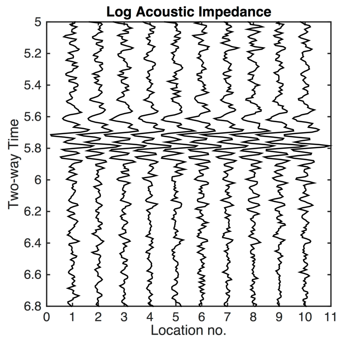
Noise auto-covariance

Example: Inversion of seismic reflection data



Histogram of reflection coefficients from igneous intrusions of Rum, Scotland, and Great Dyke, Zimbabwe

Example: Inversion of seismic reflection data



Maximum a posteriori model of the zone surrounding the crust/mantle interface north of Scotland.

Conclusions

- Although consistency is a serious matter in mathematical physics, it is often ignored in inverse theory

We have looked at 3 cases of inconsistency:

- Pitfalls in sparse parameterizations
- Inconsistencies when using subjective probabilities
- Lack of invariance of conditional probability densities

END