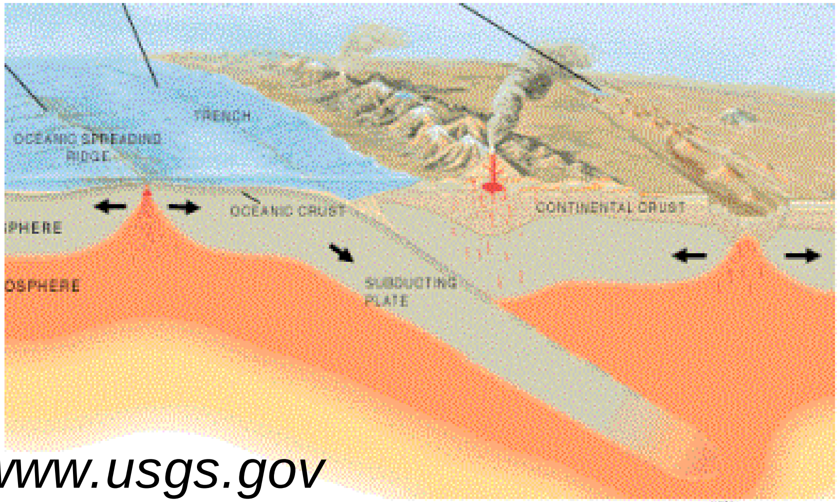
Seismic tomography: Art or science?

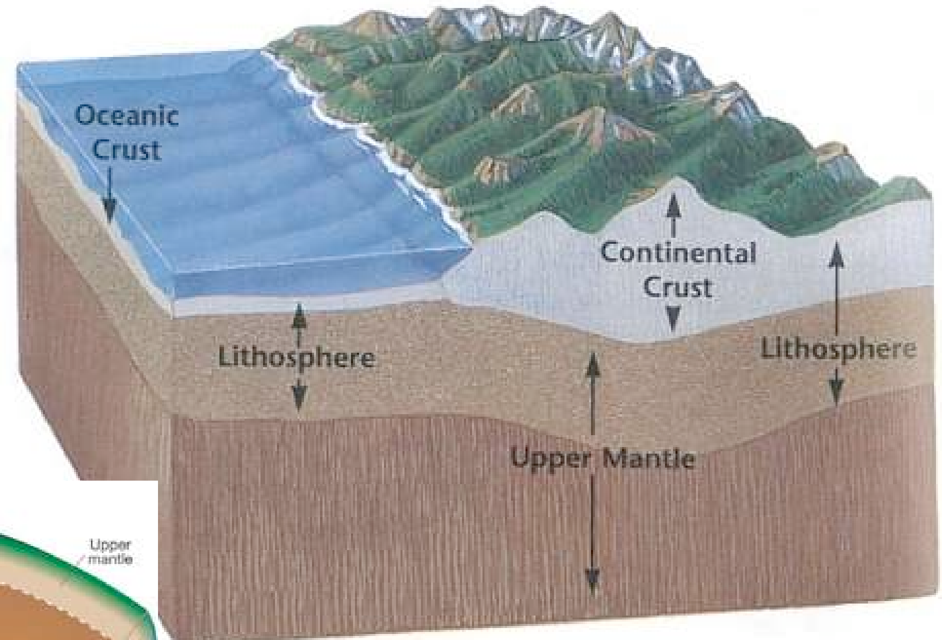
Frederik J Simons

Princeton University

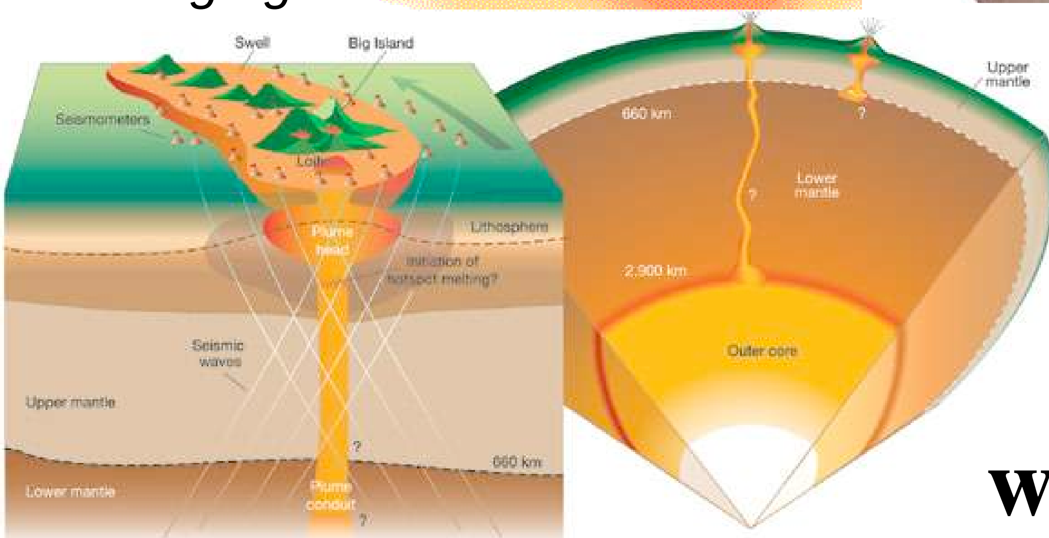
What’ s inside the Earth?



www.usgs.gov



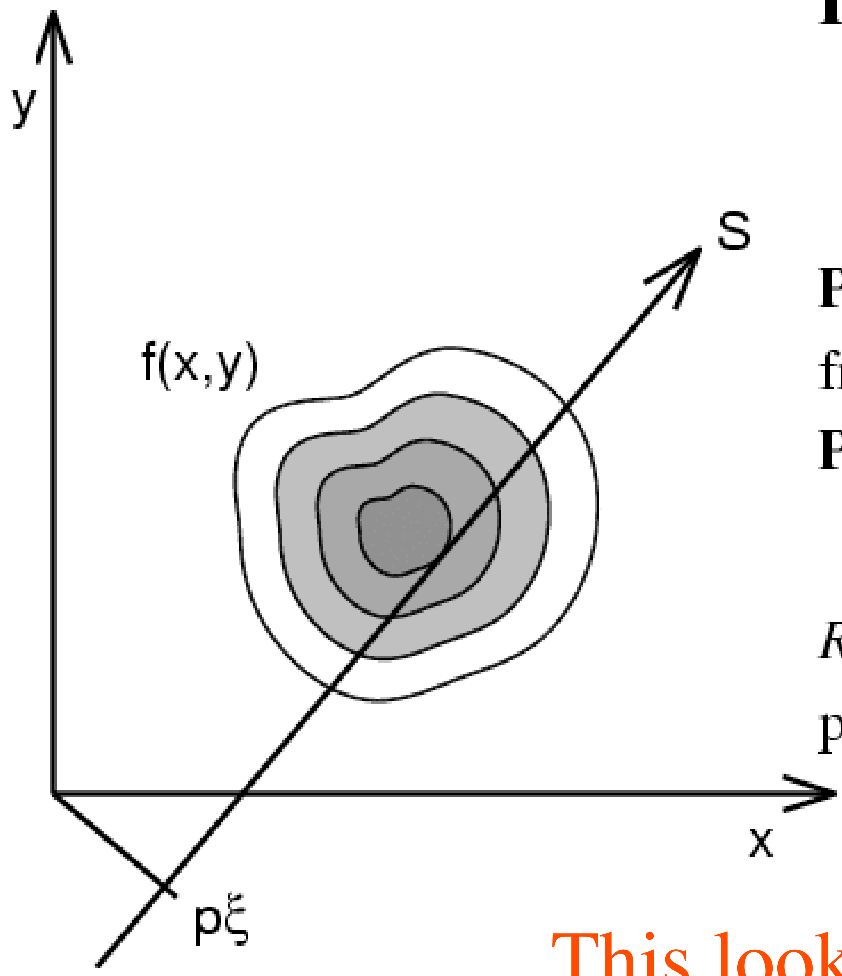
www.glencoe.com



Dalton, *Nature* 2003

Only seismic waves have actually *been there, done that*

The seismic tomography problem



Inverting the Radon transform

$$\mathcal{R}[f(x,y)](p,\boldsymbol{\xi}) = \int_S f(x,y)\,ds \qquad (1)$$

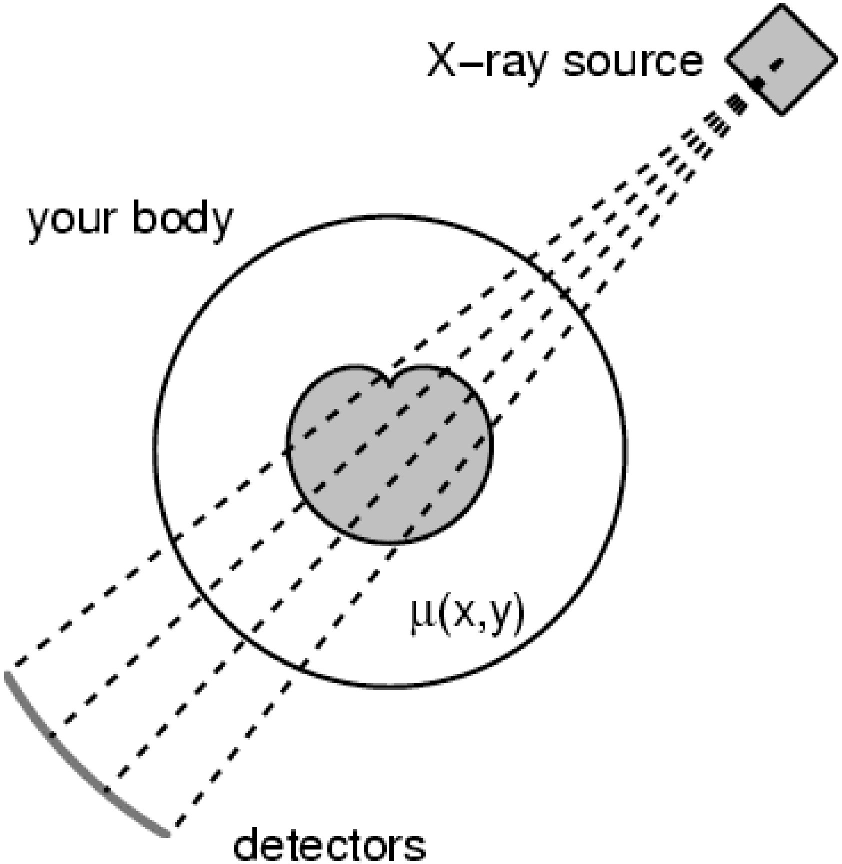
Purpose: Reconstruction of functions from their line integrals (projections).

Problem: Given $\mathcal{R}[f(x,y)](p,\boldsymbol{\xi})$, find $f(x,y)$.

Radon [1917] derived a solution to this problem, giving an expression for $\mathcal{R}^{-1}$.

This looks more complicated than it is; and that's my point.

What is $f(x, y)$? Medical applications.



X-ray absorption & scattering

Tissues and bones have $\neq$ absorption and scattering coefficients $\mu(x, y)$.

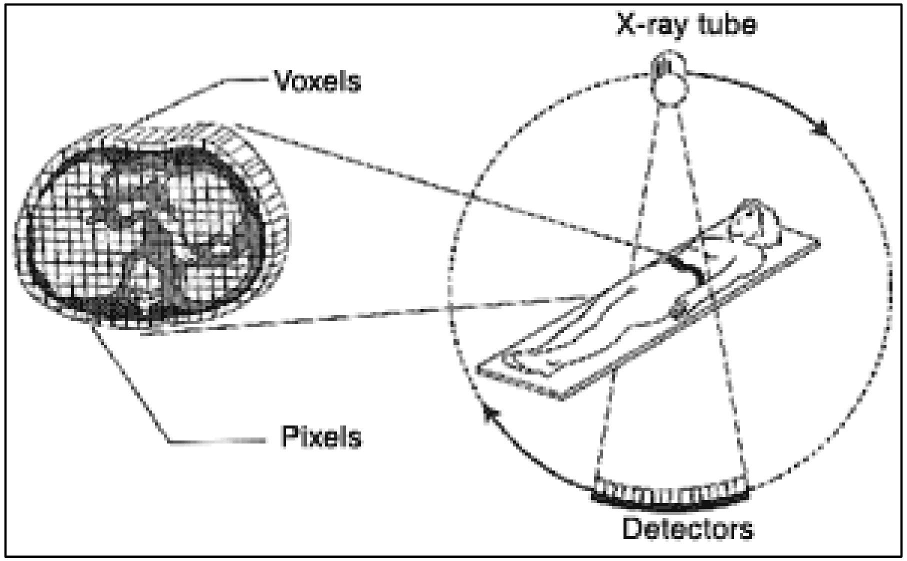
Recorded intensity goes as

$$I = I_0 \exp \left[\int_{\text{ray}} -\mu(x, y)\, ds \right] . \qquad (2)$$

Sources and detectors rotate to achieve perfect "coverage".

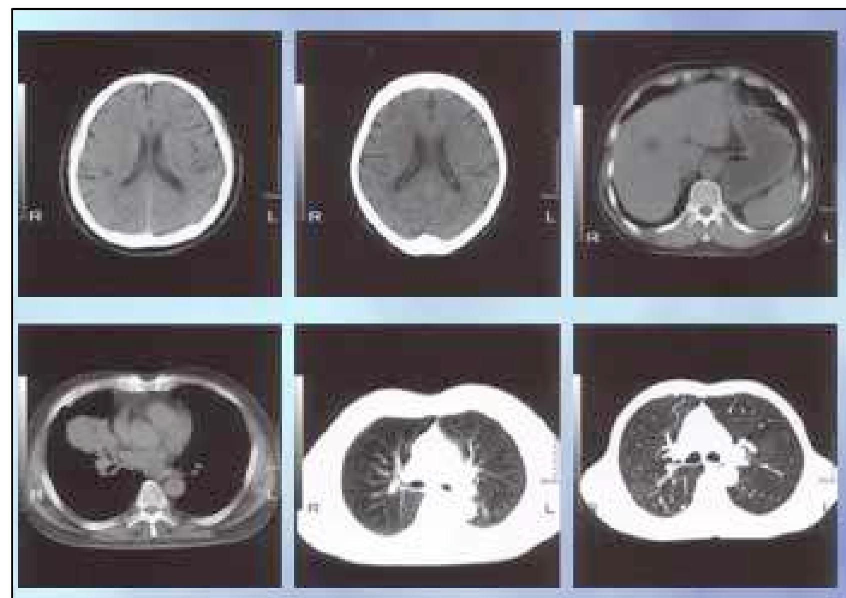
This looks simpler than it is; and that's my point.

X-Ray attenuation tomography

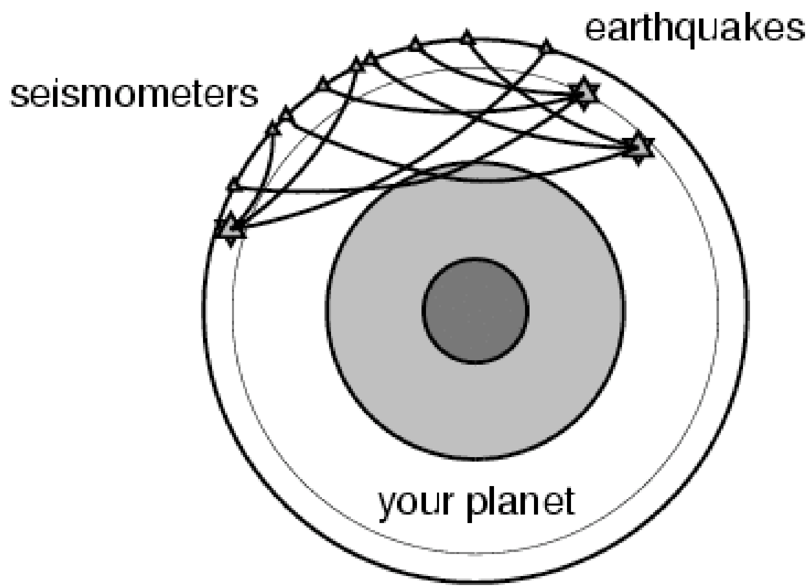


Projections from all angles:
X-ray intensity

Reconstructed image:
X-ray attenuation constants



What is $f(x, y, z)$? Seismic wavespeeds.



Travel-time tomography

The Earth is made of a heterogeneity of seismic velocities $v(x, y, z)$.

Travel-time anomalies go as

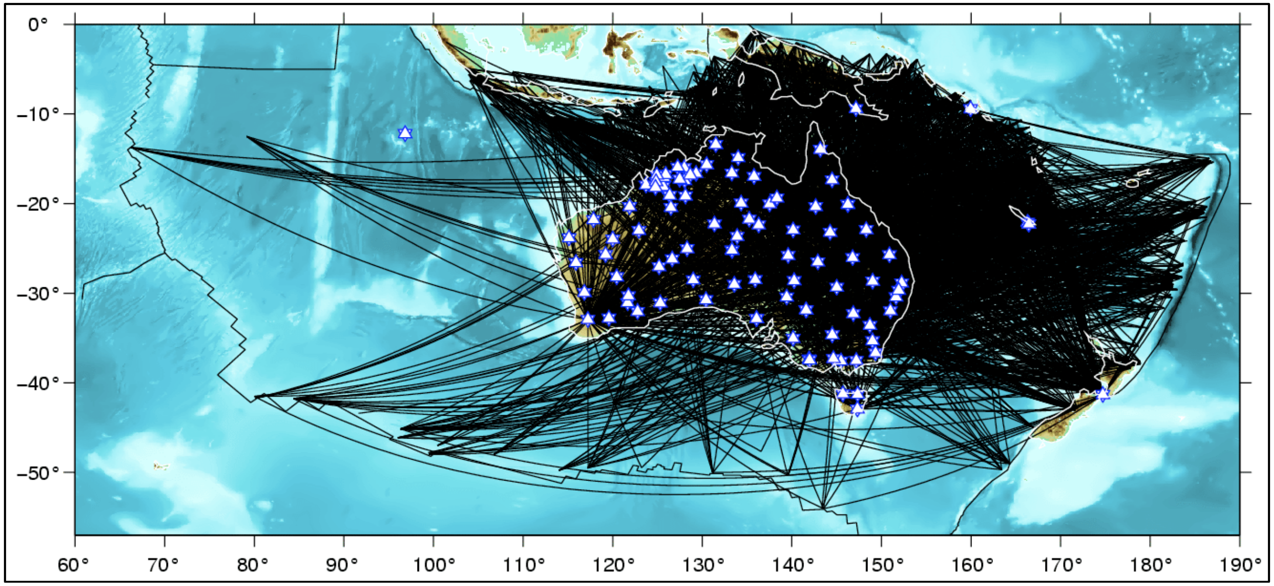
$$\delta t = \int_{\text{ray}} \frac{1}{\delta v(x, y, z)} \, ds. \tag{3}$$

Waveform tomography

Arrival times depend on the wavelength of the seismic phases.

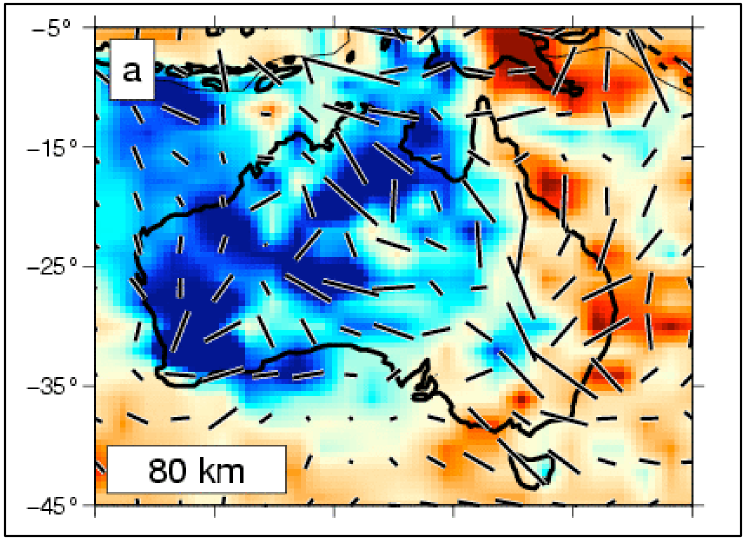
All raypaths curve and coverage is far from perfect.

Seismic wavespeed tomography



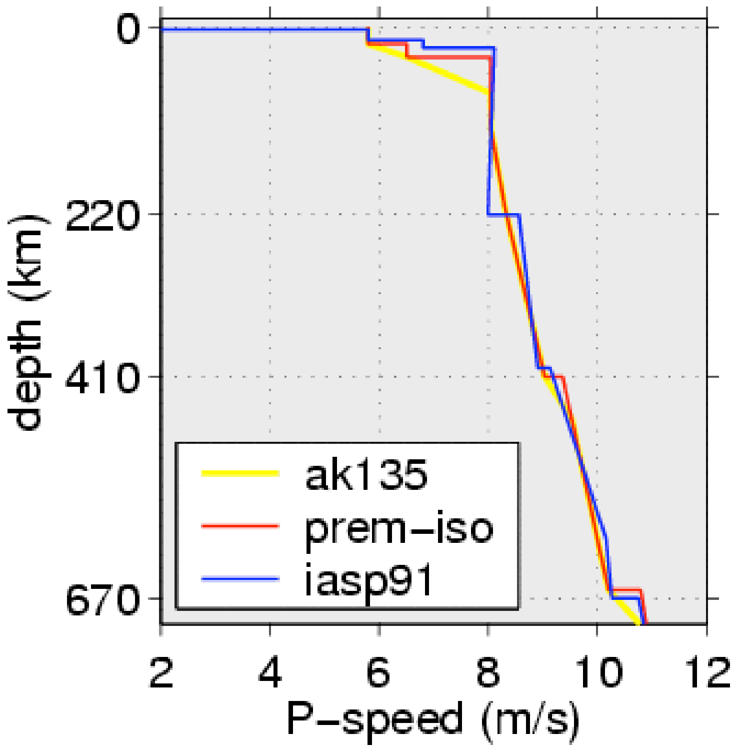
Projections from all angles:
Waveforms and arrival times

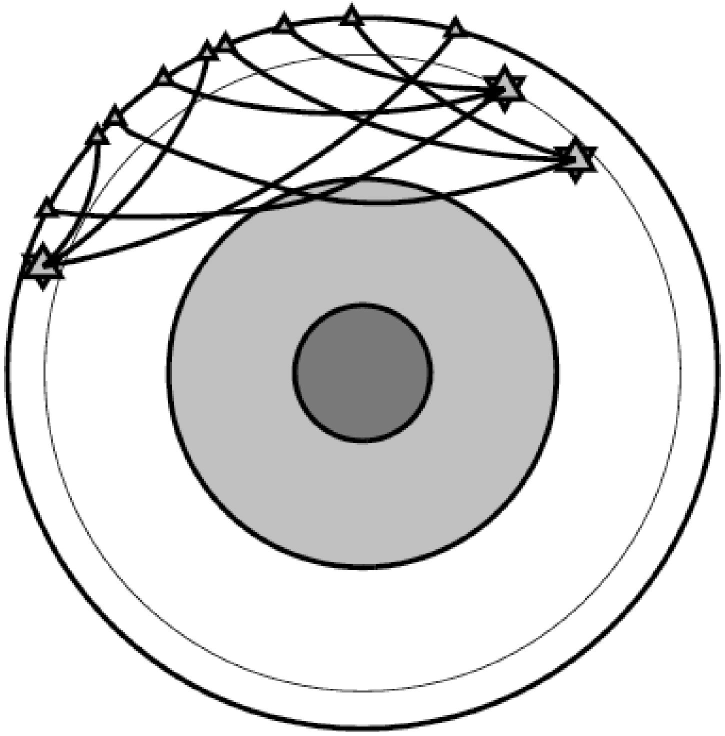
Reconstructed image:
Wavespeed variations



Forward modeling of the wave field, Part I:

Ray tracing, most 1-D





Before

After

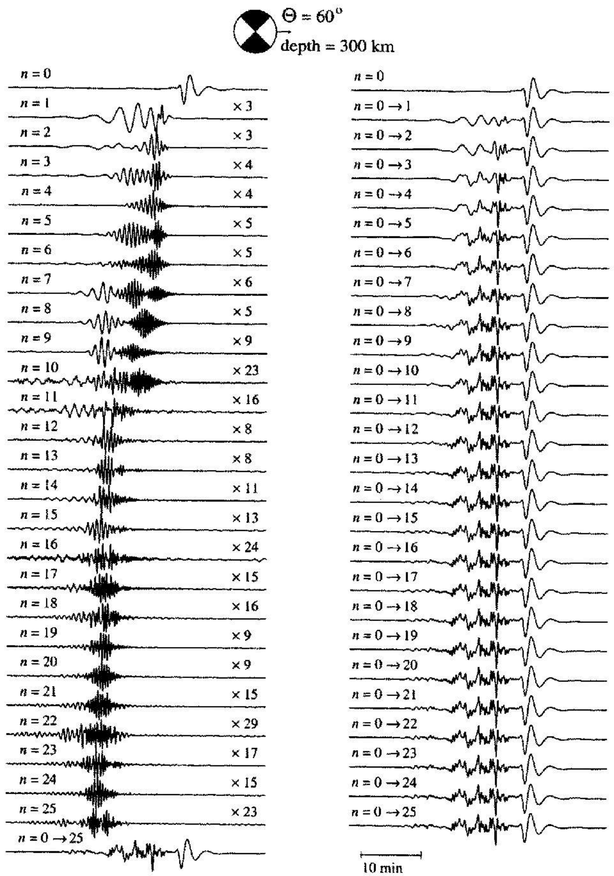
Kennett, *GJI*, 1995

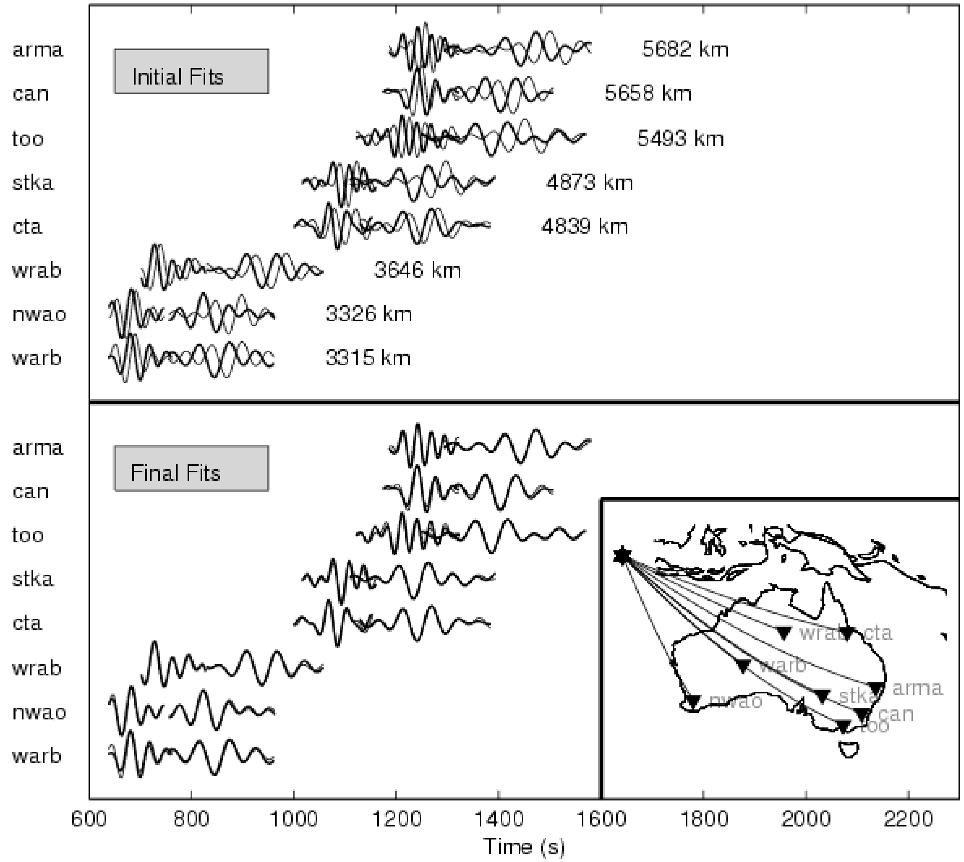
Bullen & Bolt, 1985

Buland, *BSSA*, 1983

Forward modeling of the wave field, Part II:

Normal-mode summation, 1-D





Before

After

Dahlen & Tromp, 1998

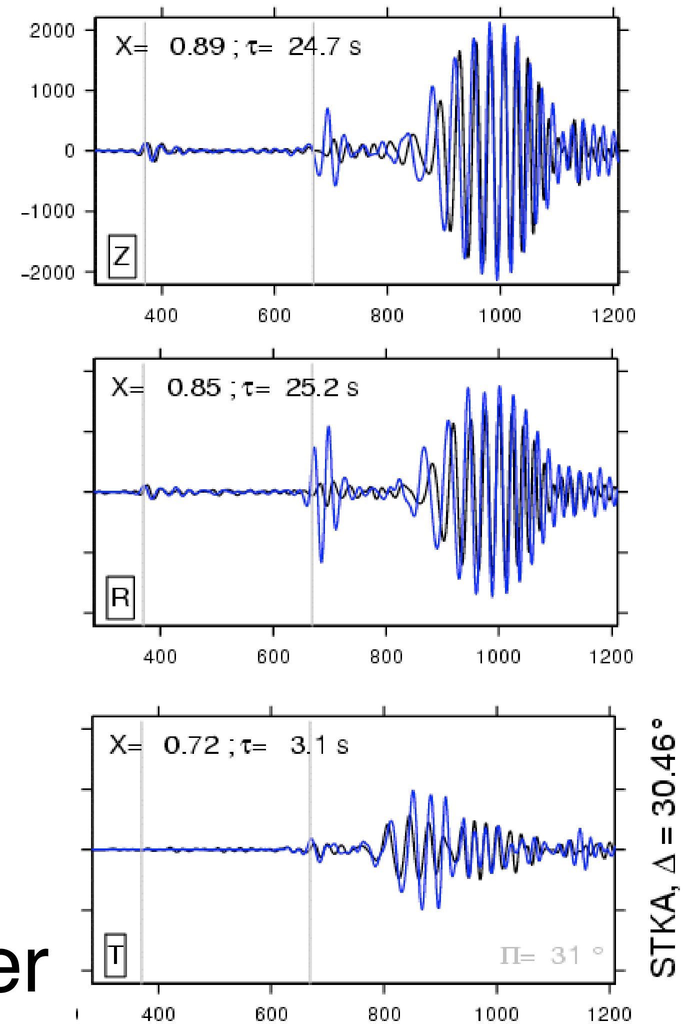
Simons, *Lithos*, 1999

Forward modeling of the wave field, Part III:

Spectral-element methods, 3-D

Before

After



Komatitsch, *GJI*, 2002

That's all there is to it. Goobye!

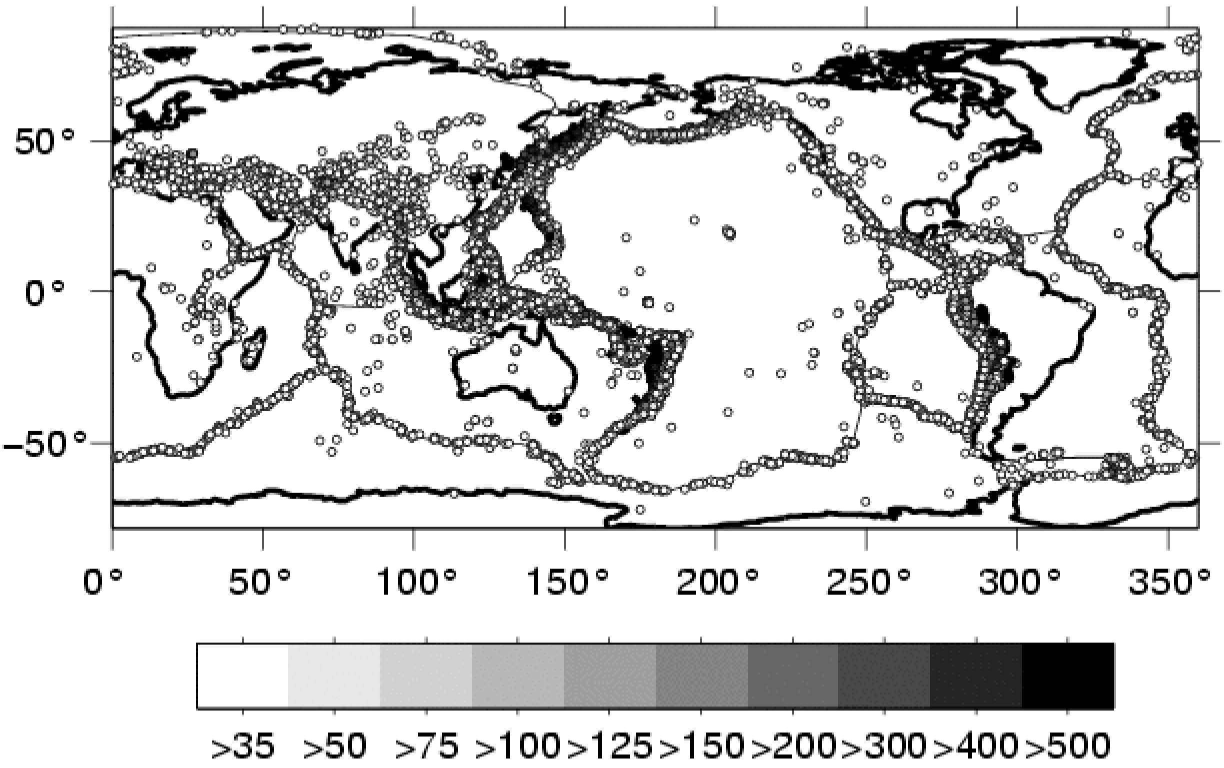
Except:

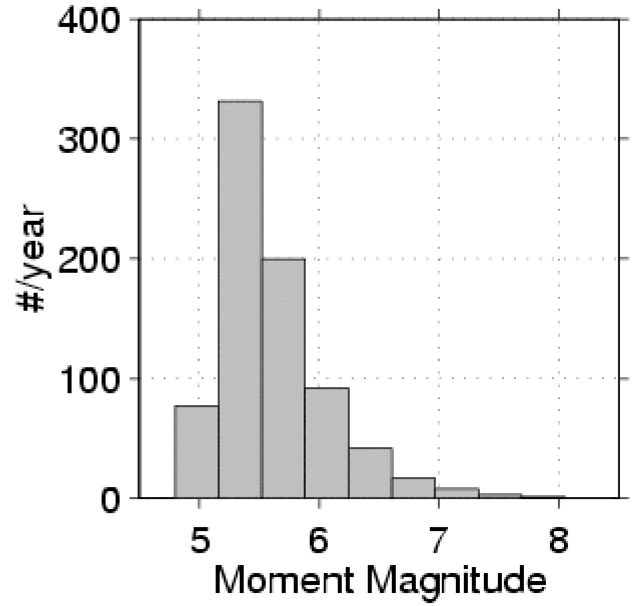
- **X-ray**: exponential of a line integral
  S-ray: raypath itself is a function of velocity
  } **non-linear** functions!
- Earth **coverage** is non-continuous
- "Experiment" is done by nature and **not repeatable**
- Earthquake **source parameters** (location, time) is uncertain

Remedy:

- Linearization
- Discretization
- Regularization (*a priori* information)

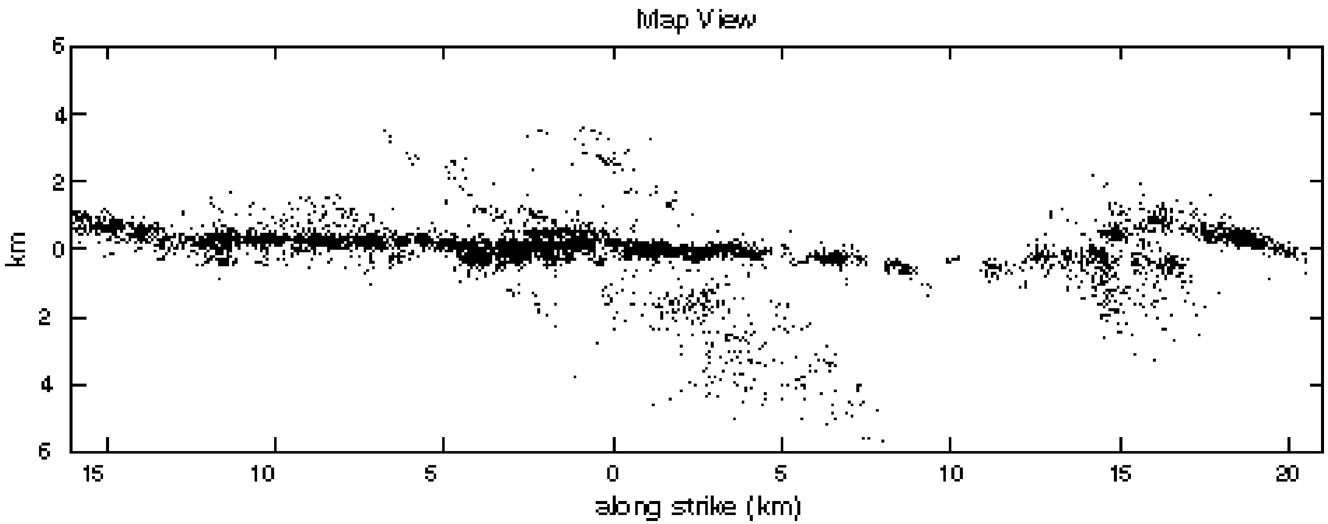
Non-continuous source coverage



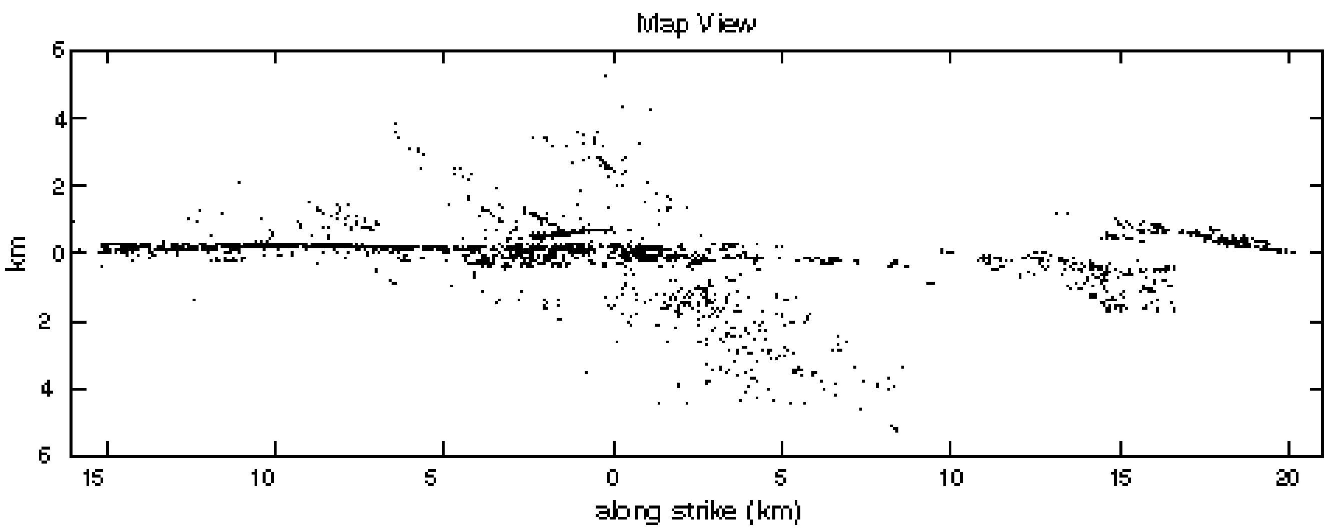


The CMT catalog of large events

Source location – (in)extricably linked



Before



After

Source relocation is big business.

Schaff, *JGR*, 2002

Recipe, Step 1: Linearize!

X-ray

Approximate $\exp(-x) \approx 1 - x$.

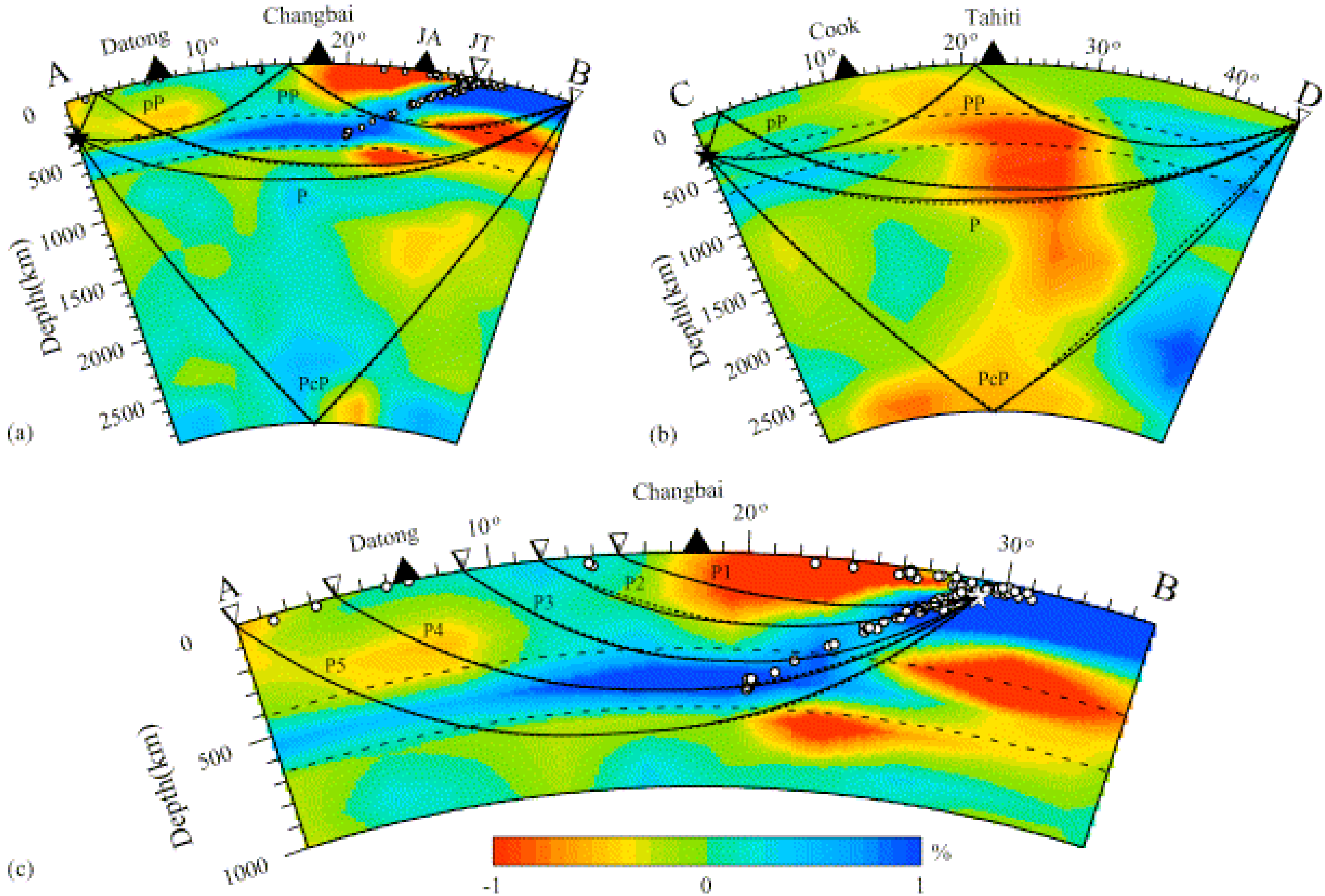
S-ray

Fermat's principle: For a small perturbation of the path, the travel-time (anomaly) is stationary. Using the *slowness*:

$$\delta s = \frac{1}{\delta v} \quad \rightarrow \quad \delta(\delta t) + \mathcal{O}[(\delta t)^2]. \tag{4}$$

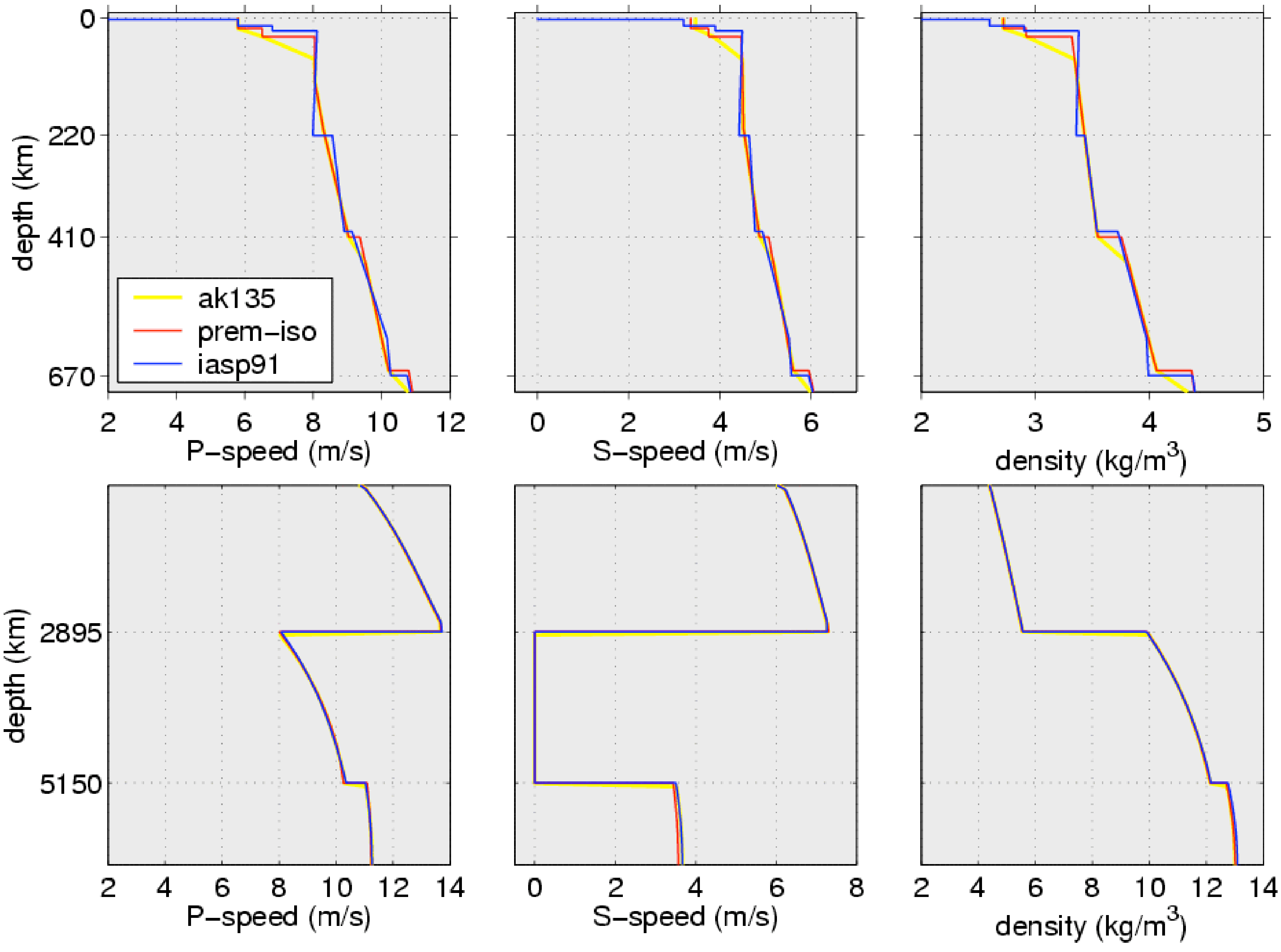
This highlights the importance of the **reference model**, usually a radial model $v(r)$, such as PREM, AK135, IASP91.

Fermat’s Principle at Work for you

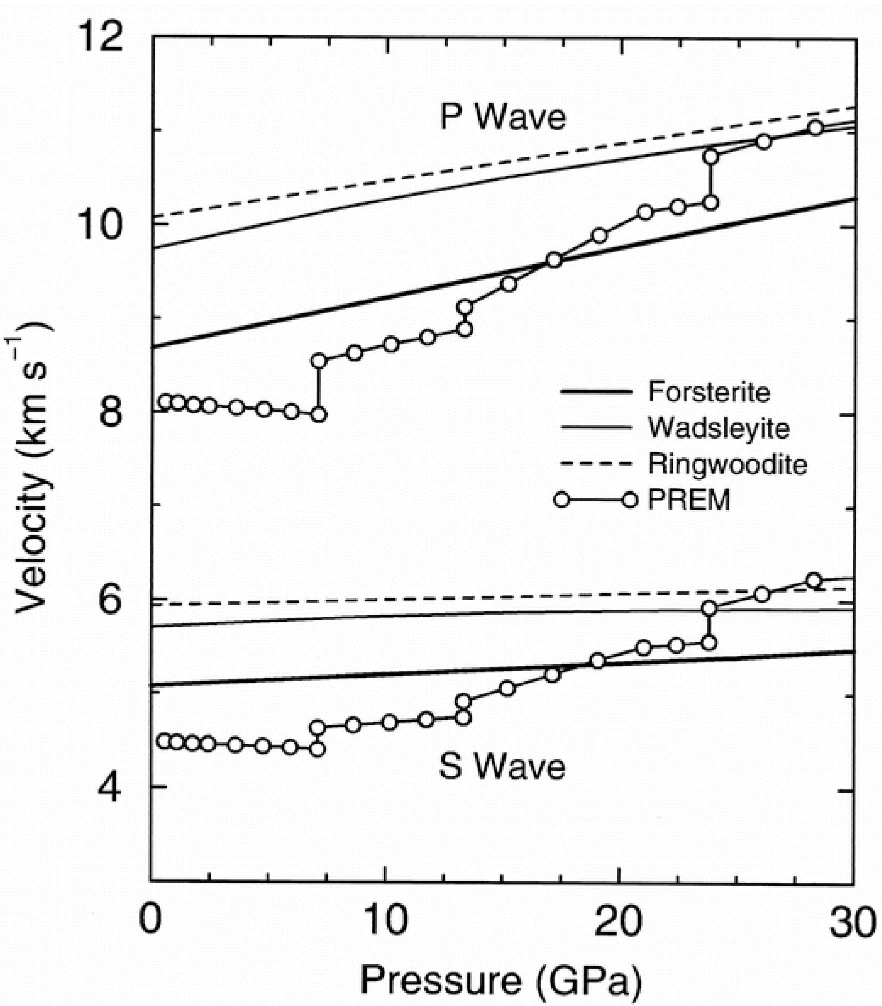


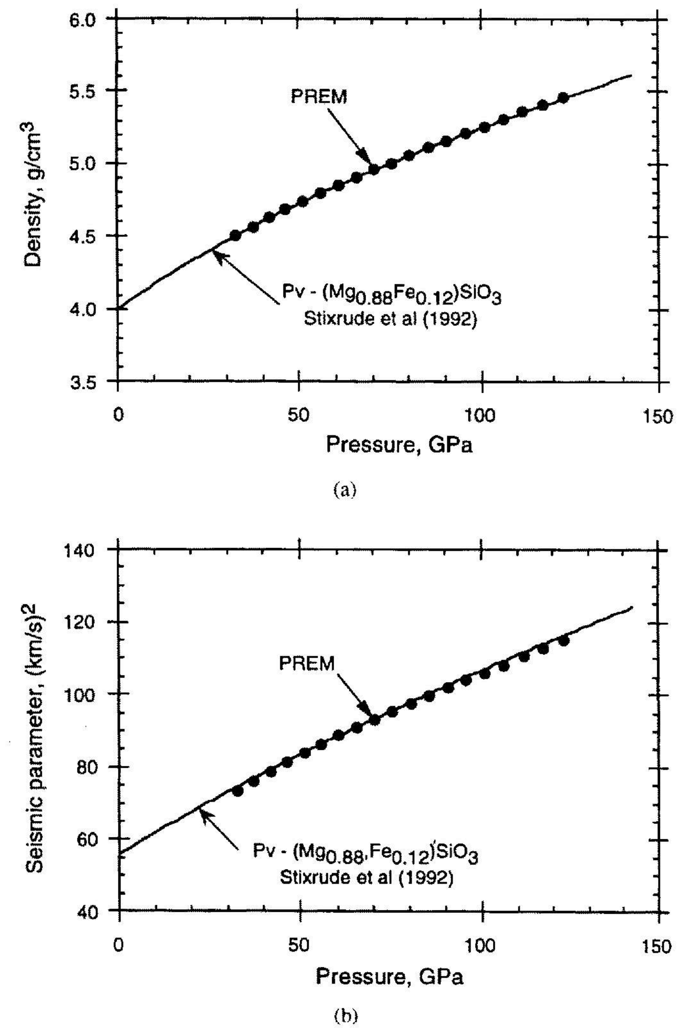
Zhao, PEPI, 2004

The reference Earth: Radial models



… and at least some of it is true…

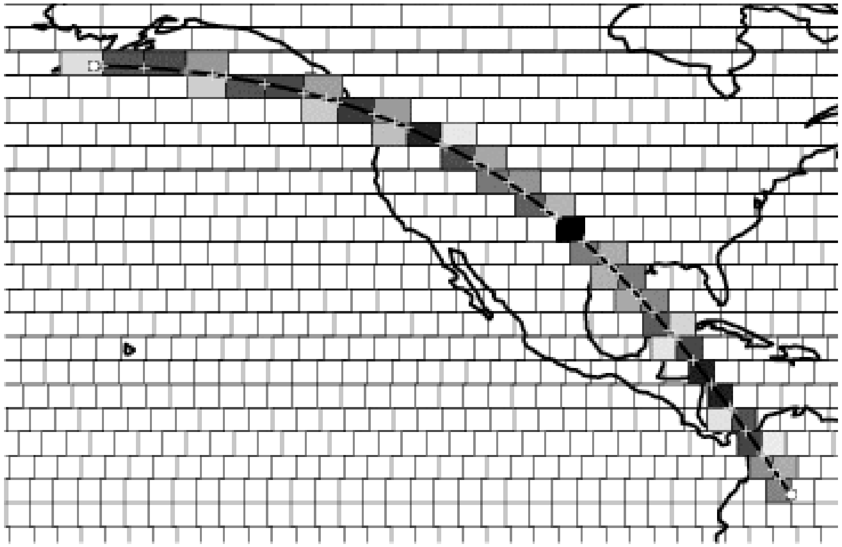




Karki et al., *Rev. Geoph.*, 2001

Jackson, 1998

Recipe, Step 2: **Discretize!**



For a set of seismic rays $i = 1 \rightarrow M$, calculate the length spent in each of $j = 1 \rightarrow N$ grid boxes, in each of which it accumulates a proportional fraction of the total travel-time anomaly δt.

$$\delta t_i = L_{ij} \delta s_j \quad \text{or} \quad \boldsymbol{\delta} \mathbf{t} = \mathbf{L} \cdot \boldsymbol{\delta} \mathbf{s} \tag{5}$$

$$\text{M travel-time anomalies} \begin{bmatrix} \vdots \\ \delta t_i \\ \vdots \end{bmatrix} = \begin{bmatrix} & \vdots & \\ \dots & L_{ij} & \dots \\ & \vdots & \end{bmatrix} \times \begin{bmatrix} \vdots \\ \delta s_j \\ \vdots \end{bmatrix} \text{N slowness perturbations} \tag{6}$$

M×N **sensitivity matrix**

Letting it simmer: Solving inverse problems

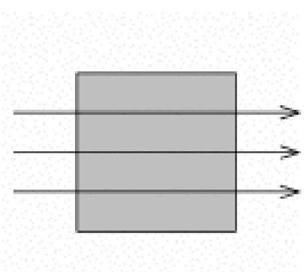
We have: $\mathbf{G} \cdot \mathbf{m} = \mathbf{d}$, which is **linear**.

You think: $\mathbf{m} = \mathbf{G}^{-1} \cdot \mathbf{d}$, but we **can't invert** a non-square $M \times N$ matrix.

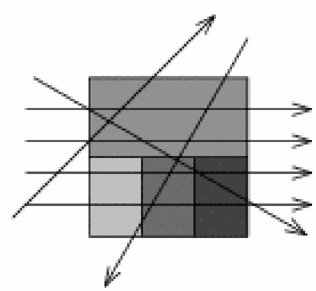
You think: $\mathbf{G}^{\mathrm{T}} \cdot \mathbf{G}$ is square, let's solve $\mathbf{G}^{\mathrm{T}} \cdot \mathbf{G} \cdot \mathbf{m} = \mathbf{G}^{\mathrm{T}} \cdot \mathbf{d}$.

You try: $\mathbf{m} = (\mathbf{G}^{\mathrm{T}} \cdot \mathbf{G})^{-1} \cdot \mathbf{G}^{\mathrm{T}} \cdot \mathbf{d}$.

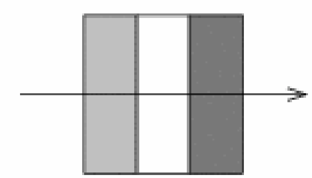
Alas! $\mathbf{G}^{\mathrm{T}} \cdot \mathbf{G}$ may be singular, ill-conditioned, under/over-determined, have (near-)zero eigenvalues, and thus be **not-invertible**.



over-determined, M>N



mixed-determined

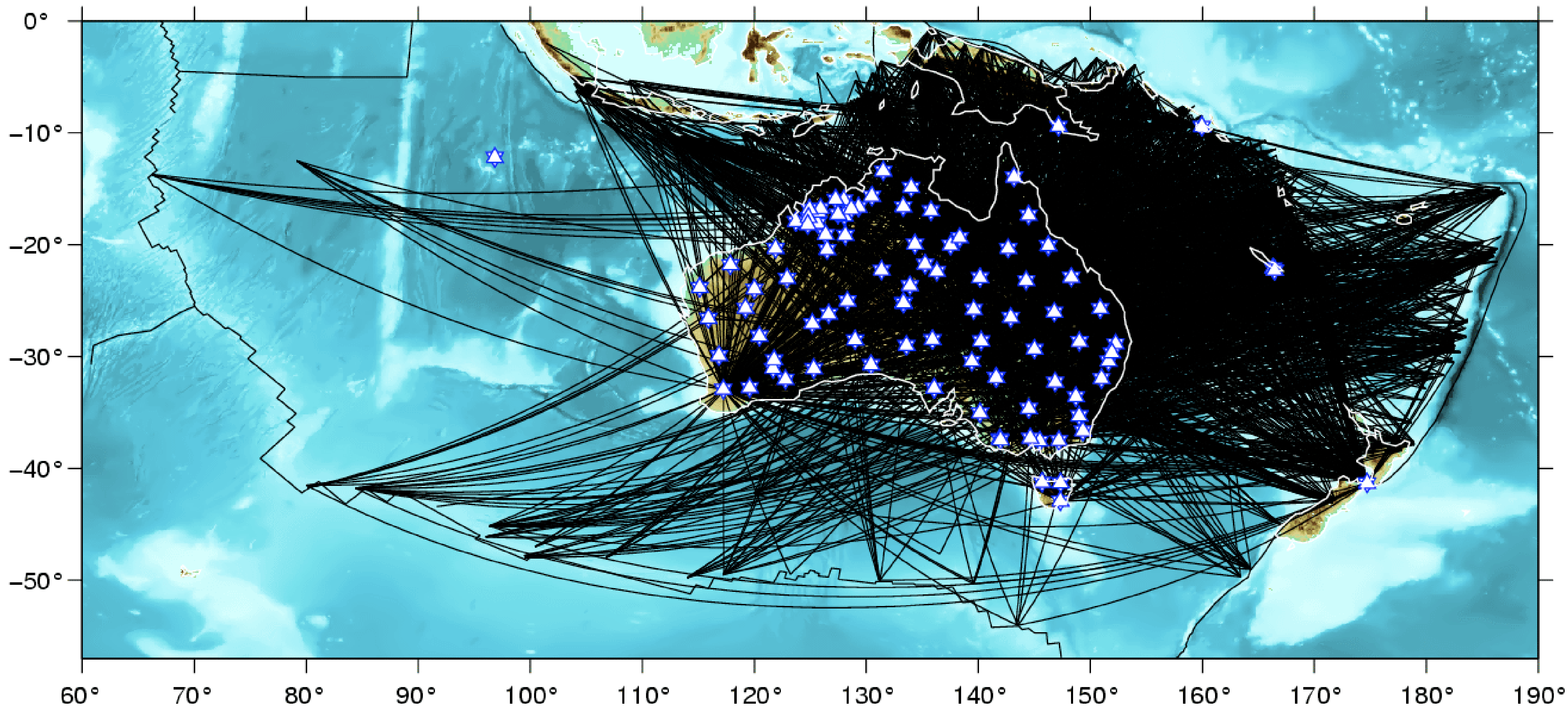


under-determined, M<N

Menke, 1989

Receiver coverage

Picking the right continent

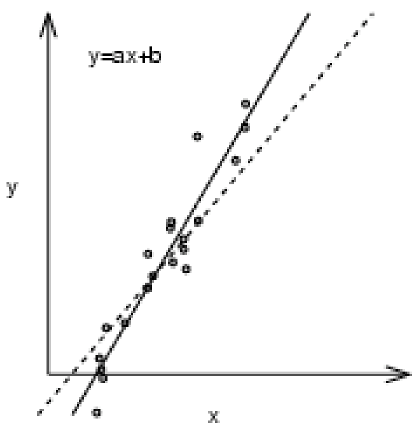


A dense path coverage minimizes the amount of a priori information needed

Simons, *GJI* 2002

Recipe, Step 3: Regularize!

Over-determined: *More data than unknowns*



Define a *penalty fuction* Φ on the *error* $\mathbf{e}$,

and minimize, by least-squares:

$$\Phi = [\mathbf{G} \cdot \mathbf{m} - \mathbf{d}]^2 = \mathbf{e}^{\mathrm{T}} \cdot \mathbf{e} \quad \text{by} \quad \frac{\partial \Phi}{\partial m_i} = 0. \tag{7}$$

This is a minimization in the *data space*.

Under-determined: *More unknowns than data*

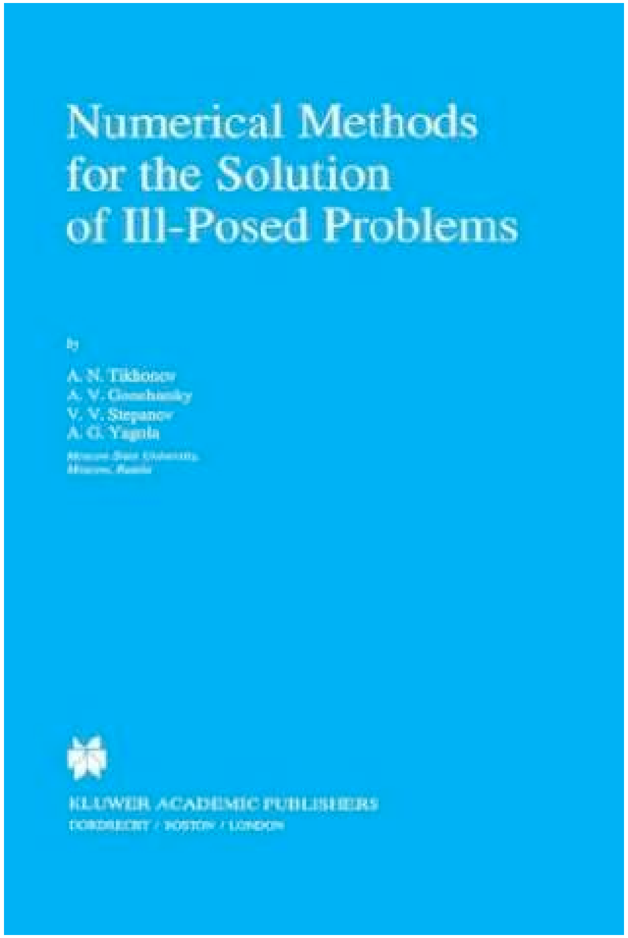
Add equations that minimize some norm in the *model space*:

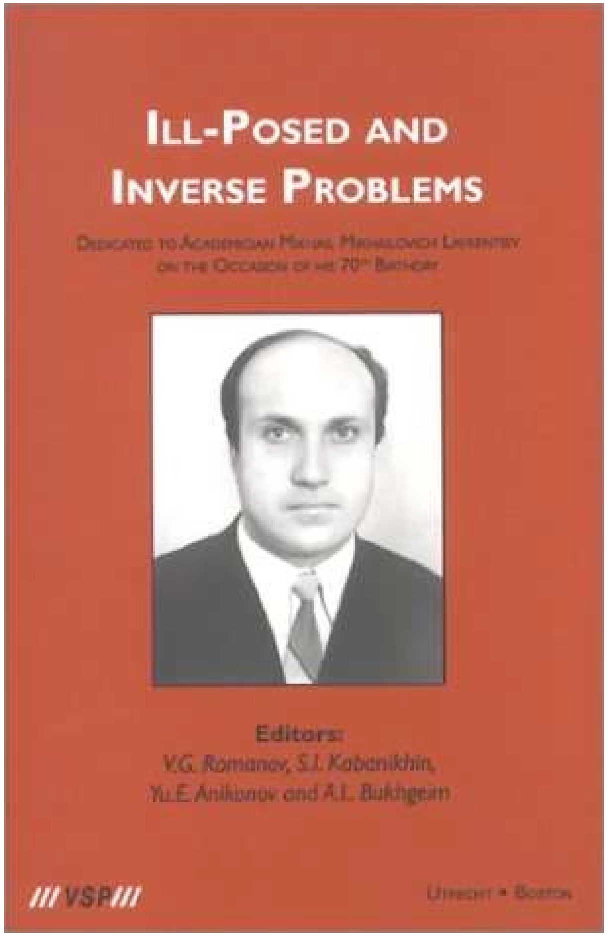
$$\Phi = \mathbf{e}^{\mathrm{T}} \cdot \mathbf{e} + \mathbf{m}^{\mathrm{T}} \cdot (\mathbf{A}^{\mathrm{T}} \cdot \mathbf{A}) \cdot \mathbf{m}. \tag{8}$$

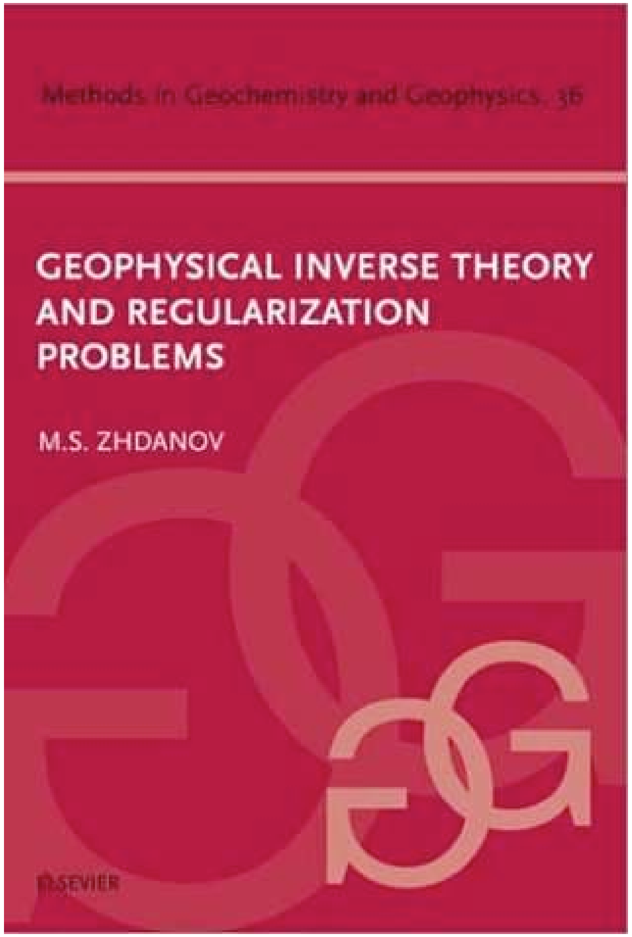
If $\mathbf{A} = \mathbf{I}$ the identity matrix $\rightarrow$ minimum model norm: **norm damping**.

If $\mathbf{A} = \mathbf{D}$ a difference matrix $\rightarrow$ minimum-roughness: **smoothing**.

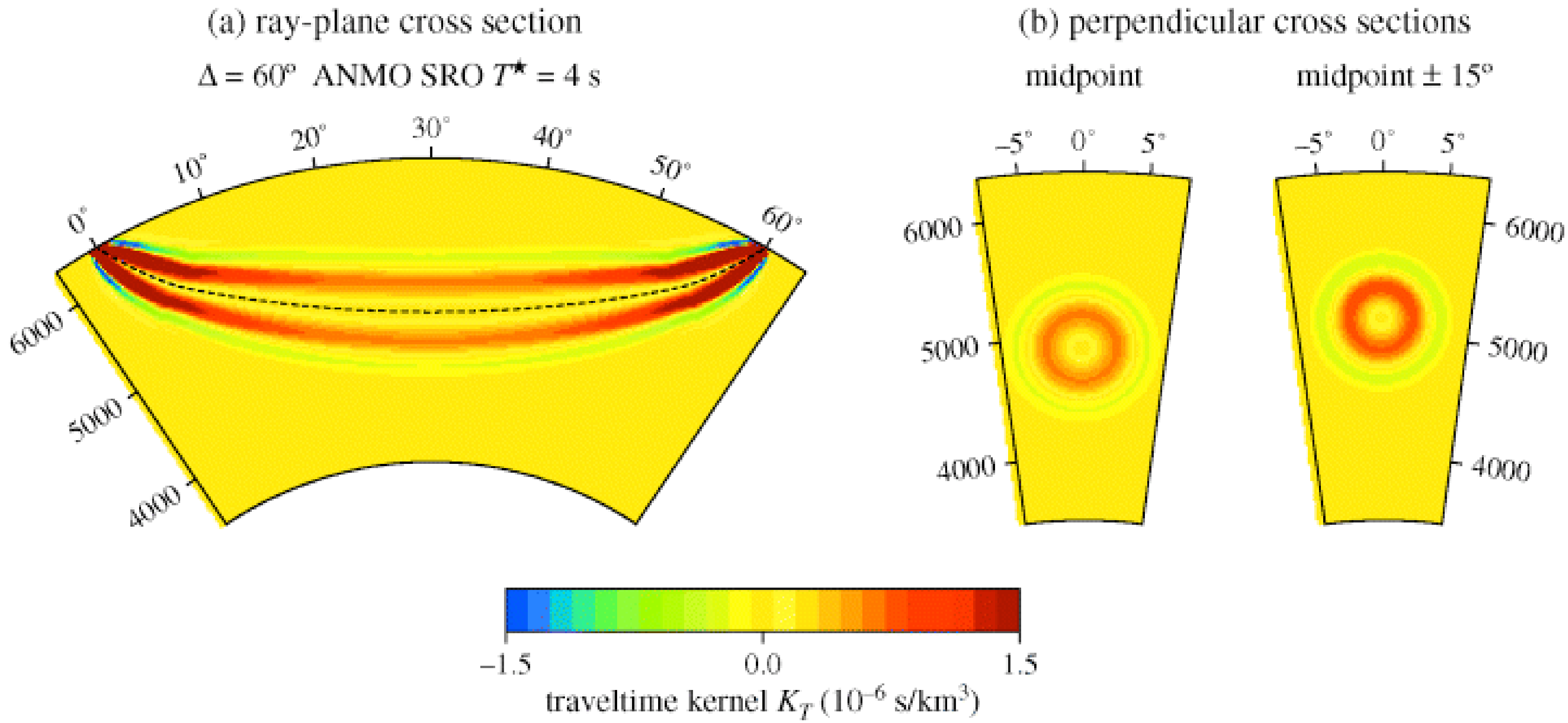
Regularization: the Mathematics







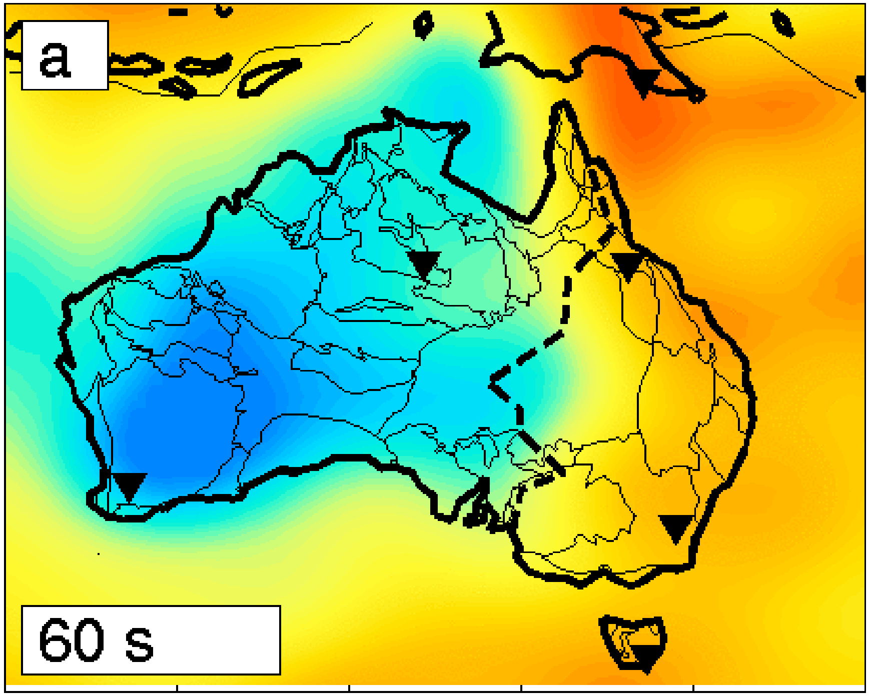
Regularization: the Physics



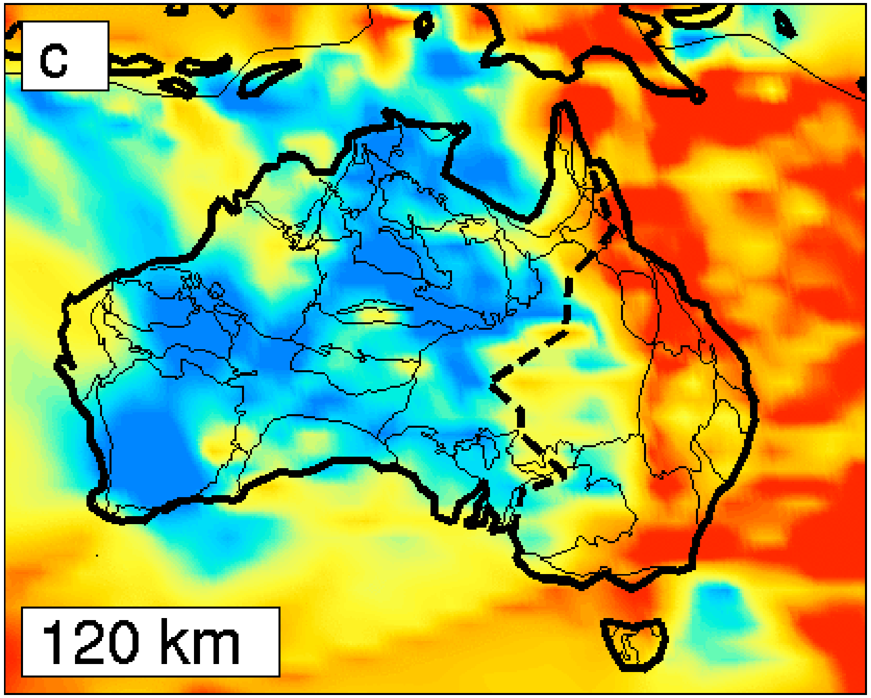
Such " fat" rays sample more of the Earth and thus we need fewer of them to have a well-constrained tomographic problem.

Dahlen, *GJI*, 2002

Regularization: the Art



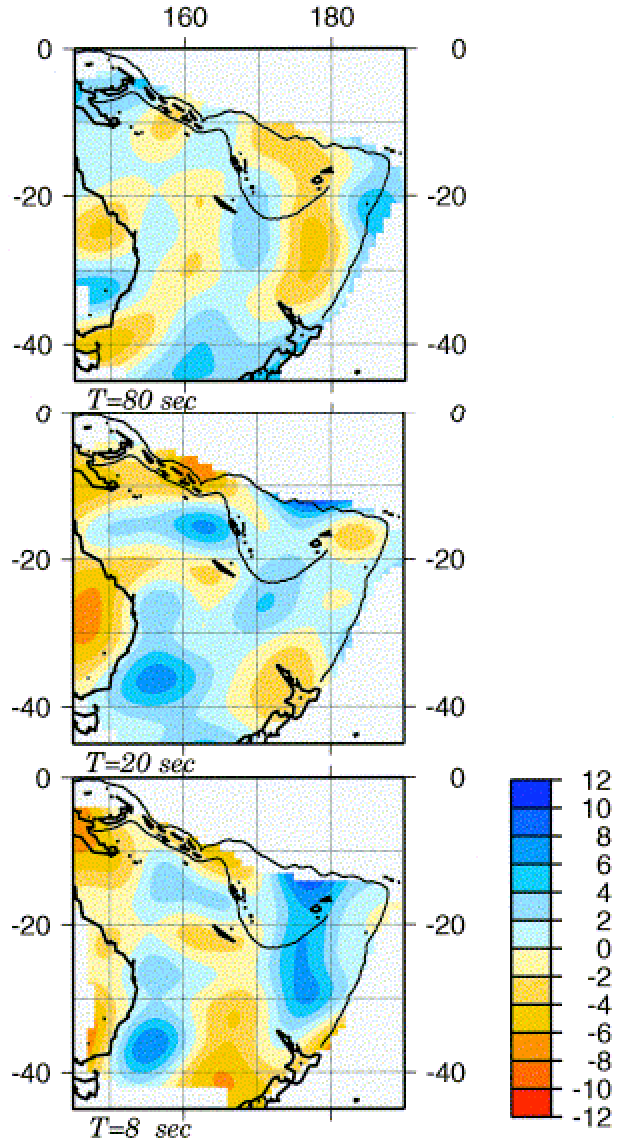
Too much?
Too smooth?



Too little?
Too rough?

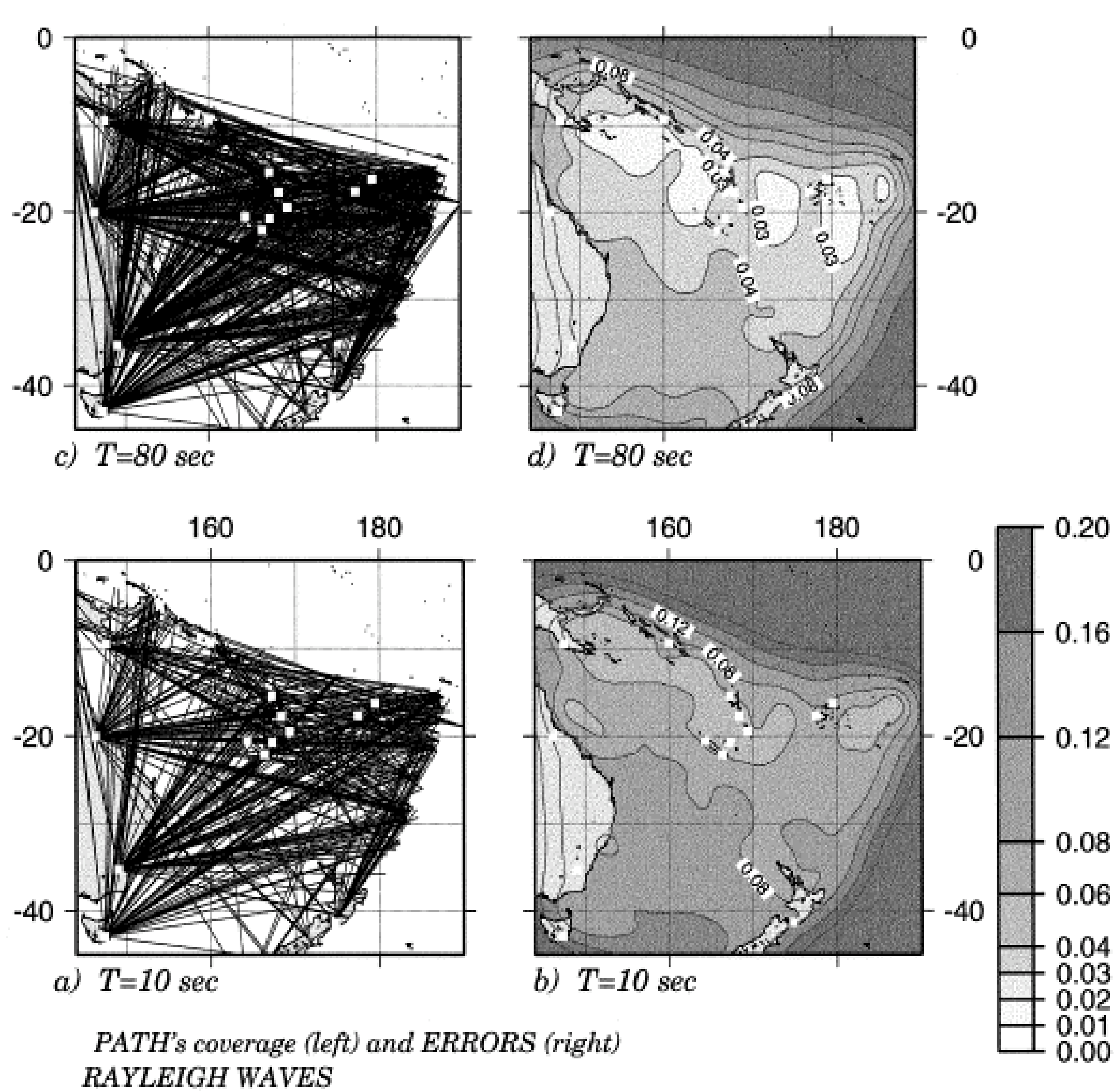
Simons, *Lithos*, 1999

How to interpret seismic models



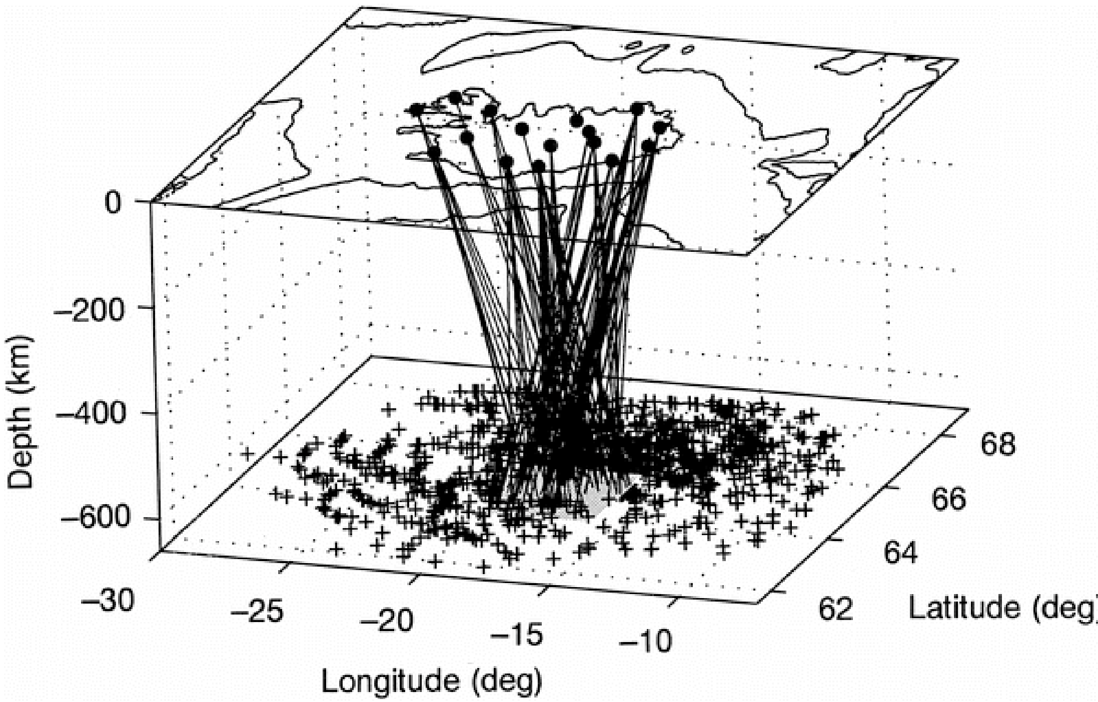
Pillet, *PEPI*, 1999

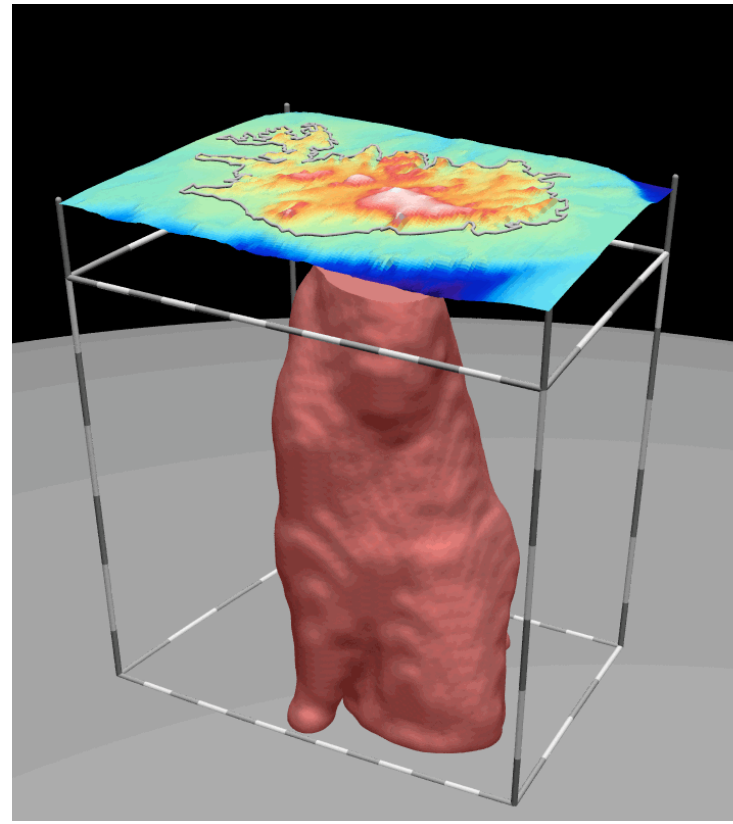
Demand to see the ray paths



Pillet, *PEPI*, 1999

Nature isn’ t always kind





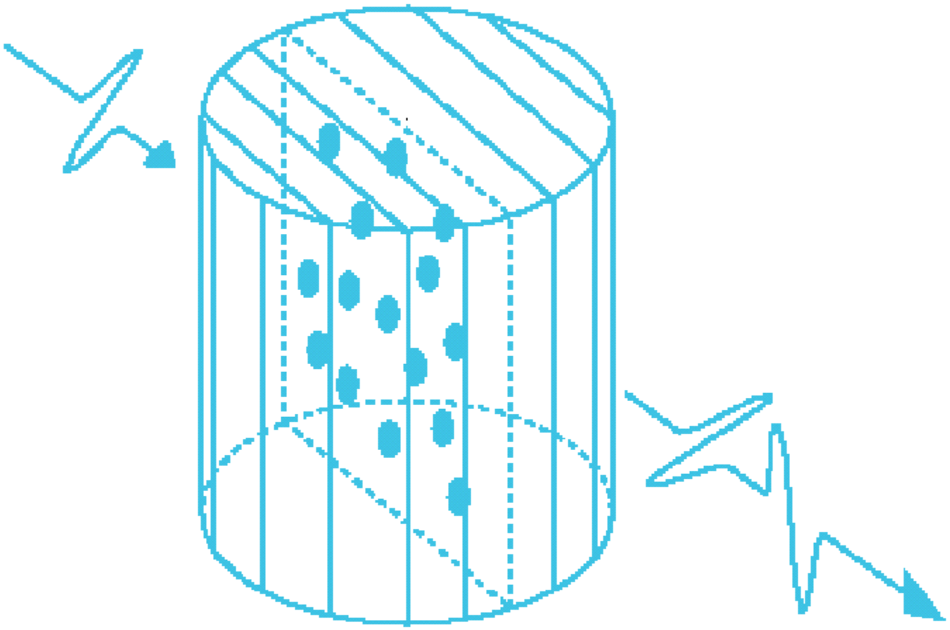
Shen, *Nature*, 1998

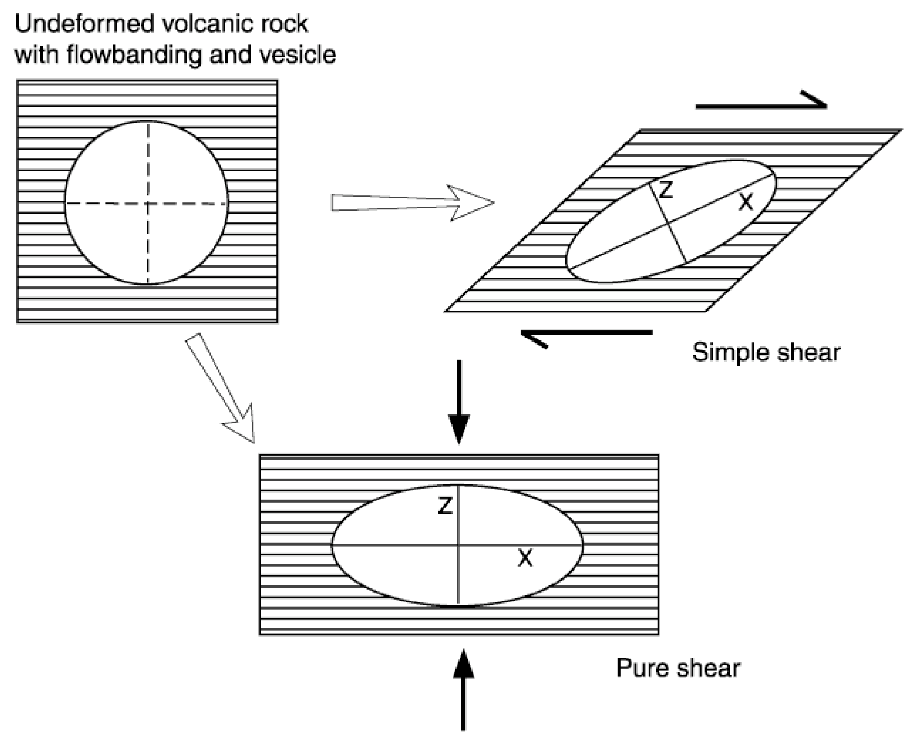
Wolfe, *Nature*, 1997

Seismic anisotropy

Wave speeds depend on

propagation direction and polarization:





No surprise: elasticity maps stress and strain, and both depend on three directions

Polarization anisotropy

- The particles of Love and Raleigh surface waves move in orthogonal directions
- SH and SV body waves sometimes exhibit clear splitting

Azimuthal anisotropy

- It’ s usually very hard to separate whether the time difference arises from an anisotropic direction or an isotropic wave speed difference (aka heterogeneity)

Why is this so hard?

For 3-D heterogeneity and slight anisotropy:

$$\delta\hat{\beta}_V = \delta\beta_V^{TI} + \frac{G_c}{2\rho\beta_V}\cos 2\theta + \frac{G_s}{2\rho\beta_V}\sin 2\theta \qquad (3)$$

Maximum direction is related to fast axis of anisotropic minerals:

$$G = \sqrt{G_c^2 + G_s^2} \quad \text{and} \quad \Psi_{\max} = \frac{1}{2}\arctan\frac{G_s}{G_c} \qquad (4)$$

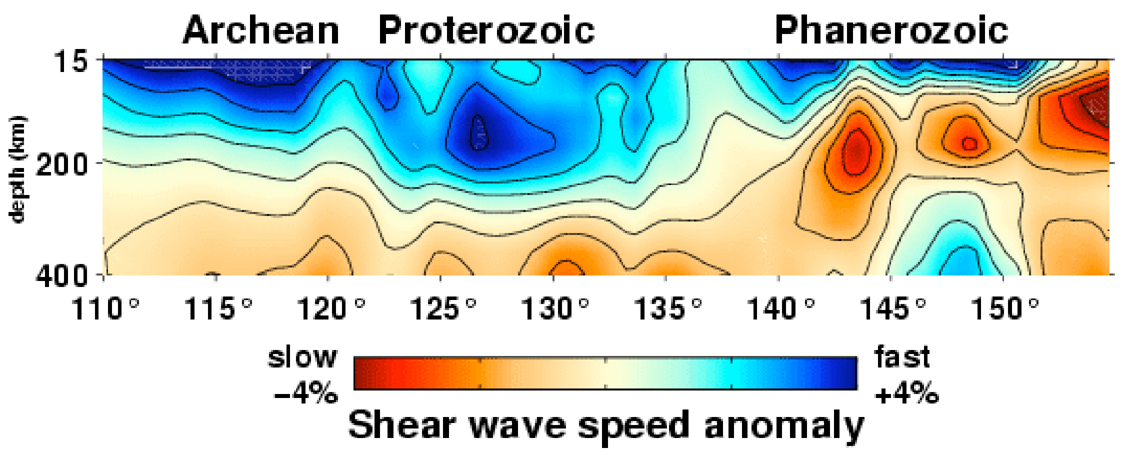
It' s very hard to tell whether a phase comes in early because it went through a fast patch or because it came in a fast direction –
heterogeneity and anisotropy " trade off."

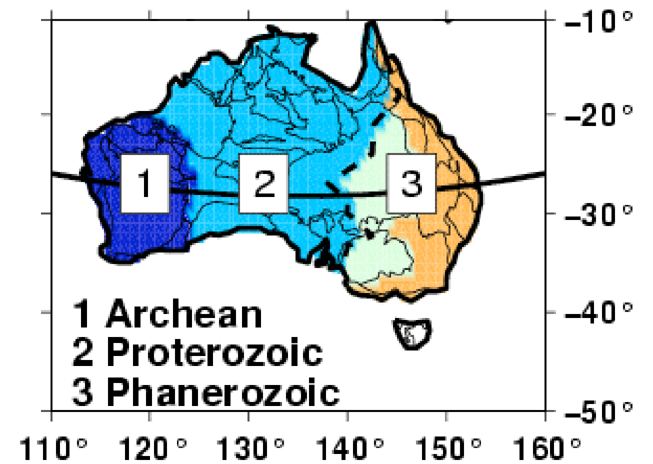
Questions to ask of the tomographer

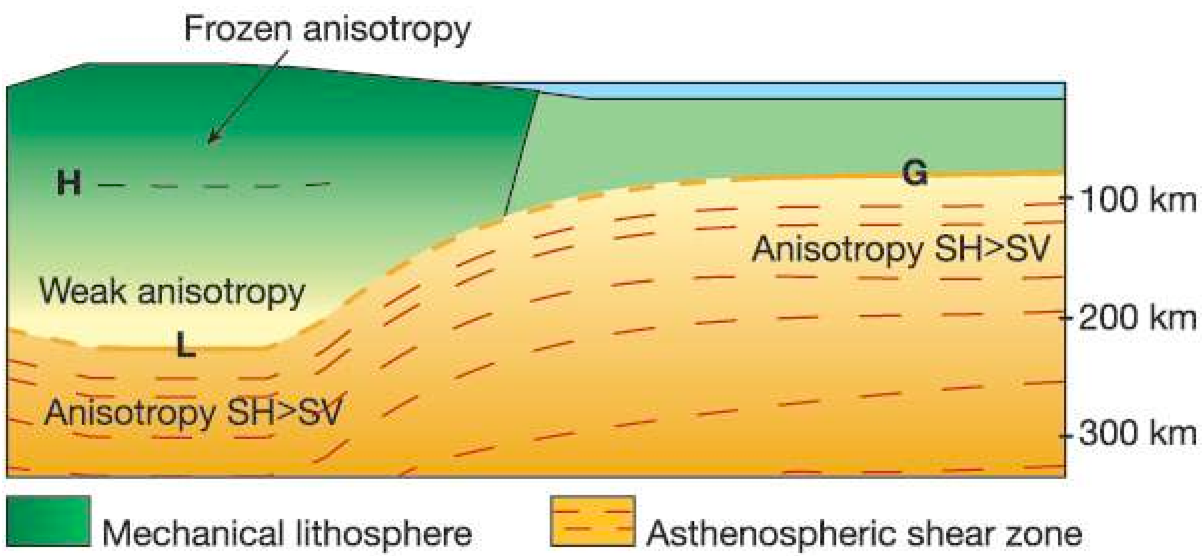
- How is the forward model computed?
- What is the ray coverage?
- What (sort of) damping did you use?
- What does velocity estimation trade off with?
- What is the grid size / the correlation length?
- How are different data sets weighted?
- How far is the final from the starting model?
- Does the starting model have discontinuities?
- How is the surface/depth parameterization
- Is your sensitivity 1-D, 2-D,or 3-D?

Journey to Middle Earth, Part I:

The continental lithosphere





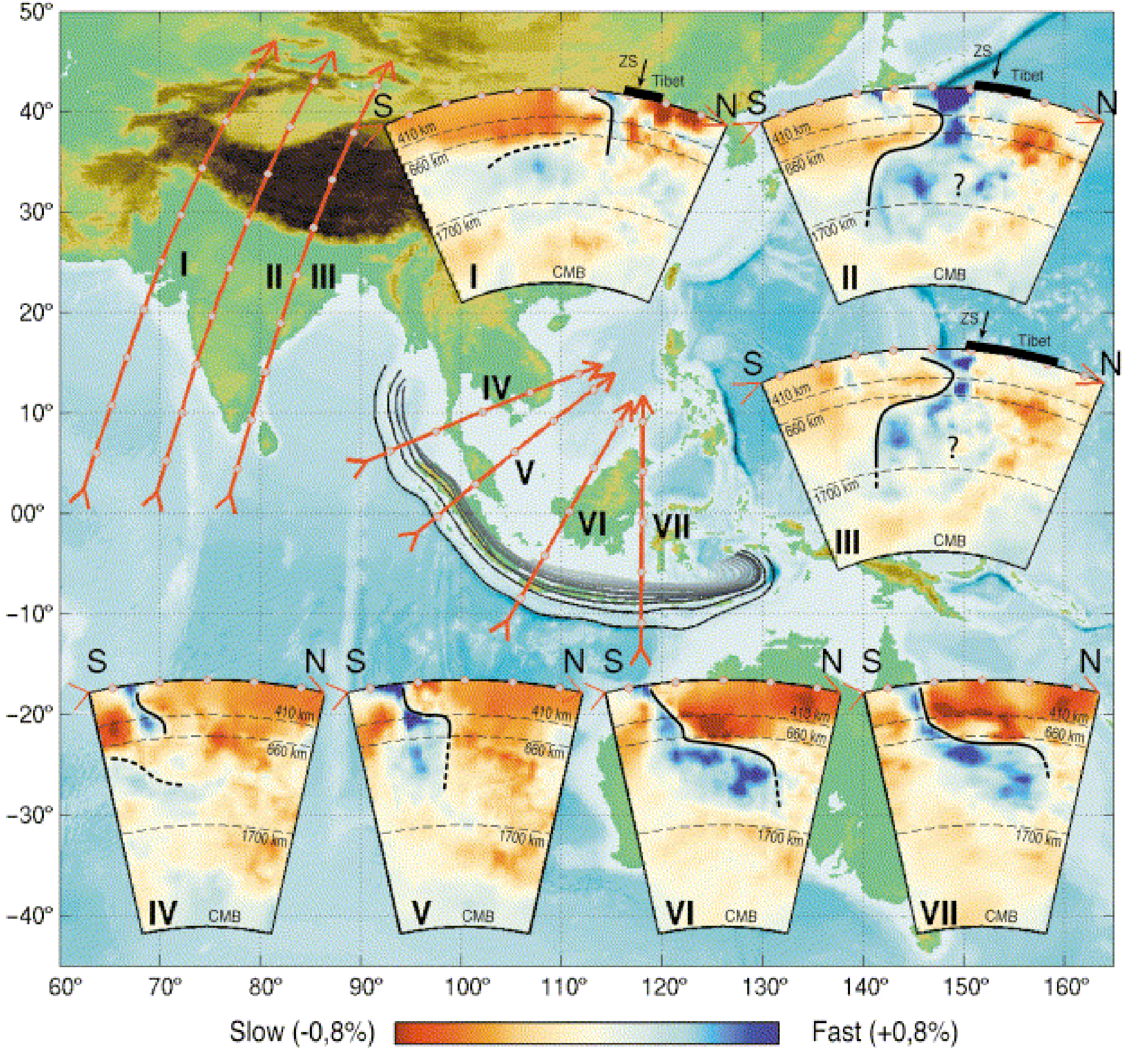


Simons, *GRL*, 2002

Gung, *Nature*, 2003

Journey to Middle Earth, Part II:

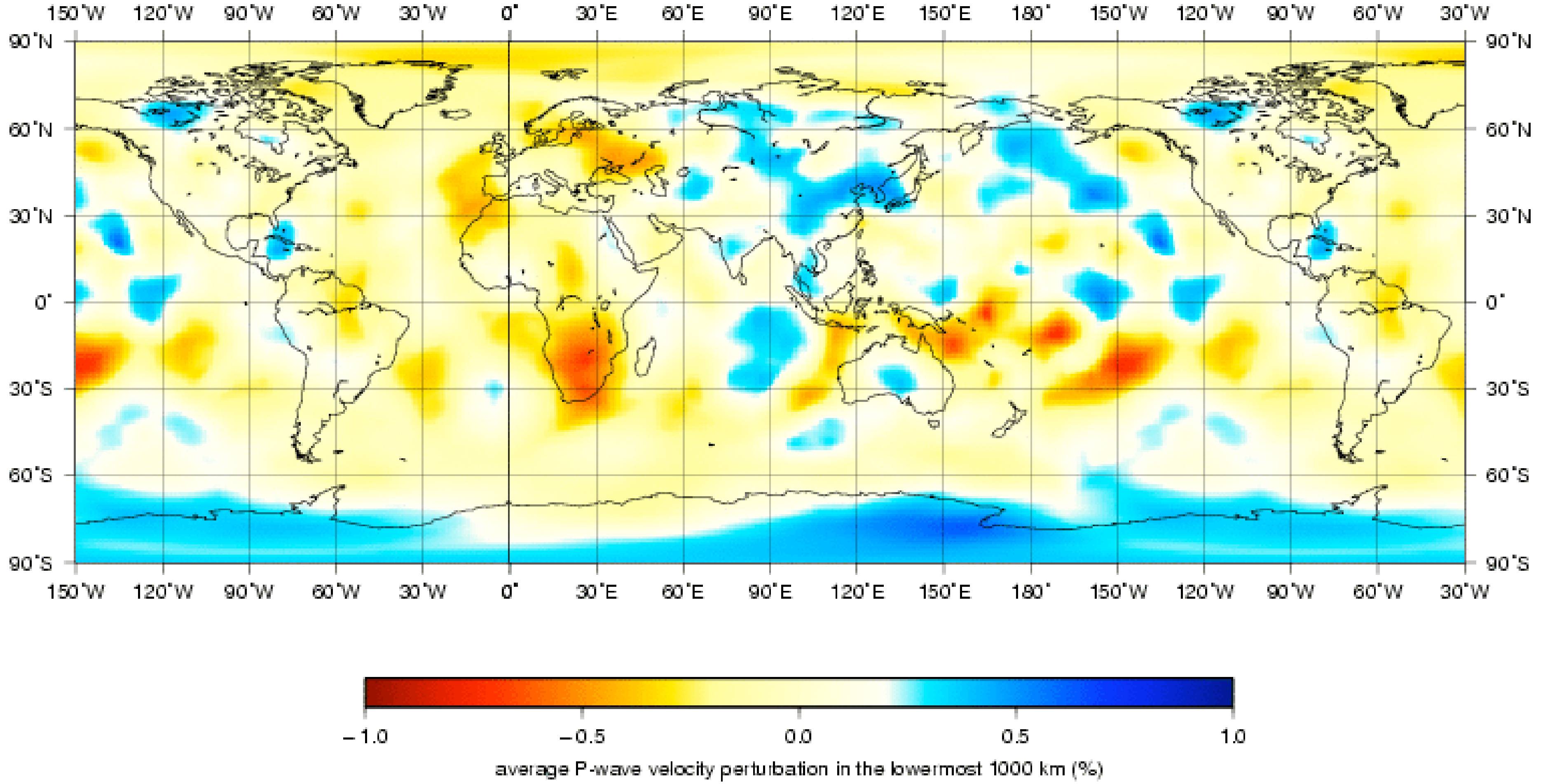
Subduction zones



Replumaz, *EPSL*, 2004

Journey to Middle Earth, Part III:

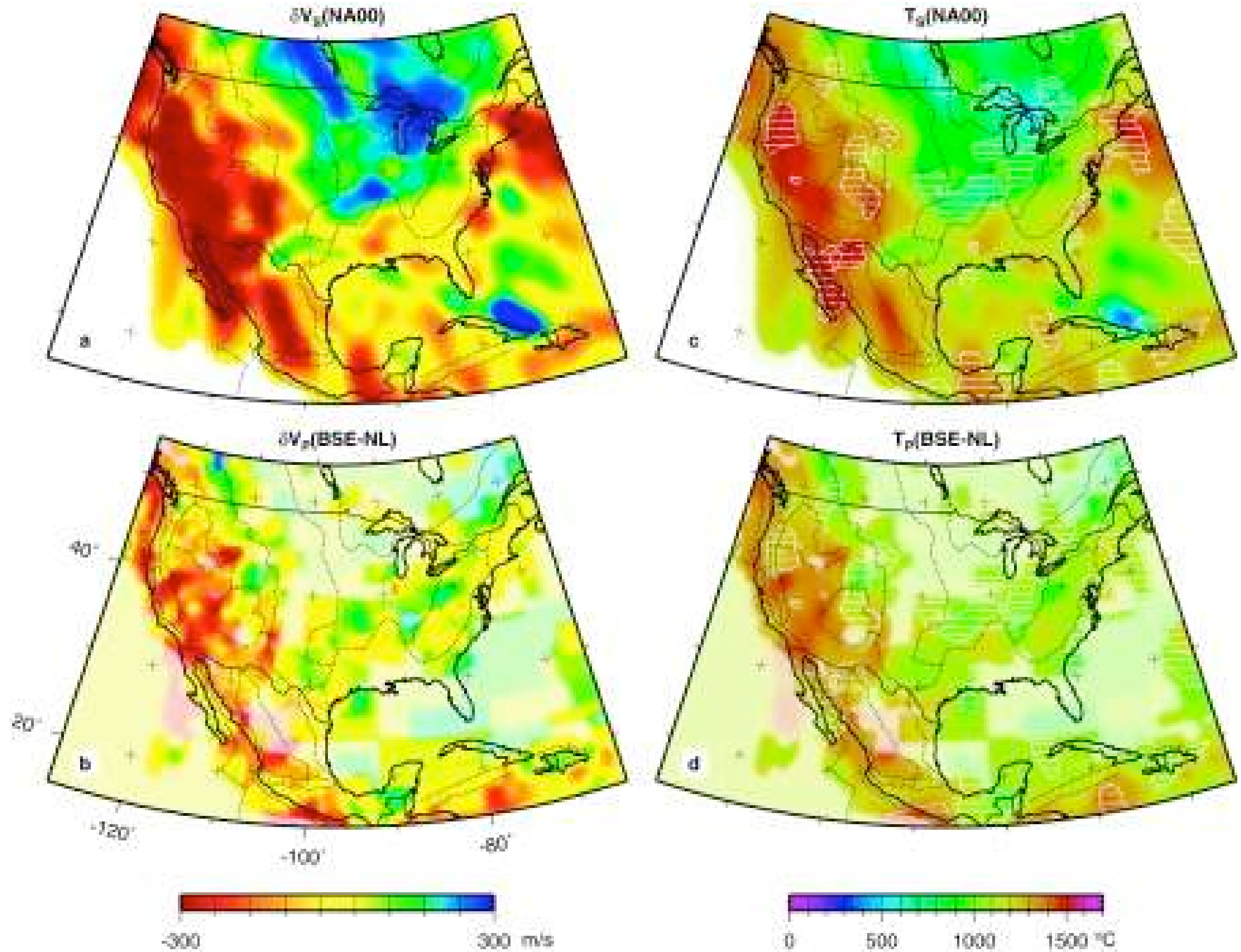
Deep mantle plumes



Montelli, *Science,* 2004

What does it all mean? Part I:

Temperature anomalies

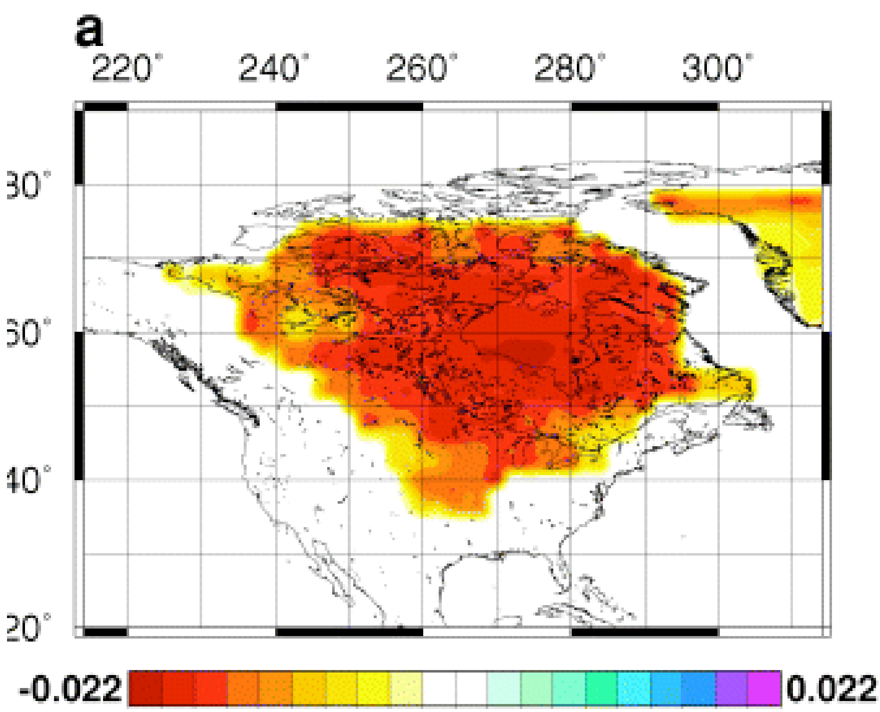


110 km

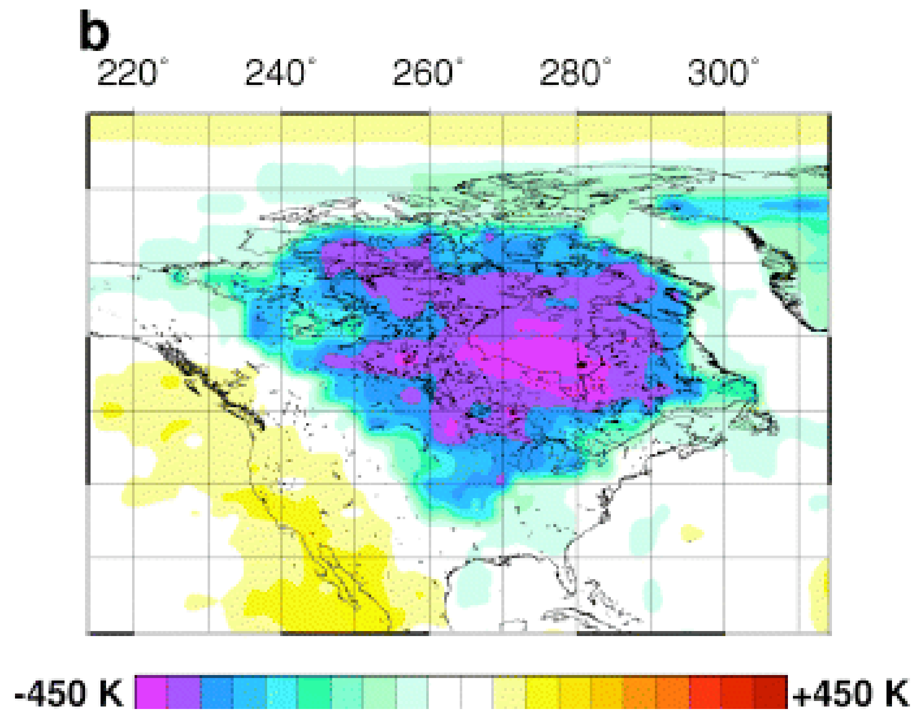
Goes, *JGR*, 2002

What does it all mean? Part II:

Compositional anomalies



ΔFe/(Fe + Mg)



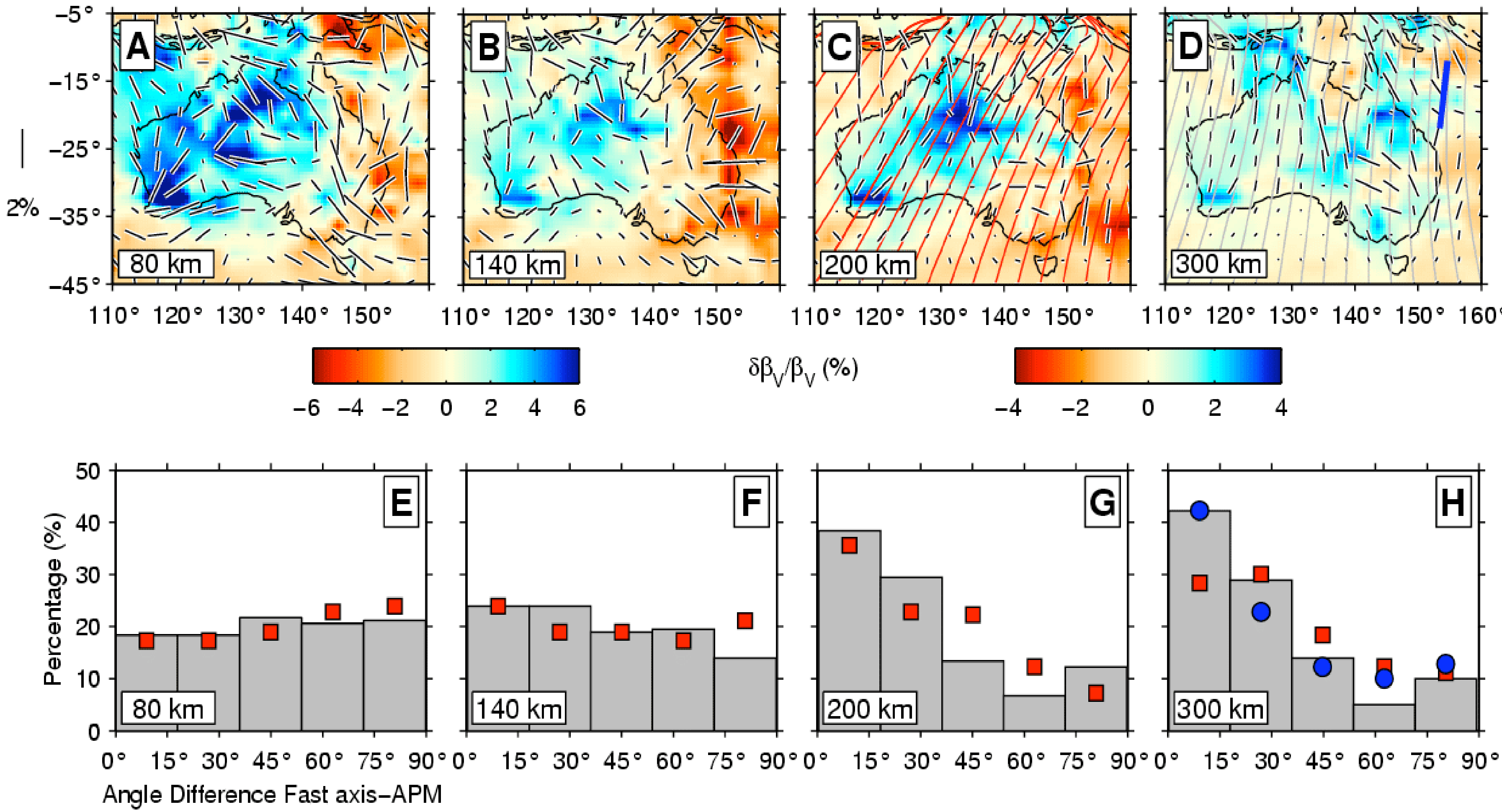
ΔT

150 km

Perry, *GJI*, 2003

What does it all mean? Part III:

Deformation in the mantle



Fossil

Contemporaneous

Simons, *EPSL*, 2003

Conclusions

- Ultimately, seismology can only tell us where, or in which direction, wave propagation is faster or slower than a reference model
- The non-seismologist has to know the basics of inverse problem modeling, understand the sometimes poor constraints, and be critical
- Improvements are being made: better data, better forward models, better inversions
- As much as with the *a posteriori* interpretation, the community needs to help defining *a priori* acceptable starting models

References

Buland, R., and C. H. Chapman, The computation of seismic travel times, *Bull. Seism. Soc. Am.*, *73*, 1271–1302, 1983.

Bullen, K. E., and B. A. Bolt, *An Introduction to the Theory of Seismology*, 4 ed., Cambridge Univ. Press, Cambridge, UK, 1985.

Dahlen, F. A., and A. Baig, Fréchet kernels for body-wave amplitudes, *Geophys. J. Int.*, *150*, 440–466, 2002.

Dahlen, F. A., and J. Tromp, *Theoretical Global Seismology*, Princeton Univ. Press, Princeton, N. J., 1998.

Dalton, R., Marine seismology — A window on the inner Earth, *Nature*, *421*, 10–12, 2003.

Dziewonski, A. M., and D. L. Anderson, Preliminary Reference Earth Model, *Phys. Earth Planet. Inter.*, *25*, 297–356, 1981.

Goes, S., and S. van der Lee, Thermal structure of the North American uppermost mantle inferred from seismic tomography, *J. Geophys. Res.*, *107*, 2050, doi:10.1029/2000JB000,049, 2002.

Gung, Y., M. Panning, and B. Romanowicz, Global anisotropy and the thickness of continents, *Nature*, *422*, 707–711, 2003.

Jackson, I. (Ed.), *The Earth's mantle, Composition, structure, and evolution*, Cambridge Univ. Press, 1998.

Karki, B. B., L. Stixrude, and R. M. Wentzcovitch, High-pressure elastic properties of major materials of earth's mantle from first principles, *Rev. Geophys.*, *39*, 507–534, 2001.

Kennett, B. L. N., E. R. Engdahl, and R. Buland, Constraints on seismic velocities in the Earth from travel-times, *Geophys. J. Int.*, *122*, 108–124, 1995.

Komatitsch, D., and J. Tromp, Spectral-element simulations of global seismic wave propagation — I. Validation, *Geophys. J. Int.*, *149*, 390–412, 2002.

Menke, W., *Geophysical Data Analysis: Discrete Inverse Theory*, vol. 45 of *International Geophysics Series*, 2nd ed., Academic Press, San Diego, Calif., 1989.

Montelli, R., G. Nolet, F. A. Dahlen, G. Masters, E. R. Engdahl, and S.-H. Hung, Finite-frequency tomography reveals a variety of plumes in the mantle, *Science*, *303*, 338–343, 2004.

Pillet, R., D. Rouland, G. Roult, and D. A. Wiens, Crust and upper mantle heterogeneities in the southwest Pacific from surface wave phase velocity analysis, *Phys. Earth Planet. Inter.*, *110*, 211–234, 1999.

Radon, J., Über die Bestimmung von Funktionen durch ihre Intergralwerte längs gewisser Mannigfaltigkeiten, *Berichte Süchsische Akademie der Wissenschaften*, *29*, 262–277, 1917.

Replumaz, A., H. Kárason, R. D. van der Hilst, J. Besse, and P. Tapponnier, 4-D evolution of SE asia's mantle from geological reconstructions and seismic tomography, *Earth Planet. Sci. Lett.*, *221*, 103–115, 2004.

Schaff, D. P., G. H. R. Bokelmann, G. C. Beroza, F. Waldhauser, and W. L. Ellsworth, High-resolution image of Calaveras Fault seismicity, *J. Geophys. Res.*, *107*, 2186, doi:10.1029/2001JB000,633, 2002.

Shen, Y., S. C. Solomon, I. T. Bjarnason, and C. J. Wolfe, Seismic evidence for a lower-mantle origin of the iceland plume, *Nature*, *395*, 62–65, 1998.

Simons, F. J., and R. D. van der Hilst, Age-dependent seismic thickness and mechanical strength of the Australian lithosphere, *Geophys. Res. Lett.*, *29*, 1529, doi:10.1029/2002GL014,962, 2002.

Simons, F. J., and R. D. van der Hilst, Seismic and mechanical anisotropy and the past and present deformation of the australian lithosphere, *Earth Planet. Sci. Lett.*, *211*, 271–286, 2003.

Simons, F. J., A. Zielhuis, and R. D. van der Hilst, The deep structure of the Australian continent from surface-wave tomography, *Lithos*, *48*, 17–43, 1999.

Simons, F. J., R. D. van der Hilst, J.-P. Montagner, and A. Zielhuis, Multimode Rayleigh wave inversion for heterogeneity and azimuthal anisotropy of the Australian upper mantle, *Geophys. J. Int.*, *151*, 738–754, 2002.

Wolfe, C. J., I. T. Bjarnason, J. C. VanDecar, and S. C. Solomon, Seismic structure of the Iceland mantle plume, *Nature*, *385*, 245–247, 1997.

Zhao, D., and J. Lei, Seismic ray path variations in a 3D global velocity model, *Phys. Earth Planet. Inter.*, *141*, 153–166, 2004.

More equations, for completeness

A linear system of equations

We're attempting to solve

$$\mathbf{d} = \mathbf{G} \cdot \mathbf{m} \tag{1}$$

Minimize penalty function of weighted error and model norms

$$\begin{aligned} \Phi = \; & (\mathbf{d} - \mathbf{G} \cdot \mathbf{m}) \cdot \mathbf{A}^{-1} \cdot (\mathbf{d} - \mathbf{G} \cdot \mathbf{m}) \\ & + \mathbf{m} \cdot \mathbf{B}^{-1} \cdot \mathbf{m} \end{aligned} \tag{2}$$

In matrix form, solve

$$\begin{bmatrix} \mathbf{A}^{-1/2} \cdot \mathbf{G} \\ \mathbf{B}^{-1/2} \end{bmatrix} \cdot \mathbf{m} = \begin{bmatrix} \mathbf{A}^{-1/2} \cdot \mathbf{d} \\ 0 \end{bmatrix} \tag{3}$$

Solution

$$\boxed{\mathbf{m} = (\mathbf{B}^{-1} + \mathbf{G}^{\mathrm{T}} \cdot \mathbf{A}^{-1} \cdot \mathbf{G})^{-1} \cdot \mathbf{G}^{\mathrm{T}} \cdot \mathbf{A}^{-1} \cdot \mathbf{d}} \tag{4}$$

Norm and first gradient regularization

For $\mathbf{A}^{-1}$, use the inverse of the data covariance matrix $\mathbf{C}_\mathrm{d}$ (BLUE)

For $\mathbf{B}^{-1}$, use the identity matrix $\mathbf{I}$ plus the squared first derivative

$$\mathbf{D_1} = \begin{pmatrix} ... & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & ... \end{pmatrix} \tag{5}$$

Minimize weighted penalty function

$$\begin{aligned} \Phi = & \ (\mathbf{d} - \mathbf{G} \cdot \mathbf{m}) \cdot \mathbf{C}_\mathrm{d}^{-1} \cdot (\mathbf{d} - \mathbf{G} \cdot \mathbf{m}) \\ & + \alpha \, \mathbf{m} \cdot \mathbf{I} \cdot \mathbf{m} + \beta \, \mathbf{m} \cdot \mathbf{D}_1^2 \cdot \mathbf{m} \end{aligned} \tag{6}$$

Solution

$$\boxed{\mathbf{m} = (\alpha \mathbf{I} + \beta \mathbf{D}_1^2 + \mathbf{G}^\mathrm{T} \cdot \mathbf{C}_\mathrm{d}^{-1} \cdot \mathbf{G})^{-1} \cdot \mathbf{G}^\mathrm{T} \cdot \mathbf{C}_\mathrm{d}^{-1} \cdot \mathbf{d}} \tag{7}$$

Bayesian inversion

Gaussian *a priori* probability function on the model parameters

$$\rho(\mathbf{m}) \propto \exp\left(-\frac{1}{2}\mathbf{m} \cdot \mathbf{C}_{\mathrm{m}}^{-1} \cdot \mathbf{m}\right) \tag{8}$$

Maximize joint distribution of data, model, subject to $\mathbf{d} = \mathbf{G} \cdot \mathbf{m}$

$$\boxed{\mathbf{m} = \mathbf{C}_{\mathrm{m}} \cdot \mathbf{G}^{\mathrm{T}} \cdot (\mathbf{C}_{\mathrm{d}} + \mathbf{G} \cdot \mathbf{C}_{\mathrm{m}} \cdot \mathbf{G}^{\mathrm{T}})^{-1} \cdot \mathbf{d}} \tag{9}$$

Equivalent to (using a trivial matrix identity)

$$\mathbf{m} = (\mathbf{C}_{\mathrm{m}}^{-1} + \mathbf{G}^{\mathrm{T}} \cdot \mathbf{C}_{\mathrm{d}}^{-1} \cdot \mathbf{G})^{-1} \cdot \mathbf{G}^{\mathrm{T}} \cdot \mathbf{C}_{\mathrm{d}}^{-1} \cdot \mathbf{d} \tag{10}$$

So in choosing norm and gradient regularization we've identified

$$\mathbf{C}_{\mathrm{m}}^{-1} = \alpha\mathbf{I} + \beta\mathbf{D}_1^2 \tag{11}$$

This imposes a particular form of the *a priori* covariance $\mathbf{C}_{\mathrm{m}}$

To Bayes or not to Bayes, what's the question?

A priori model covariance function with correlation length L

$$C_{\mathrm{m}}(\mathbf{r}_1, \mathbf{r}_2) = \sigma^2 \exp\left(-\frac{|\mathbf{r}_1, \mathbf{r}_2|^2}{2L^2}\right) \tag{12}$$

The following equivalence holds [*Yanovskaya and Ditmar, 1990*]

$$\mathbf{m} \cdot \mathbf{C}_{\mathrm{m}}^{-1} \cdot \mathbf{m} = \frac{1}{2\pi} \frac{1}{(\sigma L)^2} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{L^2}{2}\right)^n \nabla^n \mathbf{m} \cdot \nabla^n \mathbf{m} \tag{13}$$

So indeed

$$\mathbf{C}_{\mathrm{m}}^{-1} = \alpha \mathbf{I} + \beta \mathbf{D}_1^2 + \text{higher-order terms} \tag{14}$$

Exact resolution computation

For the linear problem, in a generalized sense,

$$\mathbf{m}^{\mathrm{est}} = \mathbf{G}^{-\mathrm{g}} \cdot \mathbf{d}^{\mathrm{obs}} = \mathbf{G}^{-\mathrm{g}} \cdot \mathbf{G} \cdot \mathbf{m}^{\mathrm{true}} \tag{15}$$

The resolution matrix is given by

$$\mathbf{R} = \mathbf{G}^{-\mathrm{g}} \cdot \mathbf{G} \tag{16}$$

In the Bayesian framework [*Montagner, 1986*]

$$R(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') - \frac{C_{\mathrm{p}}(\mathbf{r}, \mathbf{r}')}{C_{\mathrm{m}}(\mathbf{r}, \mathbf{r}')} \tag{17}$$

This represents the degree to which we are able to reduce the *a priori* covariance C_{m} of the model parameters (the null-state of information) by obtaining the *a posteriori* covariance structure C_{p} after the inversion.