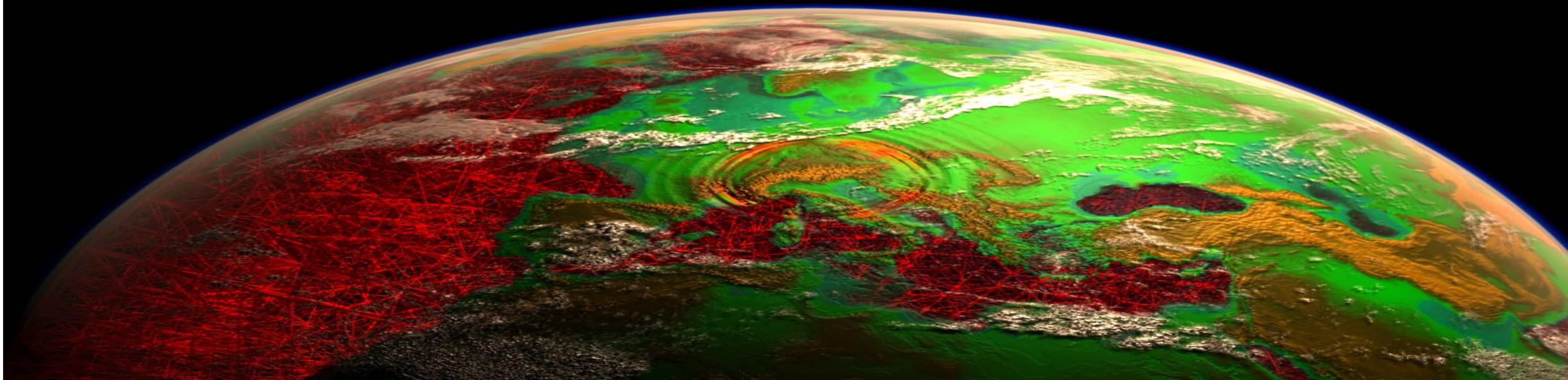
Seismic Tomography

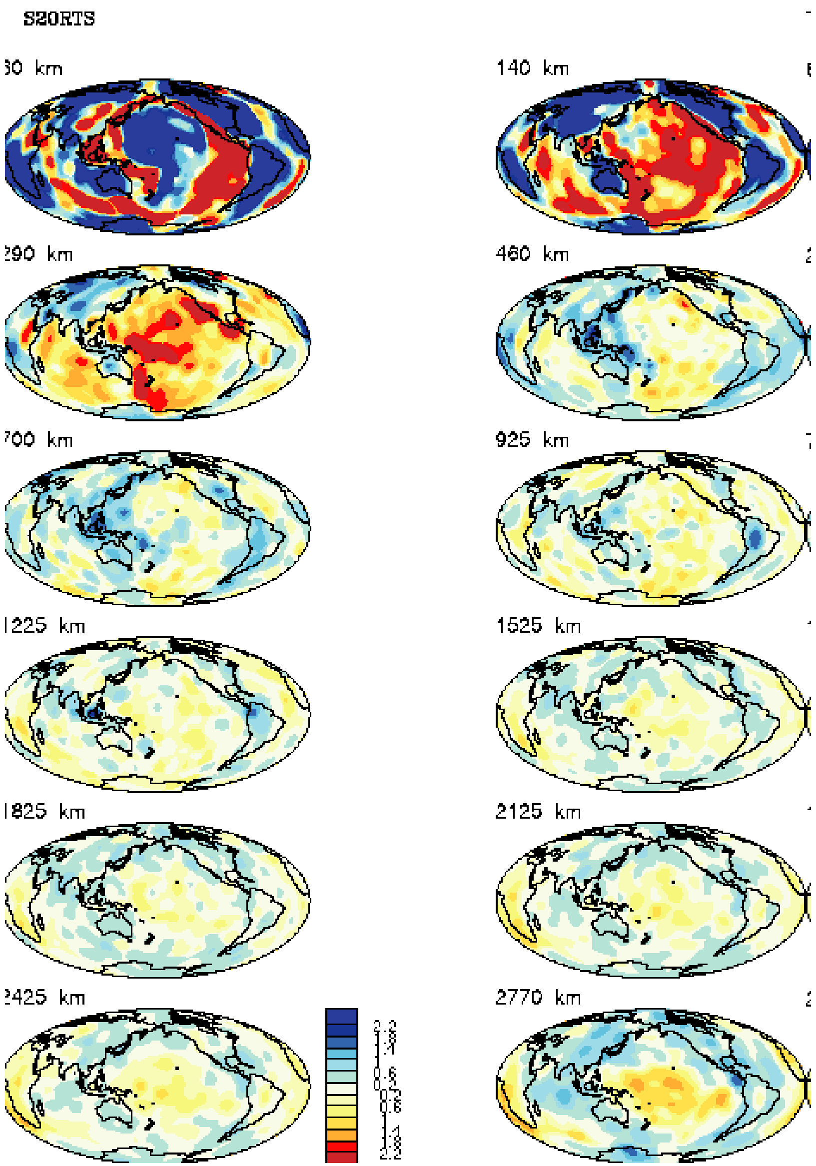
Data, Modeling, Uncertainties

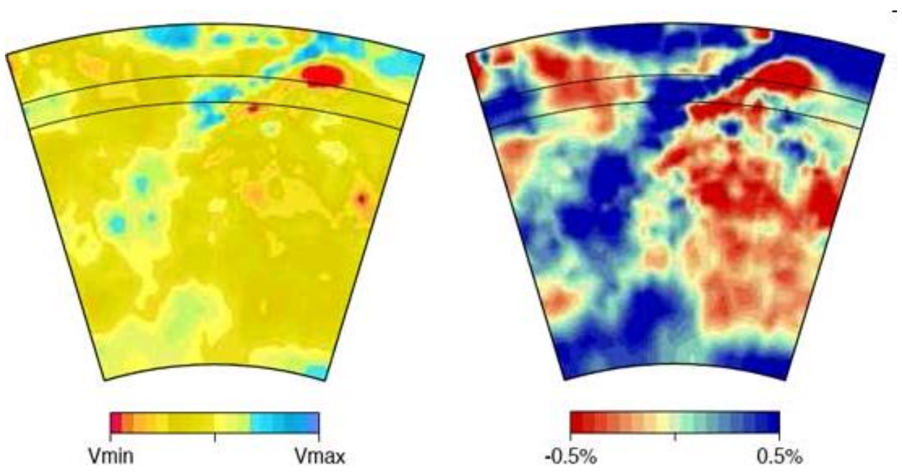
Heiner Igel, LMU Munich



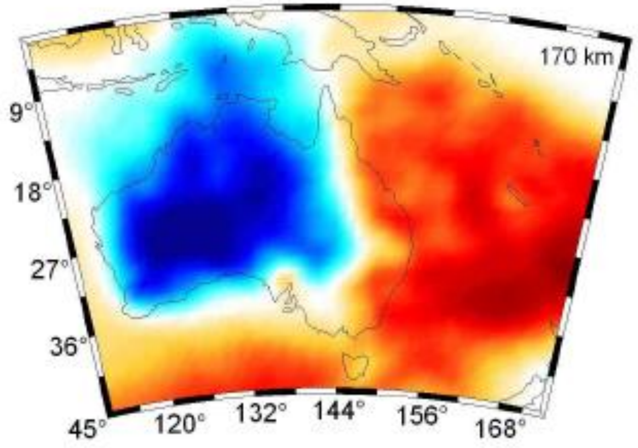
Seismic tomography

global and continental scales





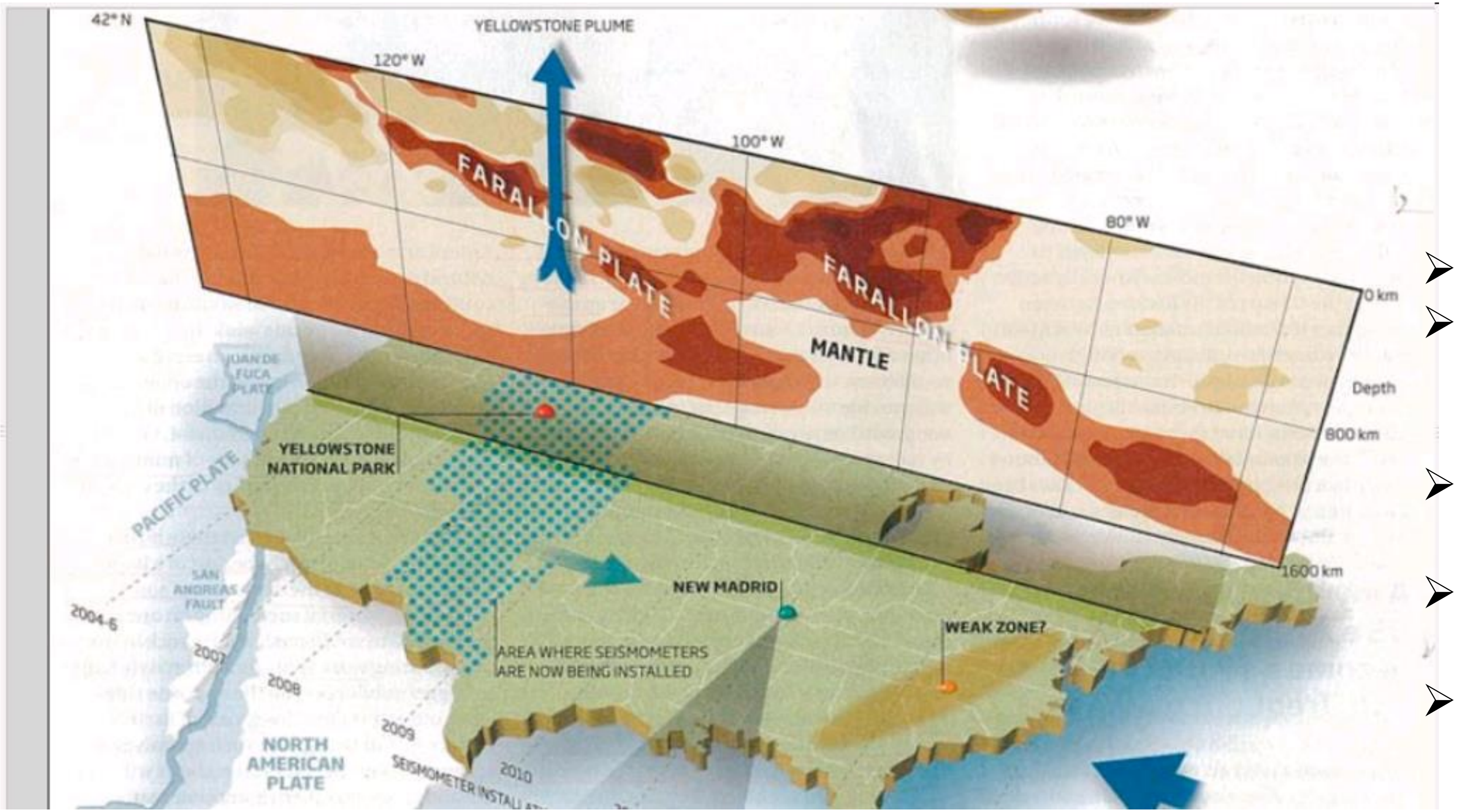
Ricard et al., 2005



Fichtner et al., 2009

Ritsema et al., 2004

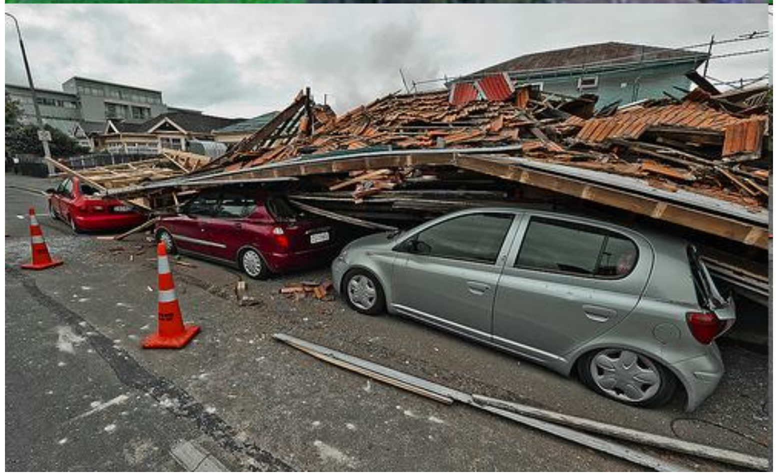
Science



- Fact or fiction?
- Significant geodynamic feature?
- Amplitude correct?
- Spatial scale correct?
- Depth correct?

What went so horribly wrong?

Christchurch, February 2011

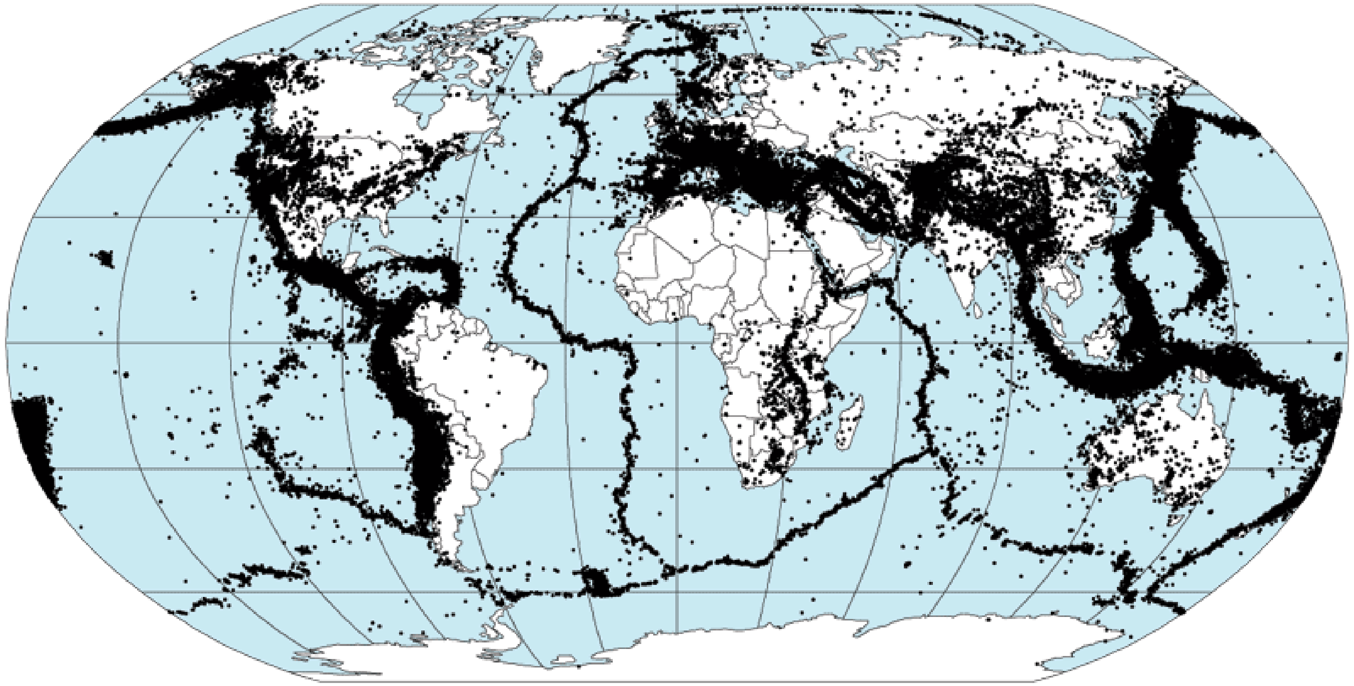


Tohoku-Oki, March 2011

Outline

- Introduction: ***earthquakes***, seismic observations, the seismo-tomographic problem
- „Classic“ tomography using ***seismic rays***
- Full ***waveform inversion*** using 3-D simulation technology – adjoint approach
- Summary and Outlook

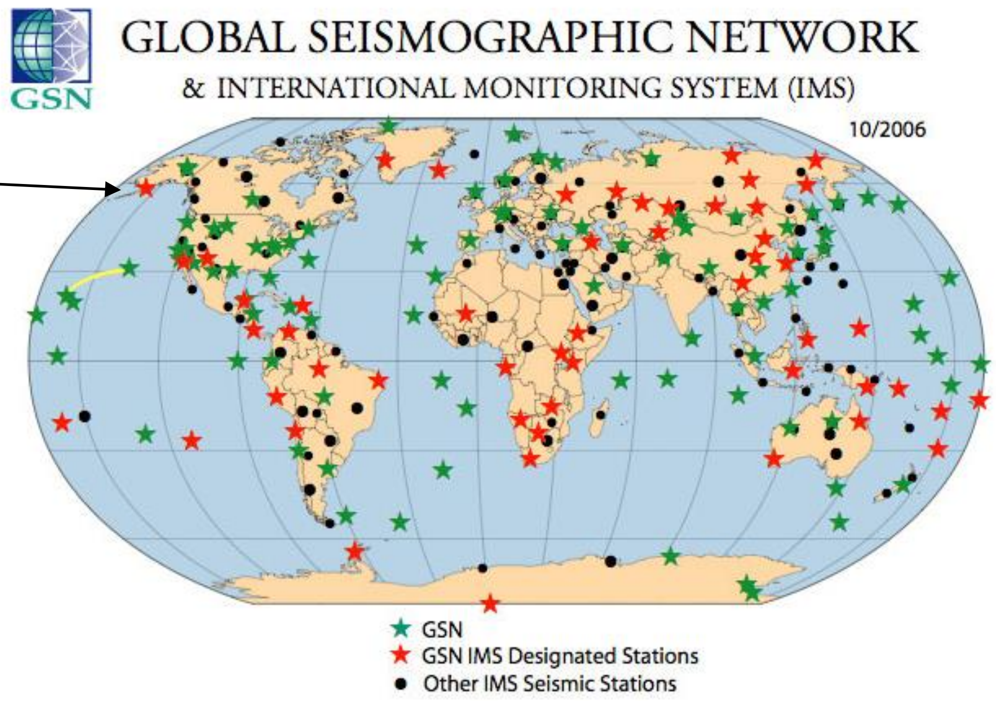
Sources of seismic energy

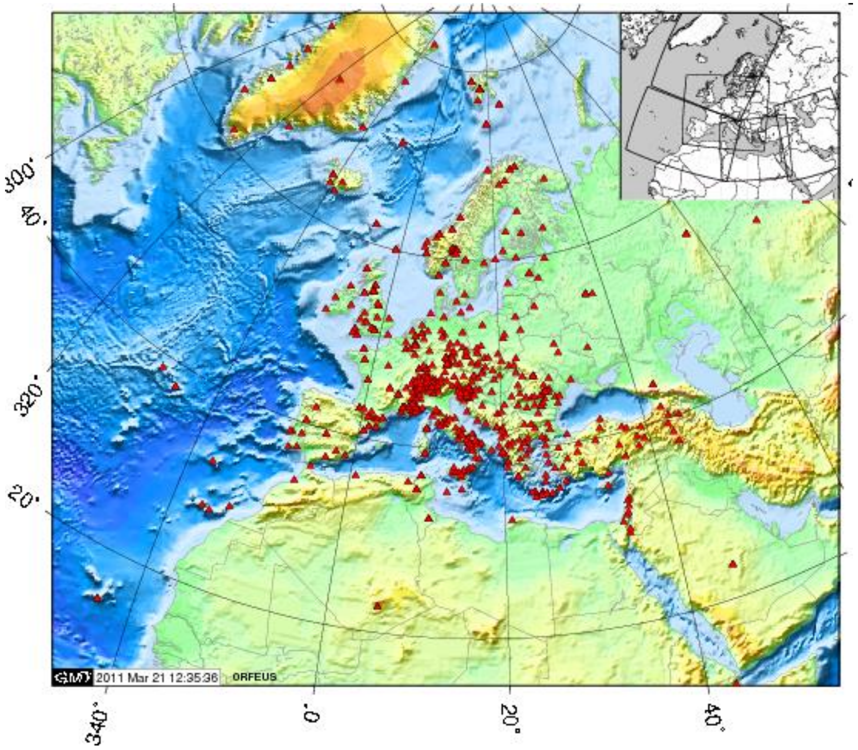


Epicenters 1963 - 1998

358,214 Events

Observational networks

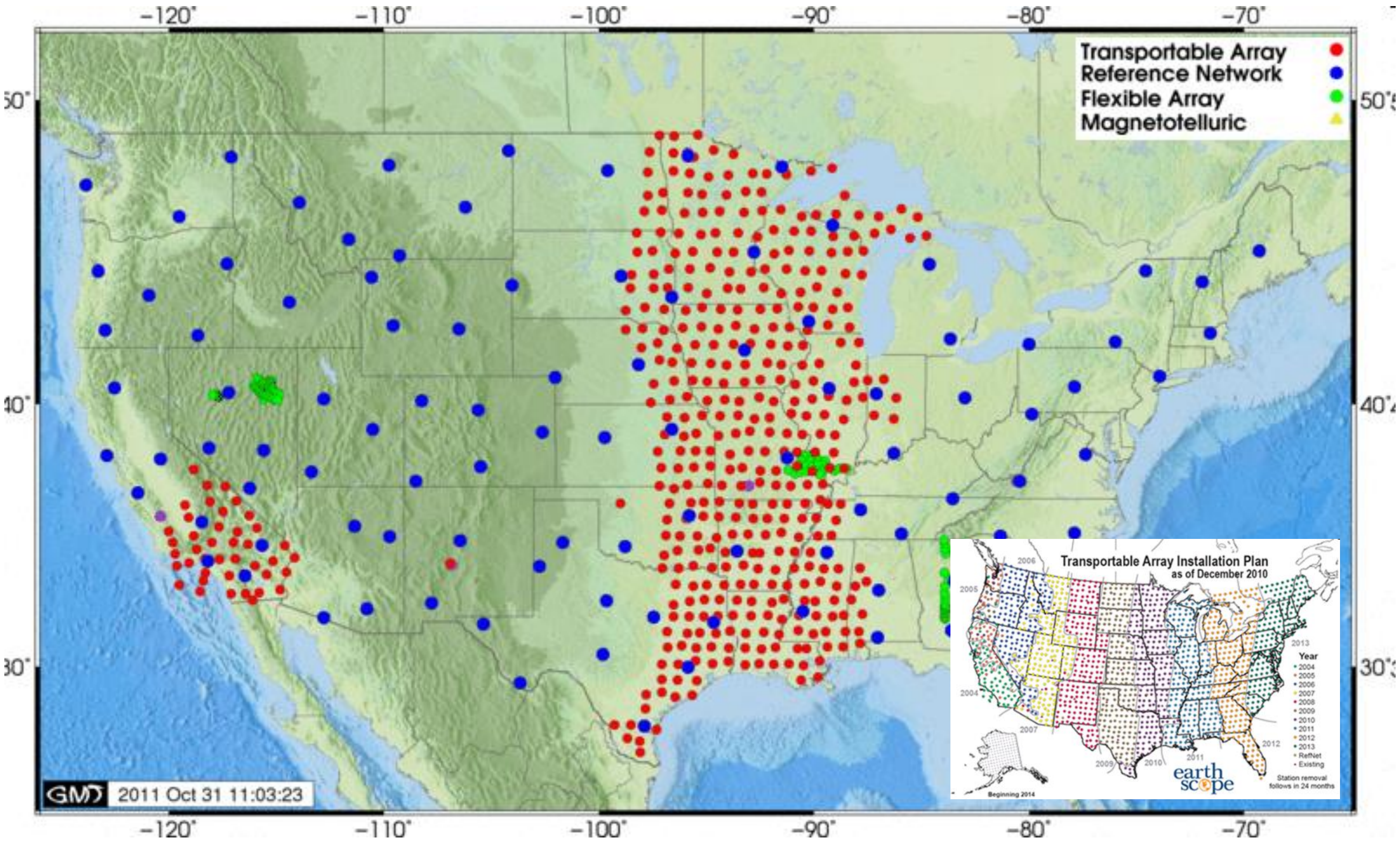




Approx. 1000 instruments in Europe alone

It is unlikely that we populate the oceans with seismometers in the near future!

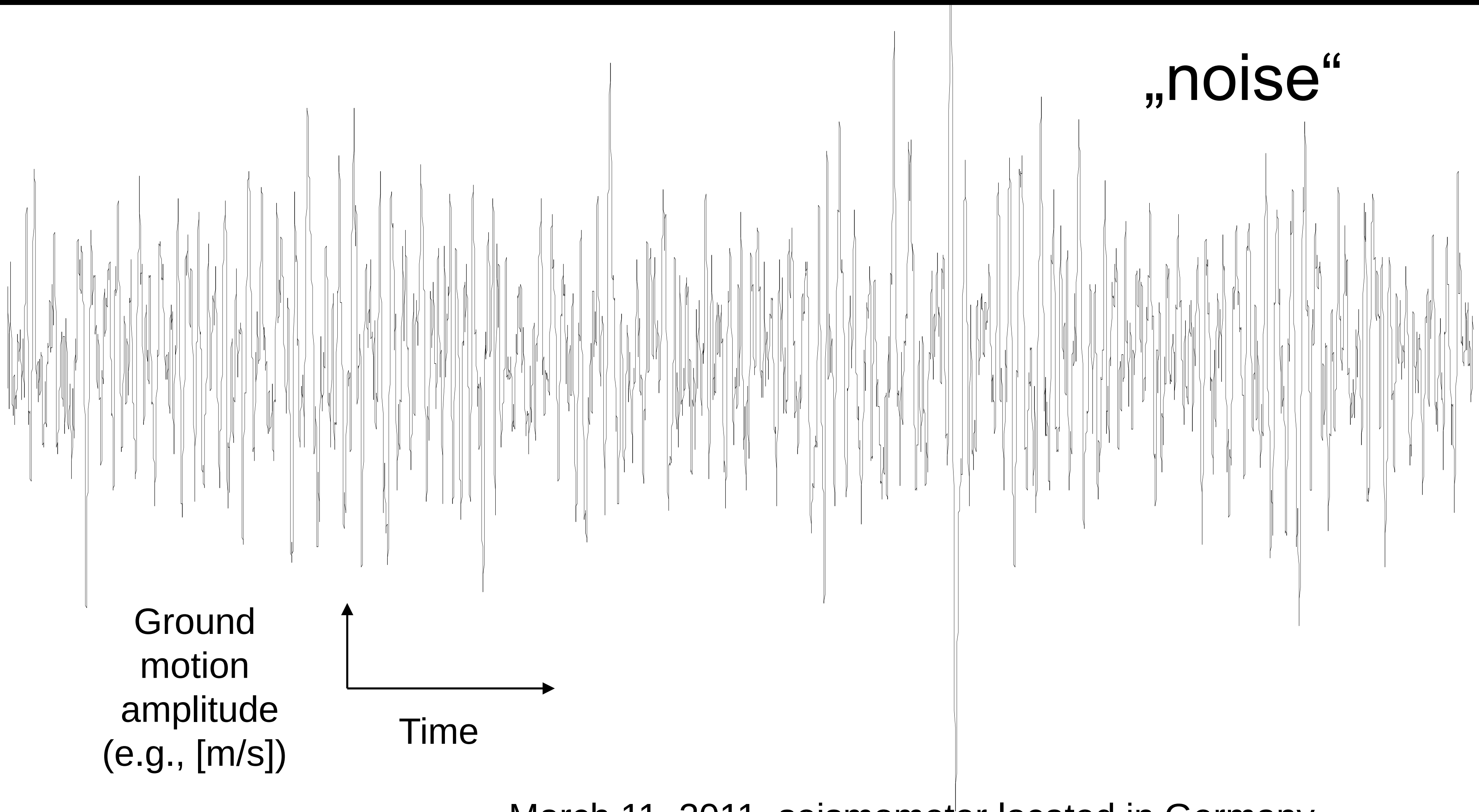
US Array



… new classes of **continental scale** tomographic models are around the corner …

What is the nature of observations and their sensitivities to Earth‘s structure in seismology?

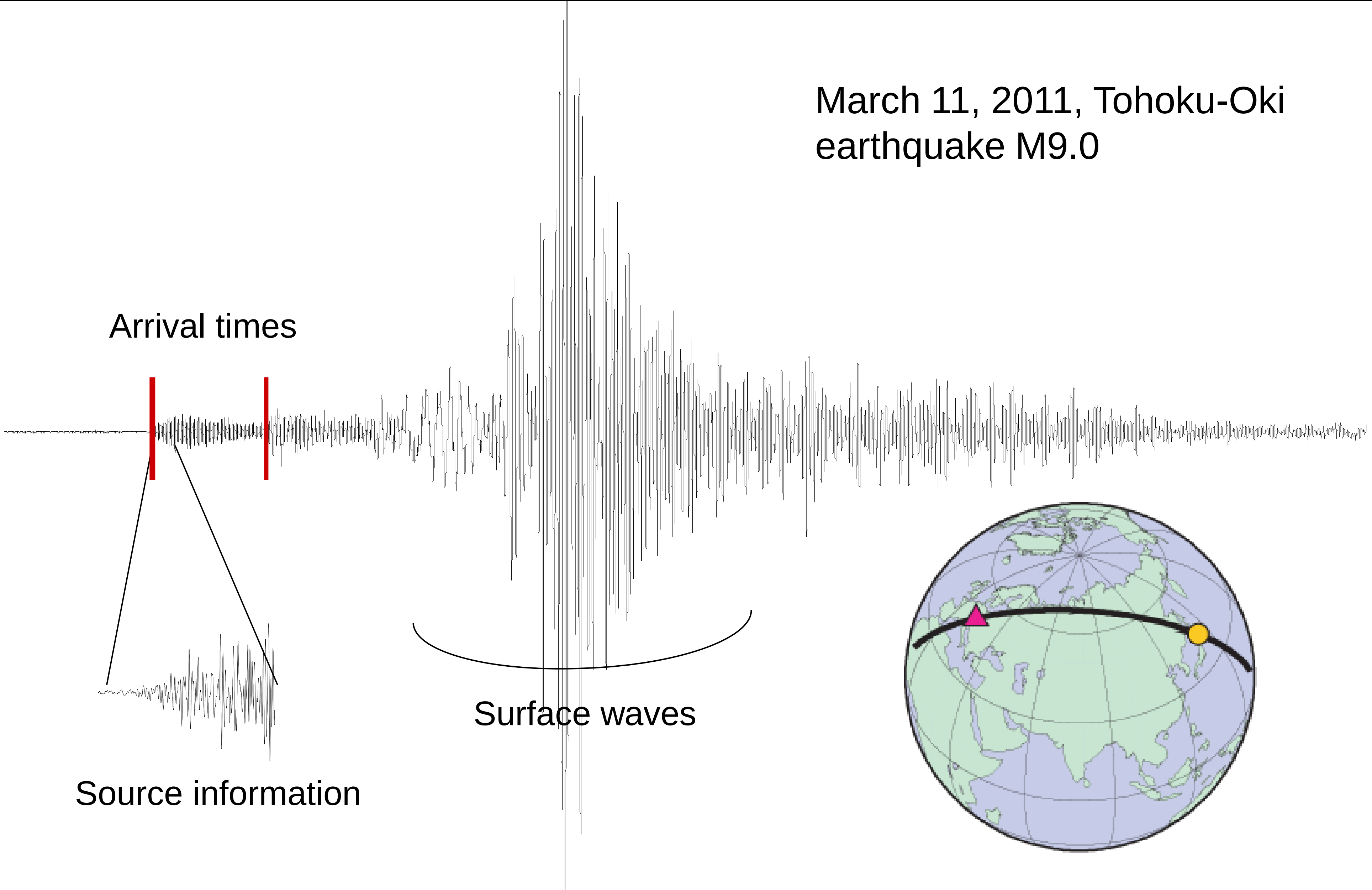
… on a seismically quiet day …



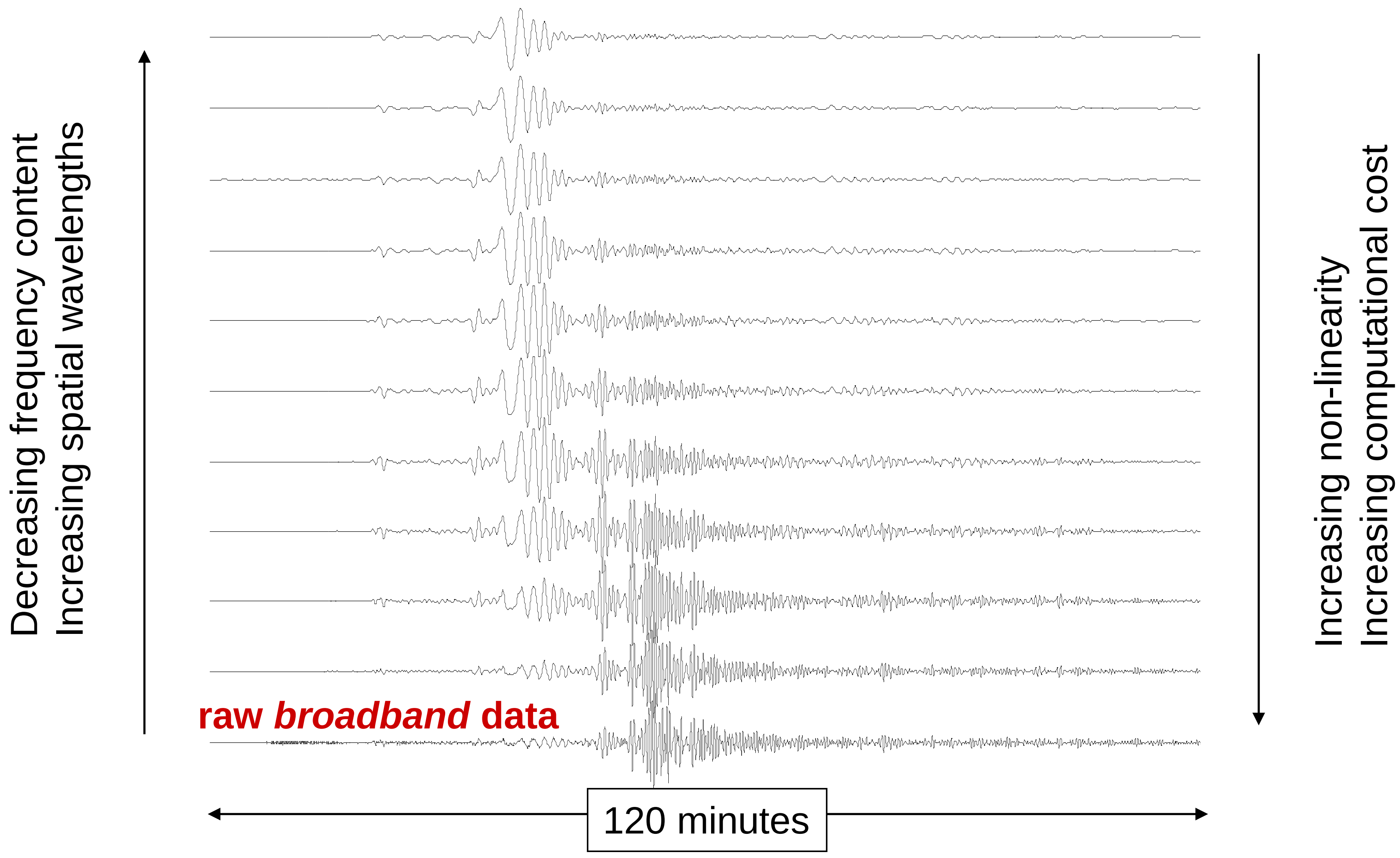
March 11, 2011, seismometer located in Germany

… that turns catastrophic …

March 11, 2011, Tohoku-Oki earthquake M9.0



Temporal scales (vertical ground motion)



Simplified convolutional model

The (noise free) seismic observation is a convolution of the source signal with a Green‘s function …

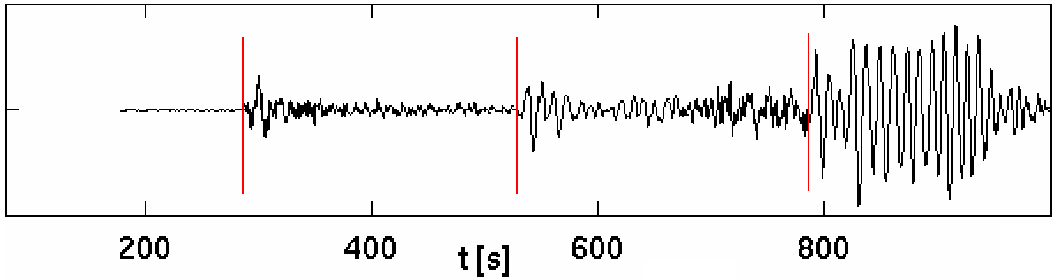
$$U(\omega,\underline{r}) = S(\omega)\, \color{red}{\mathbf{G(\omega,\underline{r})}}\, I(\omega,\underline{r})$$

- $U(\omega,\underline{r})$: Raw data in seismic archives, usually ground velocities in three orthogonal directions
- $S(\omega)$: Source mechanism, the magnitude, the source time behaviour
- $G(\omega,\underline{r})$: Impulse response of the Earth, contains all information on Earth‘s internal structure, site conditions -> **tomography**
- $I(\omega,\underline{r})$: Seismic instrument response affects amplitudes and phase information

The problem is **linear** w.r.t.sources (see talk by M. Mai)

Let‘s briefly summarize …

- Seismograms are affected by structure **and** source
- The seismic tomography problem requires (in principle) the source to be known (or assumed to be known)
- There are **two** strategies to solve the inverse problem



Classic seismic tomography

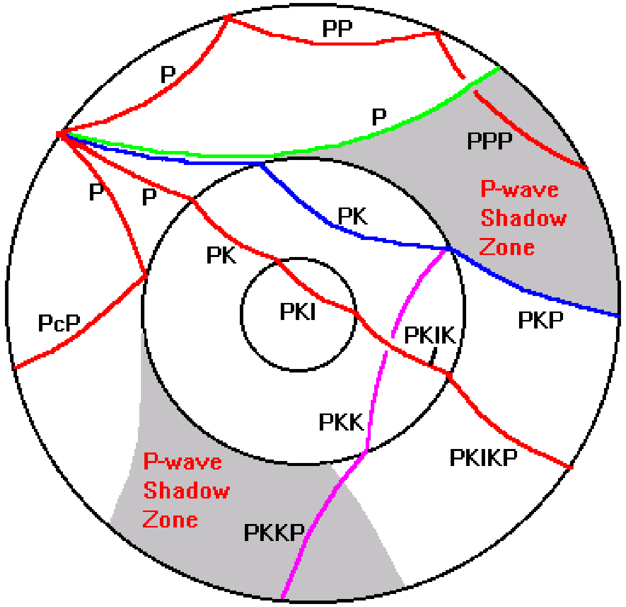
- Reduce information drastically (**travel times**)
- Reduce physics to a high-frequency approximation (ray theory)
- Identify specific signals in seismic data (P and S wave arrivals, reflections, etc.)
- Use linear inverse theory to solve for 3-D velocity structure

Full waveform inversion (FWI)

- Use (low-passed) **full waveforms** as data
- Solve *complete* forward problem (3-D elastic wave propagation)
- Apply adjoint techniques to relate data perturbation to Earth model perturbation
- Iteratively minimize overall misfit between data and synthetics

Seismic tomography

using rays



We ignore surface wave inversion and inversion of free oscillation spectra as the mathematical structure is similar

Seismic ray theory

… is a non linear problem as the ray path depends on the seismic velocity model … after linearization …

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

Travel time ***perturbations*** with respect to an initial model

Dimension **m**

Sensitivity of the i-th measurement to the j-th model parameter (basis function, pixel)

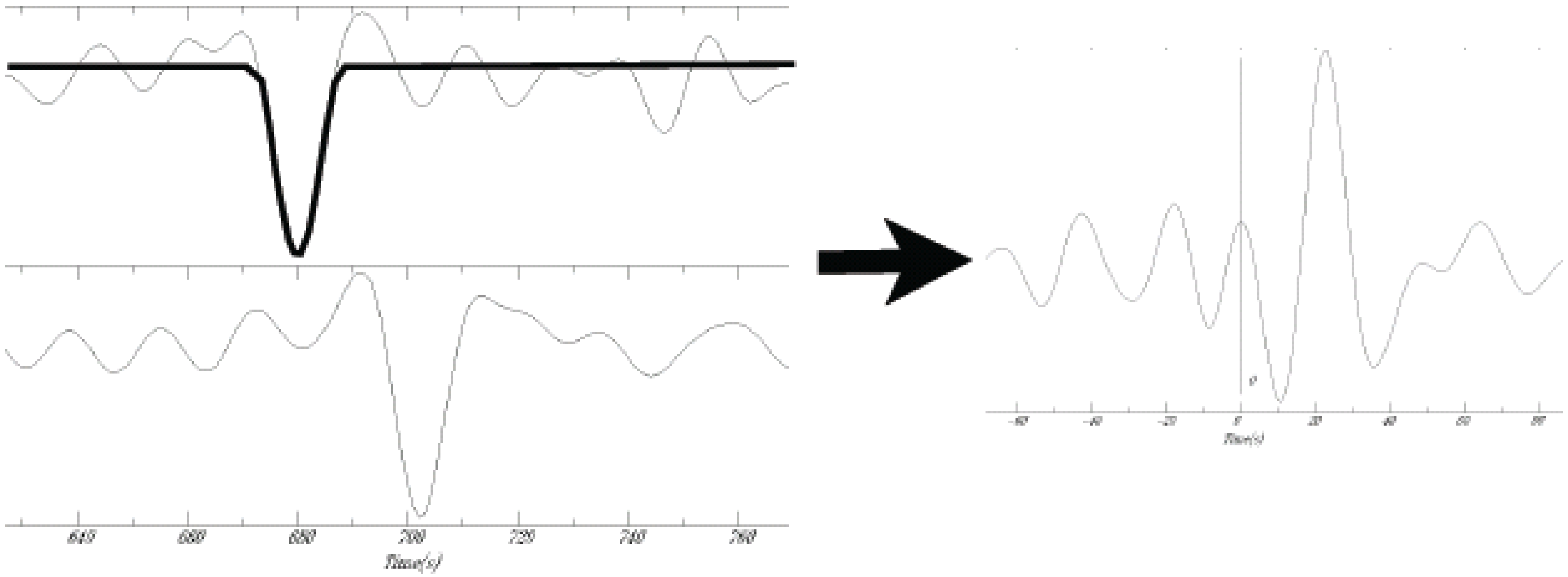
Dimension **m x n**

Solution model (***seismic velocities***)

Dimension **n**

d

What is a travel time perturbation?

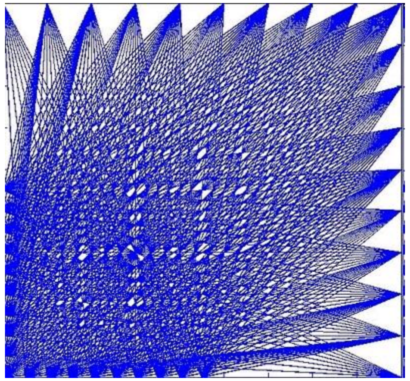


“Picking the onset is at best ambiguous or inaccurate, sometimes impossible.” (Nolet)

G

Operator that relates the model (perturbation) to the observable (travel time perturbation). In general it is an **integral** over the ray path (volume in case of finite frequencies)

$$T = \int_{raypath} \frac{ds}{v}$$



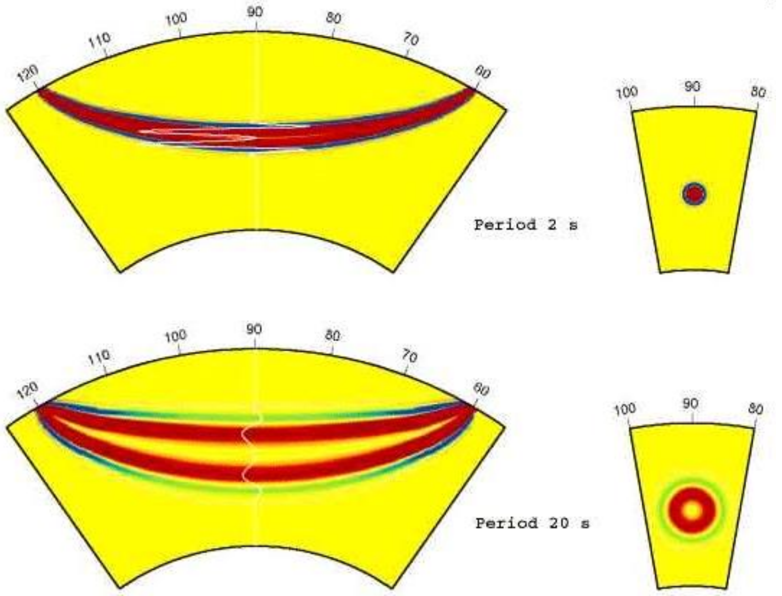
The ij entries to **G** correspond to the i-th ray path affected by the j-th slowness value (pixel or basis function).

The choice of the basis functions strongly affects the density of **G**

G - sensitivities

We can describe the effect of model perturbations on an observable (e.g., travel time dT) by a sensitivity kernel K_X for Earth model parameters seismic velocities (V_P, V_S) and density

$$\delta T = \int \left[K_P \frac{\delta V_P}{V_P} + K_S \frac{\delta V_S}{V_S} + K_\rho \frac{\delta V_\rho}{V_\rho} \right] d^3 r$$



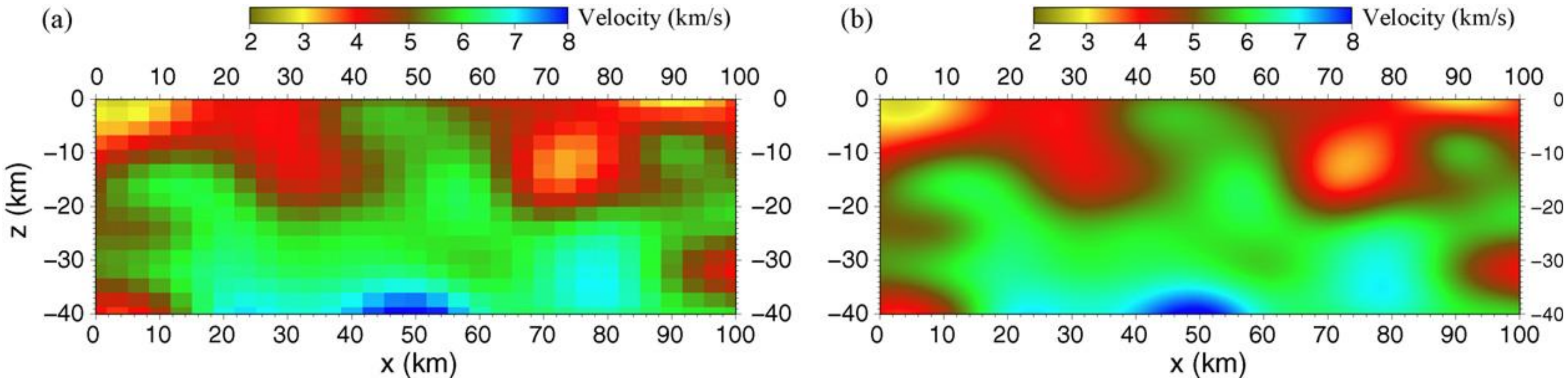
Issues:

- Trade offs
- Amplitude information
- Little sensitivity on density
- Low velocity anomalies

m

Ray-based tomographic problems have (only) P and/or S velocities as unknowns (not density, impedance, etc).

Possible parametrizations: blocks, complex volumes, splines, spherical harmonics, irregular tetrahedra, etc.



Blocks

Splines

Solution to the Inverse Problem

Basic least squares (LS) solution of the linear (-ized) inverse problem with **D** containing the cumulative effects of the regularization (smoothing) constraints (e.g., Tikhonov regularization)

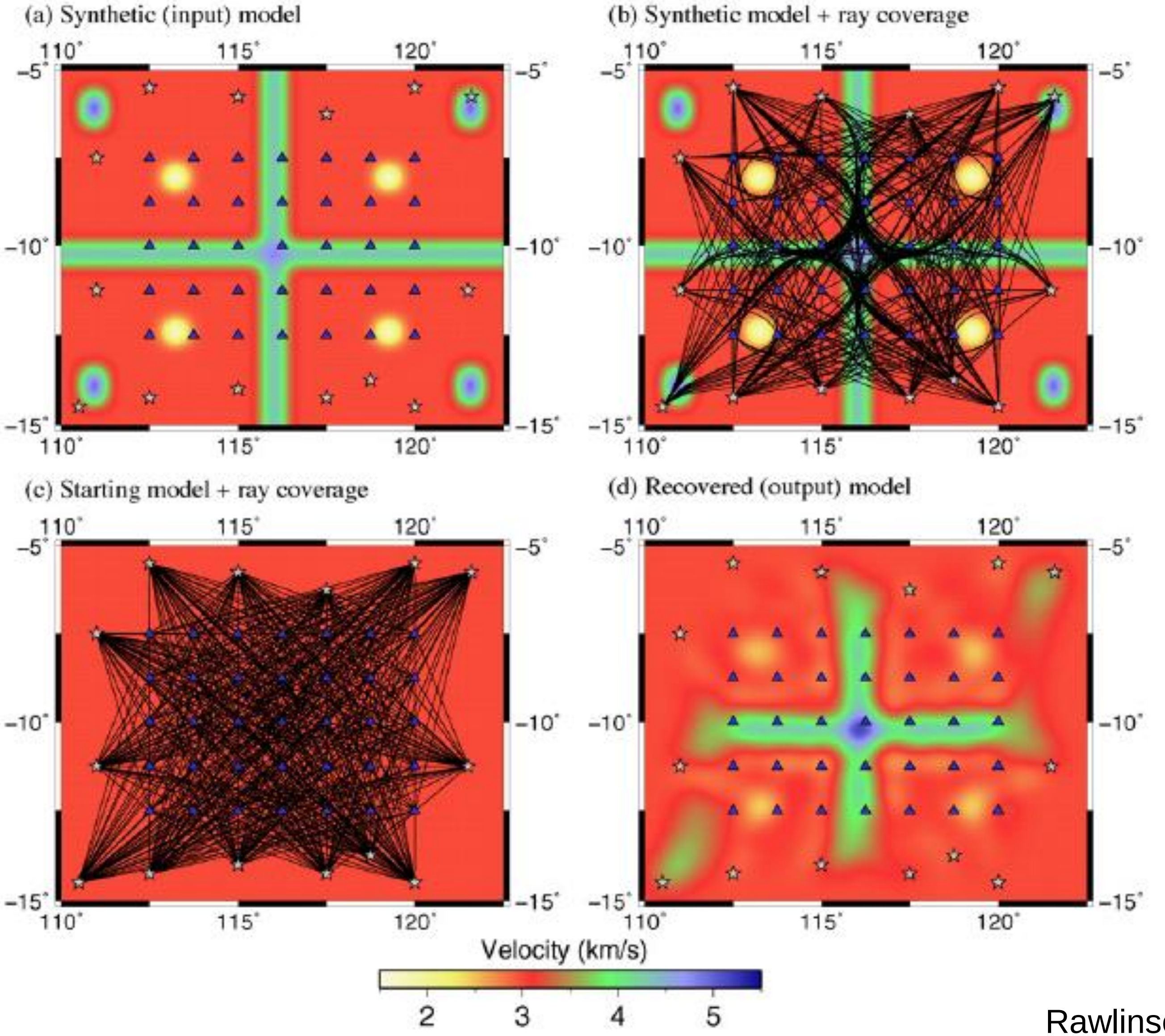
$$\Delta\mathbf{m}_{\mathbf{LS}} = (\mathbf{G}^{\mathbf{T}}\mathbf{G} + \mathbf{D})^{-1}\mathbf{G}^{\mathbf{T}}\Delta\mathbf{d}$$

Solution of this equation with conjugate gradient, LSQR, or other.

Typical dimensions:

Δd -> 10^7 travel time perturbations
Δm -> 10^5 - 10^6 unknowns

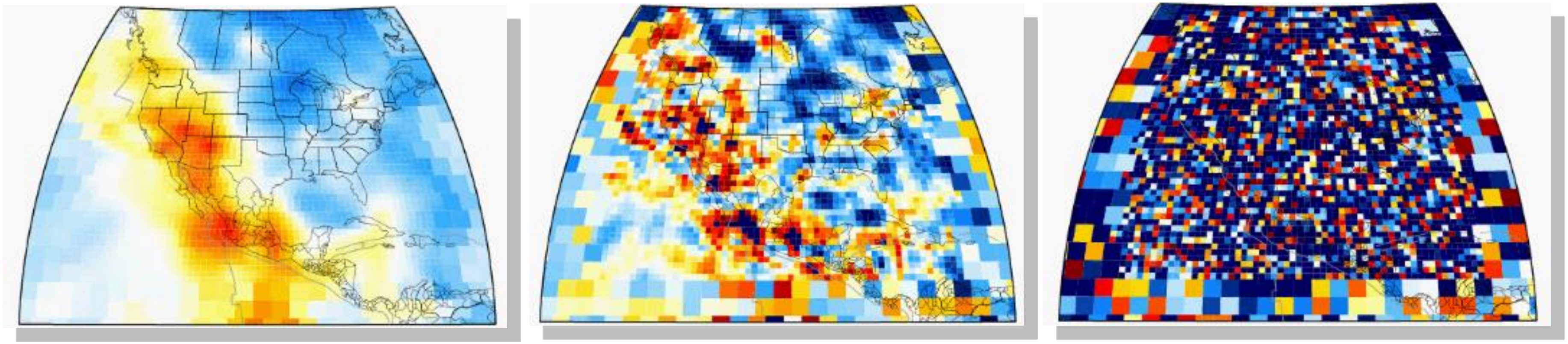
Example



Rawlinson et al., 2010

Regularization and smoothing

Decreasing misfit

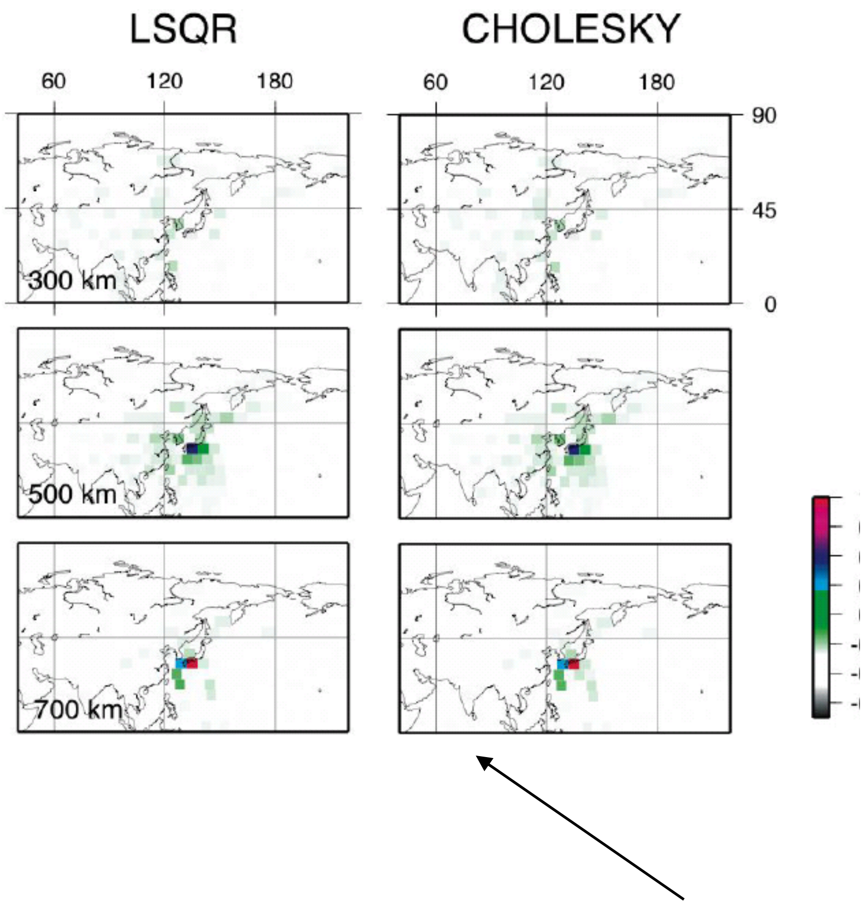


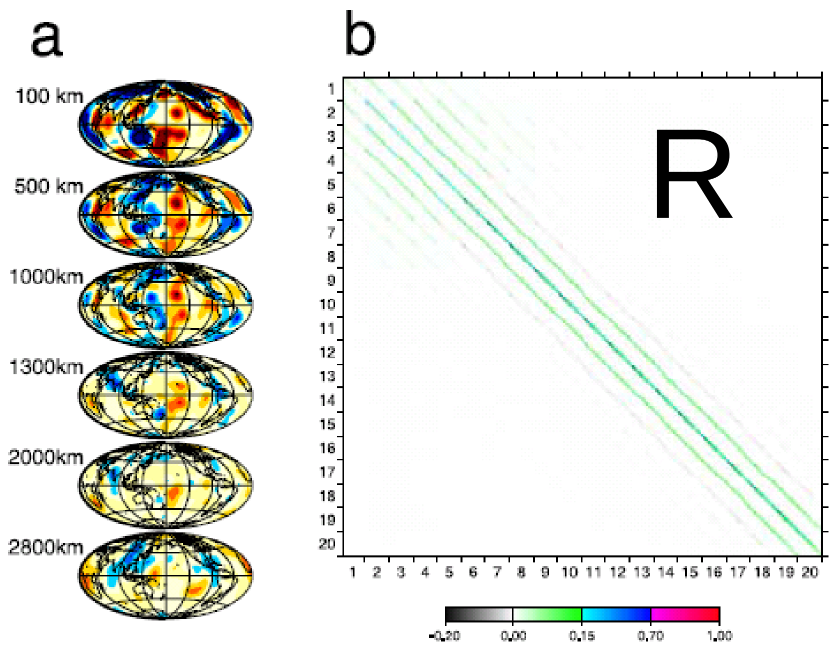
Increasing model complexity

Increasing number of degrees of freedom

Courtesy: L. Boschi

Examples



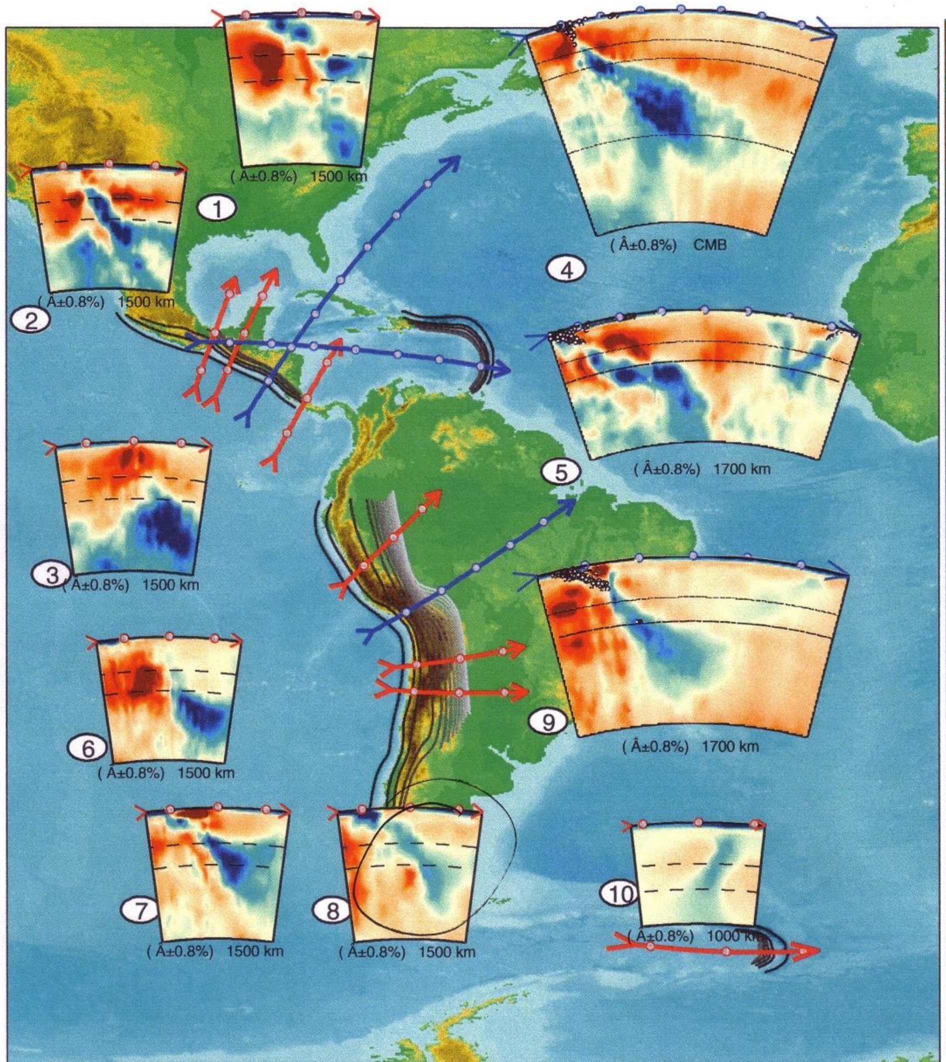


Boschi (2003)

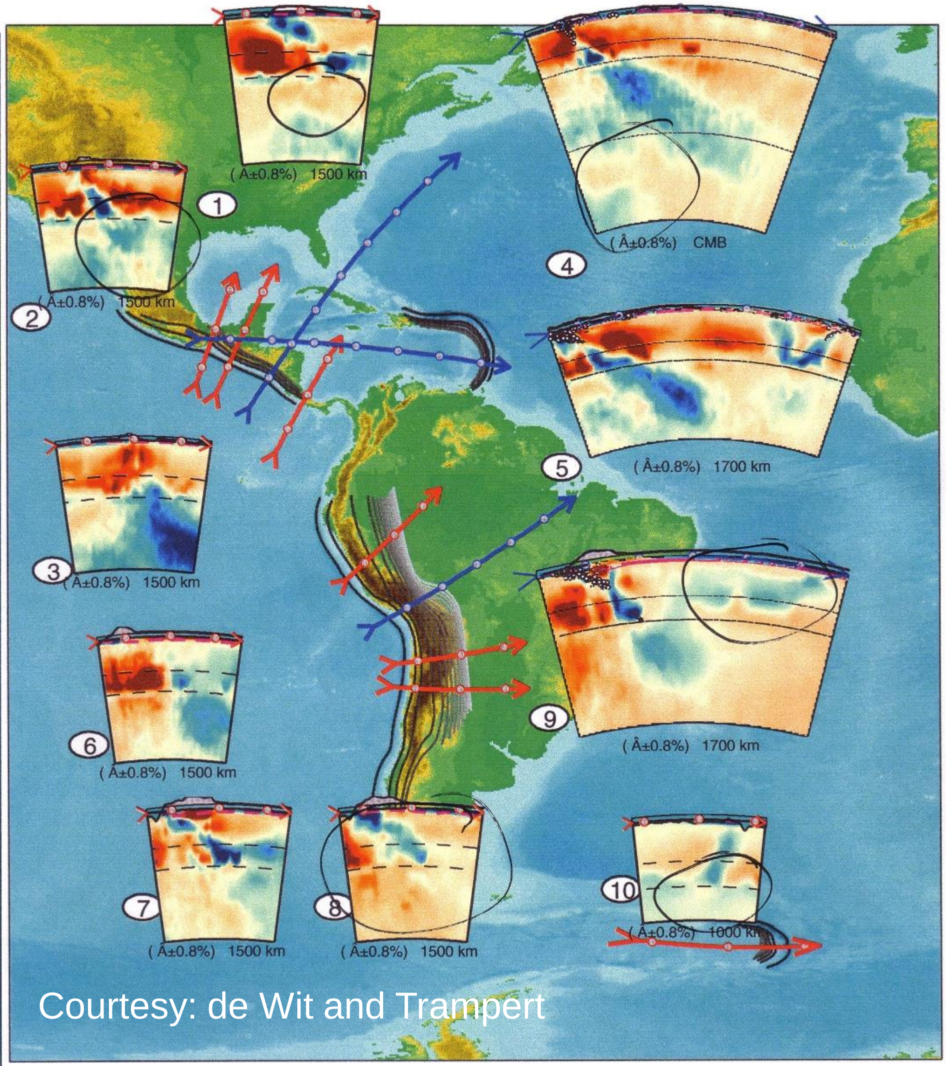
Rows of R for a well resolved pixel at 700 km depth

Exploring null spaces using SVD
misfit remains the same (< ε)

Original



Modified

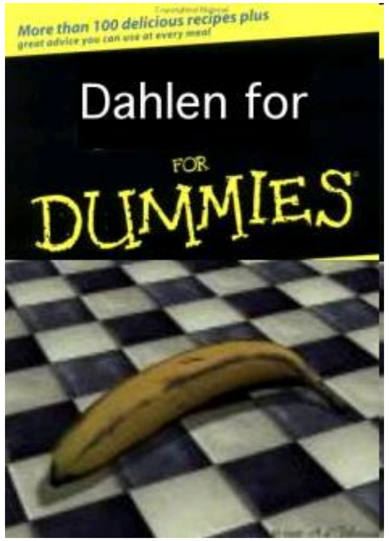


Courtesy: de Wit and Trampert

Ray-based tomography – future directions

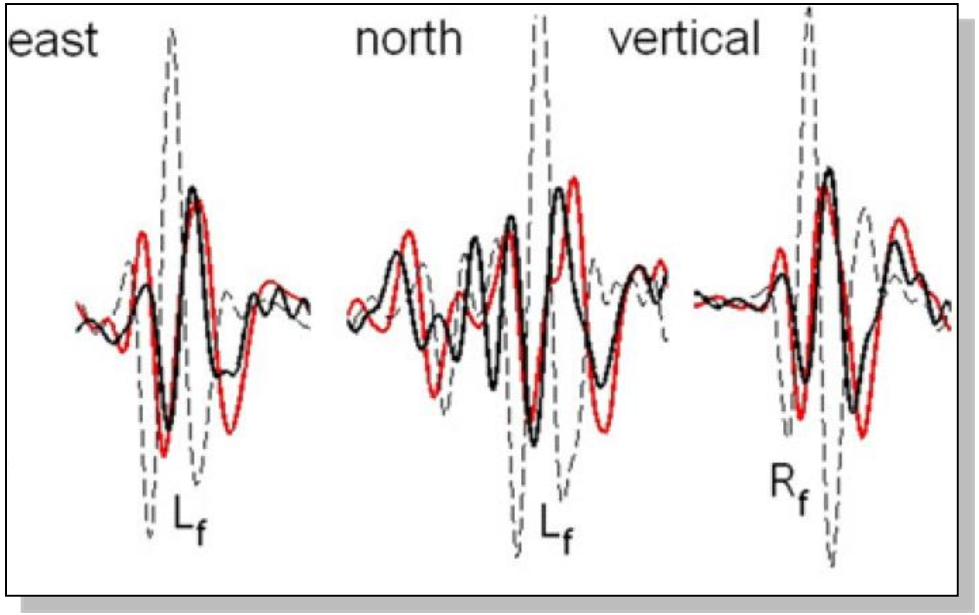
… from infinite to finite frequencies …

- Extracting travel times at different frequencies facilitates the solution of the system and adds information on the model (?)
- Finite-frequency tomography using complete kernels calculated with 3-D wave propagation tools
- Using Monte Carlo type techniques to quantify resolution (see talk by R. Zhang) in a Bayesian framework
- Calculating resolution matrix R for **really** big systems (not done yet)



The *real* thing:

Full waveform inversion

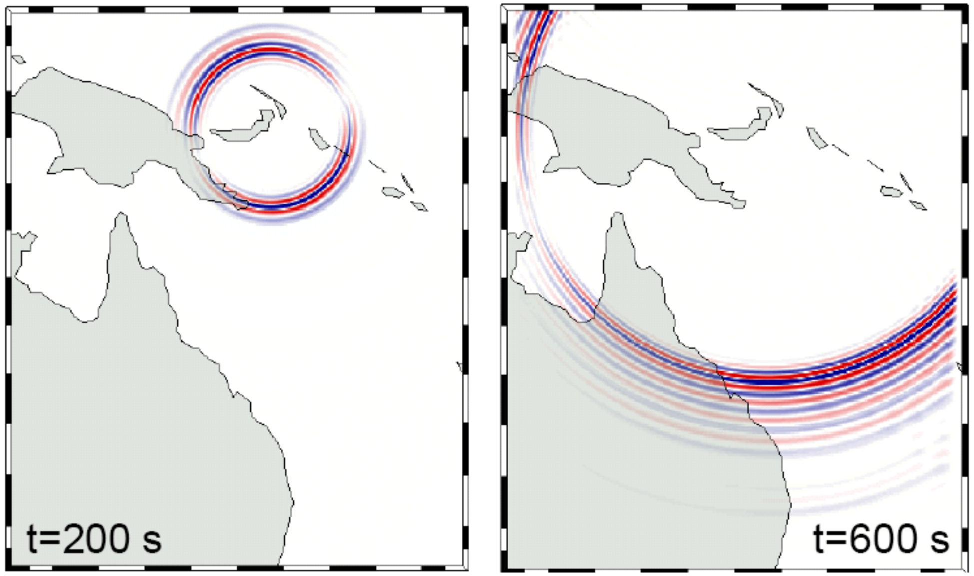


Forward problem

density — elastic tensor — force density

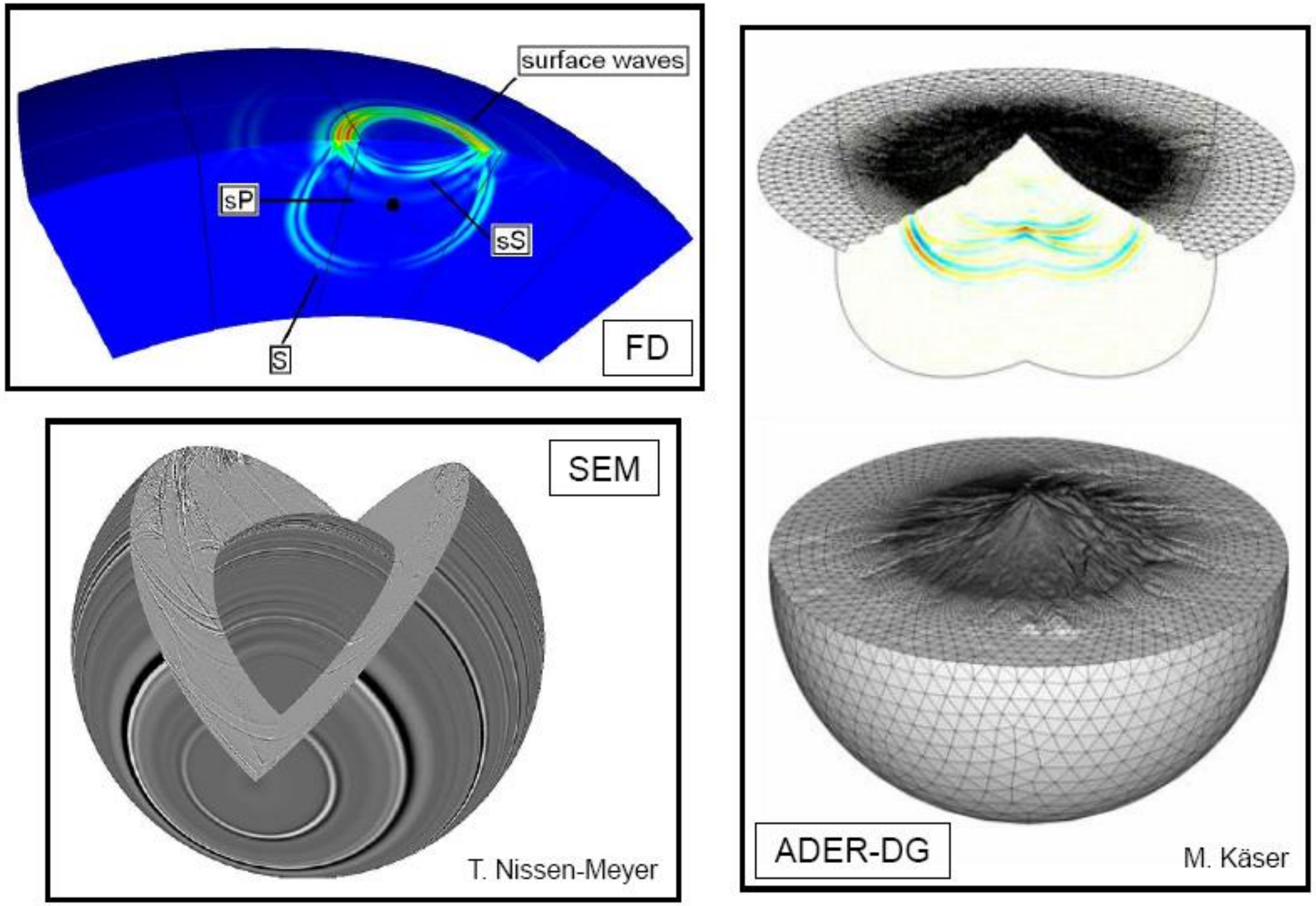
$$\rho\, \ddot{u}_i - \frac{\partial}{\partial x_j}\left(C_{ijkl} * \frac{\partial}{\partial x_k} u_l \right) = f_i$$

elastic displacement field, **u**



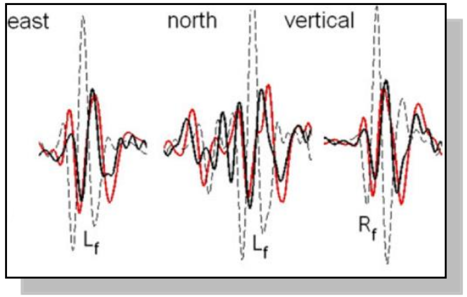
wave field @ 100 km depth

Forward problem



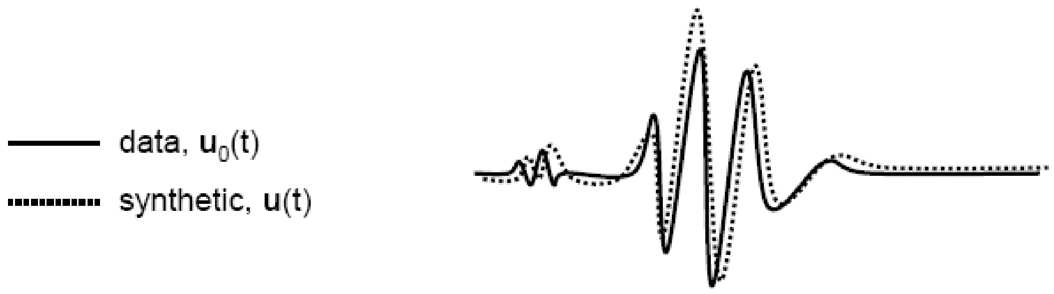
Seismology (waves and rupture) has a good **benchmarking** culture!

Three stages of FWI



forward problem	sensitivities	inversion
seismic wave propagation through heterogeneous Earth models	Quantify misfit between theory and observations	Find appropriate step length
dissipation & anisotropy	Relate data perturbation to model perturbation (adjoint -> gradient)	Calculate model update
spectral-element discretisation of the seismic wave equation	Improve gradient (preconditioning)	Adapt temporal and spatial scales (multigrid)
		Iterate until satisfactory fit
		Estimate uncertainties?

Misfit calculation



L_2 waveform misfit: $\chi = \sqrt{\int_t [u(t) - u_0(t)]^2 dt}$

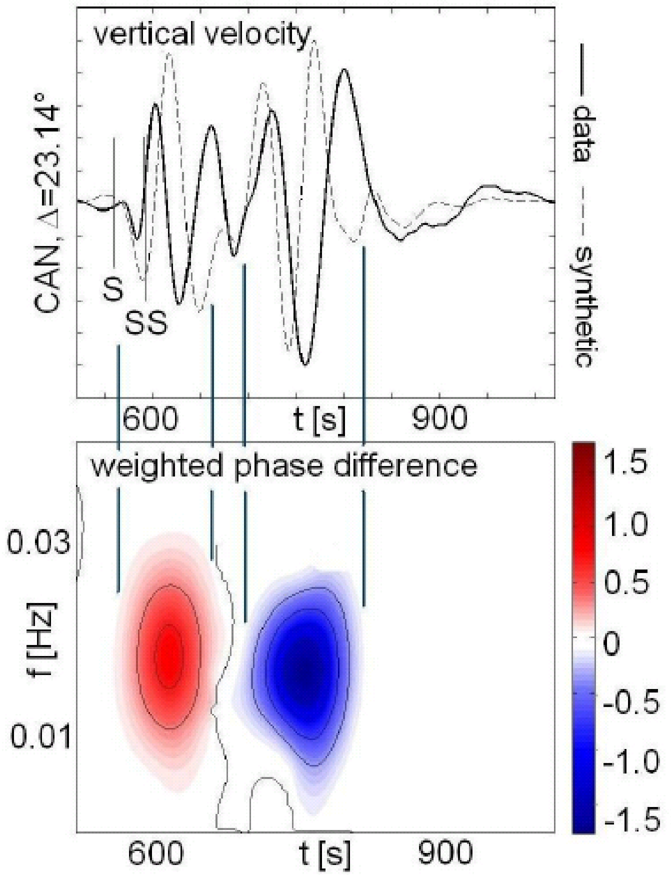
advantages

- easy and fast to implement
- uses the complete waveform

disadvantages

- not robust
- very nonlinearly related to long-wavelength structure
- over-emphasises large-amplitde waves

Time – frequency misfits



Time-frequency misfits

phase differences as functions of time and frequency

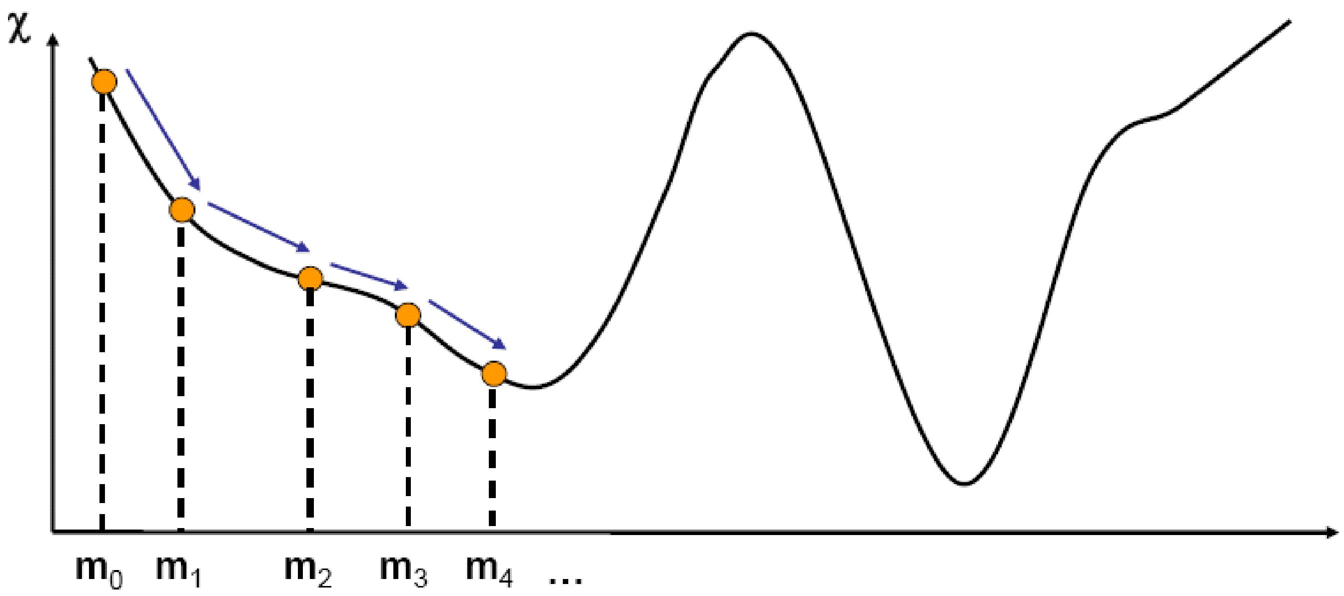
- **quasi-linearly related to Earth structure**
  improves convergence
- **independent of amplitudes**
  reliably measurable, deep structure information
- **applicable to complex waveforms**
  interfering waves, unidentifyable waves
- **continuous in frequency**
  no discrete frequency bands

Gradient-based inversion

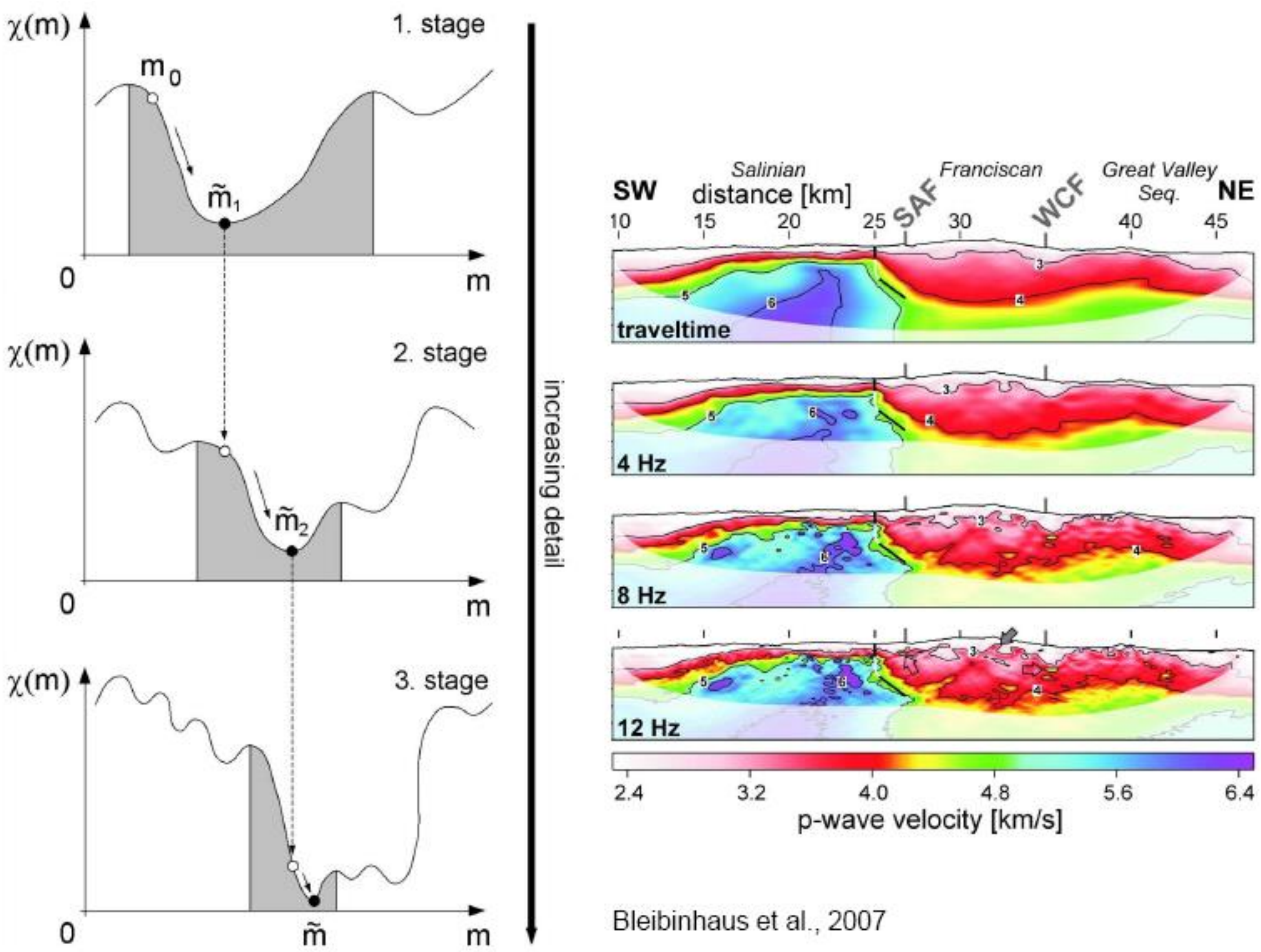
1. Start from initial Earth model $\mathbf{m}_0$

2. Update according to $\mathbf{m}_{i+1} = \mathbf{m}_i + \gamma_i \mathbf{h}_i$, with $\chi(\mathbf{m}_{i+1}) < \chi(\mathbf{m}_i)$

step length (γ_i) — descent direction ($\mathbf{h}_i$)



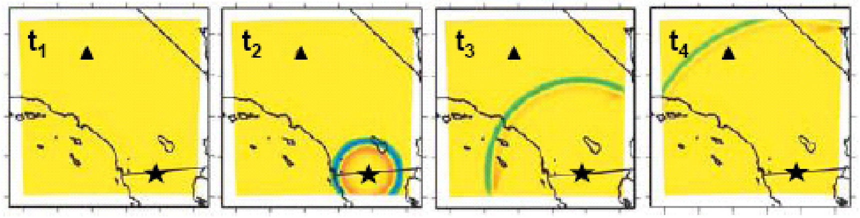
Multi-scale approach



Bleibinhaus et al., 2007

The gradient (adjoint based)

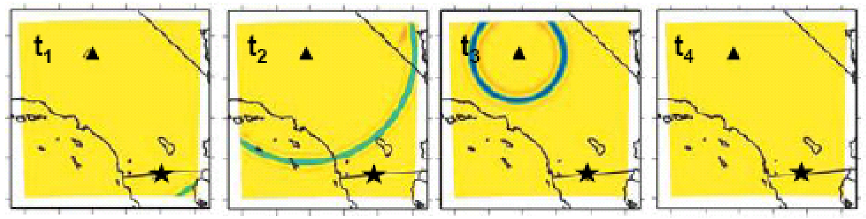
1. Solve the forward problem



→ forward field u
synthetic seismograms

2. Evaluate the misfit χ

3. Solve the adjoint problem
- also a wave equation
- runs backwards in time away from the receiver
- source determined by the misfit



→ adjoint field u^t

4. Compute the gradient by correlating u and u^t

$$\frac{\partial \chi(\mathbf{m})}{\partial \mathbf{m}_i} = \int_{\text{Earth}} \int_{\text{time}} [\mathbf{u} * \mathbf{u}^t]\, dt\, d^3x$$

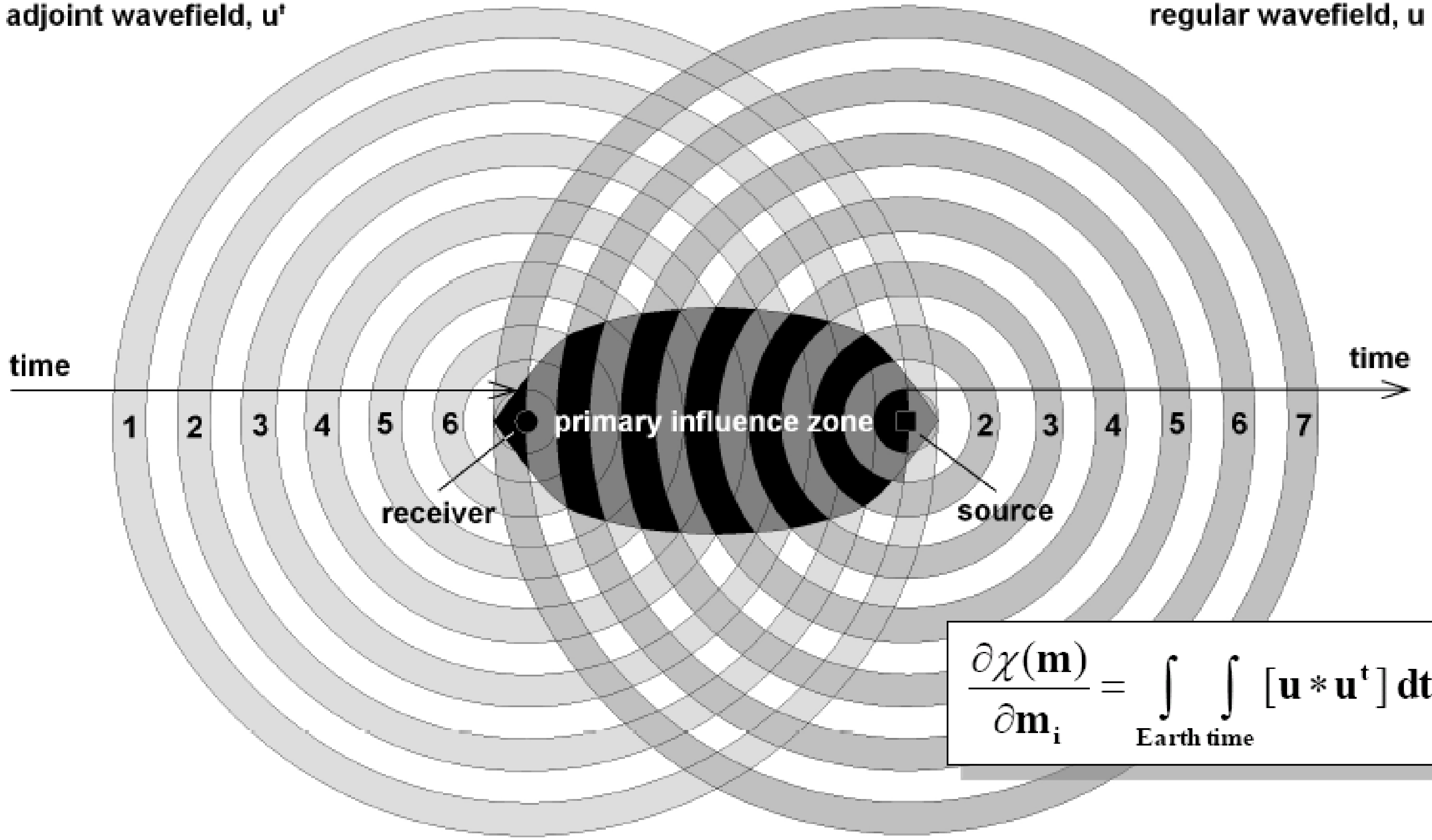
Fréchet kernel
sensitivity kernel
sensitivity density

Tape et al., 2007

The sensitivity kernel

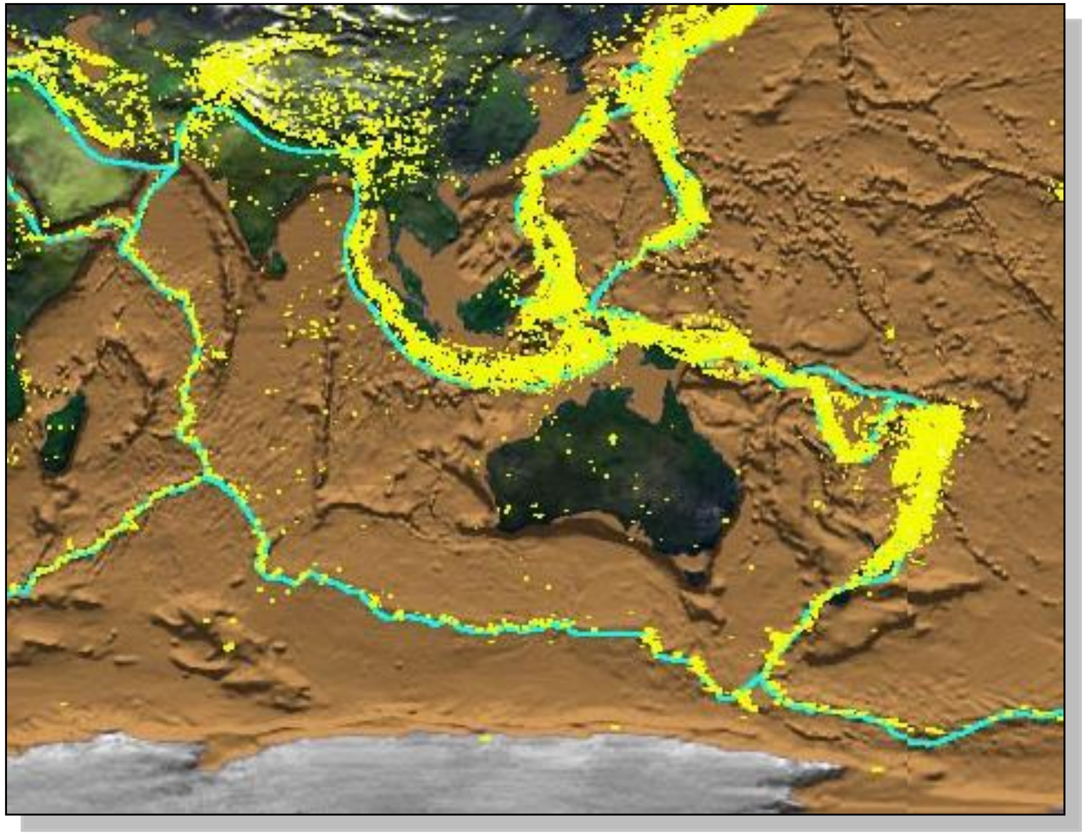
The interaction of the regular and the adjoint fields generates a primary influence zone.

First-order scattering from within the primary influence zone affects the measurement.

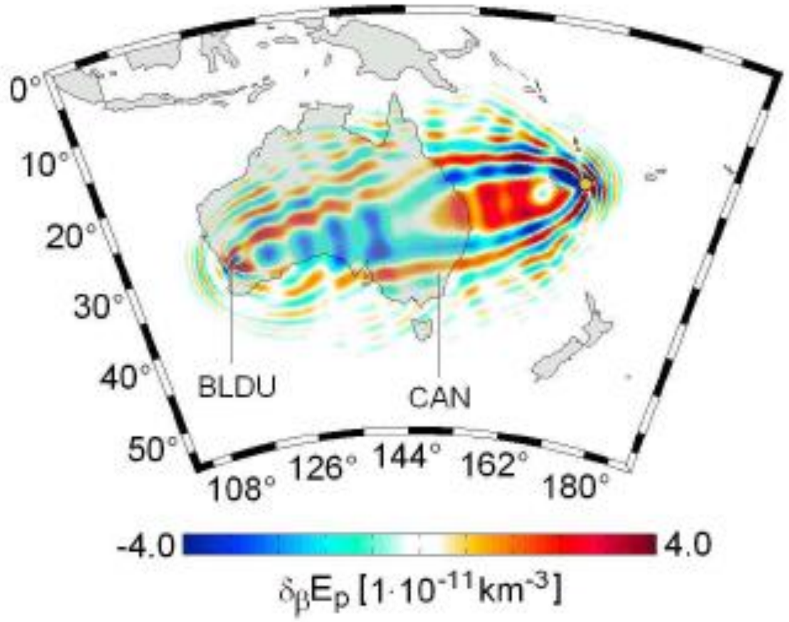


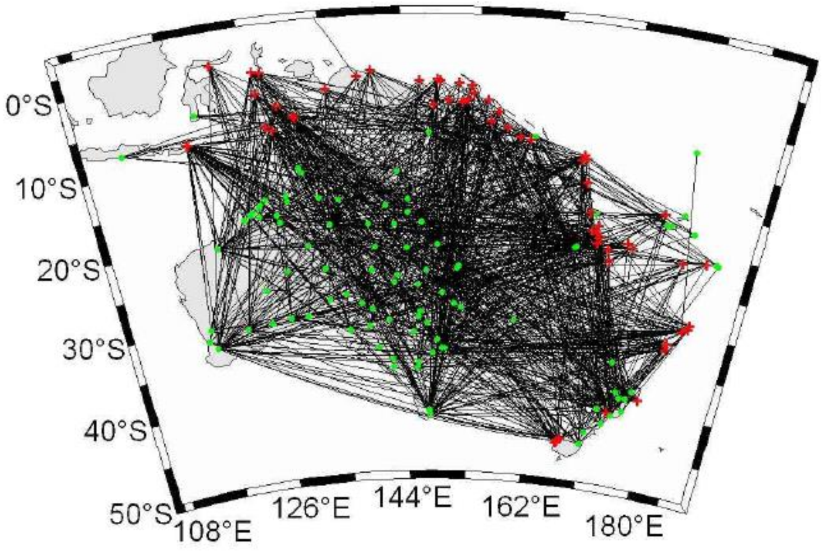
$$\frac{\partial \chi(\mathbf{m})}{\partial \mathbf{m}_i} = \int_{\text{Earth}} \int_{\text{time}} [\mathbf{u} * \mathbf{u}^t]\, dt\, d^3\mathbf{x}$$

An example of full waveform inversion on a continental scale



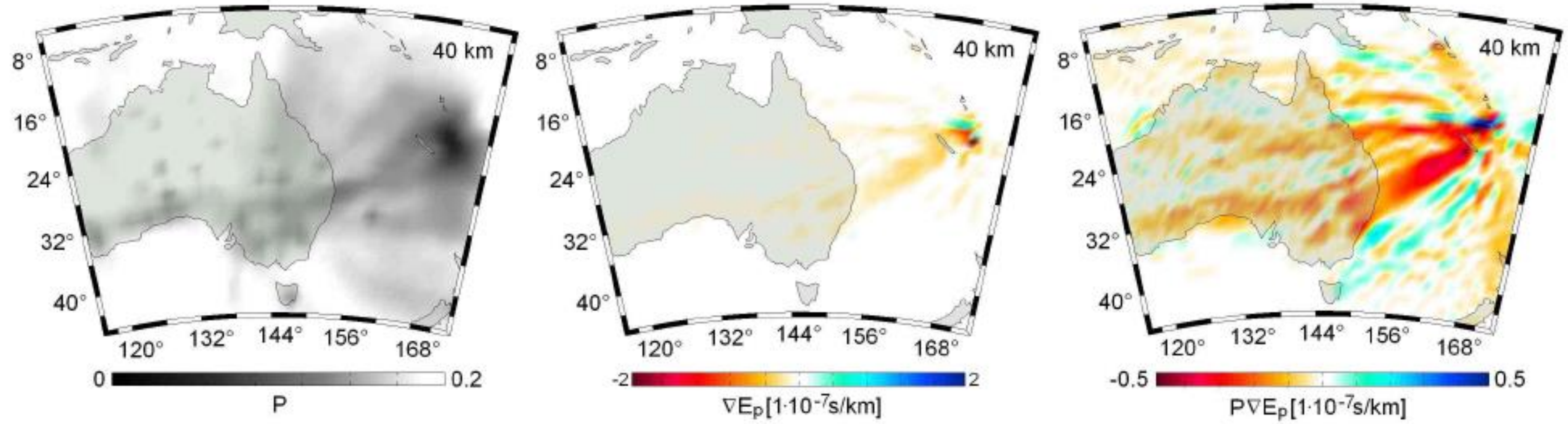
FWI sensitivity kernels





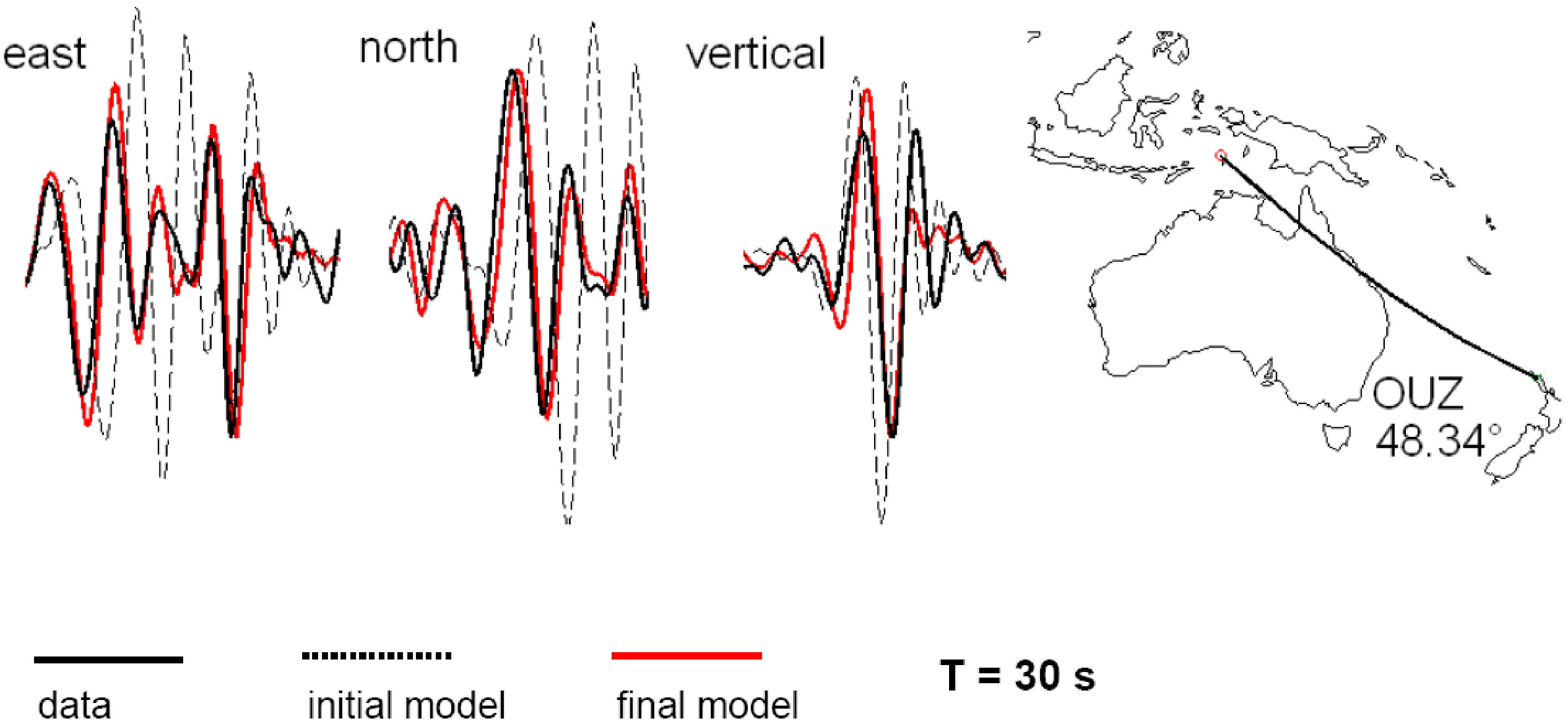
Gradient is calculated by back propagating adjoint sources (differences between theory and observations at receivers) separately for each of the approx. 40 earthquakes

Preconditioning

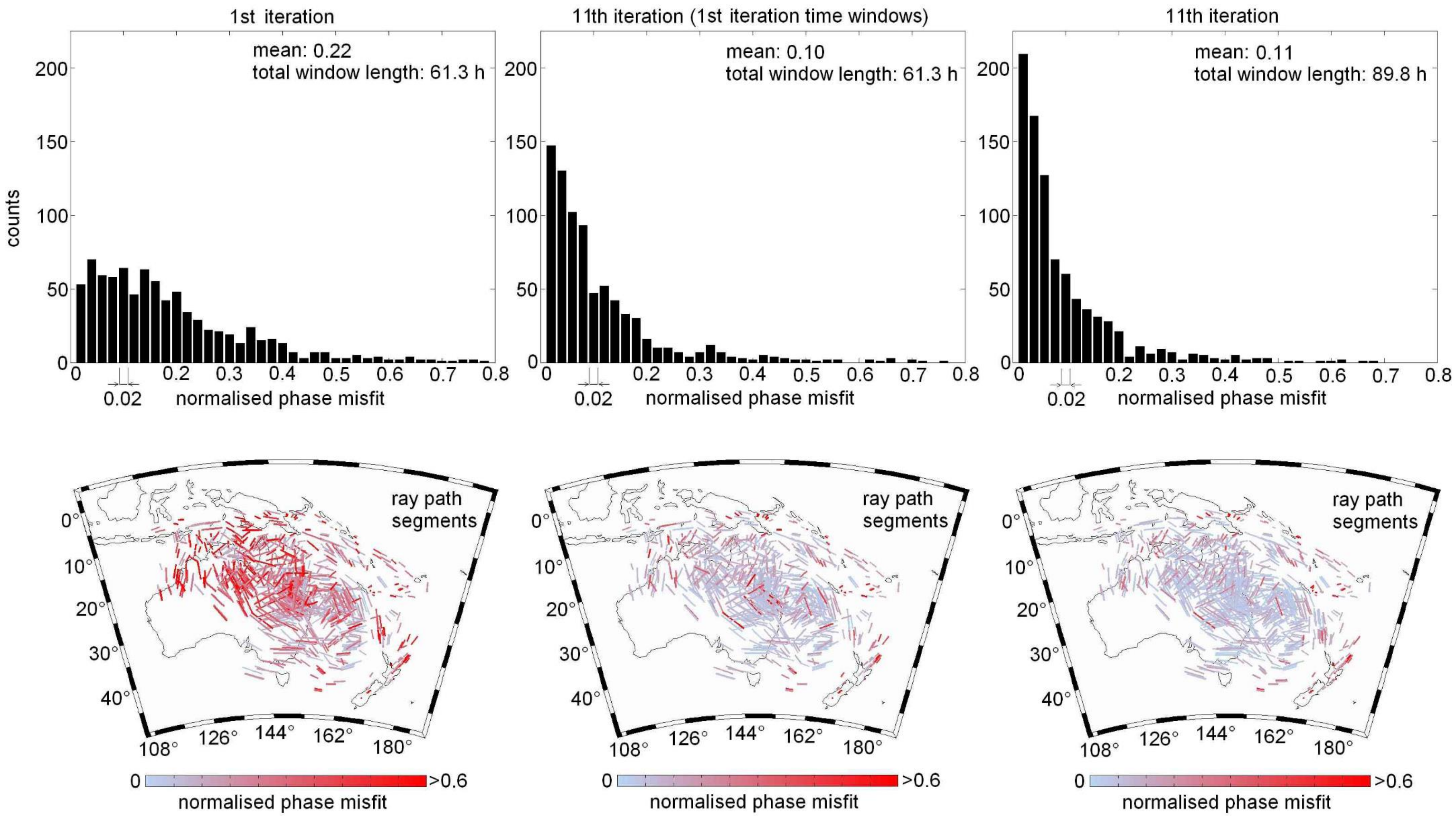


Corrections for geometric spreading effects and reduces the sensitivity with respects to structures near source and receiver

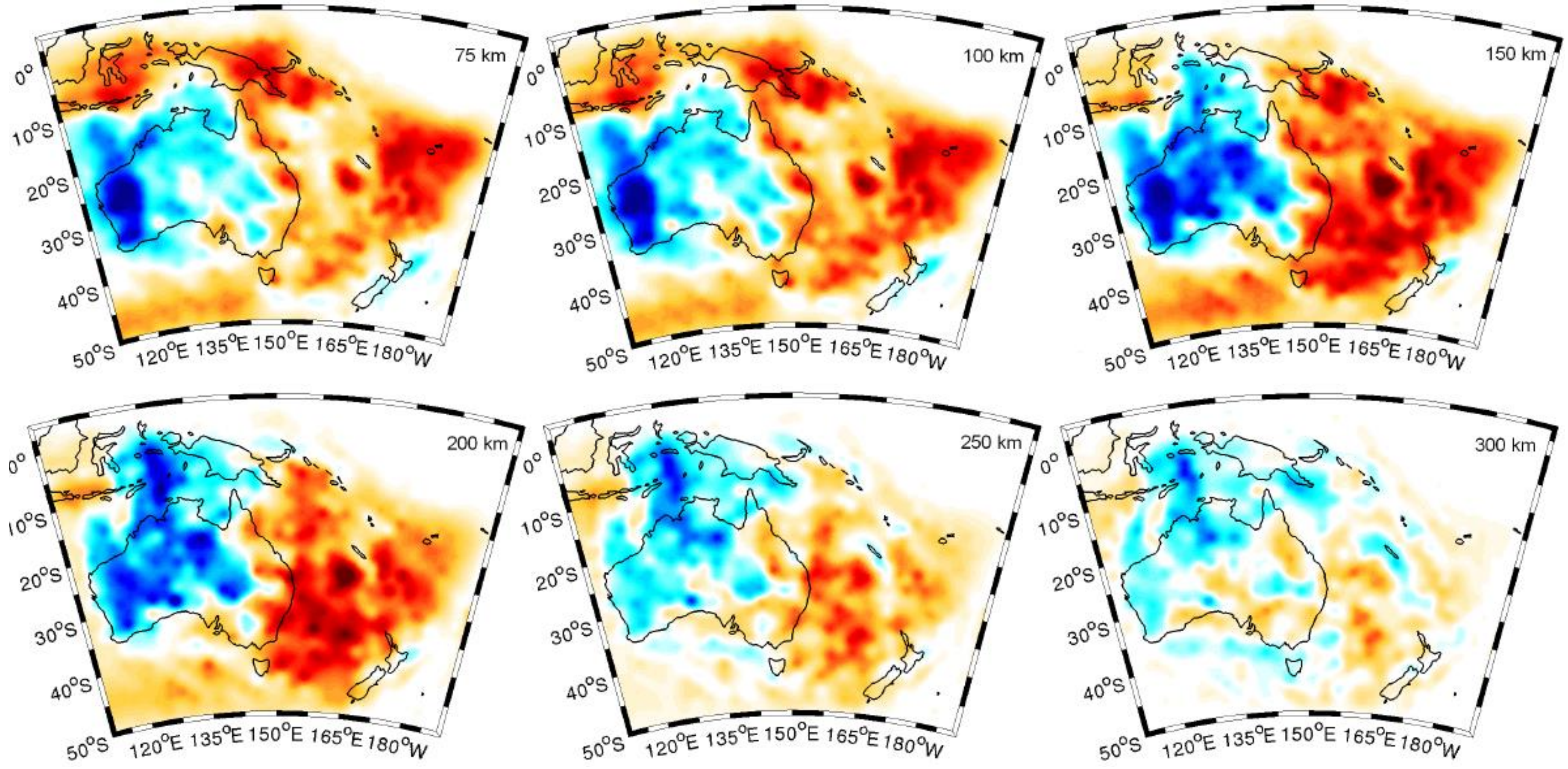
Misfit improvement



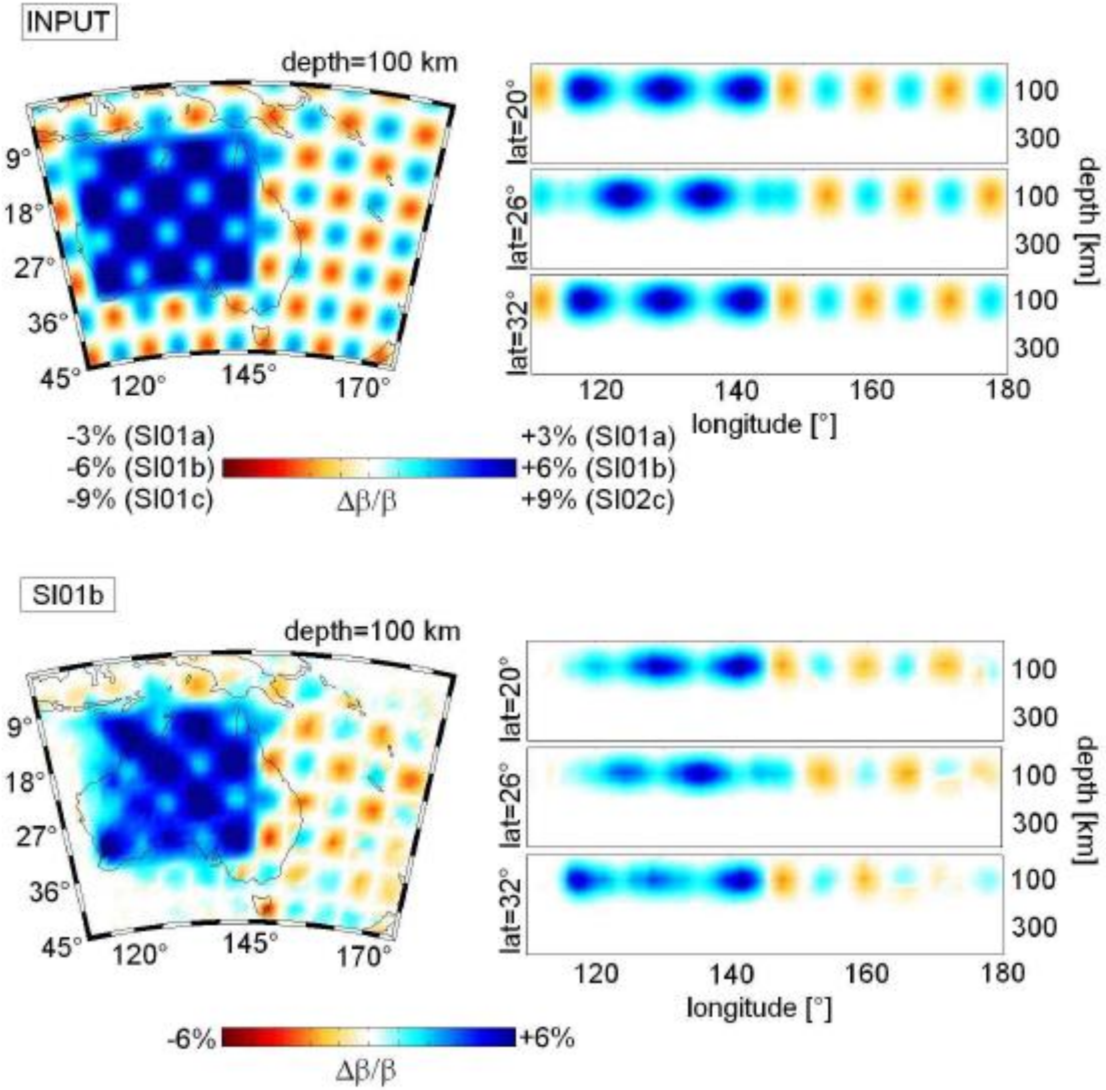
Global misfit improvement



Reconstructed Earth model



Checkerboard test – Resolution?



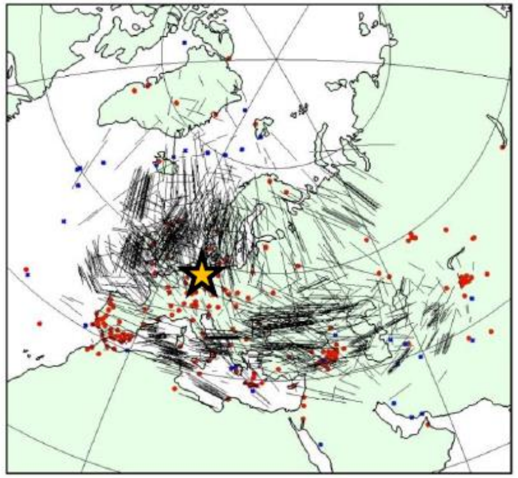
So what?

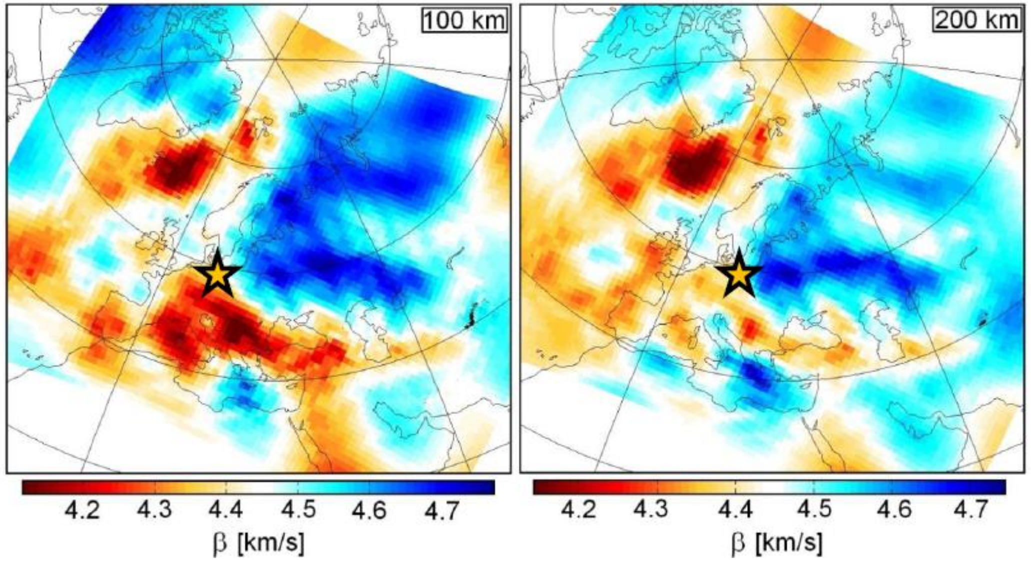
strategies to quantify resolution

Why so difficult for FWI?

- Non linear dependence of data on model parameters
- Sensitivity matrix can not be computed explicitly (as in linear problems for moderately large problems)
- Forward problem too expensive to allow fully probabilistic approaches or neural networks (except for lower-dimensional problems, see poster by Käufl et al.)

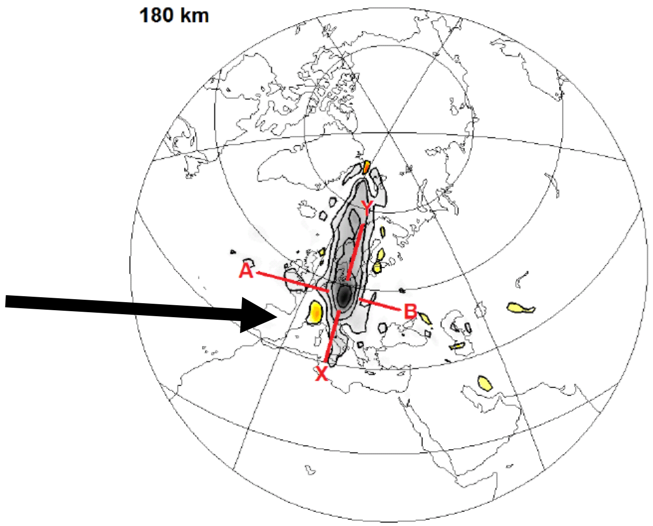
Point spread functions



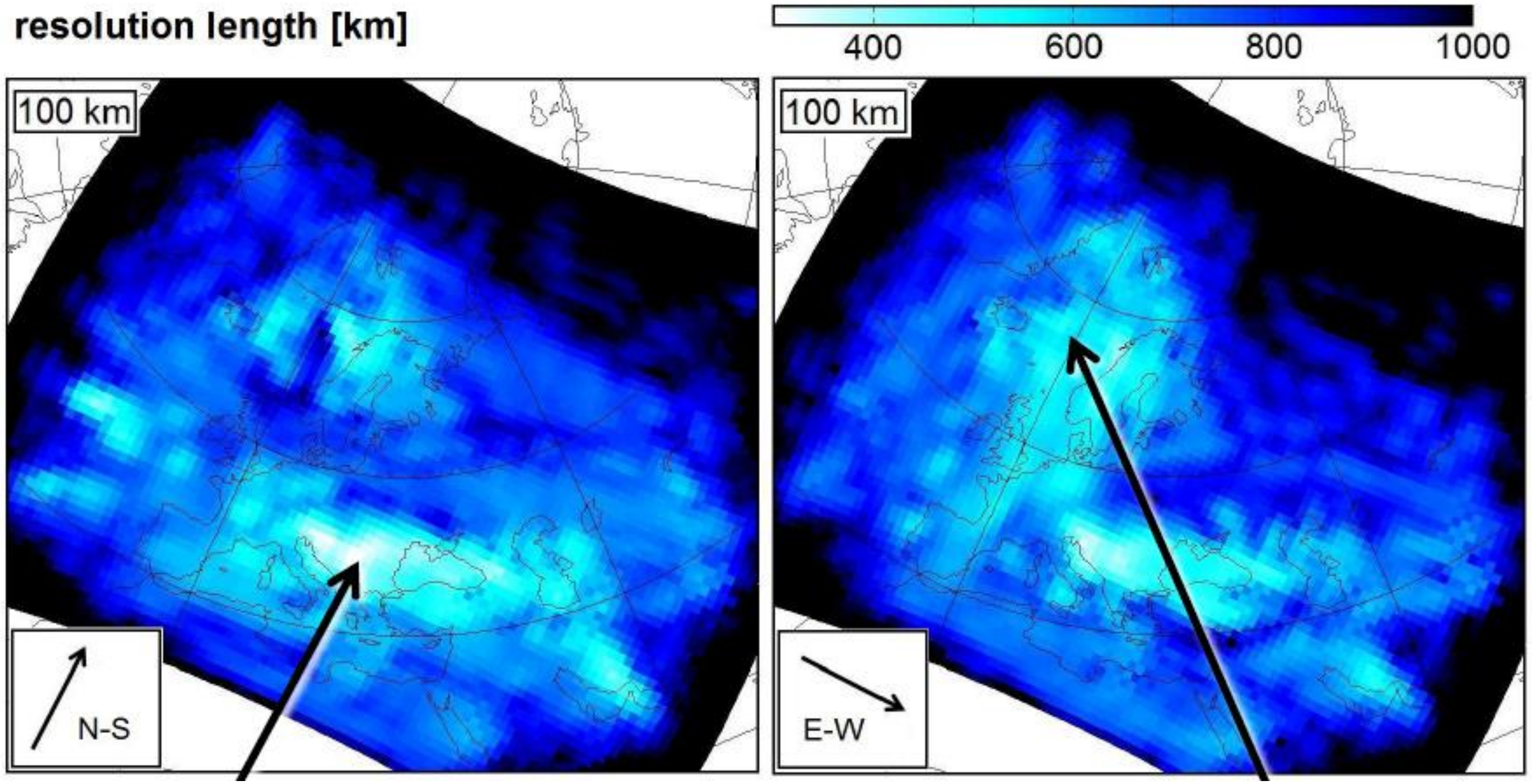


Trade off between S velocity perturbation at the yellow star
and the S velocity in the neighbouring regions (at certain depth)

Compare with **R** in previous slides (Boschi, 2003)!



Resolution length

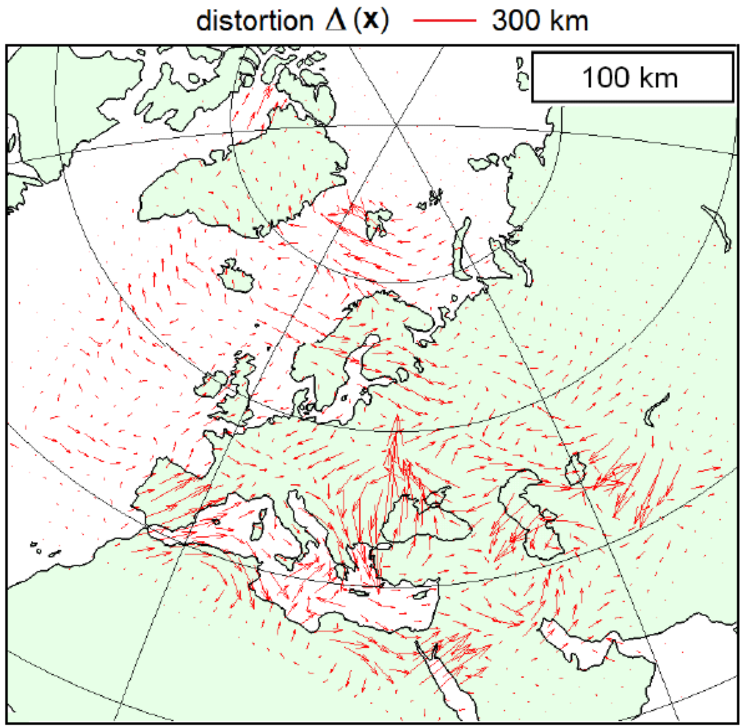


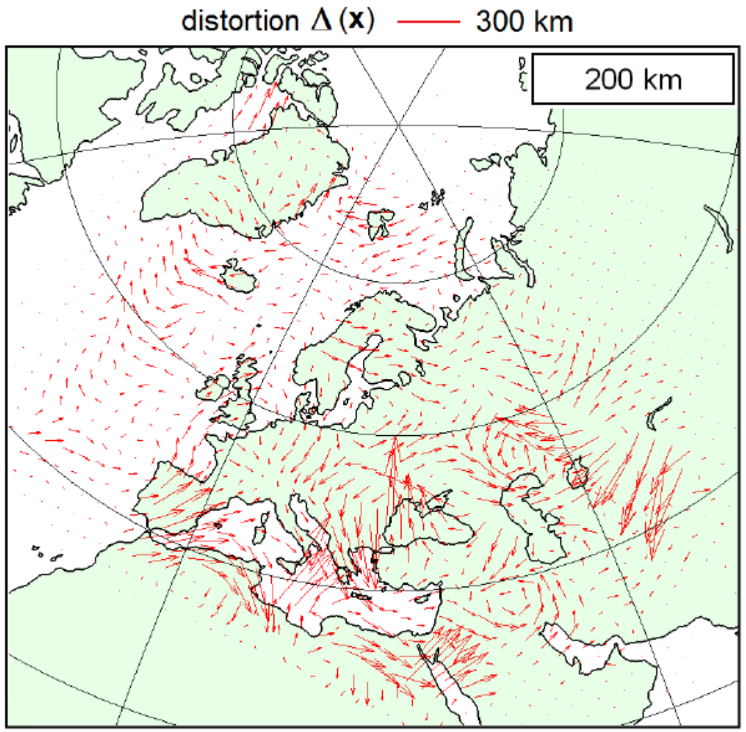
High resolution NS direction

High resolution EW direction

Image distortion

- Point-perturbations displaced by imaging
- Distortion = [position of point perturbation] – [centre of mass of its blurred image]

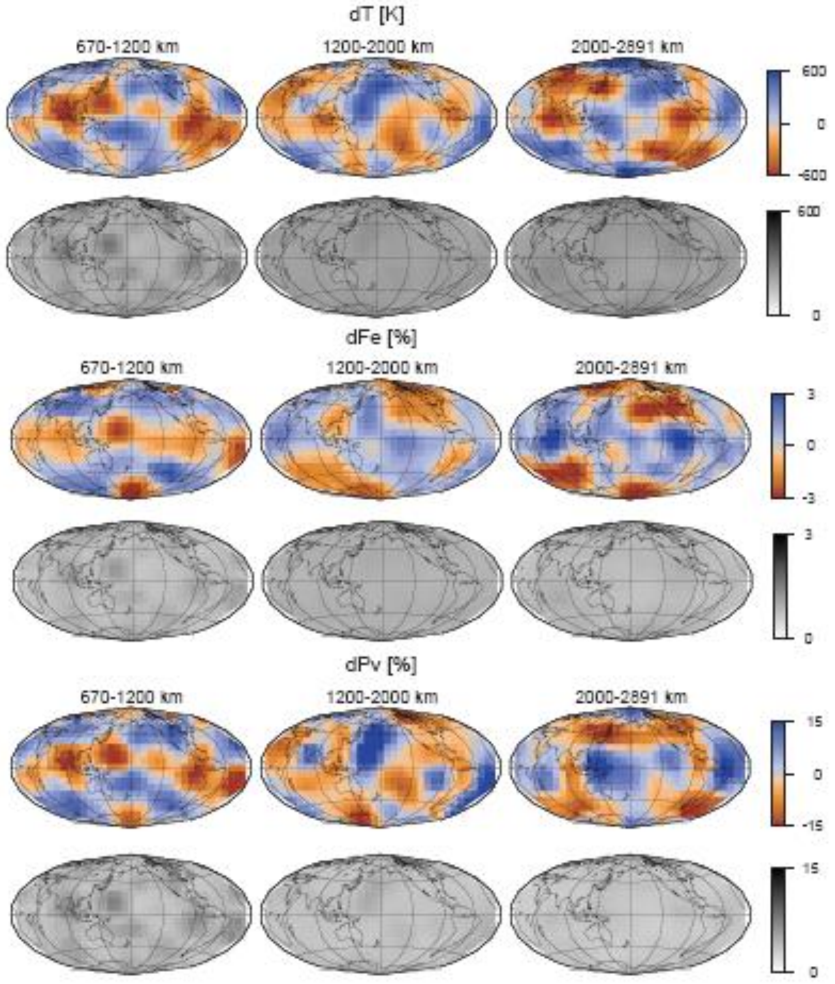




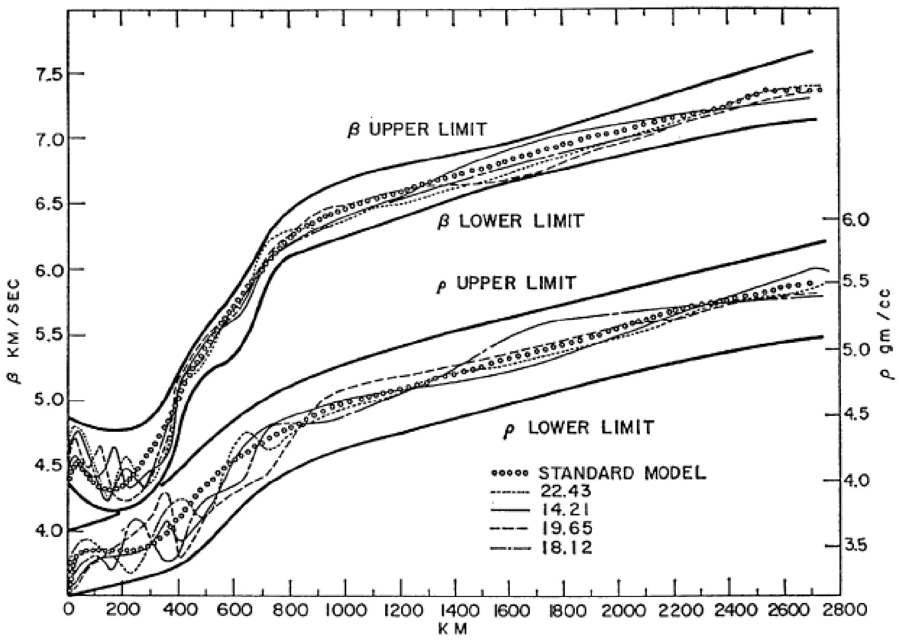
What you see may be somewhere else!

Tomography using Monte Carlo methods

-



Mosca

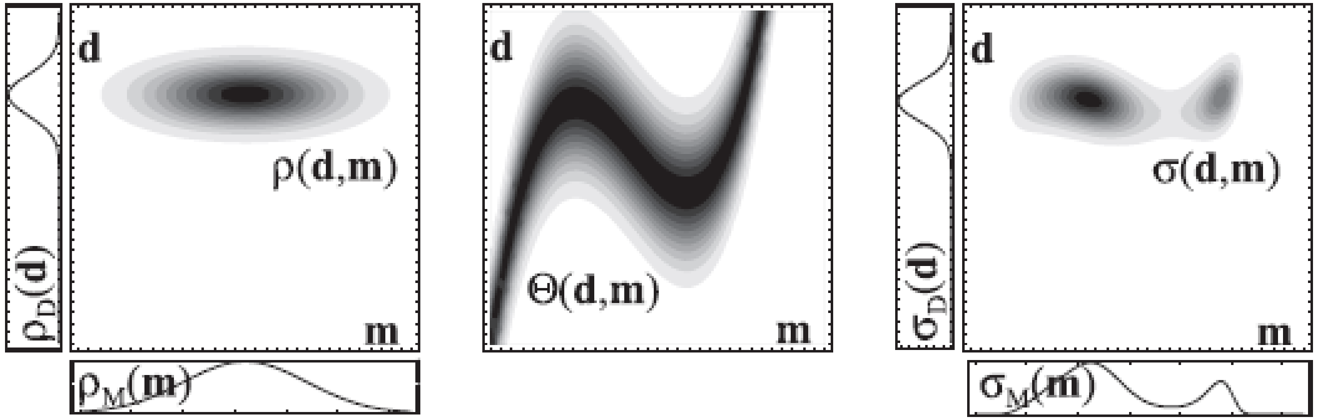


The use of MC methods is restricted to systems with limited degrees of freedom (dozens for generally nonlinear problems)

What we really should be doing …

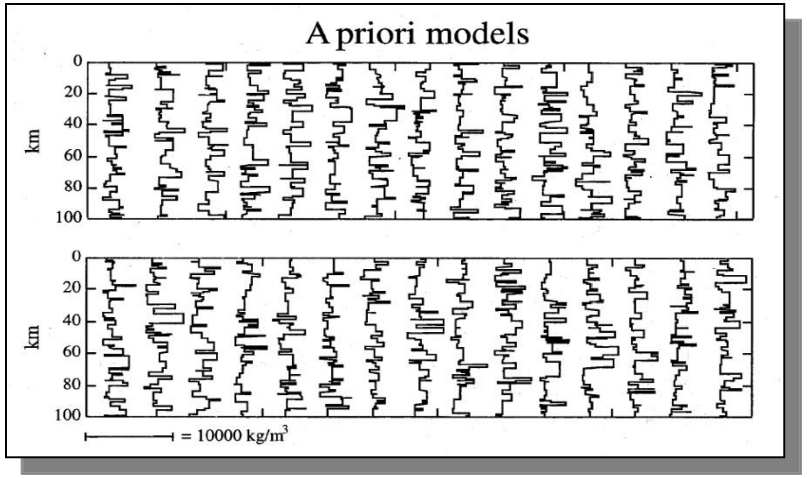
$$\sigma(d,m) = k\,\frac{\rho(d,m)\Theta(d,m)}{\mu(d,m)}$$

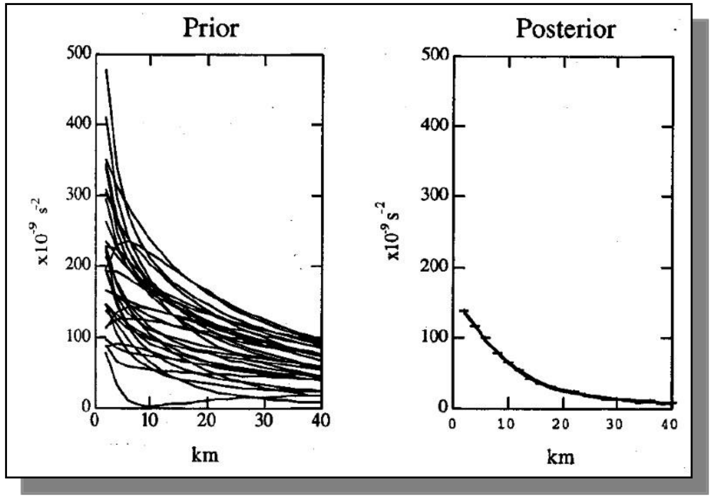
$$posterior = \frac{prior \text{ x } likelihood}{\text{evidence}}$$



Open issues with the probabilistic approach

- How can we properly describe **prior** information?
- How should we describe **data uncertainties**, errors (if not Gaussian)?
- How should we describe defficiencies in our **theory**?
- What are optimal **parametrization schemes** of the Earth model and the model space search

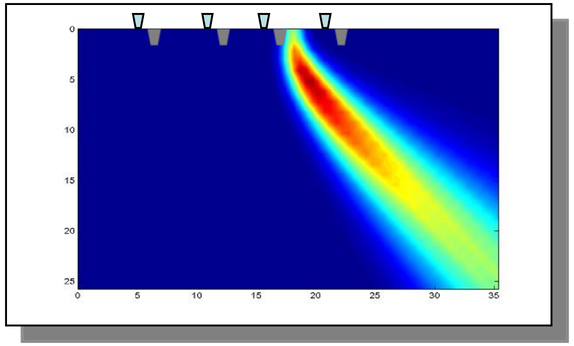


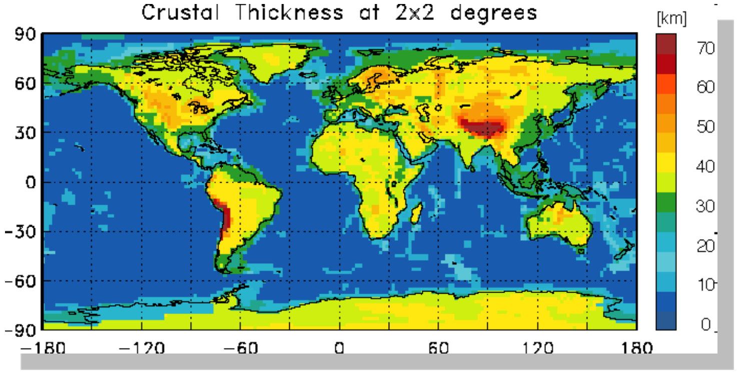


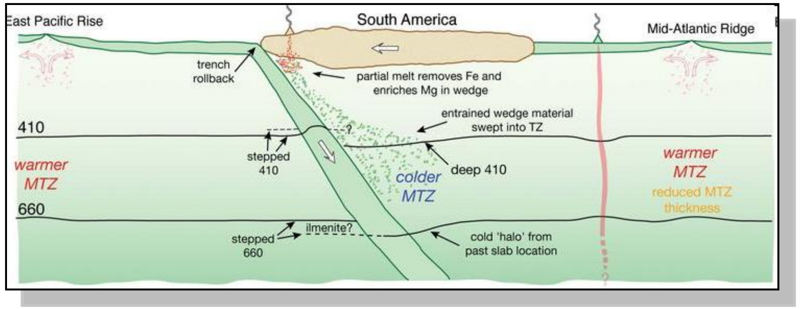
Moosegaard and Tarantola, 1995

Summary and Outlook

- Model space is huge
- Source and receivers unevenly distributed (no fix in sight!)
- Source parameters uncertain (depth, mechanism)
- Forward model inadequate (general anisotropy, Q)
- Trade-offs between Earth properties
- Near surface (crustal) structure inadequately known
- Topography of internal interfaces may be important

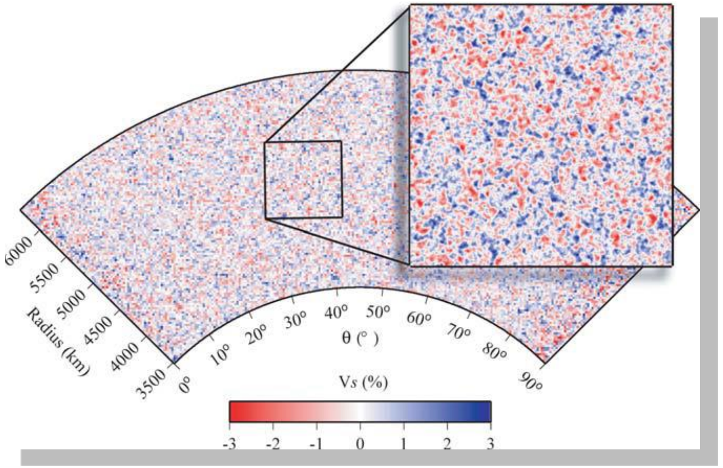


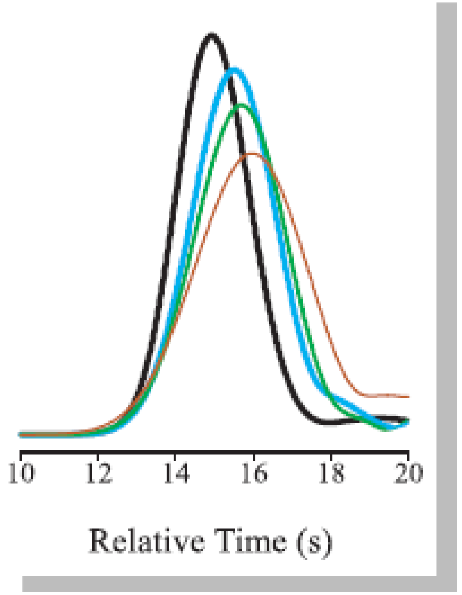




Summary and Outlook (cont'd)

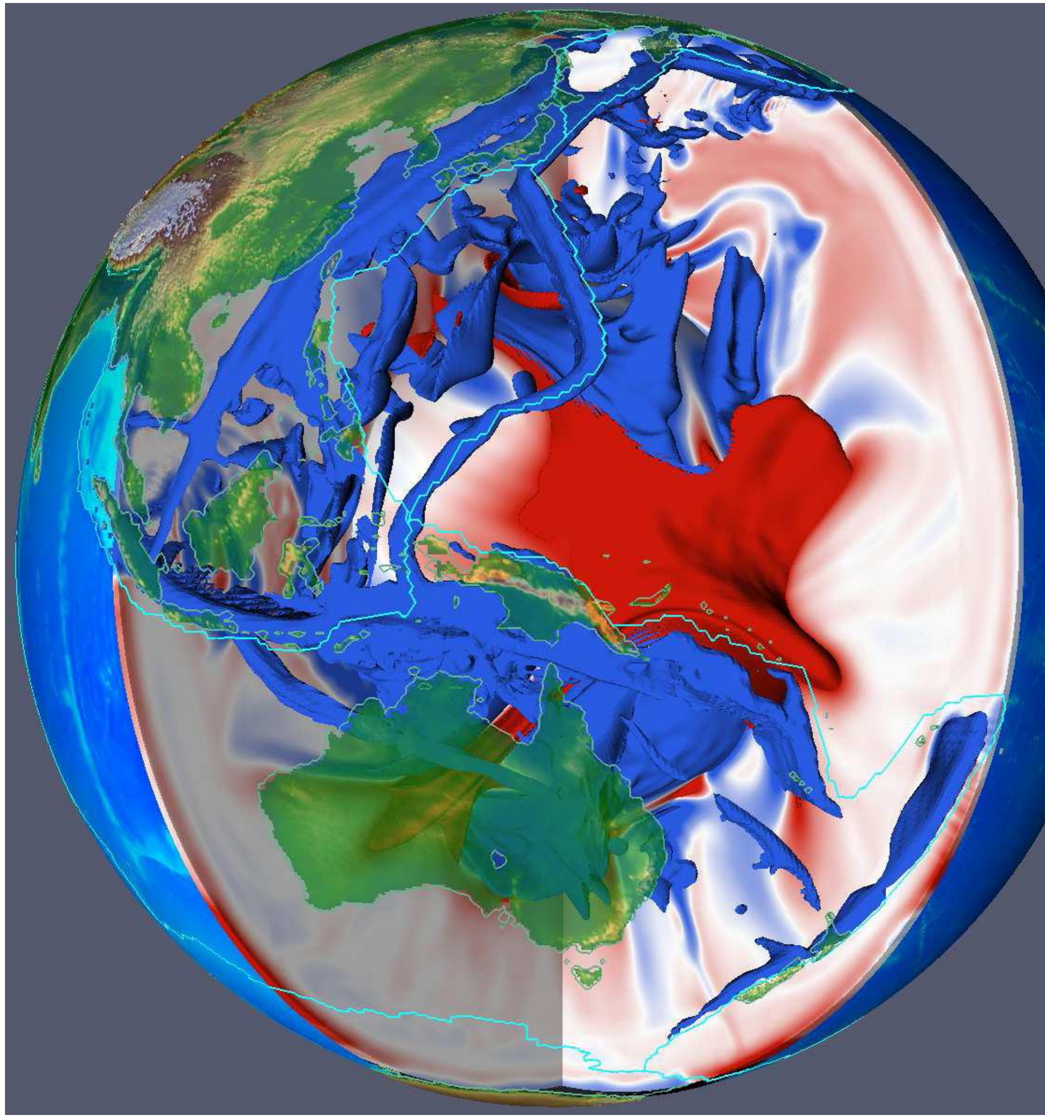
- Errors in the measurements (instrument orientation, instrument response, flipped polarity, timing errors)
- Modelling deficiencies (e.g., numerical dispersion, topography)
- Scattering (effects of small scale structures -> mantle is actually faster!)
- Noise statistics unknown



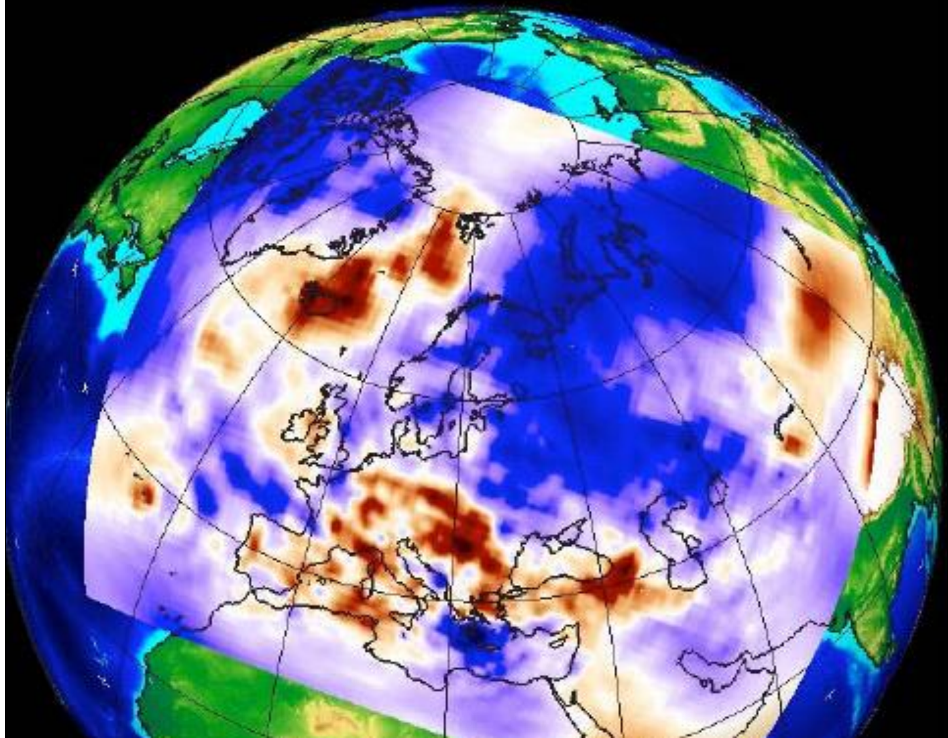


Summary: final comments

- Quantifying uncertainties is a research question and not a standardized procedure
- Many of our SCIENCE stories are told without sufficient uncertainty quantification
- Even if we can calculate uncertainties … how do we convey that information (visually, acoustically)?
- Will Exascale really help??



Thank you!



Strategies to estimate resolution

$$\Delta\mathbf{m}_{out} = (\mathbf{G}^T\mathbf{G}+\mathbf{D})^{-1}\mathbf{G}^T\Delta\mathbf{d}$$

$$\Delta\mathbf{m}_{out} = (\mathbf{G}^T\mathbf{G}+\mathbf{D})^{-1}\mathbf{G}^T\mathbf{G}\Delta\mathbf{m}_{\mathbf{in}}$$

Synthetic data for a test model

$$\mathbf{R} = (\mathbf{G}^T\mathbf{G}+\mathbf{D})^{-1}\mathbf{G}^T\mathbf{G} \approx \mathbf{I} \qquad ??$$

Resolution matrix R

Hessian and covariance

Earth model m(x) and misfit functional

$$\mathbf{m}(\mathbf{x}) = [m_1(\mathbf{x}), m_2(\mathbf{x}), \ldots, m_N(\mathbf{x})]^T$$

$$\chi(\mathbf{m}) = \chi(\tilde{\mathbf{m}}) + \frac{1}{2}\int_G \int_G [\mathbf{m}(\mathbf{x}) - \tilde{\mathbf{m}}(\mathbf{x})]^T \, \boxed{\mathbf{H}(\mathbf{x}, \mathbf{y})} \, [\mathbf{m}(\mathbf{y}) - \tilde{\mathbf{m}}(\mathbf{y})] \, d^3\mathbf{x} \, d^3\mathbf{y}$$

Hessian

… and the equivalence with probabilistic approach …

$$\sigma(\mathbf{m}) = \text{const.} \, e^{-\chi_g(\mathbf{m})}$$

$$\chi_g(\mathbf{m}) = \frac{1}{2}\int_G \int_G [\mathbf{m}(\mathbf{x}) - \tilde{\mathbf{m}}(\mathbf{x})]^T \, \boxed{\mathbf{S}^{-1}(\mathbf{x}, \mathbf{y})} \, [\mathbf{m}(\mathbf{y}) - \tilde{\mathbf{m}}(\mathbf{y})] \, d^3\mathbf{x} \, d^3\mathbf{y}$$

Variances

Following strategy suggested by Fichtner and Trampert, GJI, 2011