A Probabilistic Approach to Inverse Problems - A Bayesian Perspective

A. Khan

Institute of Geochemistry and Petrology, ETH Zurich, Switzerland

QUEST Workshop, Iceland, July 16, 2011

A Bayesian Tutorial

A Bayesian tutorial

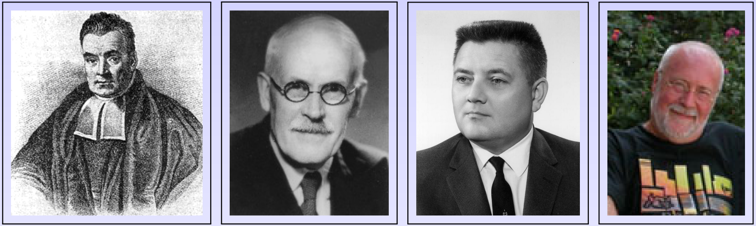


Figure: Expounders of the Bayesian viewpoint - Bayes, Jeffreys, Jaynes and Tarantola.

Albert like Jeffreys and Jaynes are Bayesians who hold that probabilities encode degrees of belief and do not exist except as a representation of information about the world. For some, this position means that a Bayesian view of probability is hopelessly, fatally subjective - "unscientific".

A Bayesian tutorial

Probabilities are conditional and assigned based on experimental and theoretical information:

$$P(A|I)$$

where A is a proposition and I is background (prior) infomation.

Objectivity arises from the requirement that the same information I will lead to the same probability assignment and thus the same inference (Jaynes, 2003).

A Bayesian tutorial

Probabilities are conditional and assigned based on experimental and theoretical information:

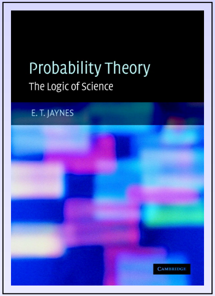
$$P(A|I)$$

where A is a proposition and I is background (prior) infomation.

Objectivity arises from the requirement that the same information I will lead to the same probability assignment and thus the same inference (Jaynes, 2003).

A Bayesian tutorial

Jaynes (2003) has made the case for the Bayesian interpretation



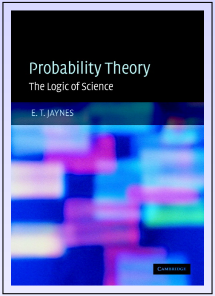
which ultimately leads to Bayes' theorem:

$$\text{posterior} = \text{prior} \times \text{likelihood}$$

$$P(\mathbf{m}|\mathbf{d}) = k\, p(\mathbf{m}) P(\mathbf{d}|\mathbf{m})$$

A Bayesian tutorial

Jaynes (2003) has made the case for the Bayesian interpretation



which ultimately leads to Bayes' theorem:

$$\text{posterior} = \text{prior} \times \text{likelihood}$$

$$P(\mathbf{m}|\mathbf{d}) = k\, p(\mathbf{m}) P(\mathbf{d}|\mathbf{m})$$

A Bayesian tutorial

The Bayesian interpretation of probability is a model for learning:

the output from one experiment (the posterior probability),

$$P(\mathbf{m}|\mathbf{d}) = k\, p(\mathbf{m}) P(\mathbf{d}|\mathbf{m})$$

can be used as input (the prior probability) in the next

$$P(\mathbf{m}|\mathbf{d}') = k\, P(\mathbf{m}|\mathbf{d}) P(\mathbf{d}'|\mathbf{m})$$

A Bayesian tutorial

The Bayesian interpretation of probability is a model for learning:

the output from one experiment (the posterior probability),

$$P(\mathbf{m}|\mathbf{d}) = k\,p(\mathbf{m})P(\mathbf{d}|\mathbf{m})$$

can be used as input (the prior probability) in the next

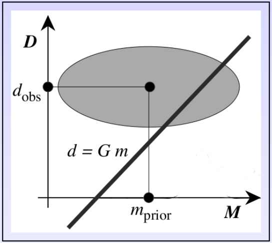
$$P(\mathbf{m}|\mathbf{d}') = k\,P(\mathbf{m}|\mathbf{d})P(\mathbf{d}'|\mathbf{m})$$

Quest for Information -

Albert Tarantola's reformulation of Inverse Theory

Solving the linear inverse problem

In the late 70s inverse theory was dominated by methods aimed at linear/linearized problems (regularization, Backus-Gilbert).

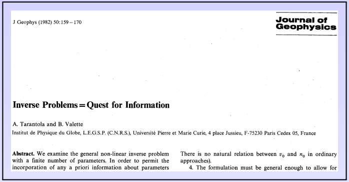


The solution to the linear inverse problem $\mathbf{d} = \mathbf{G}\mathbf{m}$ can be found through regularization (e.g., Tikhonov, 1963) and in its simplest form is

$$\mathbf{m} = (\mathbf{G}^t\mathbf{G} + \lambda\mathbf{I})^{-1}\mathbf{G}^t\mathbf{d}$$

where λ is a regularization parameter.

Quest for information


J Geophys (1982) 50:159 – 170

Journal of Geophysics

Inverse Problems = Quest for Information

A. Tarantola and B. Valette

Institut de Physique du Globe, L.E.G.S.P. (C.N.R.S.), Université Pierre et Marie Curie, 4 place Jussieu, F-75230 Paris Cedex 05, France

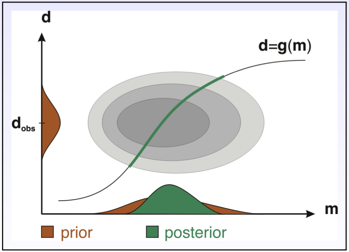
Abstract. We examine the general non-linear inverse problem with a finite number of parameters. In order to permit the incorporation of any a priori information about parameters

There is no natural relation between v_0 and n_0 in ordinary approaches).

4. The formulation must be general enough to allow for

This seminal paper presented a radical departure from established theory: parameters were represented by probability distributions instead of numbers.

- The solution was formulated as a combination of different states of information of a set of (model) parameters emanating from observations and prior information.



Quest for information

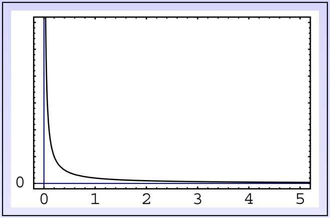
• Tarantola and Valette (1982) proposed that two independent states of information, $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$, on a set of parameters $\mathbf{x}$ could be combined to produce a new probability density according to (conjunction of states of information)

$$f(\mathbf{x}) = \frac{f_1(\mathbf{x}) f_2(\mathbf{x})}{\mu(\mathbf{x})},$$

where $\mu(\mathbf{x})$ is the homogeneous prior probability distribution (formerly the 'non-informative').

Quest for information

Examples of homogeneous priors



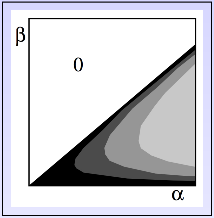


Figure: Examples of $\mu(\mathbf{x})$. Left - acoustic velocity (slowness); right - joint seismic velocities V_p and V_s. From Mosegaard & Tarantola (2002).

- Use of $x' = \log(x/x_0)$ for Jeffreys parameters (e.g., $v - s$, $\rho - \sigma$) results in a constant homogeneous probability density.
- Moreover, the probability distributions so obtained are invariant under transformations.

Quest for information

Examples of homogeneous priors

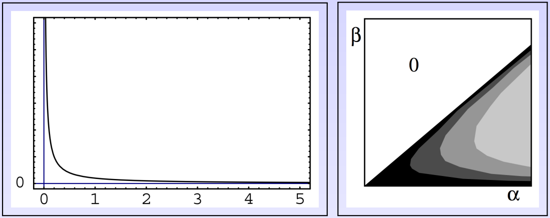


Figure: Examples of $\mu(\mathbf{x})$. Left - acoustic velocity (slowness); right - joint seismic velocities V_p and V_s. From Mosegaard & Tarantola (2002).

- Use of $x' = \log(x/x_0)$ for Jeffreys parameters (e.g., $v - s$, $\rho - \sigma$) results in a constant homogeneous probability density.
- Moreover, the probability distributions so obtained are invariant under transformations.

Quest for information

Examples of homogeneous priors

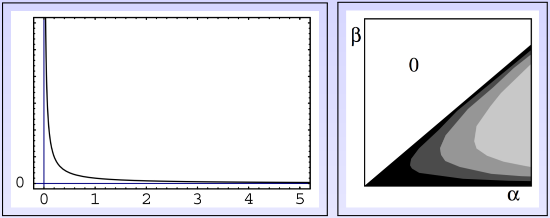


Figure: Examples of $\mu(\mathbf{x})$. Left - acoustic velocity (slowness); right - joint seismic velocities V_p and V_s. From Mosegaard & Tarantola (2002).

• Use of $x' = \log(x/x_0)$ for Jeffreys parameters (e.g., $v - s$, $\rho - \sigma$) results in a constant homogeneous probability density.
• Moreover, the probability distributions so obtained are invariant under transformations.

Quest for information

Examples of priors: Symmetry considerations

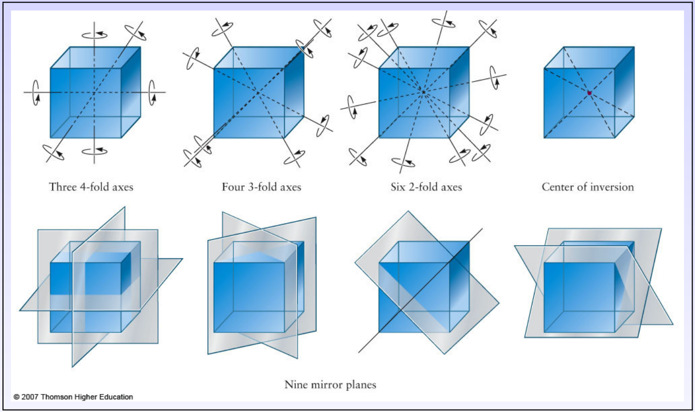


Figure: Crystal symmetry.

Quest for information

Examples of priors: Data (e.g., laboratory measurements, previous geophysical studies, etc.)

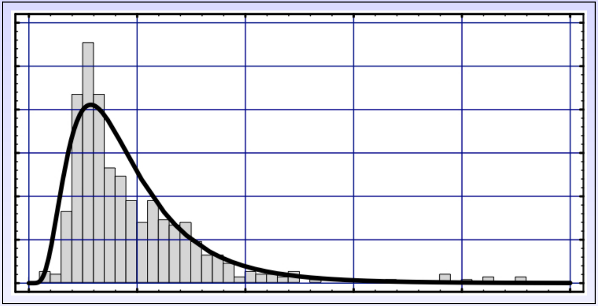


Figure: Laboratory measurements of rock density.

Quest for information

The inverse problem formulated as a combination of independent states of information:

1. prior information on $\mathbf{m}$ obtained independently of data $\rho_m(\mathbf{m})$.
2. information obtained from (uncertain) observations $\rho_d(\mathbf{d})$.
3. the joint prior $\rho(\mathbf{d}, \mathbf{m}) = \rho_d(\mathbf{d})\rho_m(\mathbf{m})$.
4. a distribution $\theta(\mathbf{d}, \mathbf{m})$ describing an uncertain theory $\mathbf{d} \approx \mathrm{g}(\mathbf{m})$ over the joint data/model space.

combining 1-4 using the conjunction of states of information results in

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

Quest for information

The inverse problem formulated as a combination of independent states of information:

1. prior information on $\mathbf{m}$ obtained independently of data $\rho_m(\mathbf{m})$.
2. information obtained from (uncertain) observations $\rho_d(\mathbf{d})$.
3. the joint prior $\rho(\mathbf{d},\mathbf{m}) = \rho_d(\mathbf{d})\rho_m(\mathbf{m})$.
4. a distribution $\theta(\mathbf{d},\mathbf{m})$ describing an uncertain theory $\mathbf{d} \approx \mathrm{g}(\mathbf{m})$ over the joint data/model space.

combining 1-4 using the conjunction of states of information results in

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

Quest for information

The inverse problem formulated as a combination of independent states of information:

1. prior information on $\mathbf{m}$ obtained independently of data $\rho_m(\mathbf{m})$.
2. information obtained from (uncertain) observations $\rho_d(\mathbf{d})$.
3. the joint prior $\rho(\mathbf{d}, \mathbf{m}) = \rho_d(\mathbf{d})\rho_m(\mathbf{m})$.
4. a distribution $\theta(\mathbf{d}, \mathbf{m})$ describing an uncertain theory $\mathbf{d} \approx \mathrm{g}(\mathbf{m})$ over the joint data/model space.

combining 1-4 using the conjunction of states of information results in

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

Quest for information

The inverse problem formulated as a combination of independent states of information:

1. prior information on $\mathbf{m}$ obtained independently of data $\rho_m(\mathbf{m})$.
2. information obtained from (uncertain) observations $\rho_d(\mathbf{d})$.
3. the joint prior $\rho(\mathbf{d},\mathbf{m}) = \rho_d(\mathbf{d})\rho_m(\mathbf{m})$.
4. a distribution $\theta(\mathbf{d},\mathbf{m})$ describing an uncertain theory $\mathbf{d} \approx \mathrm{g}(\mathbf{m})$ over the joint data/model space.

combining 1-4 using the conjunction of states of information results in

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

Quest for information

The inverse problem formulated as a combination of independent states of information:

1. prior information on $\mathbf{m}$ obtained independently of data $\rho_m(\mathbf{m})$.
2. information obtained from (uncertain) observations $\rho_d(\mathbf{d})$.
3. the joint prior $\rho(\mathbf{d}, \mathbf{m}) = \rho_d(\mathbf{d})\rho_m(\mathbf{m})$.
4. a distribution $\theta(\mathbf{d}, \mathbf{m})$ describing an uncertain theory $\mathbf{d} \approx \mathrm{g}(\mathbf{m})$ over the joint data/model space.

combining 1-4 using the conjunction of states of information results in

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

Quest for information

The inverse problem formulated as a combination of independent states of information:

1. prior information on $\mathbf{m}$ obtained independently of data $\rho_m(\mathbf{m})$.
2. information obtained from (uncertain) observations $\rho_d(\mathbf{d})$.
3. the joint prior $\rho(\mathbf{d}, \mathbf{m}) = \rho_d(\mathbf{d})\rho_m(\mathbf{m})$.
4. a distribution $\theta(\mathbf{d}, \mathbf{m})$ describing an uncertain theory $\mathbf{d} \approx \mathrm{g}(\mathbf{m})$ over the joint data/model space.

combining 1-4 using the conjunction of states of information results in

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

Quest for information

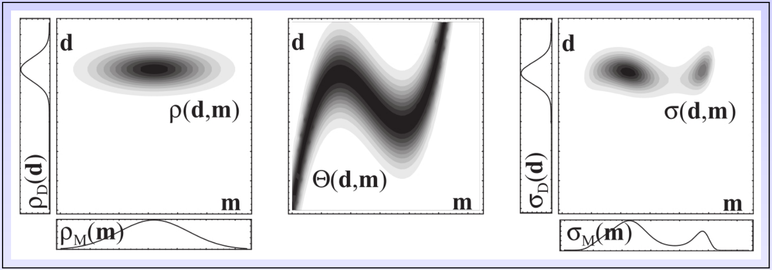


Figure: Combination of states of information. From Tarantola (2005).

which for many applications is typically written

$$\sigma_m(\mathbf{m}) = k\rho_m(\mathbf{m})L(\mathbf{m})$$

$$L(\mathbf{m}) = k \exp\left[-S(\mathbf{m})\right]$$

$$S(\mathbf{m}) = \frac{1}{2}\left[(\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs})^t \mathbf{C}_{\mathrm{D}}^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs})\right]$$

Quest for information

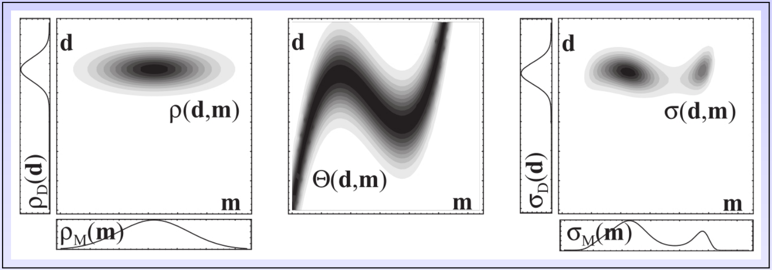


Figure: Combination of states of information. From Tarantola (2005).

which for many applications is typically written

$$\sigma_m(\mathbf{m}) = k\rho_m(\mathbf{m})L(\mathbf{m})$$

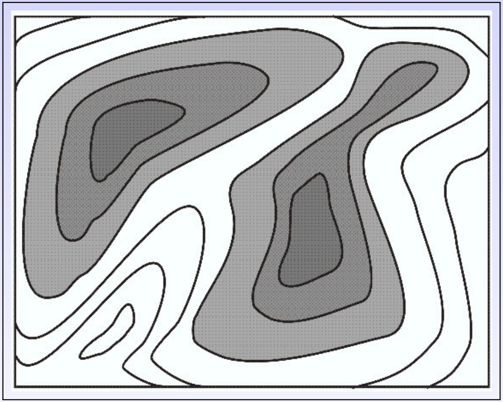
$$L(\mathbf{m}) = k\exp\left[-S(\mathbf{m})\right]$$

$$S(\mathbf{m}) = \frac{1}{2}\left[(\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs})^t \mathbf{C}_{\mathrm{D}}^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs})\right]$$

Sampling the Model Space

Sampling the posterior

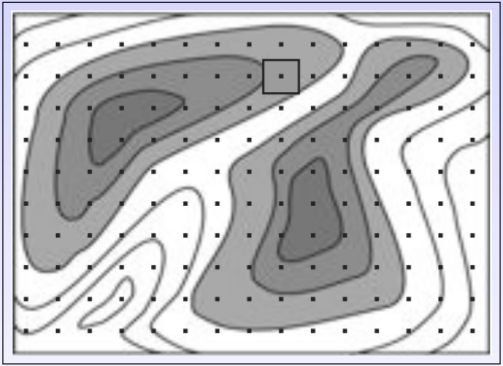
$\sigma(\mathbf{m})$ is typically pathological (e.g., multimodal, non-normalizable)



How do we obtain the posterior distribution in the model space ?

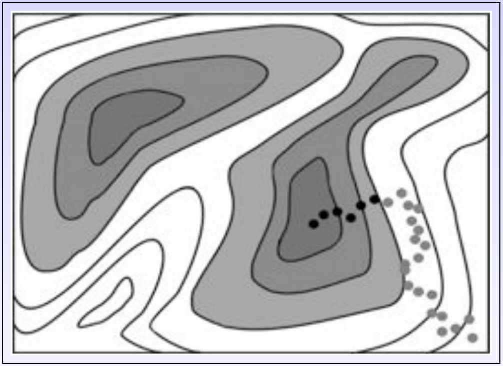
Sampling the posterior

Let us perform a grid search over the entire model space.



Sampling the posterior

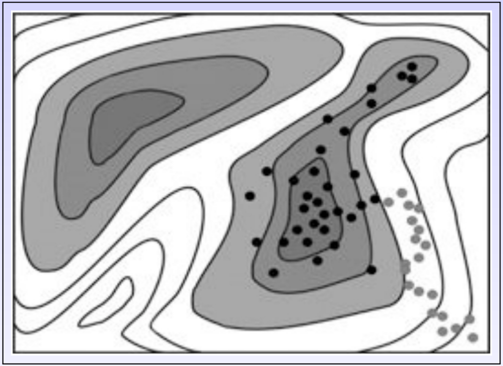
Instead, let us try a sampling-based method such as importance sampling (Metropolis, Gibbs).



There are several algorithms available: crude Monte Carlo (random search), genetic algorithms, Importance sampling methods (Gibbs, Metropolis), Neighbourhood algorithm, simulated annealing, etc.

Sampling the posterior

Importance sampling using Metropolis algorithm

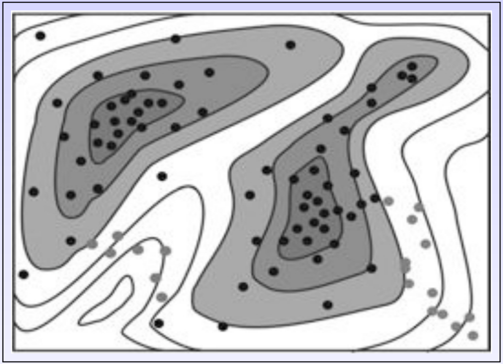


- if $L(\mathbf{m}_j) \geq L(\mathbf{m}_i)$, accept proposed transition $i \to j$.
- if $L(\mathbf{m}_j) < L(\mathbf{m}_i)$, accept proposed transition with probability

$P_{i \to j} = \frac{L(\mathbf{m}_j)}{L(\mathbf{m}_i)}$

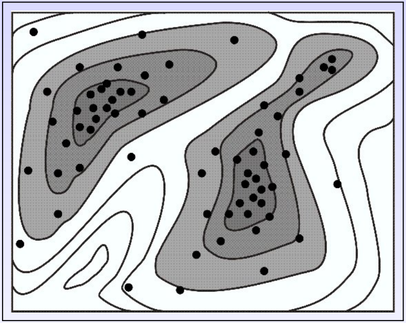
Sampling the posterior

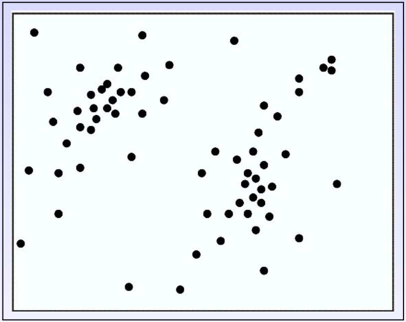
Importance sampling



Sampling the posterior

This became Albert's preferred idea - representing probability densities with samples from the probability distribution.





The solution is not one model but a collection of models that are consistent with both prior information and data

Displaying models

The 'movie' strategy applied to seismic tomography: inversion of fundamental-mode and higher order Rayleigh- and Love-wave dispersion data

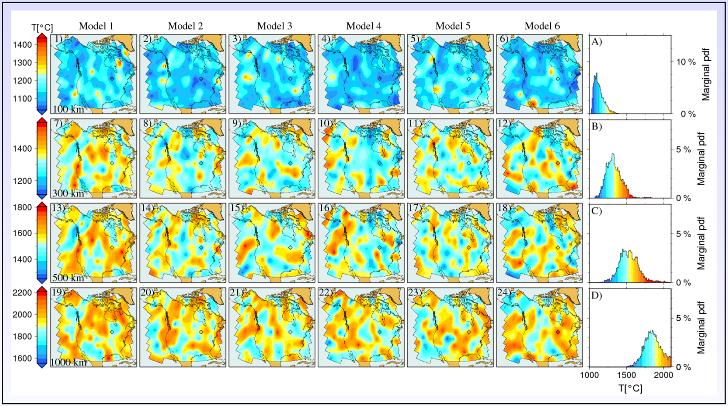


Figure: Six thermal models taken randomly from the prior pdf. From Khan et al. (2011).

Displaying models

Posterior compositional models.

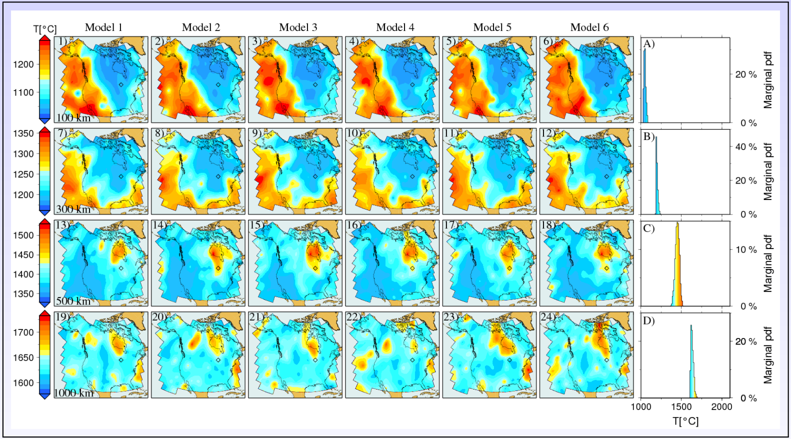


Figure: Six thermal models taken randomly from the posterior pdf (Khan et al., 2011).

Displaying models

Prior compositional models.

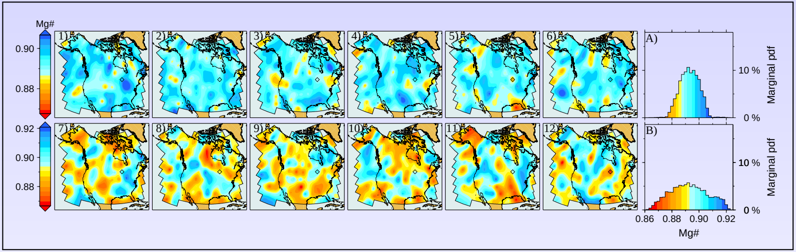


Figure: Six compositional models taken randomly from the posterior pdf (Khan et al., 2011).

Displaying models

Posterior compositional models.

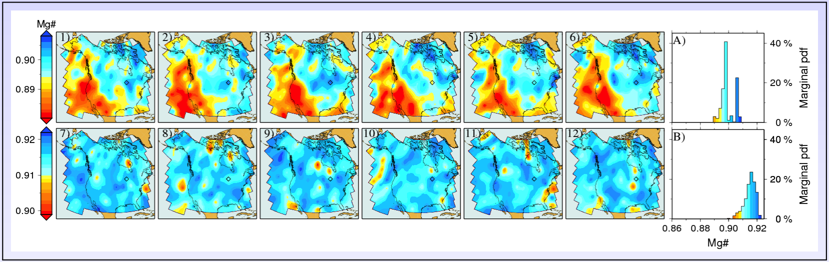


Figure: Six compositional models taken randomly from the posterior pdf (Khan et al., 2011).

Displaying models

Prior V_s models.

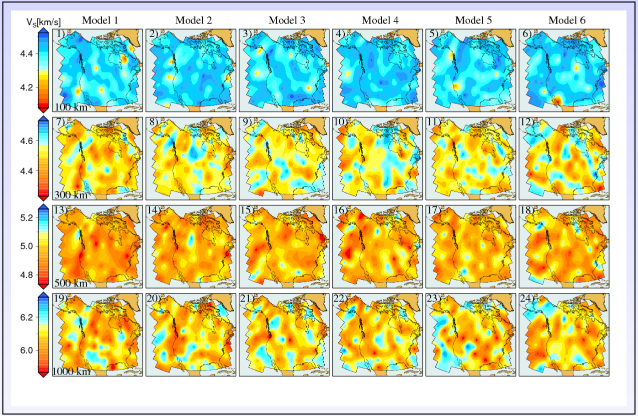


Figure: Six thermal models taken randomly from the prior pdf. From Khan et al. (2011).

Displaying models

Posterior V_s models.

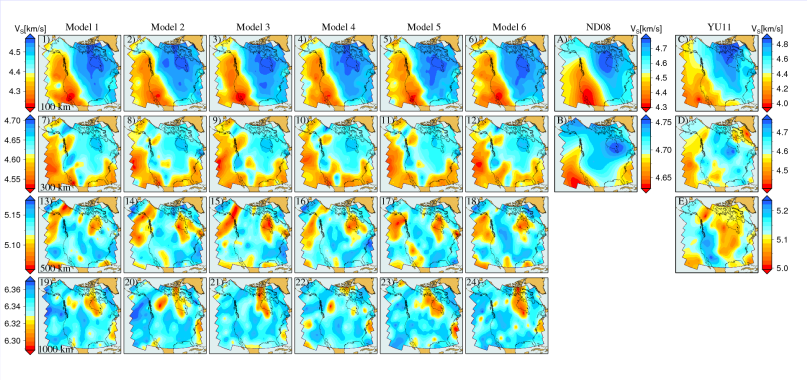


Figure: Six thermal models taken randomly from the posterior pdf (Khan et al., 2011).

Analysis of the posterior distribution

1. Calculation of resolution measures

$$\mathcal{R}(\Omega, f) = \int_{\Omega} f(\mathbf{m})\sigma(\mathbf{m})d\mathbf{m} \approx \frac{1}{N} \sum_{\{n|m_n \in \Omega\}} f(\mathbf{m}_n)$$

2. Bayesian hypothesis testing

Given two hypotheses $\mathcal{H}_i$, $\mathcal{H}_j$, the Bayes factor, $\mathcal{B}_{ij}$ in favour of $\mathcal{H}_i$ (and against $\mathcal{H}_j$) is given by the posterior to prior odds ratio.

$$\mathcal{B}_{ij}(\mathbf{d}) = \frac{\mathcal{P}(\mathbf{d}|\mathcal{H}_i)}{\mathcal{P}(\mathbf{d}|\mathcal{H}_j)} = \frac{\mathcal{P}(\mathcal{H}_i|\mathbf{d})/\mathcal{P}(\mathcal{H}_j|\mathbf{d})}{\mathcal{P}(\mathcal{H}_i)/\mathcal{P}(\mathcal{H}_j)}$$

Bayes factor provides a measure of whether the data $\mathbf{d}$ have increased or decreased the odds on $\mathcal{H}_i$ relative to $\mathcal{H}_j$. If $\mathcal{B}_{ij}(\mathbf{d}) > 1$, $\mathcal{H}_i$ is more plausible than $\mathcal{H}_j$ in the light of $\mathbf{d}$.

Analysis of the posterior distribution

1. Calculation of resolution measures

$$\mathcal{R}(\Omega, f) = \int_{\Omega} f(\mathbf{m})\sigma(\mathbf{m})d\mathbf{m} \approx \frac{1}{N} \sum_{\{n|m_n \in \Omega\}} f(\mathbf{m}_n)$$

2. Bayesian hypothesis testing

Given two hypotheses $\mathcal{H}_i$, $\mathcal{H}_j$, the Bayes factor, $\mathcal{B}_{ij}$ in favour of $\mathcal{H}_i$ (and against $\mathcal{H}_j$) is given by the posterior to prior odds ratio.

$$\mathcal{B}_{ij}(\mathbf{d}) = \frac{\mathcal{P}(\mathbf{d}|\mathcal{H}_i)}{\mathcal{P}(\mathbf{d}|\mathcal{H}_j)} = \frac{\mathcal{P}(\mathcal{H}_i|\mathbf{d})/\mathcal{P}(\mathcal{H}_j|\mathbf{d})}{\mathcal{P}(\mathcal{H}_i)/\mathcal{P}(\mathcal{H}_j)}$$

Bayes factor provides a measure of whether the data $\mathbf{d}$ have increased or decreased the odds on $\mathcal{H}_i$ relative to $\mathcal{H}_j$. If $\mathcal{B}_{ij}(\mathbf{d}) > 1$, $\mathcal{H}_i$ is more plausible than $\mathcal{H}_j$ in the light of $\mathbf{d}$.

Displaying models

Albert rejected the use of the least-squares technique, because the solution is an unlikely outcome

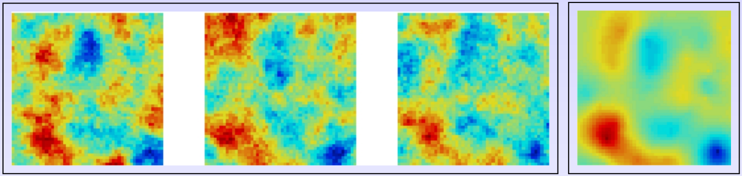
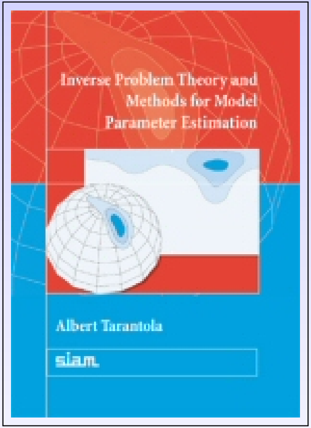


Figure: Earth models as realizations of a 2D Gaussian stochastic process, which were generated by sequential simulation and constrained by well data (left three figures). Right - least-squares solution. From Strebelle (2002).

Recommended further reading



Recommended further reading

Quest for consistency, symmetry and simplicity –
The Legacy of Albert Tarantola

Klaus Mosegaard*

ABSTRACT

On 6. December 2009, the distinguished Spanish-French physicist and geoscientist, Albert Tarantola, passed away at the age of 60. Born in Barcelona in 1949, he went to Paris where he lived most of his life and worked as a professor at Institut de Physique du

1976 wrote a Ph.D. thesis on theoretical astrophysics at Université de Paris 6. The thesis dealt with the evolution of clusters of galaxies, using the general theory of relativity, and was driven by a desire to gain a deep understanding of the structure of the universe. However, while working on his thesis, Albert realized that science is still far from achieving this goal. As a result, he turned

Geophysics, in press, 2011.

Are the Earth and Moon compositionally alike ?

Introduction

Why do we want to study the internal structure of the Moon?

Figure: The Moon-forming impact as envisioned by W. Kaufmann

Introduction

Because it holds the clue to understanding the formation and evolution of the Moon and Earth

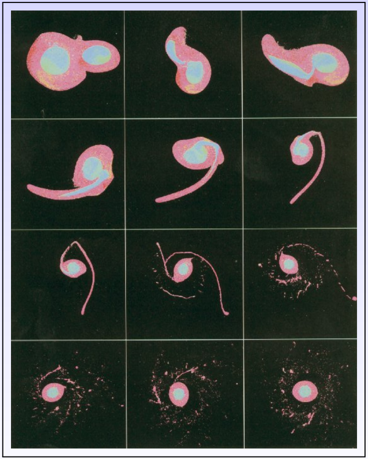
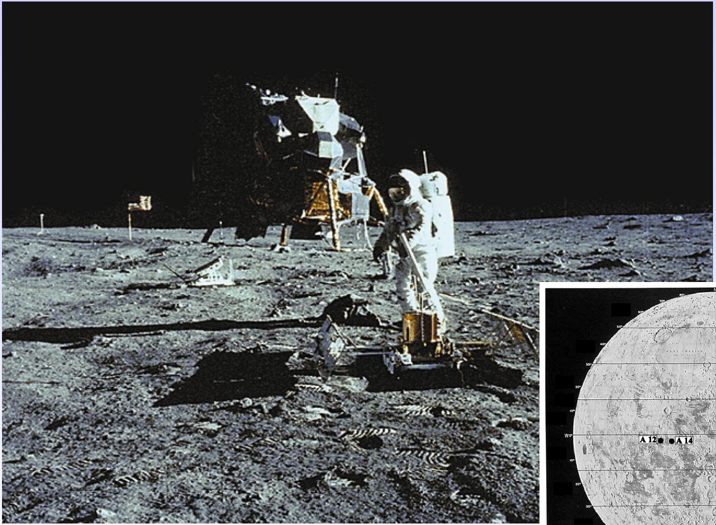


Figure: The Moon-forming impact as modeled by Canup (2000).

Are the Earth and the Moon compositionally alike ?



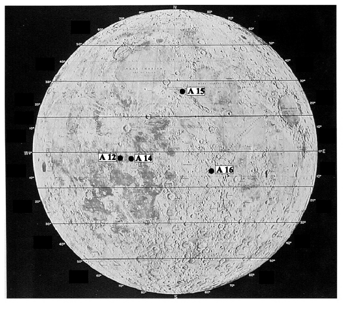


Figure: Deployment of seismometer during Apollo 11 EVA and distribution of seismic array on the lunar frontside.

Are the Earth and the Moon compositionally alike ?

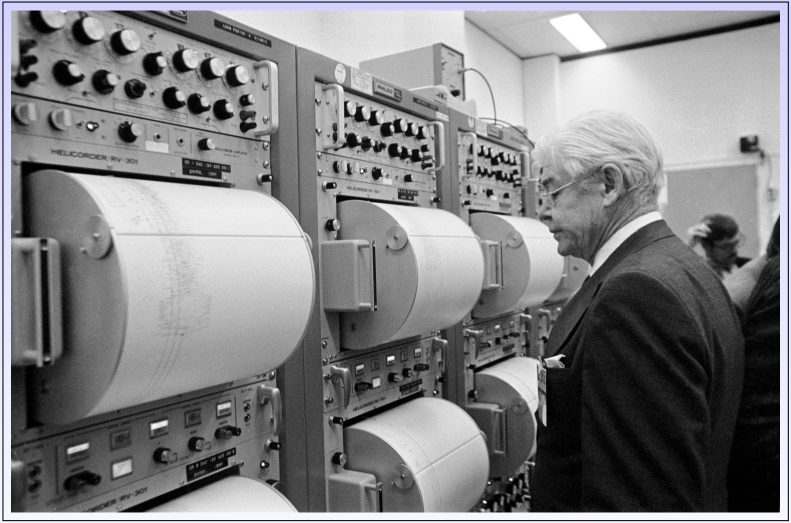


Figure: Inspection of lunar seismograms on Earth. Courtesy of NASA/JPL.

Are the Earth and the Moon compositionally alike ?

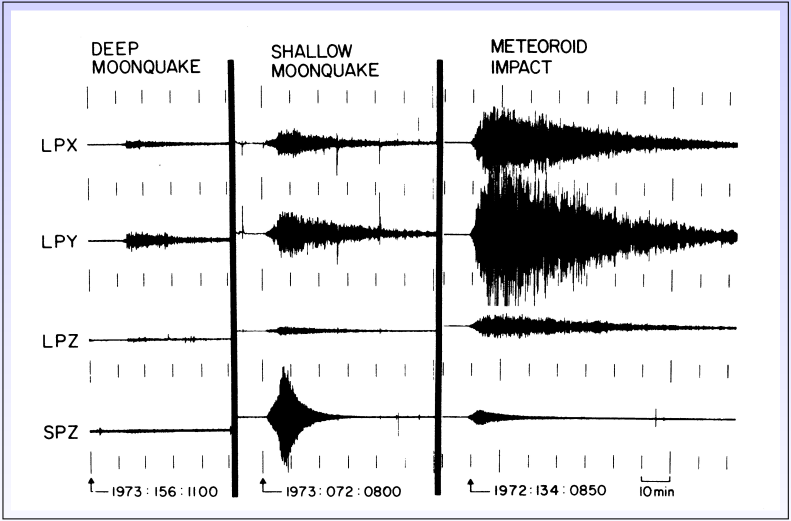


Figure: Lunar sample seismograms. From Nakamura et al. (1983).

Are the Earth and the Moon compositionally alike ?

Figure: A possible sample from the Moon ?

Are the Earth and the Moon compositionally alike ?

	v_p (km/s)
Cheeses	
Sapsego (Swiss)	2.12
Romano (Italy)	1.74
Cheddar (Vermont)	1.72
Muenster (Wisconsin)	1.57
Lunar rocks	
Basalt 10017	1.84
Basalt 10046	1.25
Near-surface layer	1.2
Terrestrial rocks	
Granite	5.9
Gneiss	4.9
Basalt	5.8
Sandstone	4.9

Table 1: Seismic velocities in selected cheeses, lunar and terrestrial rocks. The reduced velocities of lunar rocks in comparison to terrestrial rocks is due to the absence of water and the presence of porosity. (From *Schreiber & Anderson* [1970]).

Thermodynamic Modeling

We compute physical properties directly for a given chemical composition (c), pressure (P) and temperature (T) using Gibbs' free energy minimisation (e.g. Perple_X, Connolly, 2005):

$$\begin{tikzcd}
c \arrow[dr, "\mathrm{g}_1"] & & V_s \\
& M \arrow[ur, "\mathrm{g}_2"] \arrow[r, "\mathrm{g}_2"] \arrow[dr, "\mathrm{g}_2"] & V_p \\
T \arrow[ur, "\mathrm{g}_1"] & & \rho
\end{tikzcd}$$

where c is CFMAS composition, comprising the oxides of the elements CaO-FeO-MgO-Al_2O_3-SiO_2.

Are the Earth and the Moon compositionally alike ?

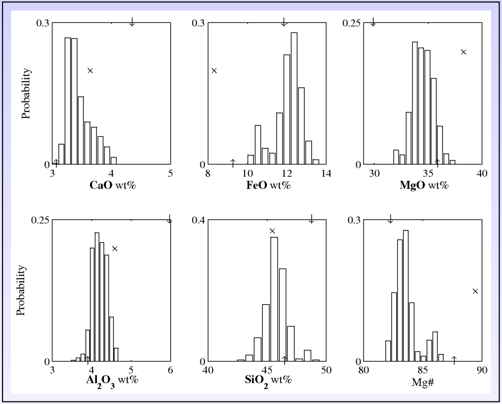


Figure: Posterior pdf's showing sampled bulk lunar compositions (silicate part). Crosses denote the Earth's PUM composition as determined by McDonough & Sun (1995). From Khan et al. (2007).

Are the Earth and the Moon compositionally alike ?

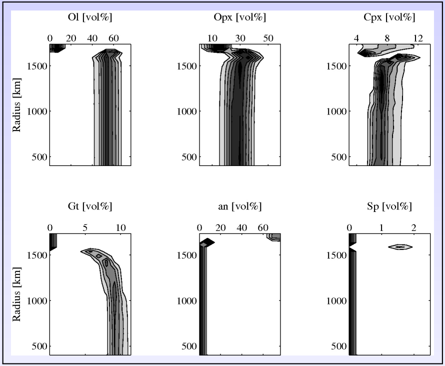


Figure: Sampled lunar modal mineralogy as a function of depth. From Khan et al. (2007).

This model is broadly consistent with constraints on mantle mineralogy derived from the experimental and observational study of the phase relationships and trace element compositions of lunar mare basalts and picritic glasses.