Geophysical Inverse Problems

Guy Masters

Cider 2006

Anatomy of a GIP

- Existence -- mainly of mathematical interest
- Uniqueness -- nope -- at least for linear inverse problems
- Construction -- what we are all obsessed with
- Evaluation -- what we should all be obsessed with

Types of GIPs

- Linear (rare -- usually uninteresting)
- Mildly non-linear (linearizable)
- Strongly non-linear (not discussed here -- model space search algorithms, GA, EP, etc)

Linear problems: an example

Mass and moment of inertia:

$$M = \int_0^R 4\pi r^2 \rho(r)\, dr$$

$$C = \int_0^R \frac{8\pi}{3} r^4 \rho(r)\, dr$$

Generic form:

$$d_i \pm \sigma_i = \int_0^R G_i(r) m(r)\, dr$$

$$d_1 = M \quad \text{and} \quad d_2 = C$$

$$G_1 = 4\pi r^2 \quad \text{and} \quad G_2 = \frac{8\pi}{3} r^4$$

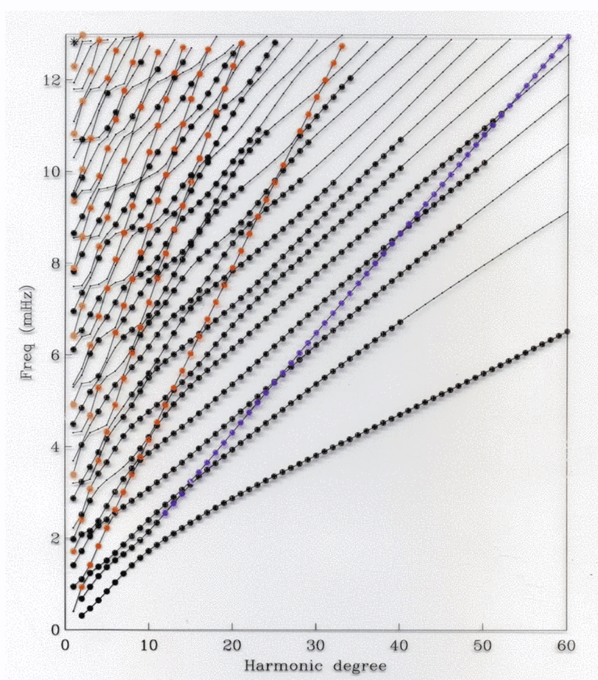
$$m(r) = \rho(r)$$

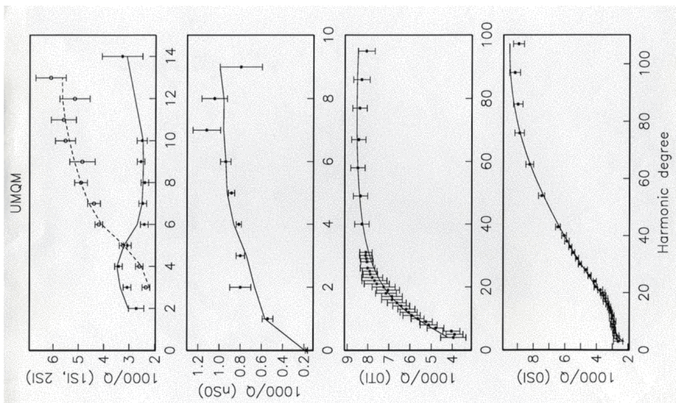
- $m(r)$ is continuous so infinite number of unknowns. Always underdetermined.
- Linear: double m and get double d.

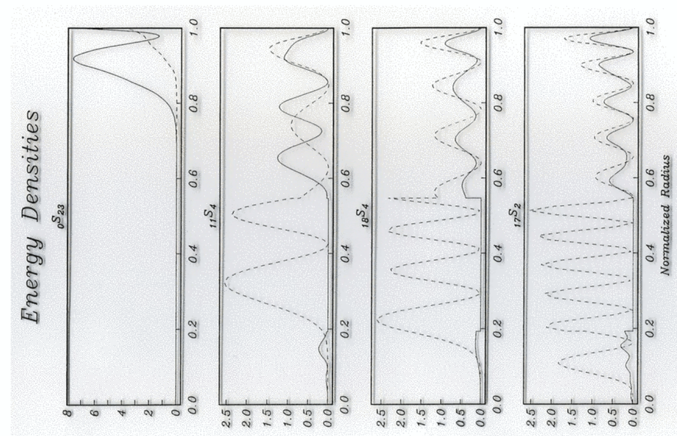
Linear problems: an example

- Inversion of attenuation of free oscillation attenuation rates for attenuation structure

$$Q_k^{-1} = \int_0^a \left[K(r)Q_\kappa^{-1} + M(r)Q_\mu^{-1}\right] r^2\, dr$$




UMQM
1000/Q (1S1, 2S1)
1000/Q ($_nS_0$)
1000/Q ($_0T_l$)
1000/Q ($_0S_l$)
Harmonic degree


Energy Densities
$_0S_{23}$
$_{11}S_4$
$_{18}S_4$
$_{17}S_2$
Normalized Radius

Linearized problems: data space solution

$$d_i \pm \sigma_i = \int_0^R G_i(r)\delta m(r)\,dr \quad i = 1 \cdots N$$

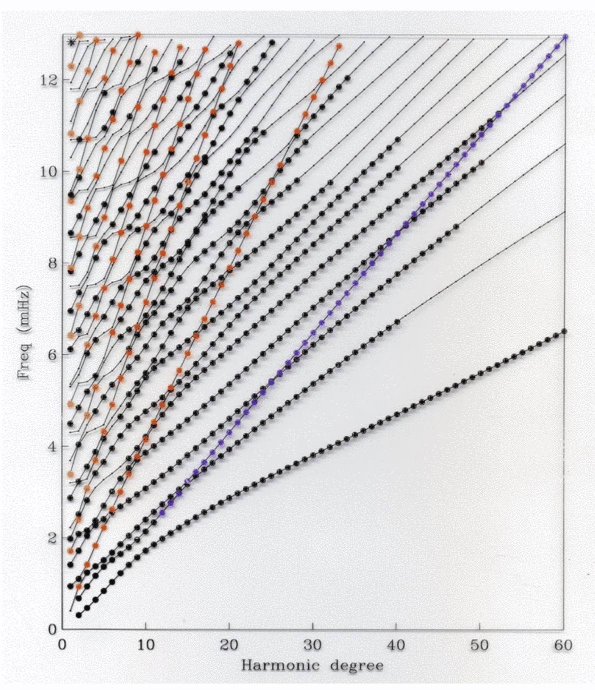
Try solution:

$$\delta m = \sum_1^N a_i G_i(r)$$

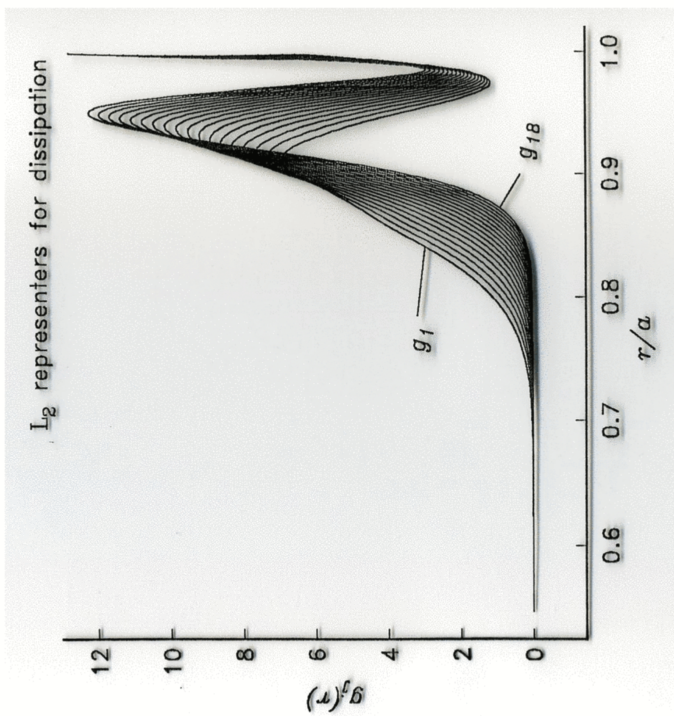
$$\mathbf{d} = \mathbf{\Gamma} \cdot \mathbf{a} \quad \text{where} \quad \Gamma_{ij} = \int_0^R G_i G_j\,dr$$

$\mathbf{\Gamma}$ is a matrix of dimension $N \times N$

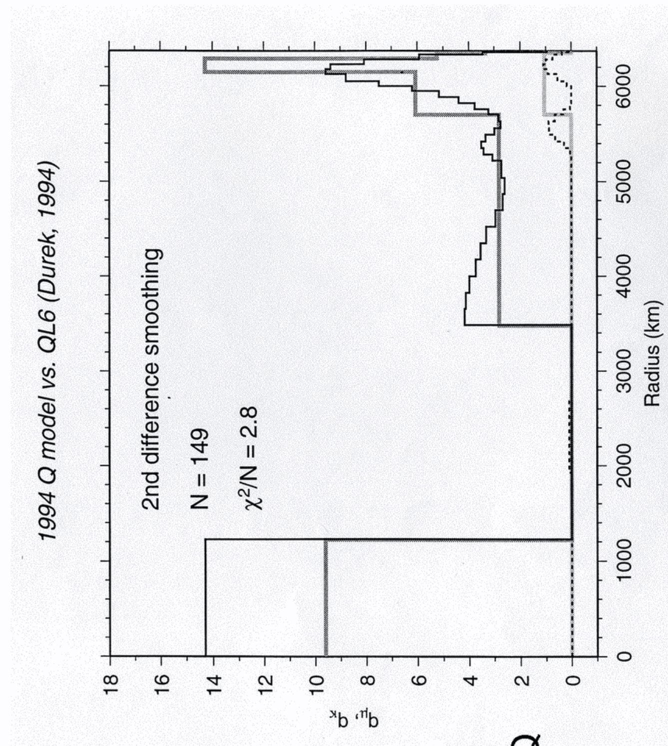
- Can solve (if N is small enough) in a variety of ways though $\mathbf{\Gamma}$ may be ill-conditioned. Leads to concept of "ranking and winnowing". Can also include smoothness conditions on the model or model perturbation.
- Resolution of model can be explored using Backus-Gilbert theory (see accompanying pdfs).



Shear energy kernels for 0S21(g1) to 0S38 (g18)



210 data -- only 20 combinations with errors less than 10%



$q = 1000./Q$

What about model evaluation?

Linear problems: Backus-Gilbrt

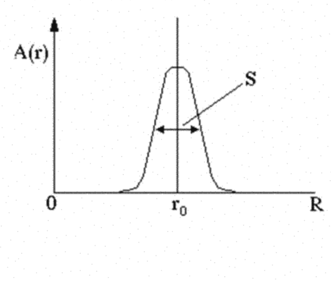
$$d_i \pm \sigma_i = \int_0^R G_i(r) m(r)\, dr$$

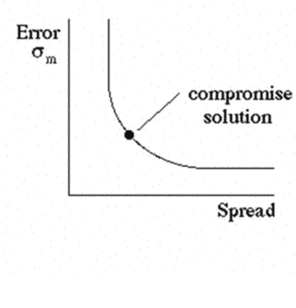
Take linear combination of data:

$$\bar{m}(r_0) = \sum_i a_i d_i = \int_0^R [\sum_i a_i G_i(r)] m(r)\, dr$$

$$= \int_0^R A(r, r_0) m(r)\, dr$$

Try to make $A(r, r_0)$ delta-like at a target radius r_0. Tradeoff between ability to make A delta-like and the precision of $\bar{m}(r_0)$ which is given by $\sqrt{\sum_i a_i^2 \sigma_i^2}$





Linearized problems: an example

Consider the free oscillation inverse problem. Suppose we let $\mathbf{m}(r)$ be the true Earth and $\mathbf{m}_0(r)$ be a starting model where $\mathbf{m}$ is usually taken to be the triplet of functions:

$$\mathbf{m}(r) = (\rho(r), \alpha(r), \beta(r))$$

then let

$$\delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$$

$$\delta\omega_i = \langle \mathbf{G}_i(\mathbf{m}_0), \delta\mathbf{m} \rangle + O|\delta\mathbf{m}|^2$$

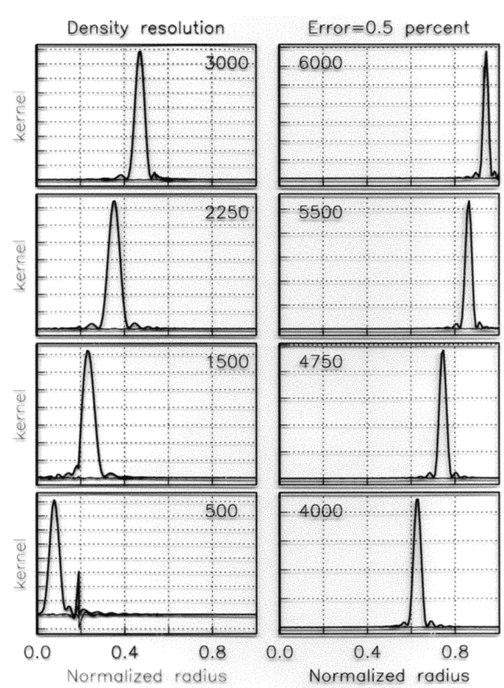
where, for the i'th mode

$$\delta\omega_i = \omega_{i_{obs}} - \omega_{i_{model}}$$

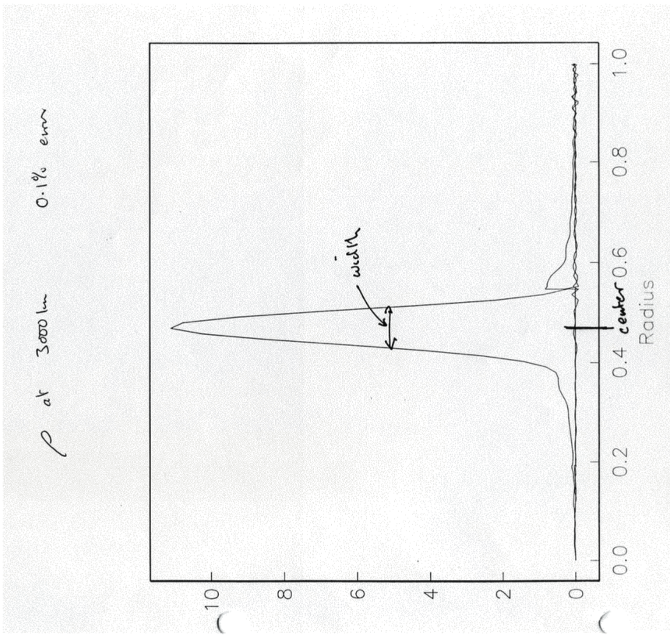
The braket notation is shorthand for:

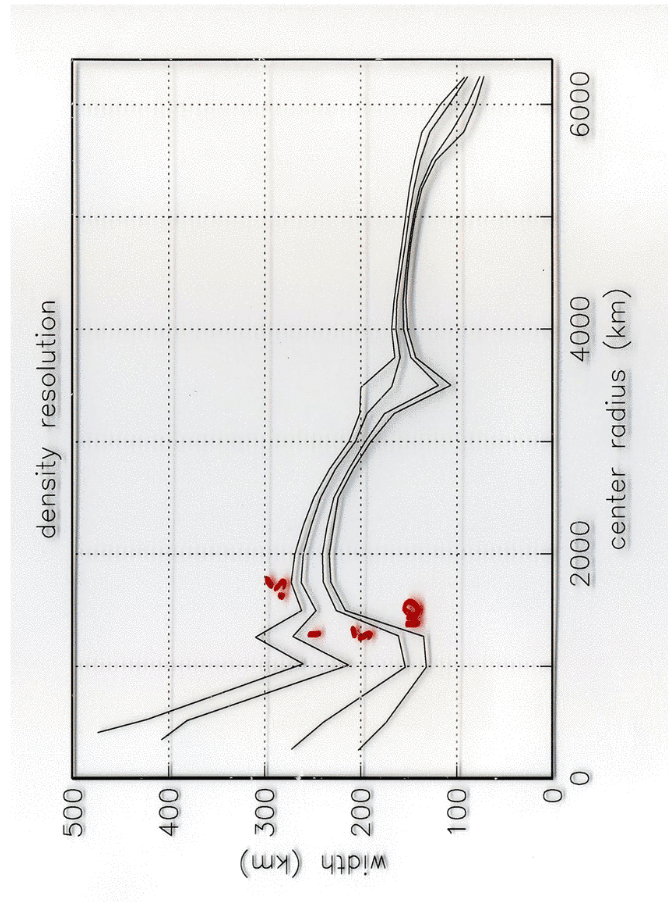
$$\langle \mathbf{G}_i(\mathbf{m}_0), \delta\mathbf{m} \rangle = \int_0^R (G_{\rho_i}\delta\rho + G_{\alpha_i}\delta\alpha + G_{\beta_i}\delta\beta)\, dr$$

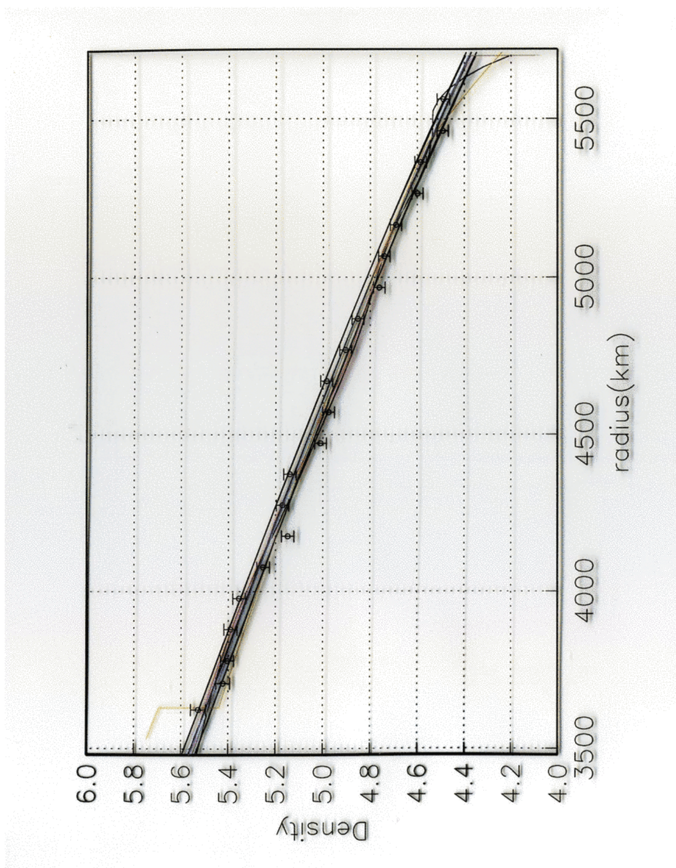
and the G_i's can be computed from the eigenfunctions of the i'th mode for the starting model.



Backus-Gilbert resolving power theory


ρ at 3000 km
0.1% err
width
center
Radius
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10


density resolution
center radius (km)
width (km)
0
2000
4000
6000
0
100
200
300
400
500



In practice -- too much data
therefore parameterize

Linearized problems: model space solution

$$d_i \pm \sigma_i = \int_0^R G_i(r)\delta m(r)\,dr \quad i = 1 \cdots N$$

Try expanding model in M basis functions, $f_i(r)$:

$$\delta m = \sum_1^M a_i f_i(r)$$

$$\mathbf{d} = \mathbf{A} \cdot \mathbf{a} \quad \text{where} \quad A_{ij} = \int_0^R G_i f_j \, dr$$

$\mathbf{A}$ is a matrix of dimension $N \times M$

- Problem may now be formally overdetemined ($N \geq M$) but will probably be ill-conditioned.
- Technically, we are now doing a "parameter estimation" problem since we have a finite number of unknows.

Parameterization

- Global bases lead to dense matrices? Spherical harmonics:

$$\delta m(r, \theta, \phi) = \sum_{l,m} \delta m_l^m(r) Y_l^m(\theta, \phi)$$

- Expansion up to degree l has $(l+1)^2$ coefficients
- With a dense matrix, limited to, say, 10,000 parameters. For a 20 layer mantle, l must be only about 20

$$l + \tfrac{1}{2} = \frac{2\pi a}{\lambda} = \frac{40000}{\lambda}$$

- Corresponds to a smallest wavelength of about 2000km
- USE LOCAL BASIS!

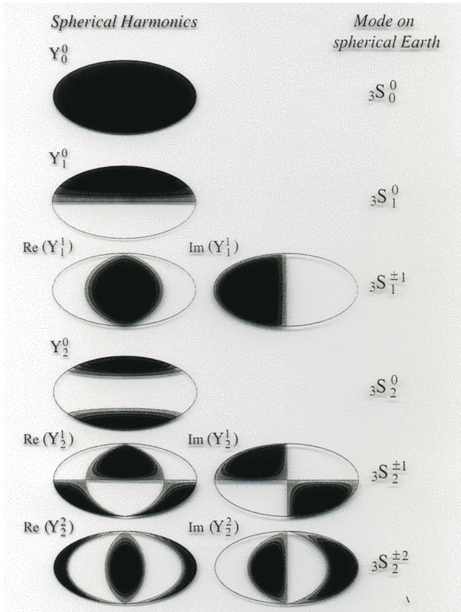


Figure 17 Some low-order spherical harmonics plotted in Hammer-Aitoff projection. Note that the singlets of a spheroidal mode have these shapes on the $\hat{r}$ component of recordings on a spherical Earth (eq. (A.43)).

Finding a model

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{d}$$

- Standard least squares: minimize $(\mathbf{A} \cdot \hat{\mathbf{x}} - \mathbf{d})^2$

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{d}$$

- Squared condition number (bad!). Use SVD

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T$$

$\mathbf{U}$ has dimension $N \times N$, $\mathbf{V}$ has dimension $M \times M$ and $\boldsymbol{\Lambda}$ is a $M \times M$ with non-zero diagonal elements. Note that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}$.

$$\hat{\mathbf{x}} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\mathbf{d} = \mathbf{A}^+\mathbf{d}$$

- $\mathbf{A}^+$ is an example of a generalized inverse of $\mathbf{A}$. If $\mathbf{A}$ is ill-conditioned, some of $\boldsymbol{\Lambda}$ will be small.
- Model covariance matrix:

$$\mathbf{A}^+\mathbf{I}(\mathbf{A}^+)^T = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{V}^T = \mathbf{V}\boldsymbol{\Lambda}^{-2}\mathbf{V}^T$$

- Small singular values lead to big model uncertainties. If you remove small singular values then resolution is degraded:

$$\hat{\mathbf{x}} = \mathbf{A}^+\mathbf{A}\mathbf{x} = \mathbf{R}\mathbf{x} = \mathbf{V}\mathbf{V}^T\mathbf{x}$$

Regularization

- Need more control on model contruction: explicit regularization. Minimize:

$$f = (\mathbf{Ax} - \mathbf{d})^T(\mathbf{Ax} - \mathbf{d}) + \lambda(\mathbf{Dx})^T\mathbf{Dx}$$

$$(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{D}^T\mathbf{D}) \cdot \mathbf{x} = \mathbf{A}^T\mathbf{d}$$

Solution is

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A} + \lambda\mathbf{D}^T\mathbf{D})^{-1}\mathbf{A}^T\mathbf{d}$$

whence $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A} + \lambda\mathbf{D}^T\mathbf{D})^{-1}\mathbf{A}^T$

- A reasonable forrm for $\mathbf{D}$ is a first or second difference operator (don't use $\mathbf{D} = \mathbf{I}$):

$$\begin{pmatrix} 1 & -1 & 0 & 0 & \ldots \\ 0 & 1 & -1 & 0 & \ldots \\ 0 & 0 & 1 & -1 & \ldots \end{pmatrix}$$

- Actually solve (to avoid squaring condition number):

$$\begin{pmatrix} \mathbf{A} \\ \lambda^{\frac{1}{2}}\mathbf{D} \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{d} \\ \mathbf{0} \end{pmatrix}$$

Matrices usually too big to use direct inversion

use iterative techniques

True iterative techniques - SIRT

Iteratively solve $\mathbf{A}\mathbf{x} = \mathbf{d}$. At iteration q, define residual:

$$\mathbf{r}^q = \mathbf{d} - \mathbf{A} \cdot \mathbf{x}^q$$

Let first iterate exactly solve first equation, second iterate the second and so on. To get a unique perturbation minimize:

$$\left(A_{ij}\Delta x_j^q - r_i^q\right)^2$$

Then

$$\Delta x_j = \frac{A_{ij} r_i}{\sum_k A_{ik}^2}$$

Doesn't work well. One modification is to compute the correction for each row, as above, then average all corrections to get a mean $\mathbf{\Delta x}$:

$$\Delta x_j = \frac{1}{M} \sum_{i=1}^{M} \frac{A_{ij} r_i}{\sum_k A_{ik}^2}$$

This is Simultaneous Iterative Reconstruction Technique (SIRT). Works fine – butdoesn't actually solve the problem you want to solve.

Gradient techniques

$$f(\mathbf{x}) = \tfrac{1}{2}\left(\mathbf{A} \cdot \mathbf{x} - \mathbf{d}\right)^2$$

$$f = \tfrac{1}{2}\left(\mathbf{A} \cdot \mathbf{x} - \mathbf{d}\right)^T\left(\mathbf{A} \cdot \mathbf{x} - \mathbf{d}\right)$$

$$= \tfrac{1}{2}\left[\mathbf{d}^T \cdot \mathbf{d} + \mathbf{x}^T \cdot \mathbf{A}^T \cdot \mathbf{A} \cdot \mathbf{x} - 2\mathbf{x}^T \cdot \mathbf{A}^T \cdot \mathbf{d}\right]$$

Now define the square symmetric matrix $\mathbf{B} = \mathbf{A}^T \cdot \mathbf{A}$ and the vector $\mathbf{b} = \mathbf{A}^T \cdot \mathbf{d}$ then

$$f = \tfrac{1}{2}\left[\mathbf{d}^T \cdot \mathbf{d} + \mathbf{x}^T\mathbf{B} \cdot \mathbf{x} - 2\mathbf{x}^T \cdot \mathbf{b}\right]$$

Define the misfit function $\phi(\mathbf{x})$ as the last two terms:

$$\phi(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^T\mathbf{B} \cdot \mathbf{x} - \mathbf{x}^T \cdot \mathbf{b}$$

The gradient of ϕ with respect to $\mathbf{x}$ is simply

$$\nabla\phi(\mathbf{x}) = \mathbf{B} \cdot \mathbf{x} - \mathbf{b}$$

At any point $\mathbf{x}_k$ on the surface, the downhill slope is given by

$$-\nabla\phi(\mathbf{x}_k) = \mathbf{b} - \mathbf{B} \cdot \mathbf{x}_k = \mathbf{r}_k$$

and is zero at a solution which fits the data ($\mathbf{B} \cdot \mathbf{x} - \mathbf{b} = 0$)

Steepest descent and conjugate gradient

- Steepest descent:

Consider sequence of solutions:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{u}_k$$

If the directions in which we go $\mathbf{u}_k$ is specified, it is easy to find λ_k which will minimise ϕ. If choose $\mathbf{u}_k = \mathbf{r}_k$ we get "steepest descent algorithm" (since we are going downhill). Not always good!

- Conjugate gradient

Choose directions to be "orthogonal" to one another in the sense that

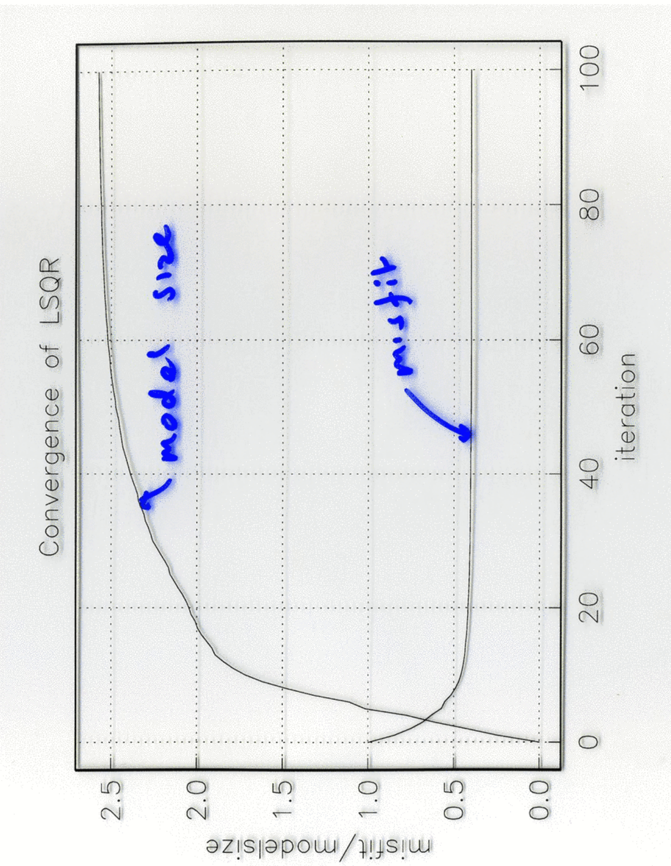
$$\mathbf{u}_k^T \cdot \mathbf{B} \cdot \mathbf{u}_j = 0$$

If there are M parameters, get convergence in M iterations. Leads to fast algorithms that usually converge in much less than M iterations and can take advantage of sparse matrices.

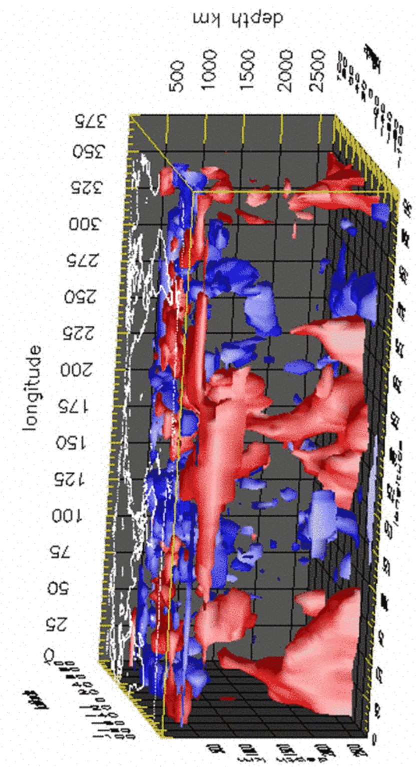
- A consequence of using such algorithms is that we no longer have the "genrealized inverse" and can't compute the resolution and model covariance matrices.

Example -- S-wave mantle tomography

- Inversion of large travel time, surface wave and free oscillation data sets
- Voxel parameterization: 4 degree lateral dimension, 100km in upper mantle and 200km in lower mantle radial dimension
- 50,000 model parameters, 300,000 data



Shear velocity -- +-1% isovelocity surfaces



Includes S and SS cluster analysis data

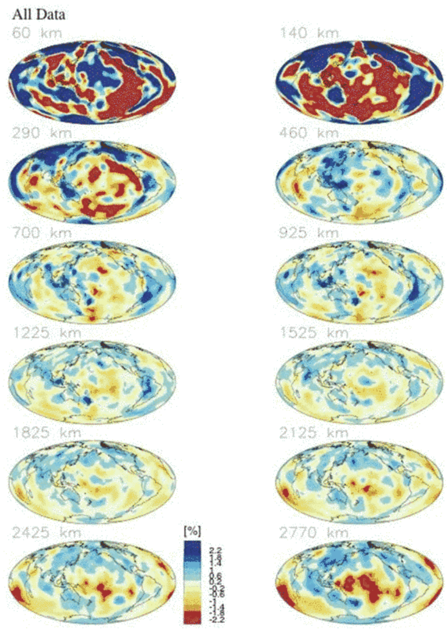


Plate 2. Cross-sections of the Earth at various depths showing perturbations of shear velocity (in percent). This high-resolution model was constructed using surface-wave and free-oscillation data as well as body wave data. Compare with plate 1. Note that the upper mantle is completely different but the anomalies in the lower mantle appear to be remarkably robust.

How well resolved?

- Biggest issue is data coverage and mixture of data types
- Less important is theory (e.g. banana-donuts) though may be important for certain kinds of structures
- What happens without surface waves?

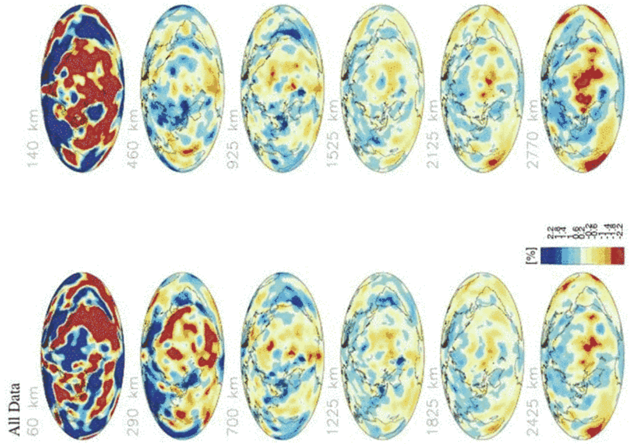


Plate 2. Cross-sections of the Earth at various depths showing perturbations of shear velocity (in percent). This high-resolution model was constructed using surface-wave and free-oscillation data as well as body wave data. Compare with plate 1. Note that the upper mantle is completely different but the anomalies in the lower mantle appear to be remarkably robust.

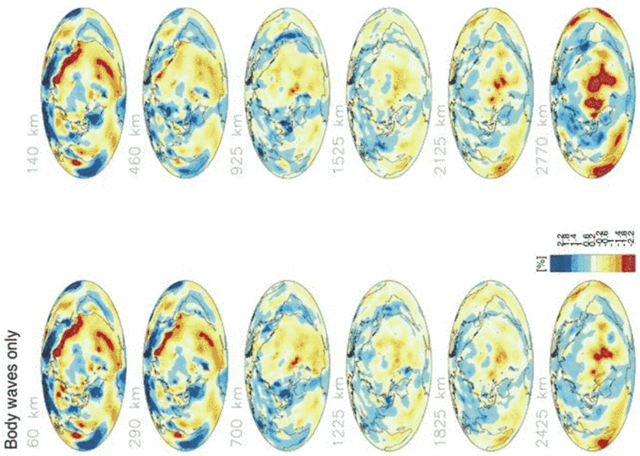
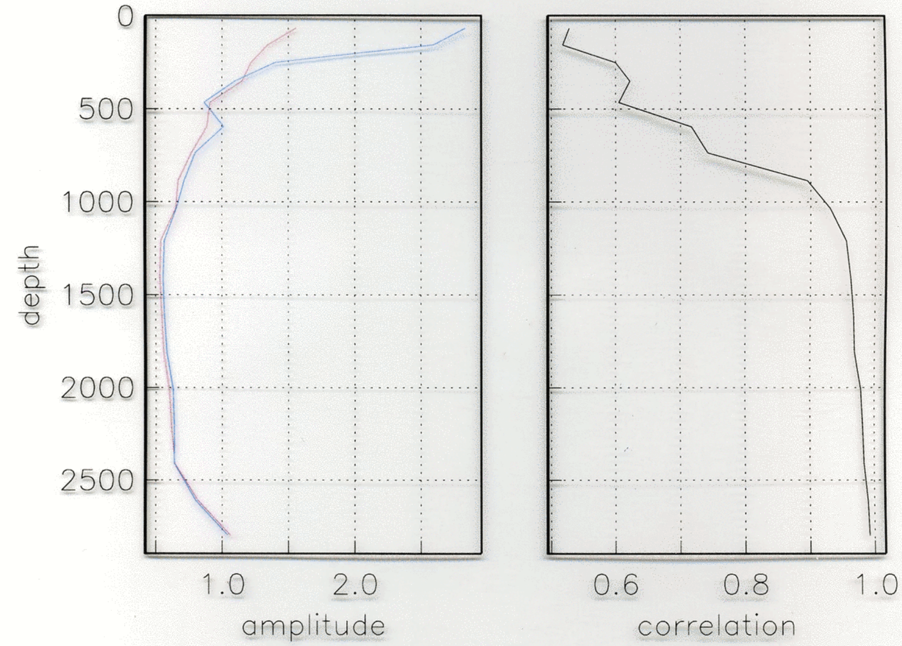


Plate 1. Cross-sections of the Earth at various depths showing perturbations of shear velocity (in percent). This high-resolution model was constructed using body wave data alone.



Resolution and Error

Ways to evaluate resolution and error when the generalized inverse is not available.

- Compute selected singular values and singluar vectors using SVDPACK (for sparse matrices)
- Compute a row (or column) of the resolution matrix using a "spike test"
- Evaluate resolution using a checkerboard test.
- Evaluate resolution by inverting for a specific type of known structure (e.g. a slab).
- Evaluate model errors using a Monte-Carlo technique

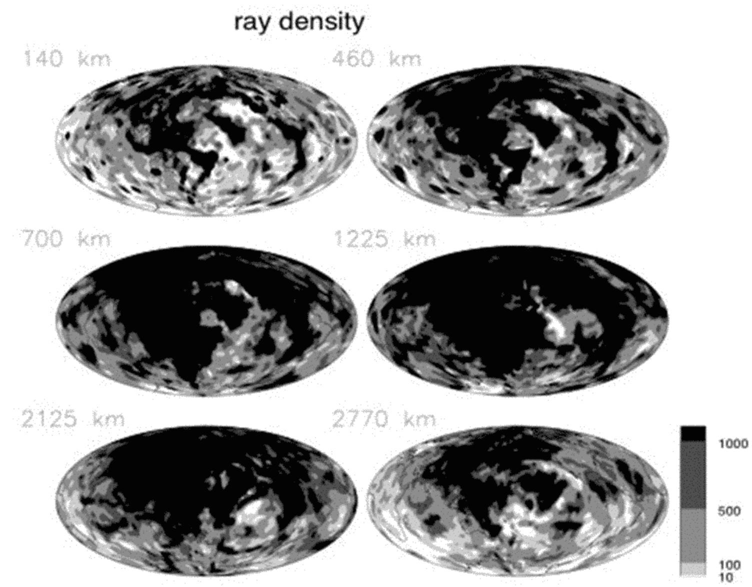
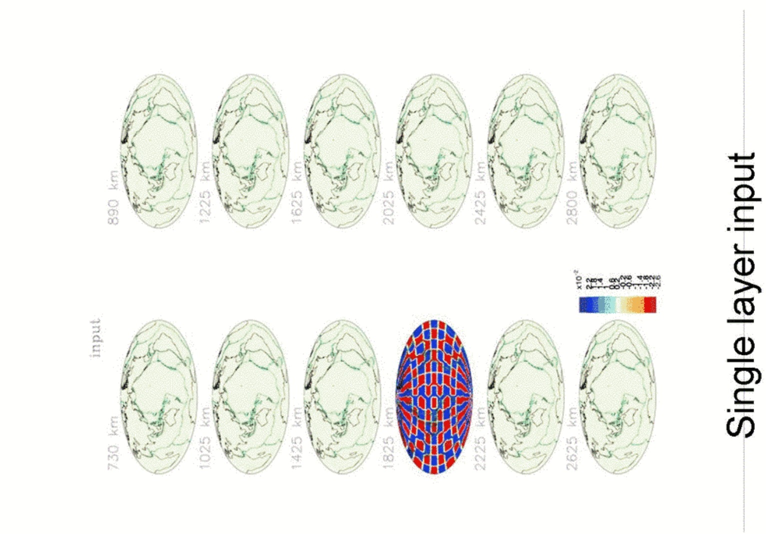
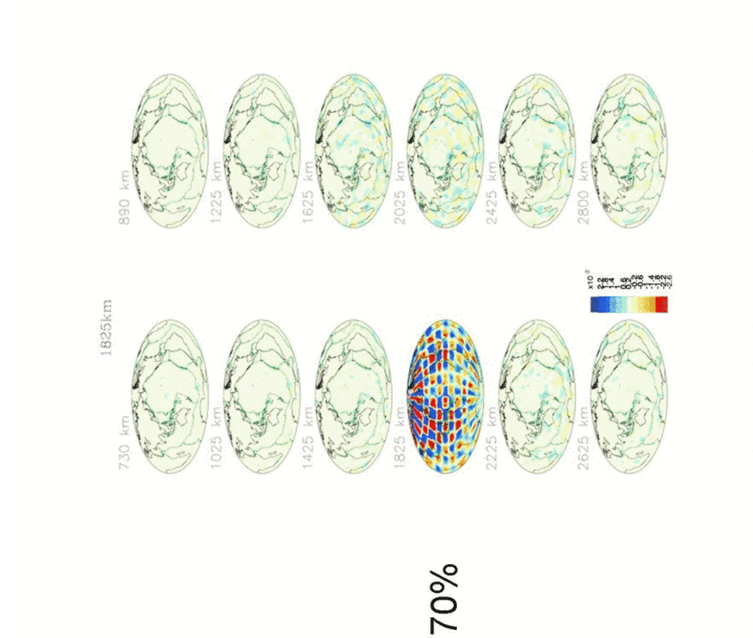
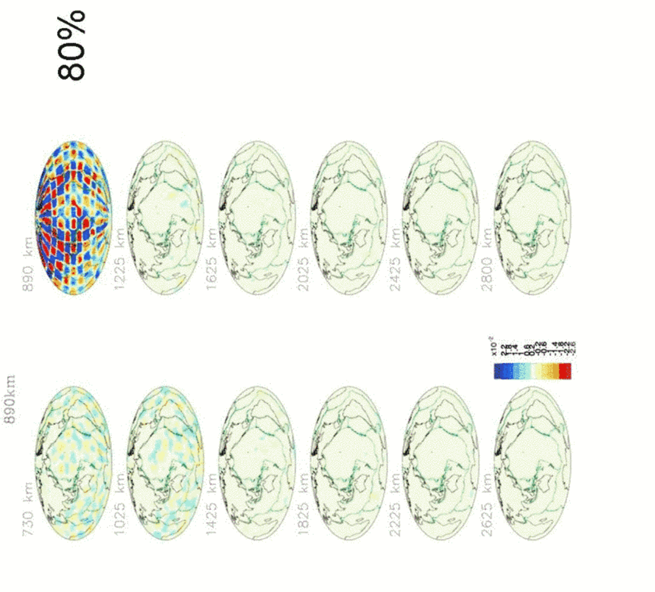


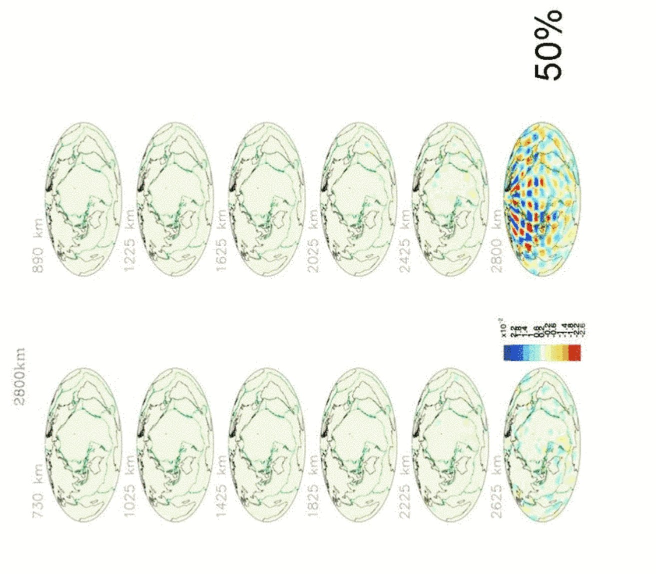
Figure 11. A graphical representation of the matrix connecting shear velocity perturbations to absolute and differential shear travel time data sets. Darker colors represent better sampling. Note that some parts of the lowermost mantle in the southern hemisphere remain poorly sampled.

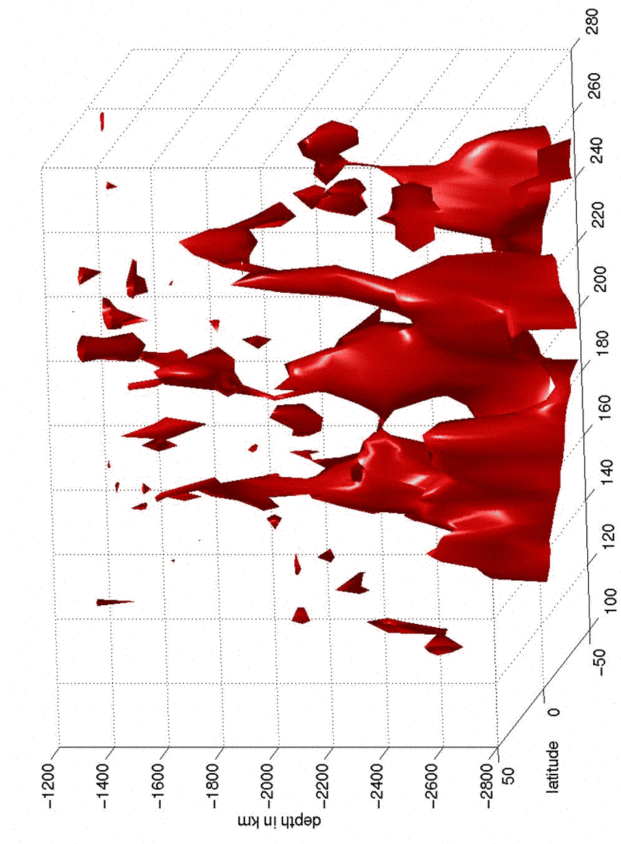
Checkerboard tests


Single layer input
input
730 km
1025 km
1425 km
1825 km
2225 km
2625 km
890 km
1225 km
1625 km
2025 km
2425 km
2800 km


70%
1825km
730 km
1025 km
1425 km
1825 km
2225 km
2625 km
890 km
1225 km
1625 km
2025 km
2425 km
2800 km

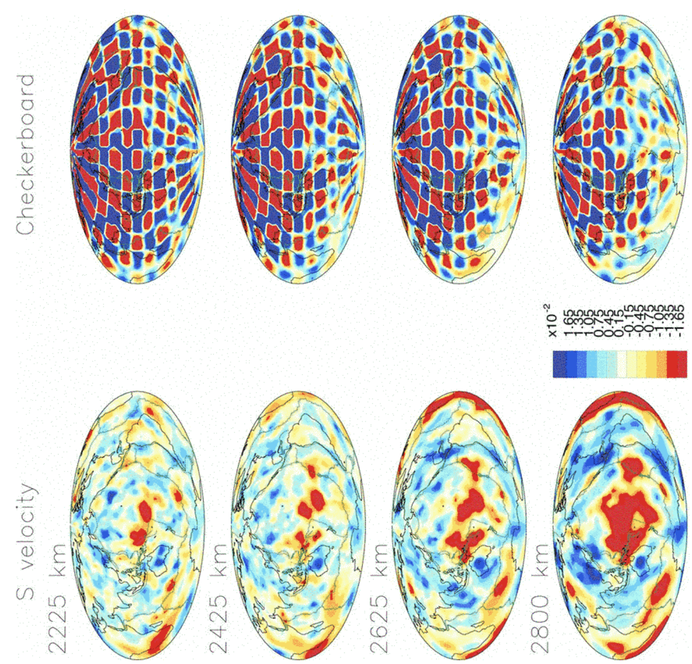

80%
890 km
1225 km
1625 km
2025 km
2425 km
2800 km
890km
730 km
1025 km
1425 km
1825 km
2225 km
2625 km

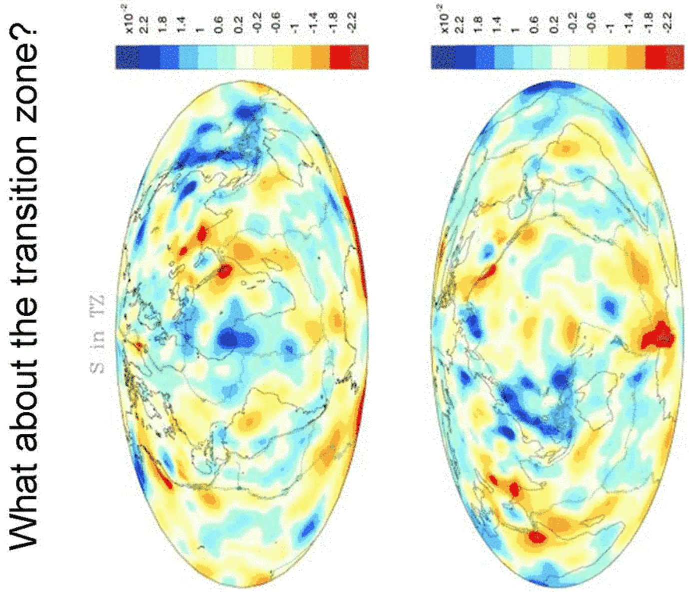

50%
890 km
1225 km
1625 km
2025 km
2425 km
2800 km
2800km
730 km
1025 km
1425 km
1825 km
2225 km
2625 km

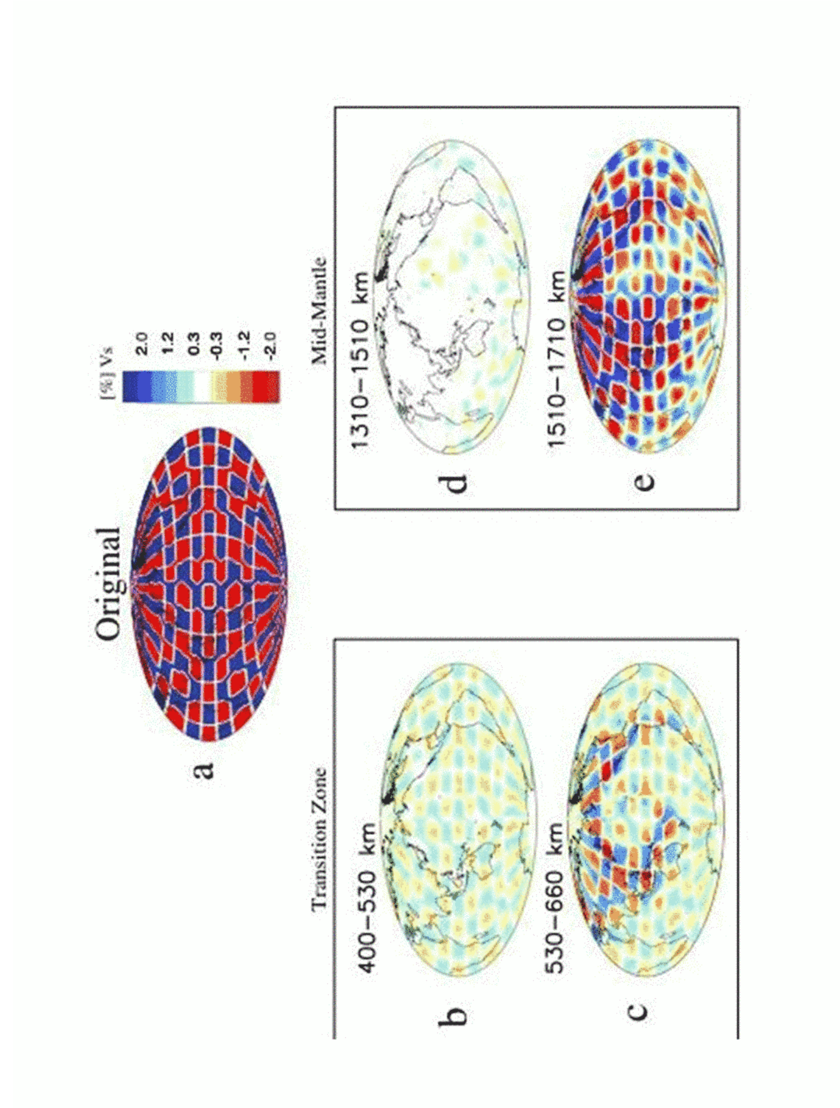


South Pacific Superplume??

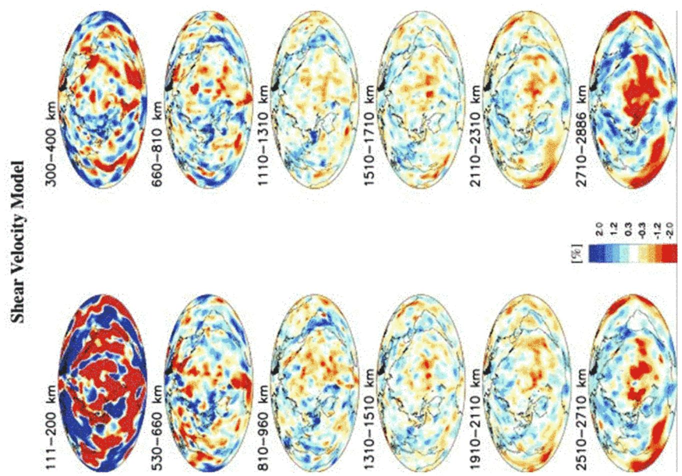
Improves resolution of structure near the base of the mantle




What about the transition zone?
S in TZ


Original
a
[%] Vs
2.0
1.2
0.3
-0.3
-1.2
-2.0
Transition Zone
400–530 km
b
530–660 km
c
Mid-Mantle
1310–1510 km
d
1510–1710 km
e


Errors from Monte-Carlo procedure
Standard Deviation CA
111–200 km
300–400 km
400–530 km
660–810 km
810–960 km
1110–1310 km
1310–1510 km
1510–1710 km
1910–2110 km
2110–2310 km
2510–2710 km
2710–2886 km


Shear Velocity Model
111–200 km
300–400 km
530–660 km
660–810 km
810–960 km
1110–1310 km
1310–1510 km
1510–1710 km
1910–2110 km
2110–2310 km
2510–2710 km
2710–2886 km
[%]

Final comments

- Choice of final model is still somewhat subjective (would be less so if we really knew the errors on our data).
- Your ability to resolve structure depends mainly on data converage. Smoothing is still necessary -- leads to tradeoff between resolution and error
- You get to investigate this for yourself tomorrow!

Beware the tomography police!