Constructing 3D Sensitivity Kernels and Working Towards 3D Tomographic Inversions Based upon Adjoint Methods

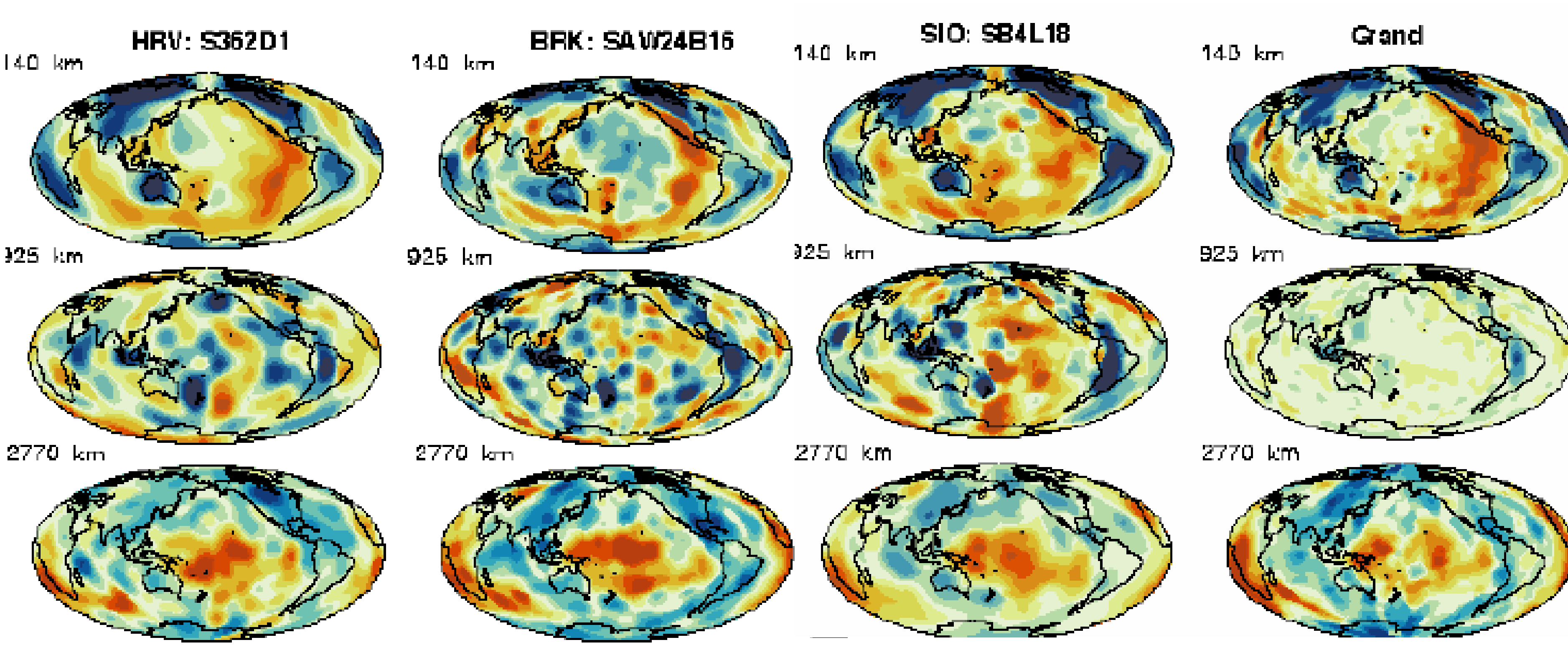
Qinya Liu

Jeroen Tromp, Carl Tape, Alessia Maggi

Urska Manners, Guy Masters

Oct 2007, CIG / SPICE / IRIS Workshop

Seismic Tomography



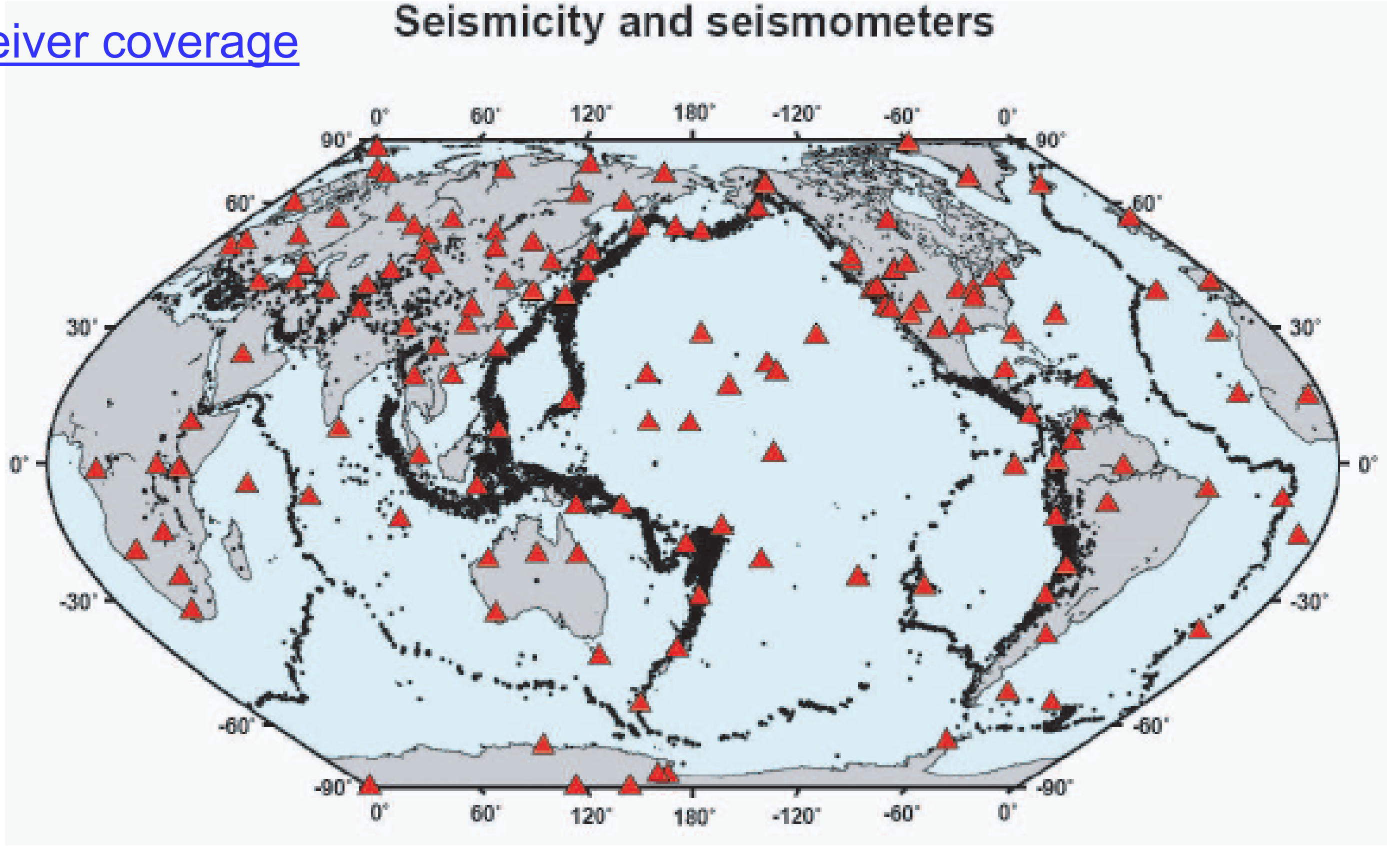
http://mahi.ucsd.edu/Gabi/rem2.dir/shear-models.html

Visual comparison of existing S-wave tomography images

resolution, Resolution, RESOLUTION !!!

Limitations of Current Inversions

- Source and receiver coverage

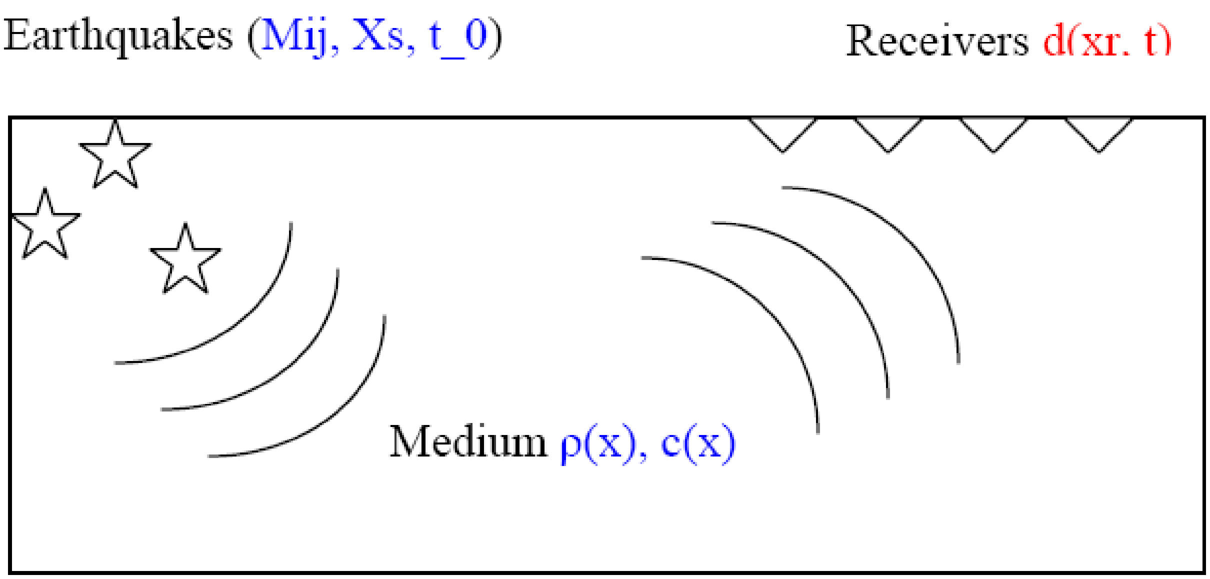


- Theory itself

-Forward problem

-Inverse problem

Forward Problem -- The Wave Equation!



$$\rho\, \partial_t^2 \mathbf{s} - \nabla \cdot \mathbf{T} = \mathbf{f},$$

$$\mathbf{T} = \mathbf{c} : \nabla \mathbf{s},$$

B.C.

$$\hat{\mathbf{n}} \cdot \mathbf{T} = 0 \quad \text{on } \partial\Omega,$$

I.C.

$$\mathbf{s}(\mathbf{x}, 0) = 0, \qquad \partial_t \mathbf{s}(\mathbf{x}, 0) = 0.$$

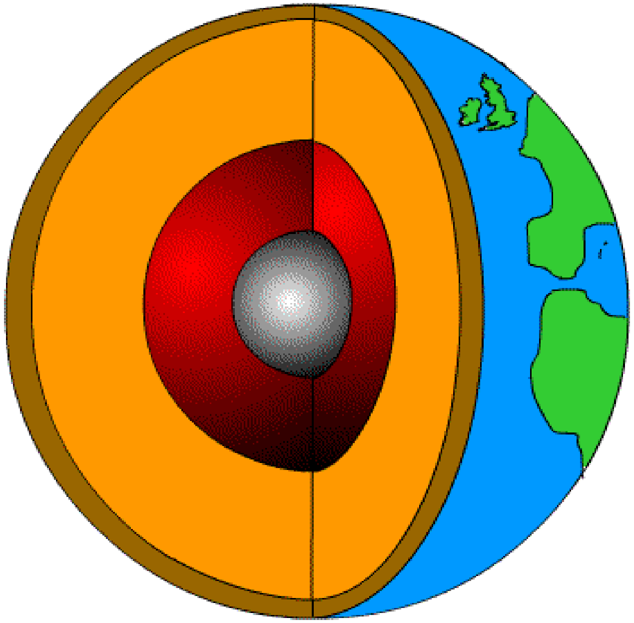
Important components:

- Earthquake source
- Earth Model
- Recorded Seismograms

Solving the Wave Equation

Complications

- Vp(r), Vs (r) increase with r
- internal discontinuities
- fluid OC
- 3D lateral heterogeneity ΔVp(r), ΔVs(r)
- Anisotropy, Q, etc



Semi-Analytical Solutions to the WE

For global wave propagation:

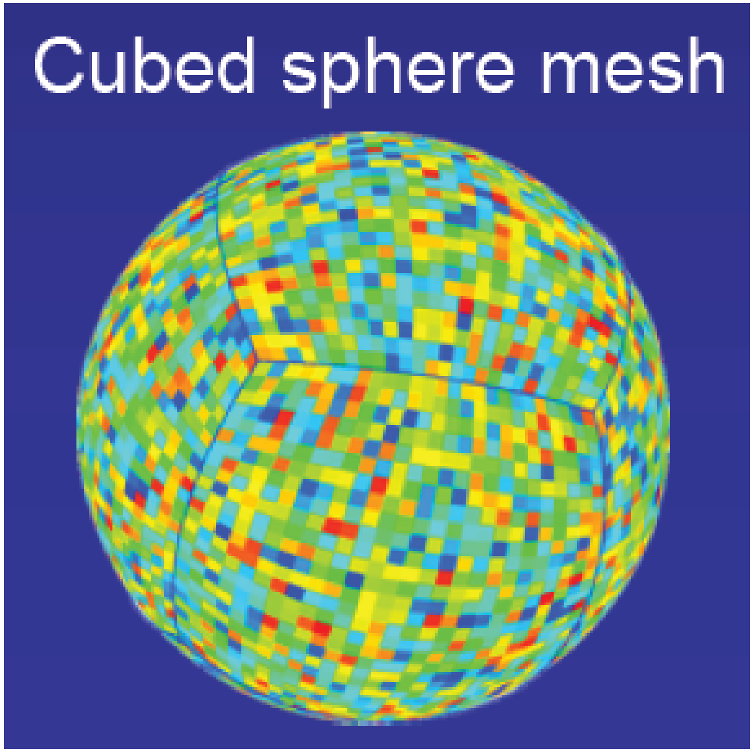
	Body Waves	Surface Waves	Normal Modes
1D	Ray Theory, WKBJ, FK, Generalized RT	Path Average Approximation	For 1-D Earth model Only, practically accurate to 5 seconds and longer
"2D, 3D"	Ray Tracing	Non-linear Asymptotical Theory	

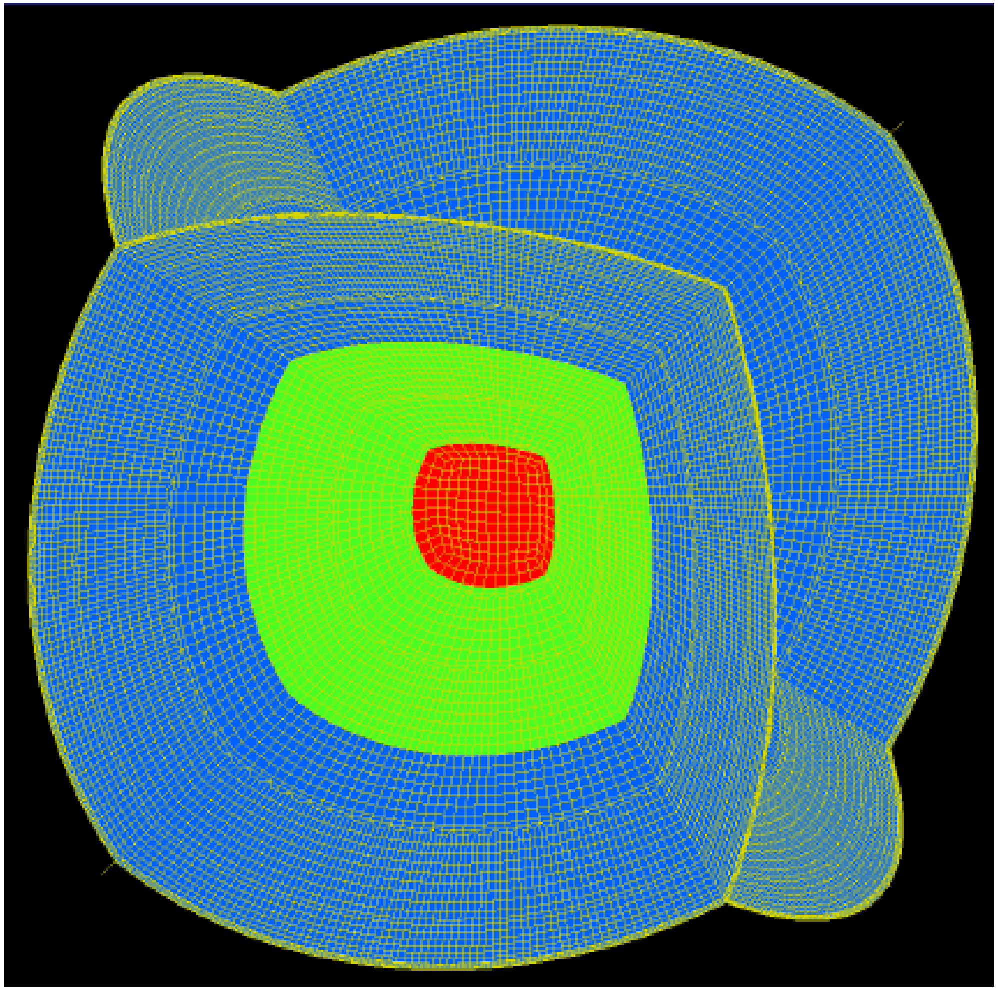
<u>Observables:</u>

- waveforms,
- travel times of body waves, Vp, Vg of surface waves, eigen f's of modes, etc

Numerical Solutions to the WE

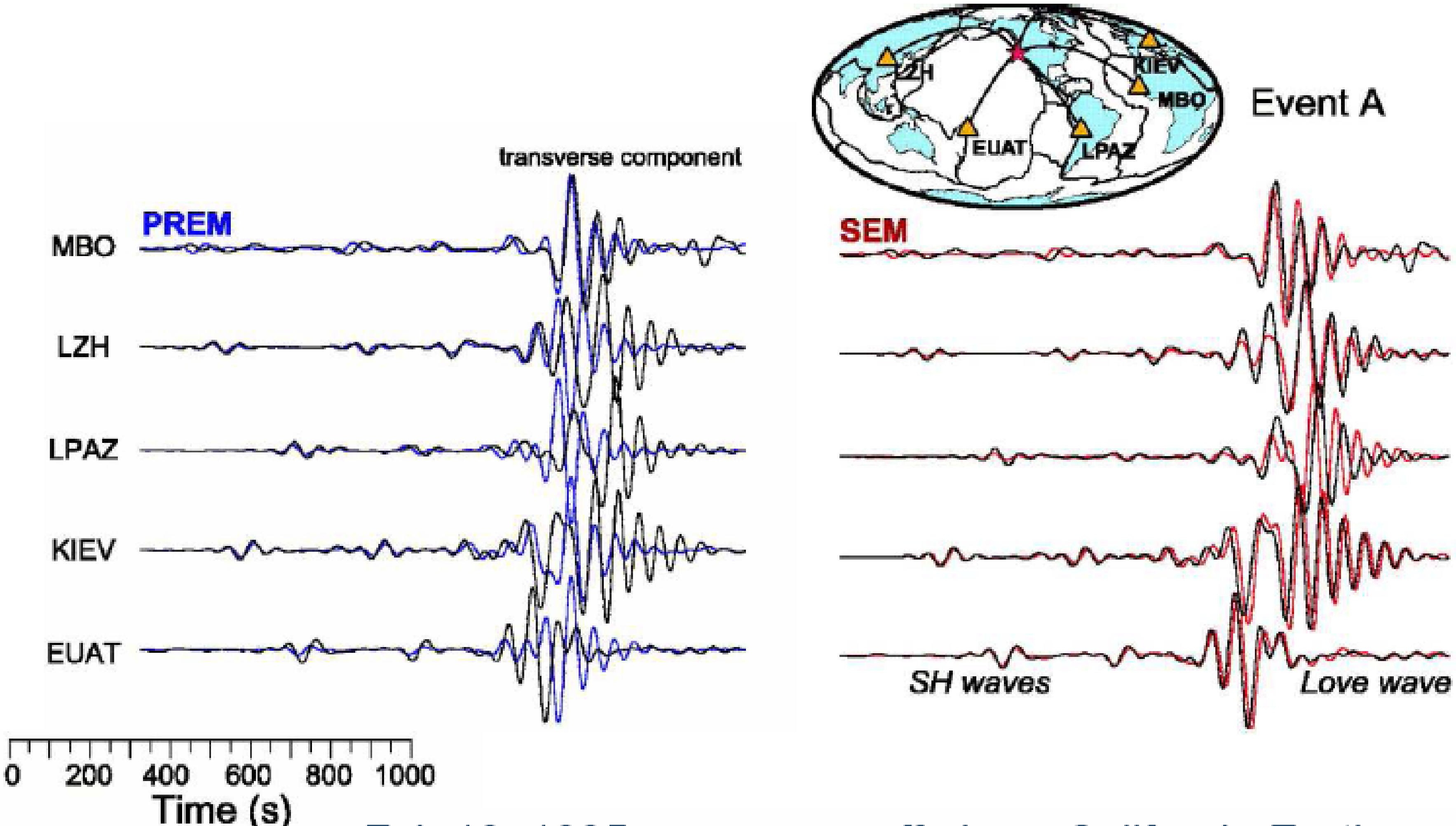
- Finite Difference Method
- Finite Element Method
- Spectral-element Method





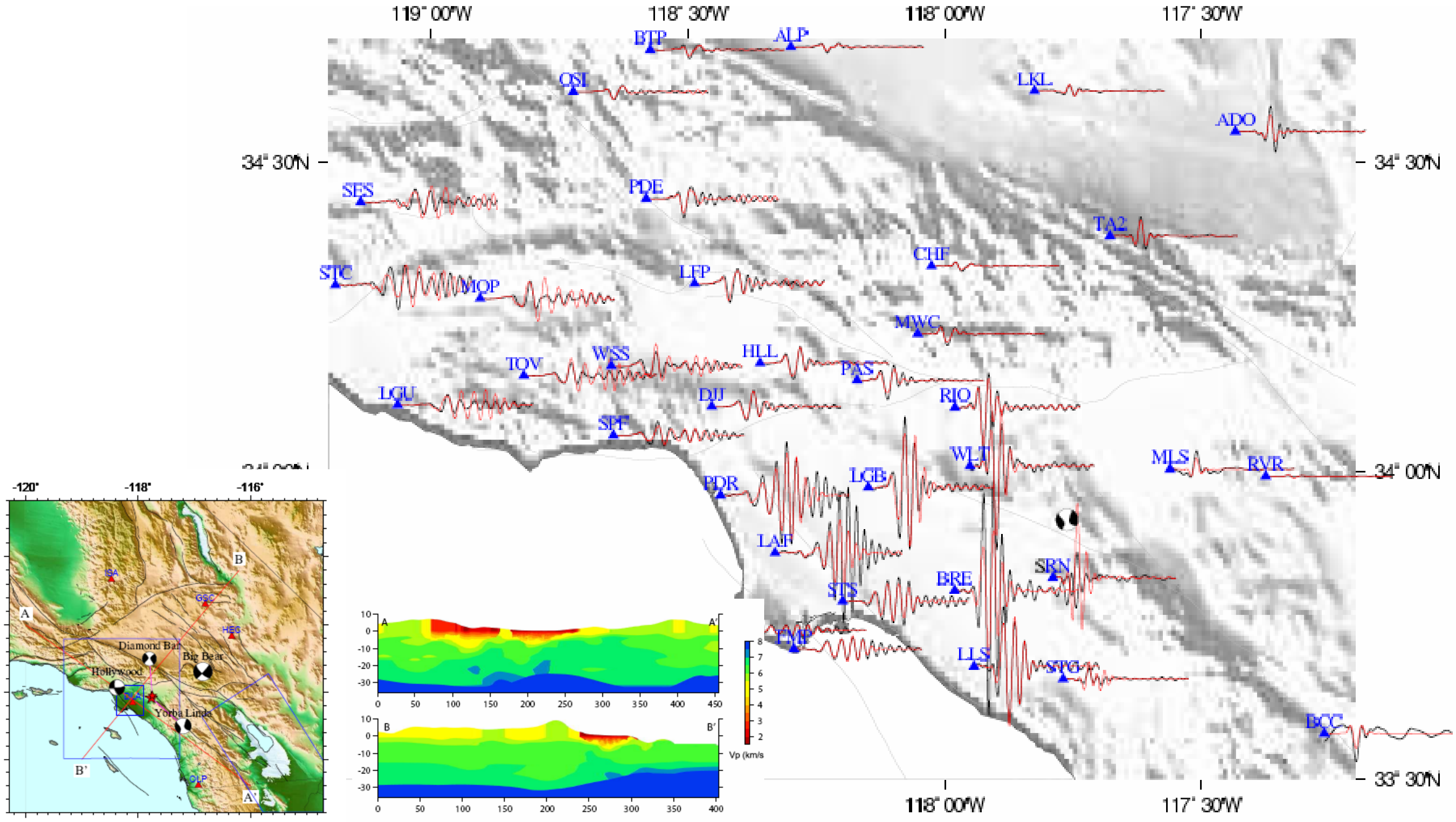
New V4.0 Mesh (Michea & Komatitsch)

Application of SEM at global scale



Feb 19, 1995, $M_w = 6.6$, off-shore California Earthquake
transvserse components, 50 seconds and longer
Komatitsch et al. (2002)

Application of SEM to 3D SC model



Sep 3, 2002 $M_w = 4.2$ Yorba Linda Earthquake

(Komatitsch et al 2004)

Seismic Inversion Problem

We minimize the difference between the observables for the data $O(\mathbf{d}(\mathbf{x}_r, t))$ and the observables for the synthetics $O(\mathbf{s}(\mathbf{x}_r, t, \mathbf{m}))$ predicted by a model $\mathbf{m}$:

$$\phi = \sum_r [O(\mathbf{s}(\mathbf{x}_r, t, \mathbf{m})) - O(\mathbf{d}(\mathbf{x}_r, t))]^2$$

Linearize at a reference model $\mathbf{m}^0$, and solve the linear system:

$$\frac{\partial^2 \phi}{\partial m_j \partial m_k}|_{\mathbf{m}^0}(m_k - m_k^0) = -\frac{\partial \phi}{\partial m_j}|_{\mathbf{m}^0}$$

where

$$\frac{\partial \phi}{\partial m_j} = \sum_r [O(\mathbf{s}) - O(\mathbf{d})] \frac{\partial O(\mathbf{s})}{\partial m_j}$$ Gradient of the misfit function

$$\frac{\partial^2 \phi}{\partial m_j \partial m_k} = \sum_r \frac{\partial O(\mathbf{s})}{\partial m_j} \frac{\partial O(\mathbf{s})}{\partial m_k} + \sum_r [O_\alpha(\mathbf{s}) - O_\alpha(\mathbf{d})] \frac{\partial^2 O_\alpha}{\partial m_j \partial m_k}$$ Hessian

Frechét Derivatives

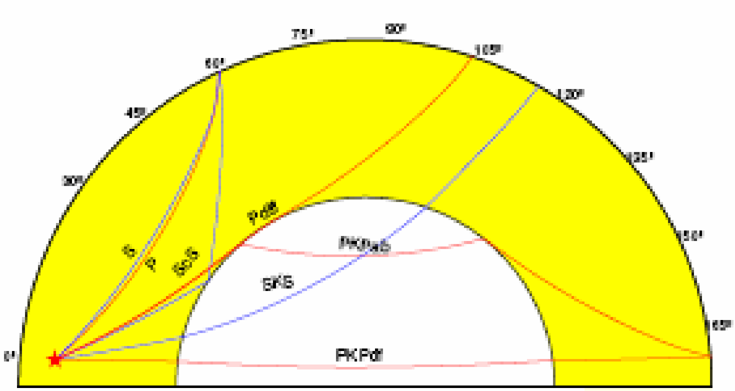
Key ingredients for the seismic inverse problem: $\frac{\partial O(\mathbf{s})}{\partial m_j}$ or $\frac{\partial \mathbf{s}}{\partial m_j}$

- Frechét Derivatives
- Sensitivity kernels, Frechét Kernels, for model $\mathbf{m}(\mathbf{x})$

Examples:

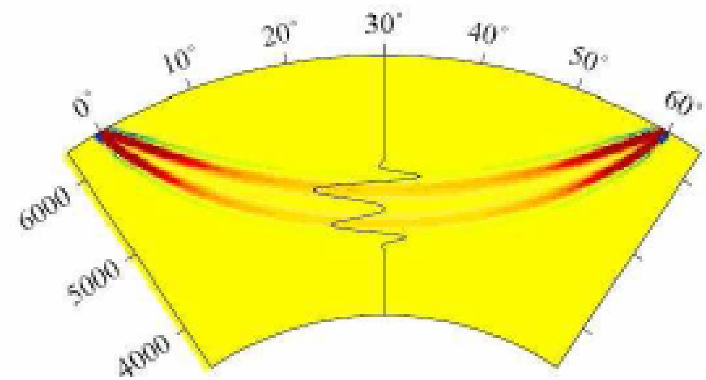
Rays for TT (high-f approximation)

$$\delta T_i = -\int_{Ray_i} \frac{1}{c}\frac{\delta c}{c} dRay$$



Finite-frequency 'banana-dougnut' kernels

$$\delta T_i = \int K^T_{\mathbf{m}}(\mathbf{x}) \frac{\delta \mathbf{m}}{\mathbf{m}} d\mathbf{x}$$



Hung & Dahlen (2000)

Gradient of the misfit function – Revisited

Again, the misfit function:

$$\phi = \sum_r [O(\mathbf{s}(\mathbf{x}_r, t, \mathbf{m})) - O(\mathbf{d}(\mathbf{x}_r, t))]^2$$

where $\mathbf{s}(\mathbf{x}, t)$ is subject to the wave equation:

$$\rho \ddot{\mathbf{s}} = \nabla \cdot (\mathbf{c} : \nabla \mathbf{s}) + \mathbf{f} \quad \text{in } V$$

$$\mathbf{s}(\mathbf{x}, 0) = \dot{\mathbf{s}}(\mathbf{x}, 0) = 0$$

$$\hat{n} \cdot (\mathbf{c} : \nabla \mathbf{s}) = 0 \quad \text{on } \Omega$$

We define the action with Lagrange multiplier as:

$$L(\mathbf{s}, \rho, \mathbf{c}) = \phi - \int_V \int_t \lambda(\mathbf{x}, t) \cdot [\rho \ddot{\mathbf{s}} - \nabla \cdot (\mathbf{c} : \nabla \mathbf{s}) - \mathbf{f}] \, d\mathbf{x} dt$$

Now we take the variation of the Lagrangian:

$$\delta L = \int \int \left\{ [***] \, \delta \ln \rho(\mathbf{x}) + \int \int [***] \, \delta \ln \mathbf{c}(\mathbf{x}) + \int \int [***] \, \delta \mathbf{s}(\mathbf{x}, t) \right\} d\mathbf{x} dt$$

(Liu & Tromp 2006)

Variation of the Lagrangian

- $\delta \mathbf{s}(\mathbf{x}, t)$ term vanishes – Lagrange multiplier $\lambda(\mathbf{x}, t)$ satisfies:

$$\rho\, \ddot{\lambda} = \nabla \cdot (\mathbf{c} : \nabla \lambda) + \sum_r [O(\mathbf{s}(\mathbf{x}, t, \mathbf{m})) - O(\mathbf{d}(\mathbf{x}, t))] \frac{\partial O}{\partial \mathbf{s}}(t) \delta(\mathbf{x} - \mathbf{x}_r) \quad \text{in } V$$

$$\lambda(\mathbf{x}, T) = \dot{\lambda}(\mathbf{x}, T) = 0$$

$$\hat{n} \cdot (\mathbf{c} : \nabla \lambda) = 0 \quad \text{on } \Omega$$

Define the adjoint wavefield $\mathbf{s}^{\dagger}(\mathbf{x}, t') = \lambda(\mathbf{x}, T - t)$, which also satisfies the FORWARD wave equation

$$\rho\, \mathbf{s}^{\dagger}_{t't'} = \nabla \cdot (\mathbf{c} : \nabla \mathbf{s}^{\dagger})$$

$$+ \sum_r [O(\mathbf{s}(\mathbf{x}, T - t, \mathbf{m})) - O(\mathbf{d}(\mathbf{x}, T - t))] \frac{\partial O}{\partial \mathbf{s}} (T - t) \delta(\mathbf{x} - \mathbf{x}_r) \quad \text{in } V$$

$$\mathbf{s}^{\dagger}(\mathbf{x}, 0) = \mathbf{s}^{\dagger}_{t'}(\mathbf{x}, 0) = 0$$

$$\hat{n} \cdot (\mathbf{c} : \nabla \mathbf{s}^{\dagger}) = 0 \quad \text{on } \Omega$$

Gradient of the misfit function – continued

Then the variation of the action becomes

$$\delta L = \int (K_\rho \delta \ln \rho + K_{\mathbf{c}} \delta \ln \mathbf{c})\, d\mathbf{x}$$

where

$$\begin{aligned} K_\rho(\mathbf{x}) &= -\int_t \ddot{\mathbf{s}}(\mathbf{x},t) \cdot \mathbf{s}^\dagger(\mathbf{x}, T-t)\, dt \\ K_{\mathbf{c}}(\mathbf{x}) &= -\int_t \nabla \mathbf{s}(\mathbf{x},t)\, \nabla \mathbf{s}^\dagger(\mathbf{x}, T-t)\, dt \end{aligned}$$

- Above kernel expressions do not change when different Observables O's are minmized
- Only the adjoint sources that are used to generate the adjoint field are different: $\sum_r [O(\mathbf{s}(\mathbf{x}, T-t, \mathbf{m})) - O(\mathbf{d}(\mathbf{x}, T-t))] \frac{\partial O}{\partial \mathbf{s}} (T-t) \delta(\mathbf{x} - \mathbf{x}_r)$
- Formally, only $2 * N_{events}$ simulations are needed to compute the gradient of the misfit function

Special cases

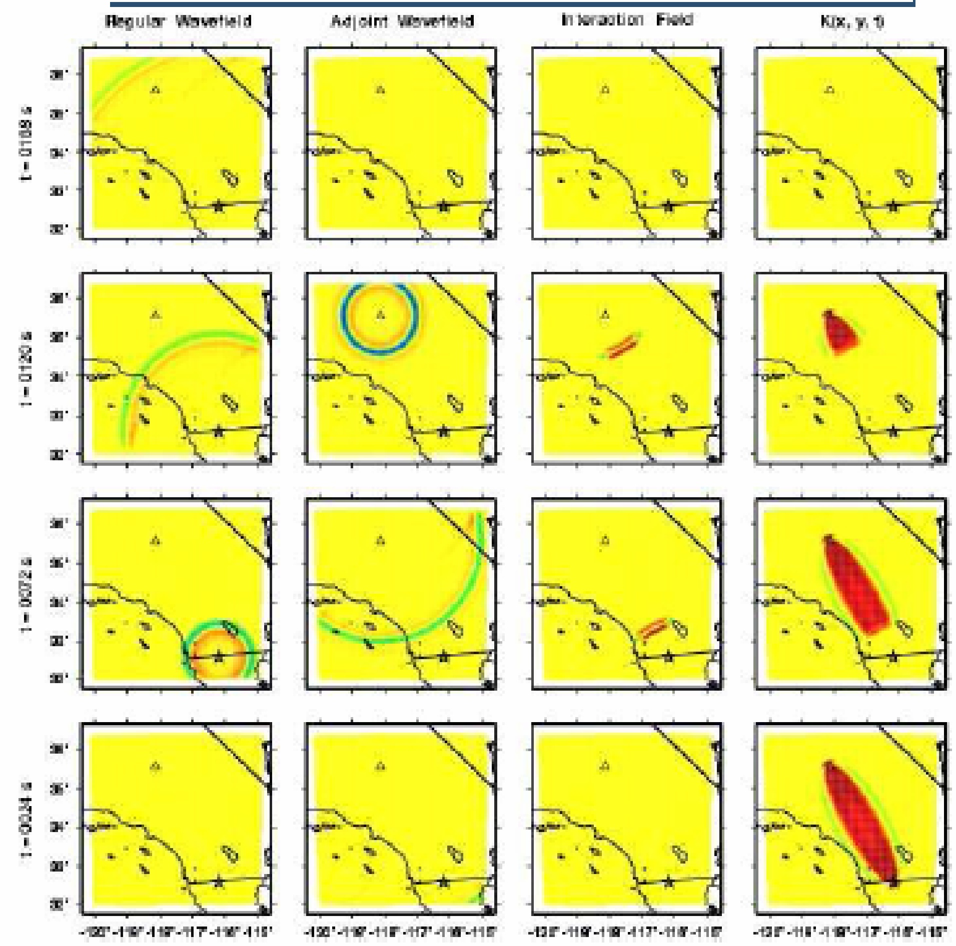
Observables	Adjoint sources	Kernel Names
Finite-frequency travel-time for a single source receiver pair	$\phi = T_r(\mathrm{s}), \quad \frac{\dot{s}}{\int \dot{s}^2 dt}$	Banana-Doughnut kernels
FF TT measurements for multiple source receiver pairs	$\sum_r [T_r(\mathrm{s}) - T_r(\mathrm{d})] \frac{\dot{s}}{\int \dot{s}^2 dt}$	Gradient of the travel-time misfit function
Waveform difference between multiple source receiver pairs	$\sum_r [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)]$	Gradient of the waveform misfit function

Same number of simulation (2) for generating

- one-source-one-receiver kernel
- one-source-multiple-receiver kernels

2-D sensitivity kernel Construction Example

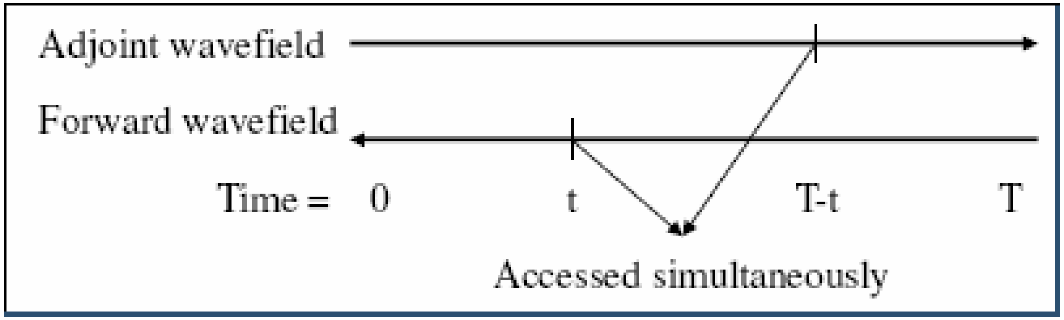
$$K(\mathbf{x}) = \int_0^T \ddot{\mathbf{s}}(\mathbf{x}, t) : \mathbf{s}^{\dagger}(\mathbf{x}, T - t)\, dt$$



(Tape et al 2006)

Numerical Implementations

- Simultaneous access to the forward wave field and the adjoint wave field at opposite times: $\mathbf{s}(\mathbf{x},t)$ and $\mathbf{s}^\dagger(\mathbf{x},T-t)$



  - Save the whole forward wave field $\mathbf{s}(\mathbf{x},t)$
  -

$$\rho\,\partial_t^2\mathbf{s} = \nabla\cdot(\mathbf{c}:\nabla\mathbf{s}) + \mathbf{f} \quad \text{in } V,$$

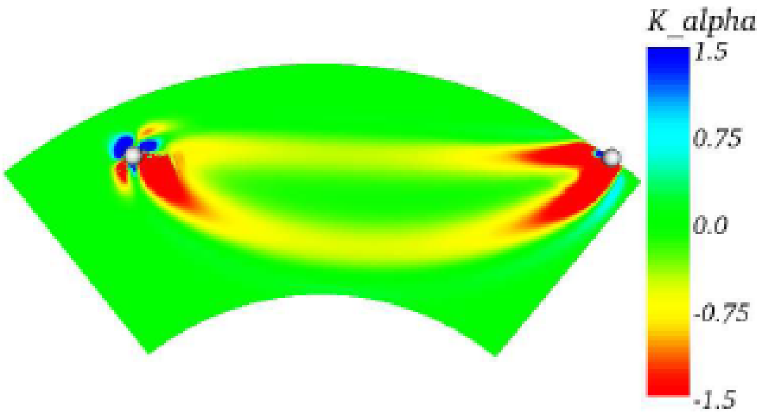
$$\mathbf{s}(\mathbf{x},T) \quad \text{and} \quad \partial_t\mathbf{s}(\mathbf{x},T) \quad \text{given}$$

$$\hat{\mathbf{n}}\cdot(\mathbf{c}:\nabla\mathbf{s}) = \mathbf{0} \quad \text{on } \Omega$$

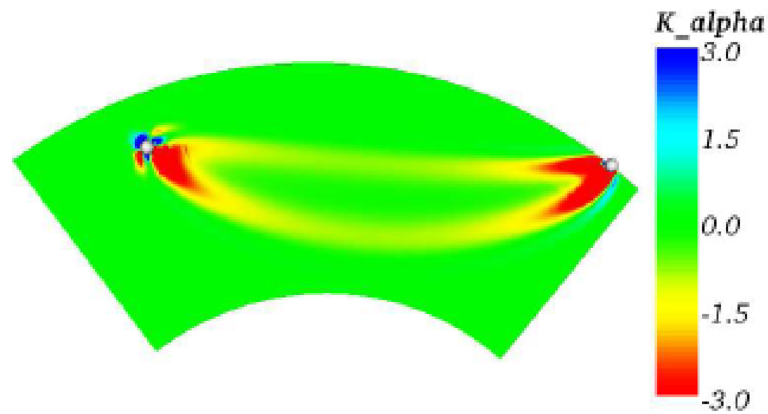
One regular forward simulation for $\mathbf{s}(\mathbf{x},t)$, saving the wave field at the last time step; One adjoint + backward simulation to carry both $\mathbf{s}(\mathbf{x},T-t)$ and $\mathbf{s}^\dagger(\mathbf{x},t)$ to construct the kernel on the fly.

Global sensitivity kernels – P phase

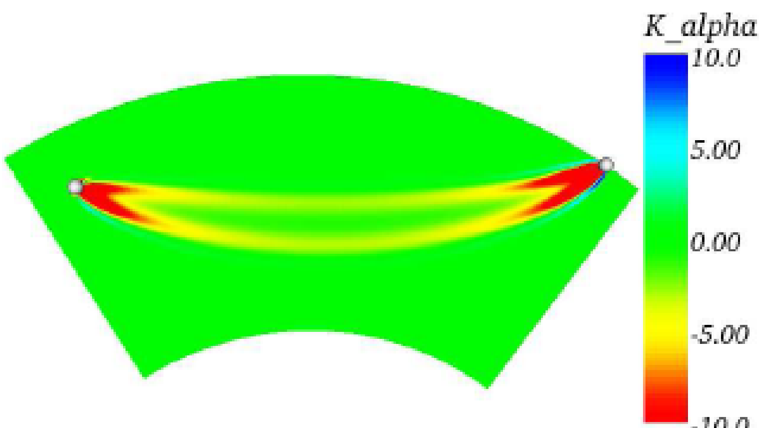
Jun 9, 1994 $M_w = 8.1$ Bolivian Earthquake at depth = 647 km



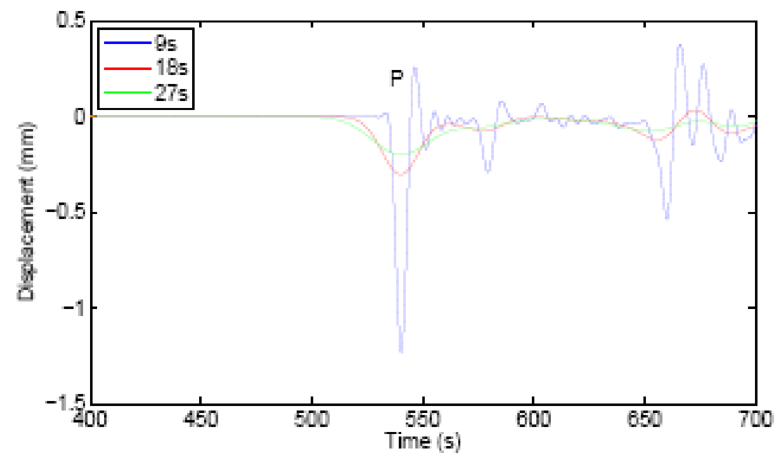
27s



18s

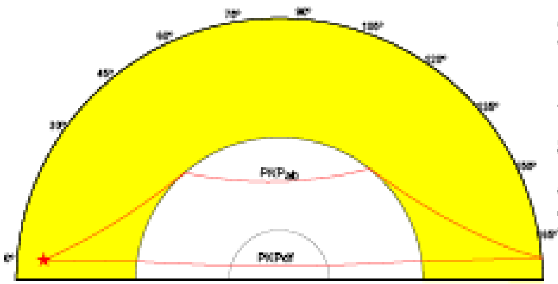


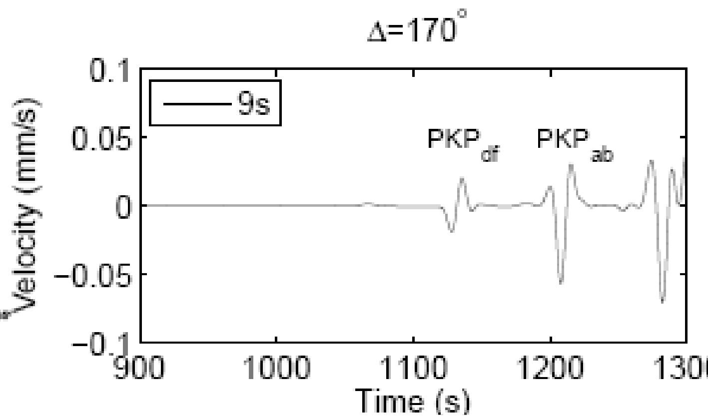
9s

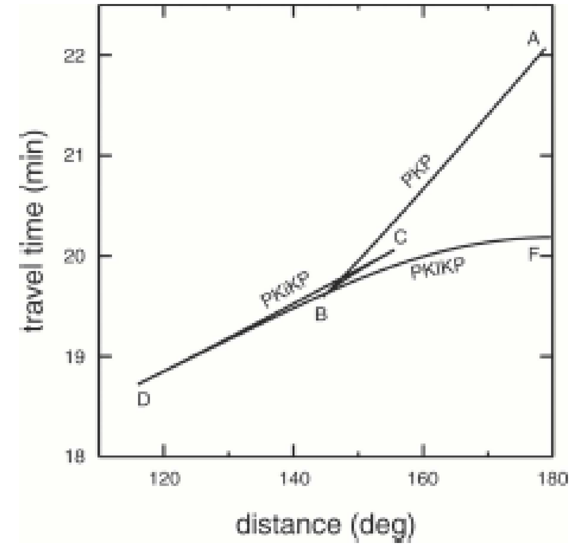


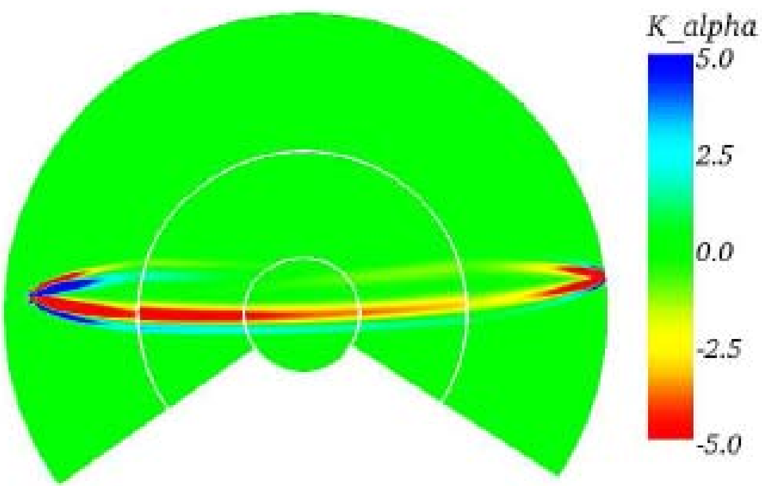
(Liu & Tromp 2007)

PKPab and PKPdf kernels

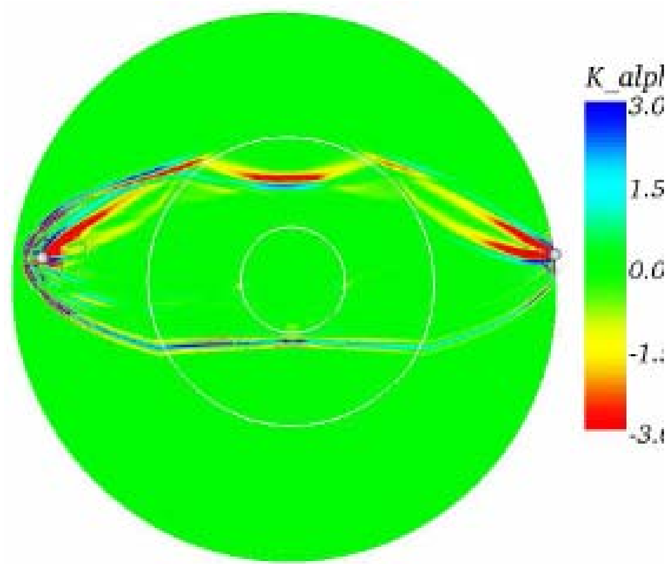






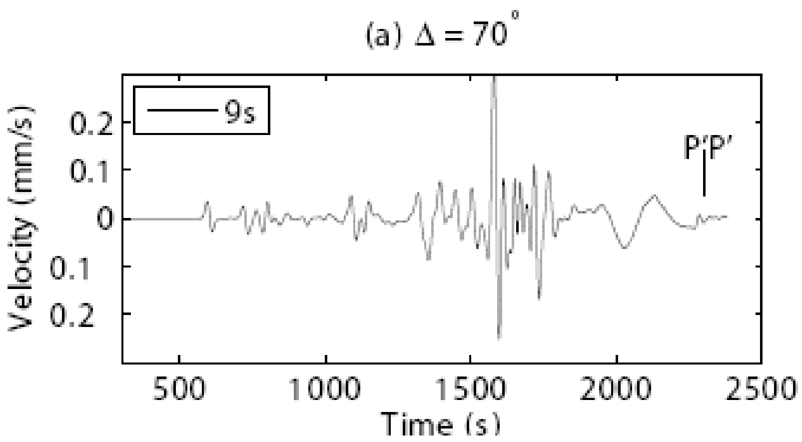


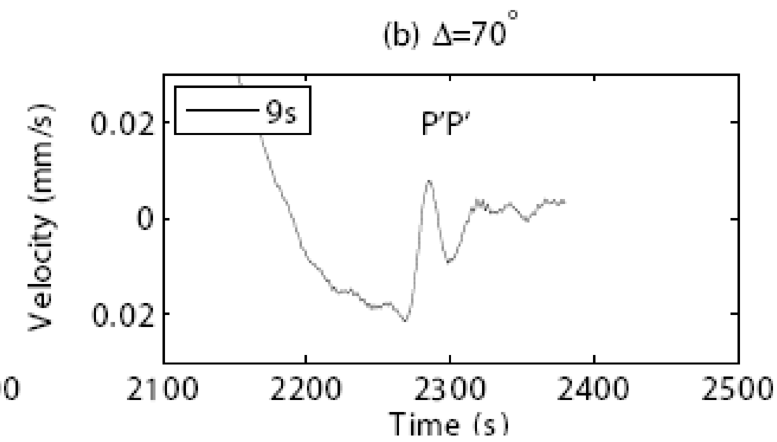
PKPdf kernel at 9s

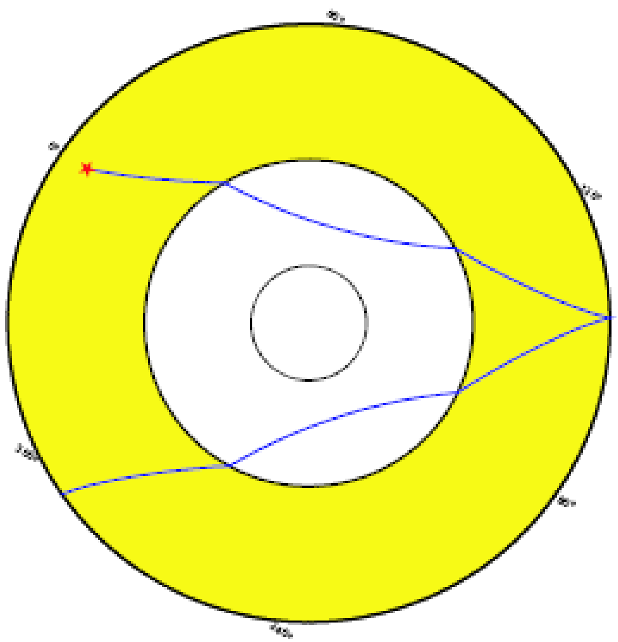


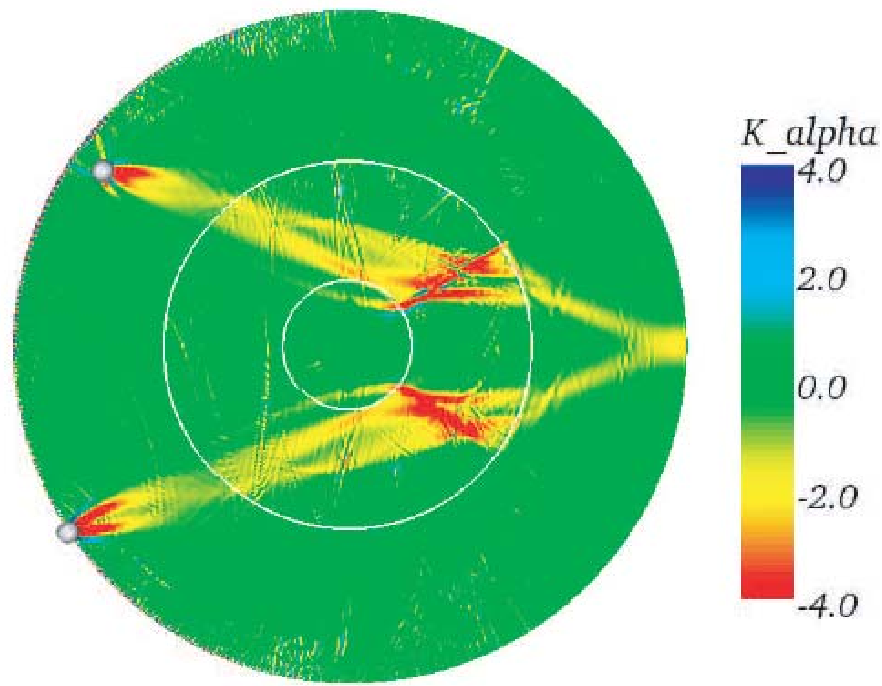
PKPab kernel at 9s

PKPPKP (P'P') kernels



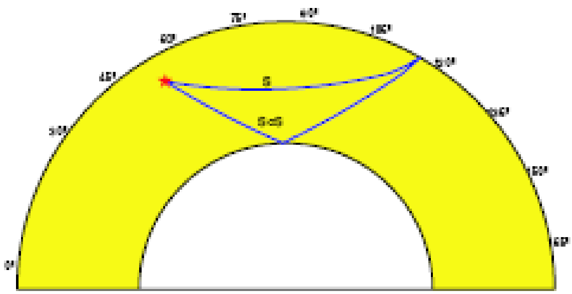


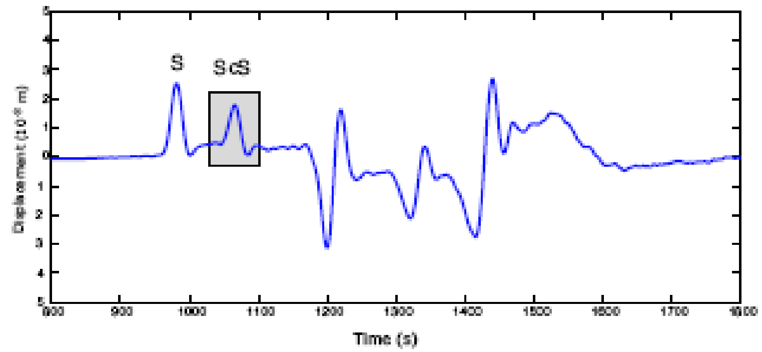




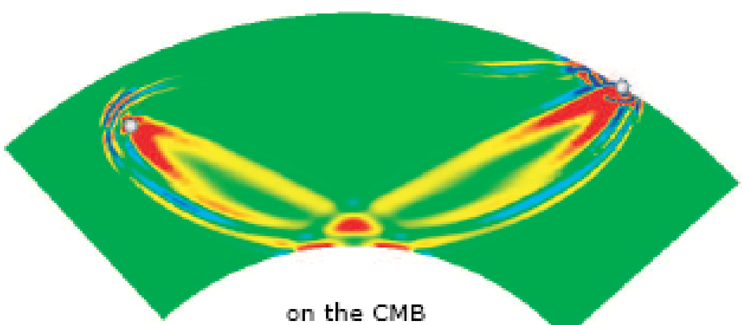
Boundary kernels – ScS at CMB

ScS at $\Delta = 60^{\circ}$



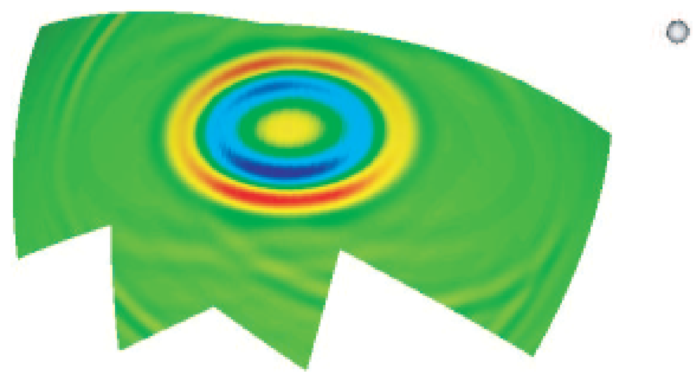


K_{β} kernel on the source-receiver cross-section



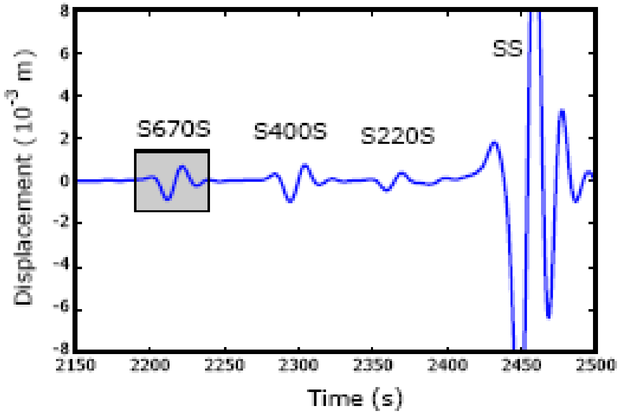
on the CMB

Boundary kernel for topography on CMB

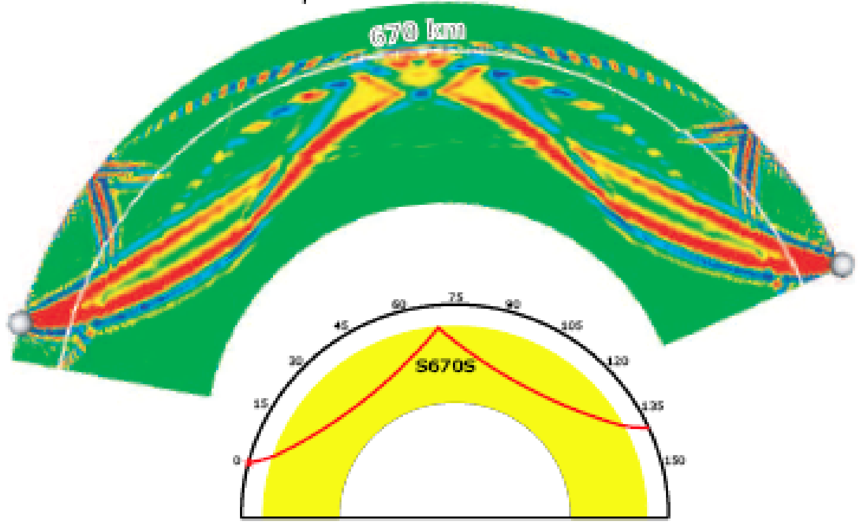


Boundary Kernels – S670S on 670-discontinuity

S670S at Δ = 140°



Kβ sensitivity kernel



Boundary kernel for topography on 670 km discontinuity
(looking down on 670 km)

Traditional Tomography

- Traditional Tomography Inversions

$$\sum_r \frac{\partial O(\mathbf{s})}{\partial m_j}\frac{\partial O(\mathbf{s})}{\partial m_k}(m_k - m_k^0) \quad = \quad -\sum_r [O(\mathbf{s}) - O(\mathbf{d})]\frac{\partial O(\mathbf{s})}{\partial m_j}$$

In practice, invert using iteratively Quasi-Newton scheme:

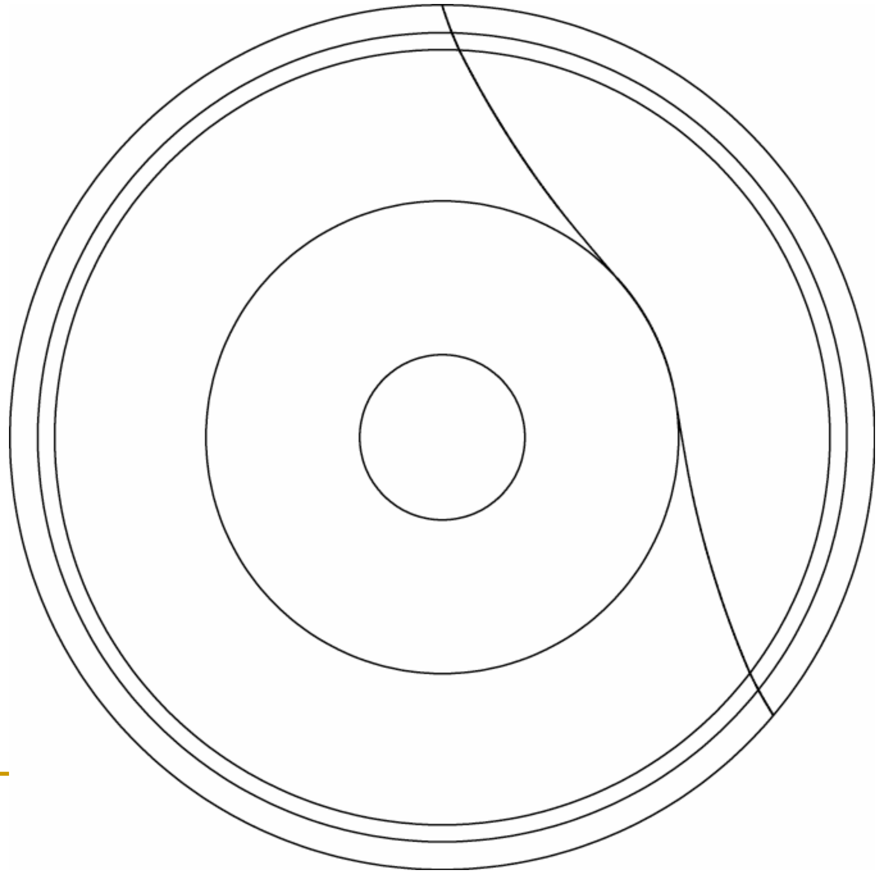
- Pick a reference model $\mathbf{m}^0$
- Compute the gradient and approximate Hessian of the misfit function
- Solve the linear system and update model $\mathbf{m}$
- Iterate using the updated model as reference
- Number of simulations per iteration: $2 * N_{receivers} * N_{events}$ to obtain $\frac{\partial O(\mathbf{s})}{\partial m_j}$

Solutions:

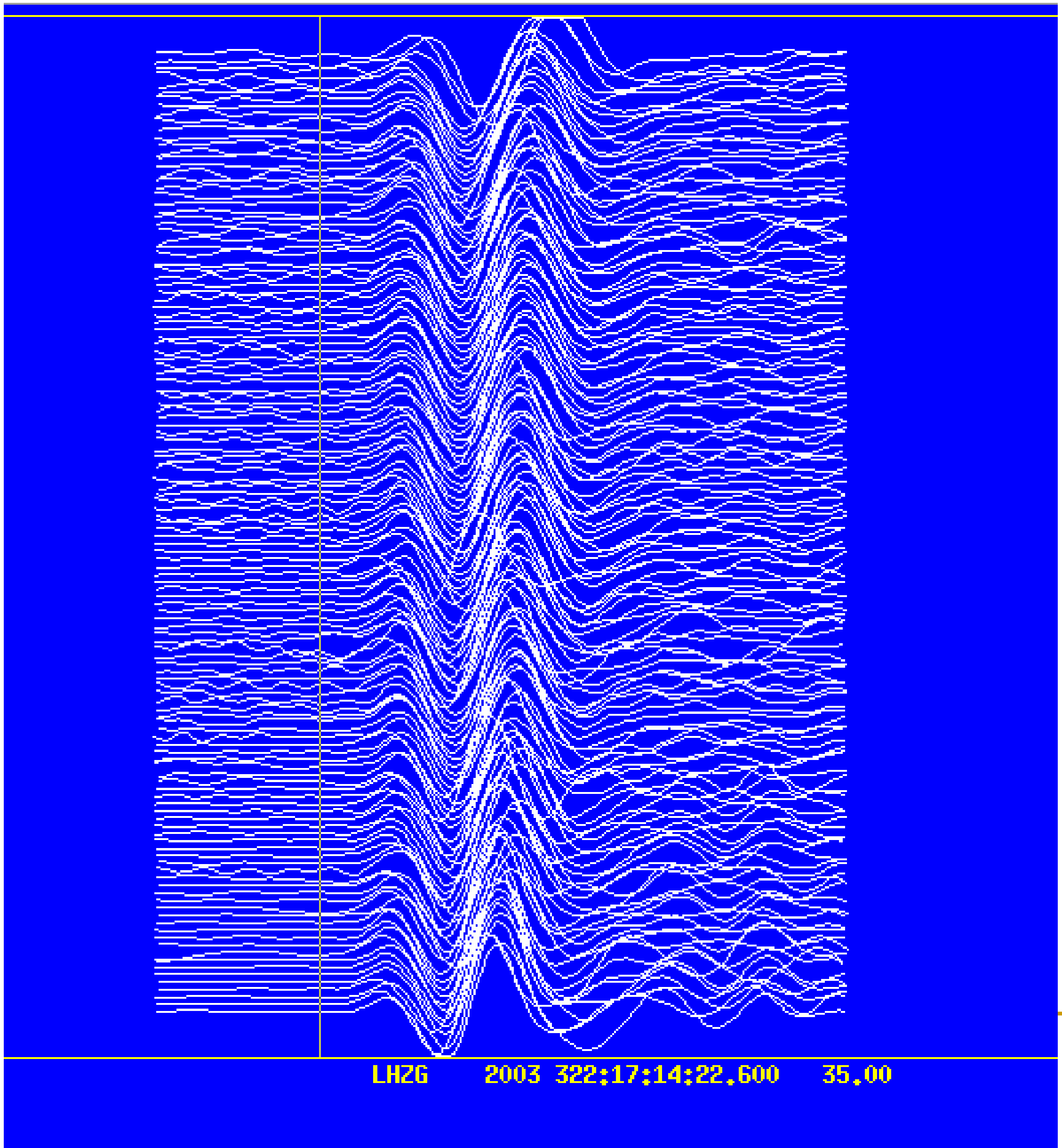
- Use 1D reference model (ray theory, FF banana-doughnut kernels)
- Use source-receiver reciprocity (Scatter Integral Method, USC Group, Chen et al 2006)

Examples: Tomographic Inversions with Diffracted waves

- Diffracted waves travel along base of mantle
- Lots of data
- Coverage of CMB poor in existing models



Traditional Tomography with Diffracted waves



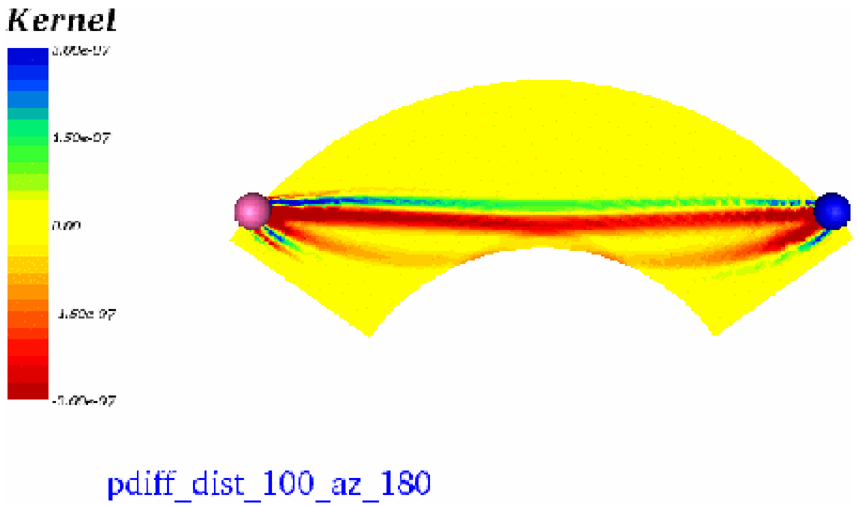
Cluster Analysis

- 20,000 Sdiff
- 31,000 Pdiff
- 1986-2005
- 100° - 170°

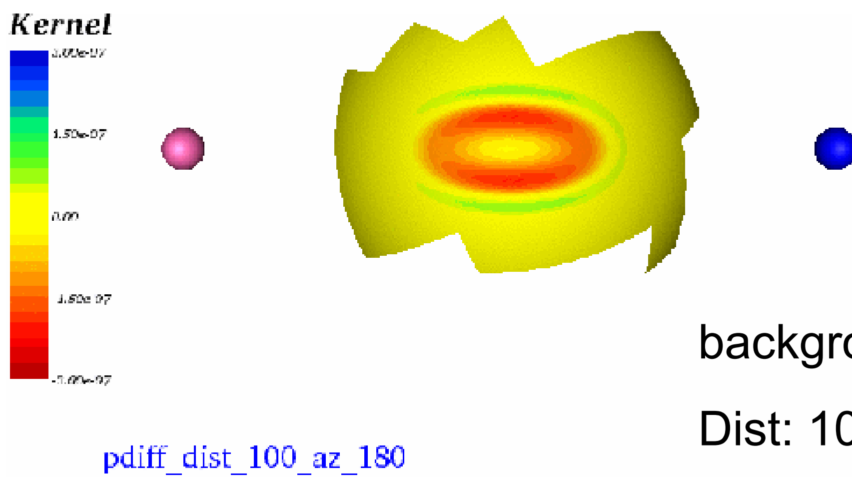
Manners & Masters (2007)

Pdiff kernels (Movie)

source-receiver
cross-section



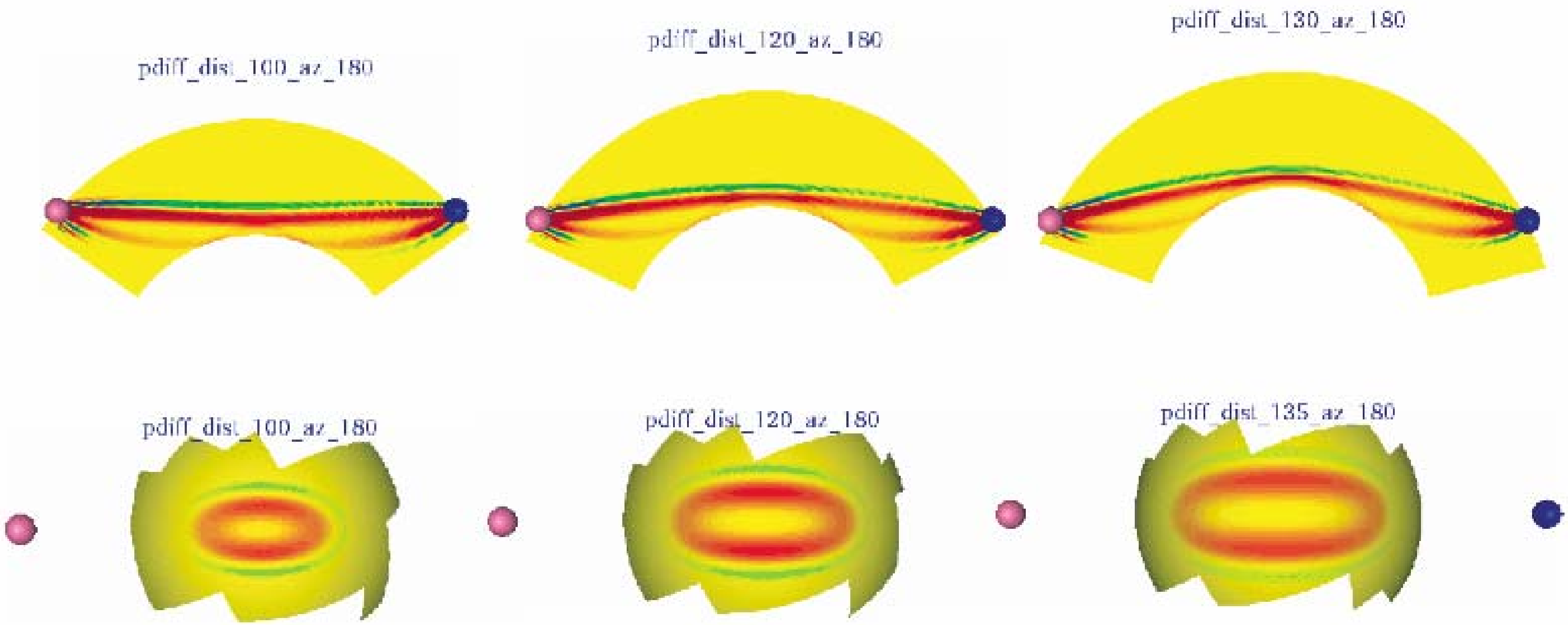
CMB
cross-section



background model: AK135

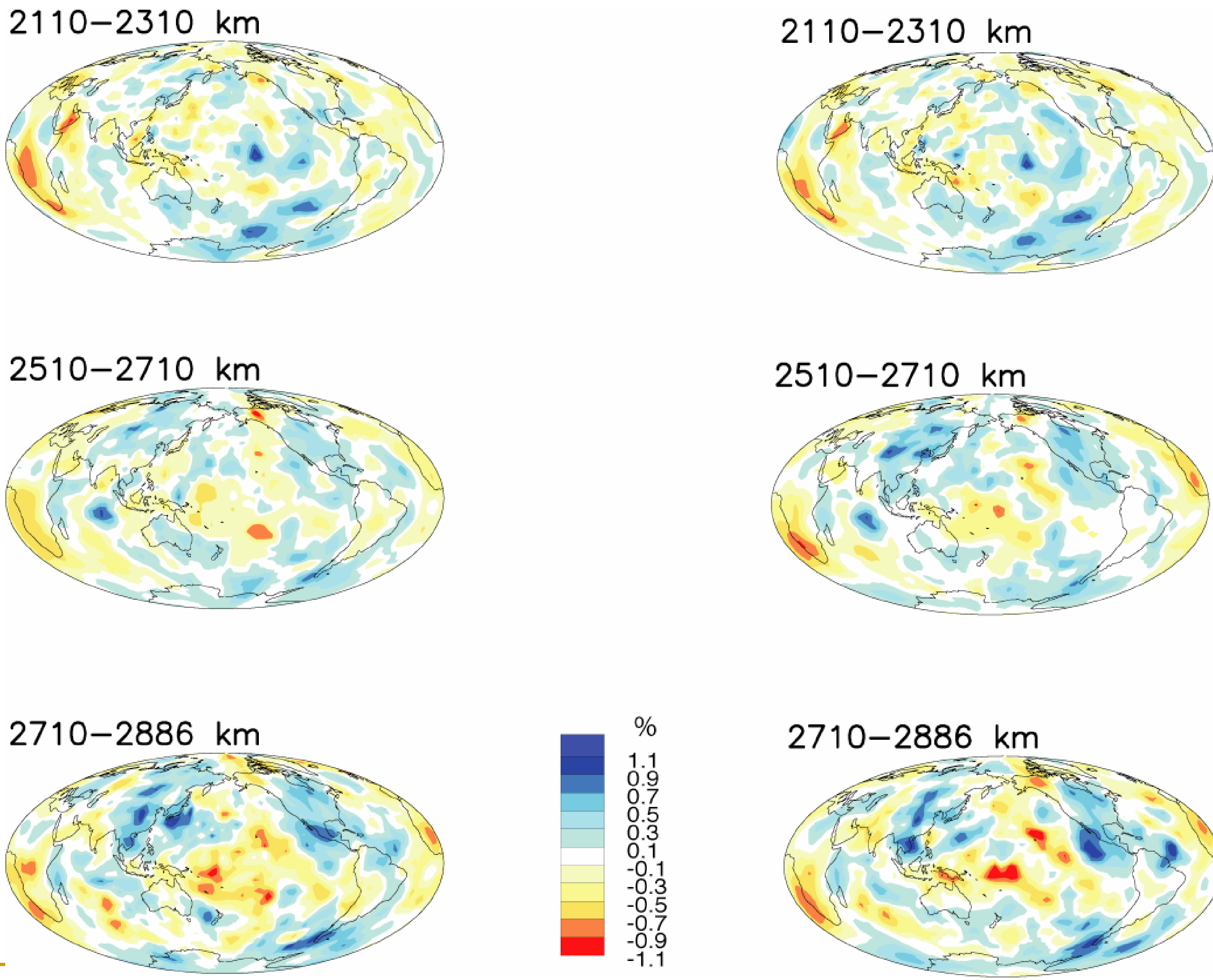
Dist: 100 – 140 degrees

at 18 seconds and longer



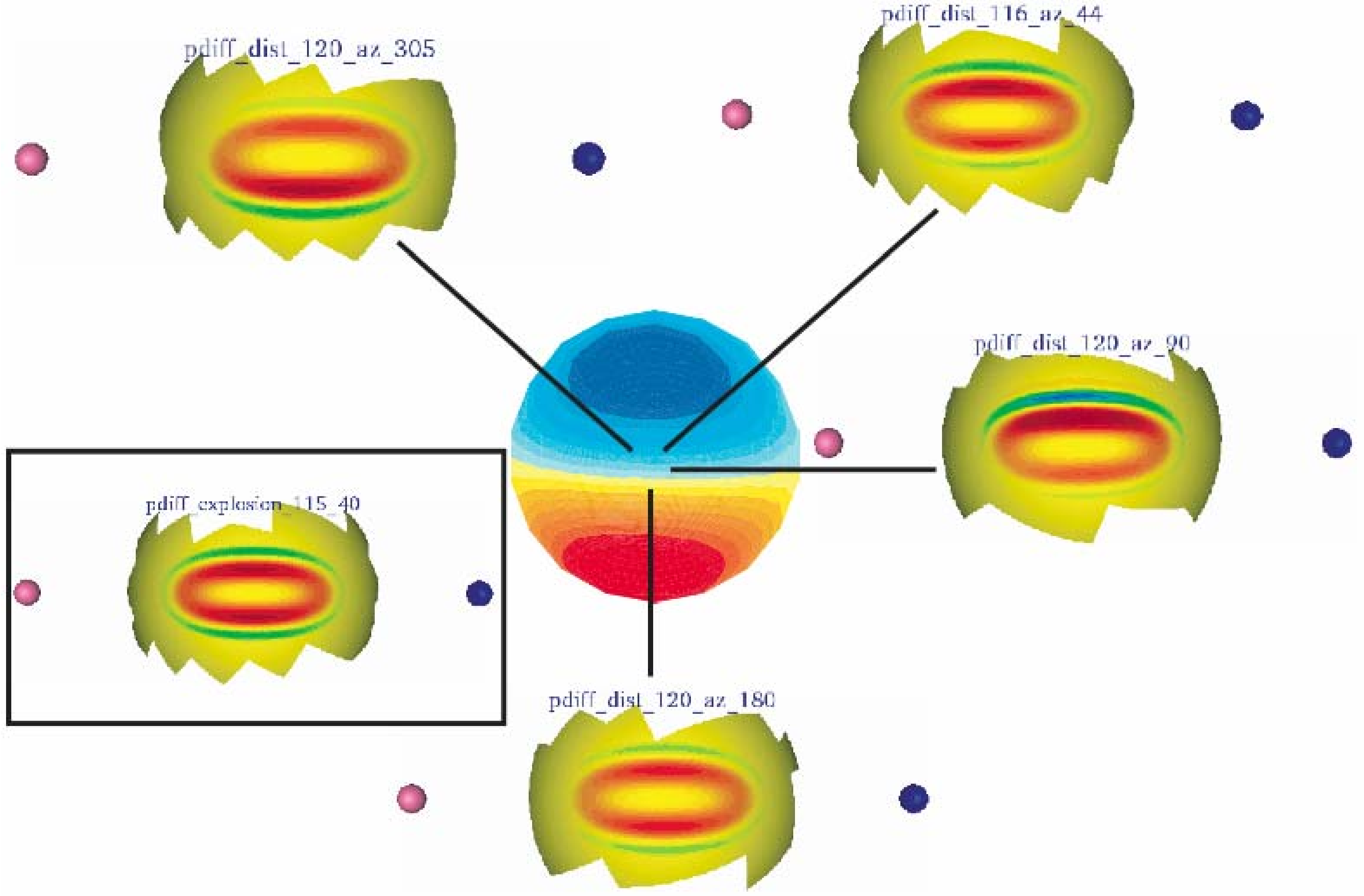
- Kernels constructed for a 1D background model vary smoothly as a function of epicentral distance so an interpolatable library of kernels can be easily constructed -- makes inversion of large datasets practical

Ray theory | Finite frequency



Preliminary P models (Manners et al)

Effect of radiation pattern on kernels



Adjoint Method & Conjugate Gradient Algorithm

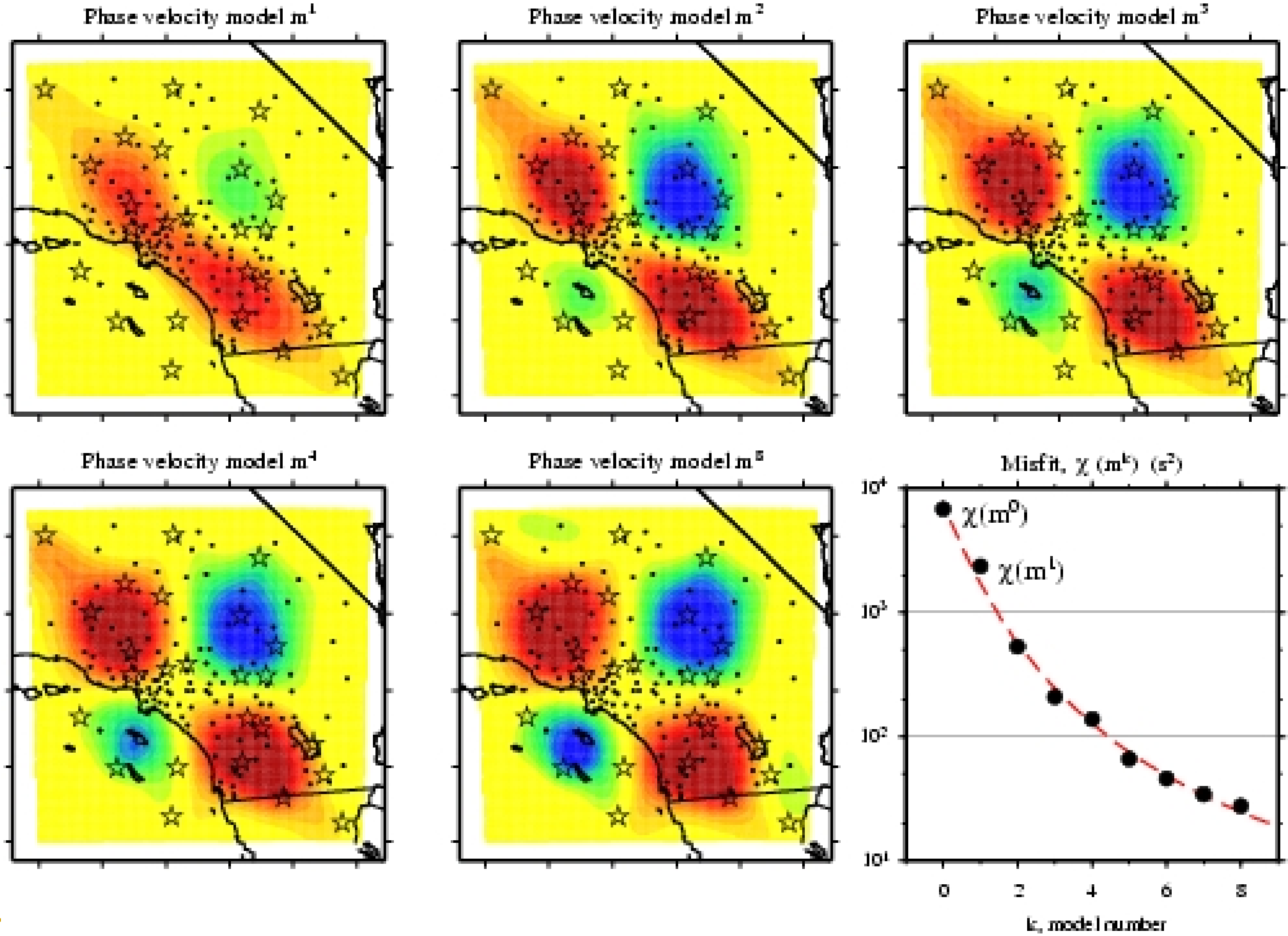
$$\frac{\partial^2 \phi}{\partial m_j \partial m_k}\Big|_{\mathbf{m}^0}(m_k - m_k^0) = -\frac{\partial \phi}{\partial m_j}\Big|_{\mathbf{m}^0}$$

$$\sum_r \frac{\partial O(\mathrm{s})}{\partial m_j}\frac{\partial O(\mathrm{s})}{\partial m_k}(m_k - m_k^0) = -\sum_r [O(\mathrm{s}) - O(\mathbf{d})]\frac{\partial O(\mathrm{s})}{\partial m_j}$$

Minimize ϕ based only on the gradient of the misfit function $\frac{\partial \phi}{\partial \mathbf{m}}$, not the Hessian

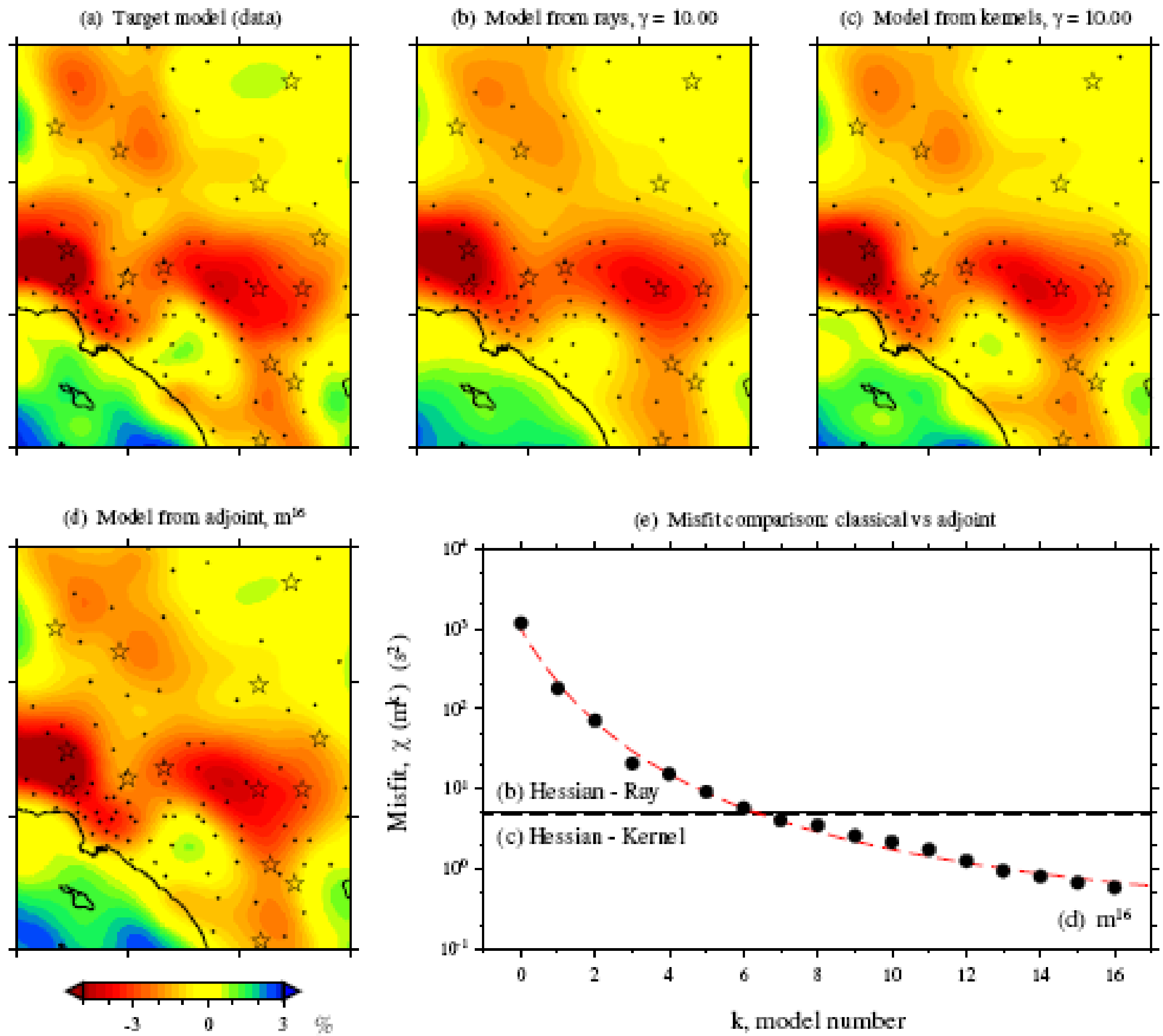
- every iteration, calculate the gradient direction $\frac{\partial \phi}{\partial \mathbf{m}}$ (2 simulations)
- update search direction according to CG algorithm
- find the minimum model along the search direction (1 more simulation)
- Number of simulation per iteration: $3 * N_{events}$

Conjugate Gradient Iterative Algorithm



(Tape et al 2006)

Comparisons

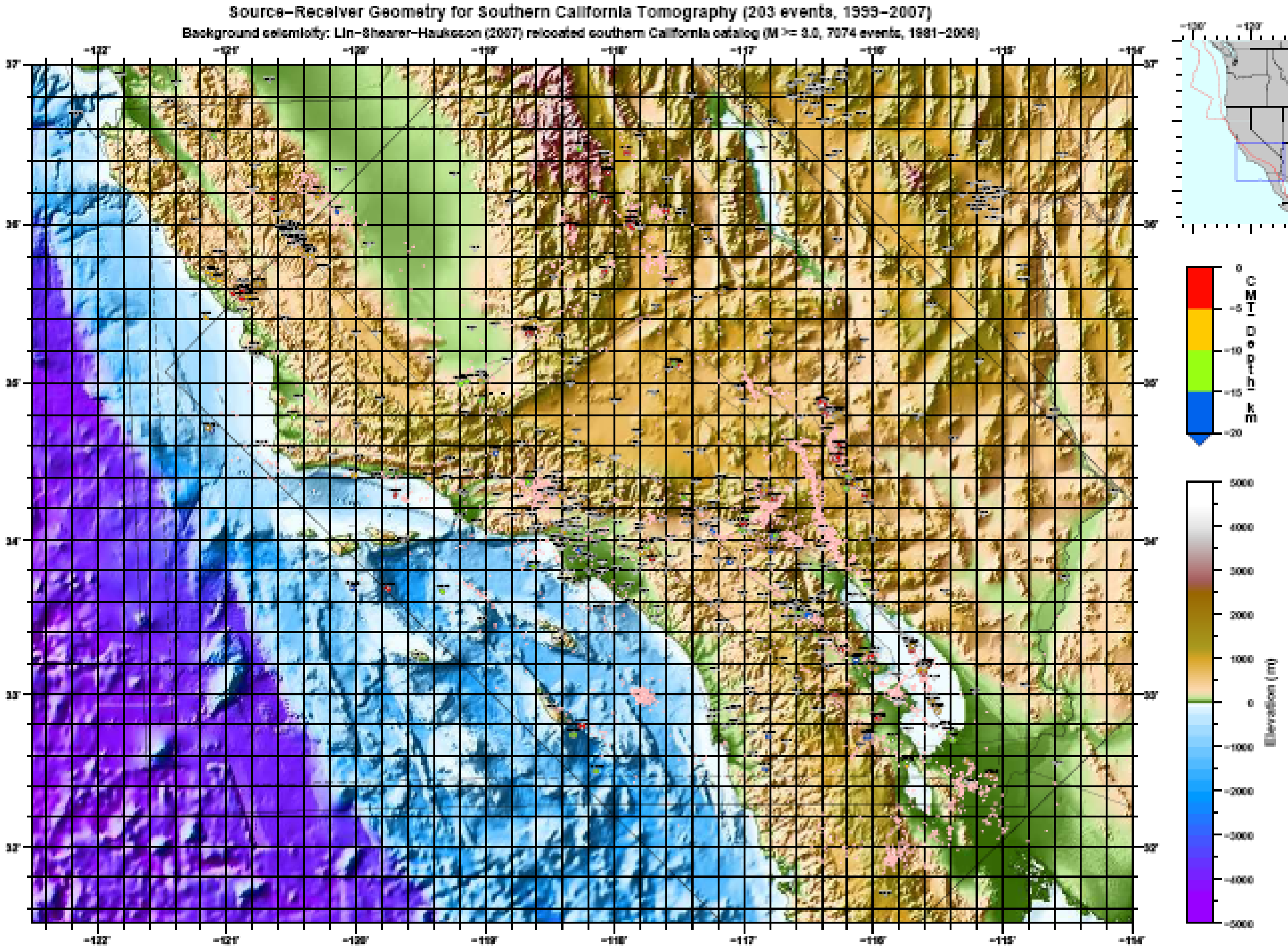


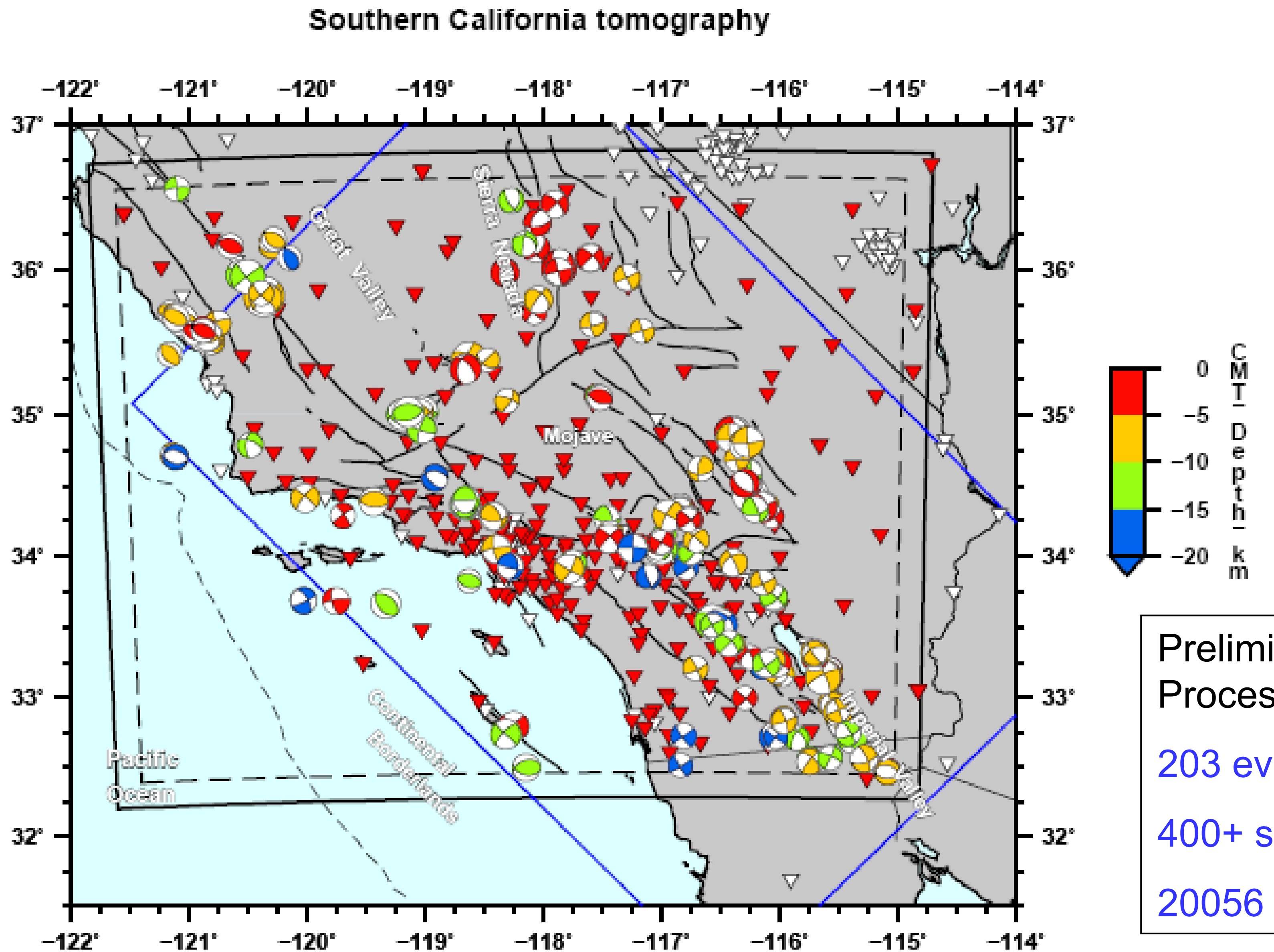
(Tape et al 2006)

Adjoint Method – advantages and disadvantages

- **We can finally use a 3-D reference model!**
- **We make NO approximations in computing the gradient of the misfit function!**
- **However, we can not afford to compute the 2nd order derivative of the misfit function, i.e., it is impractical to computed individual 'banana-doughnut' kernels.**
- **So we minimize the misfit function only based upon its gradient using conjugate-gradient schemes.**
- **And it costs the same amount of computation to assimilate as many measurements as possible for one event, which allows us to include many more phases than previous studies.**
- **Since we use 3-D numerical simulations as our forward solver, we don't need to know what phases we are picking in the seismograms.**

Adjoint Tomography in Southern California





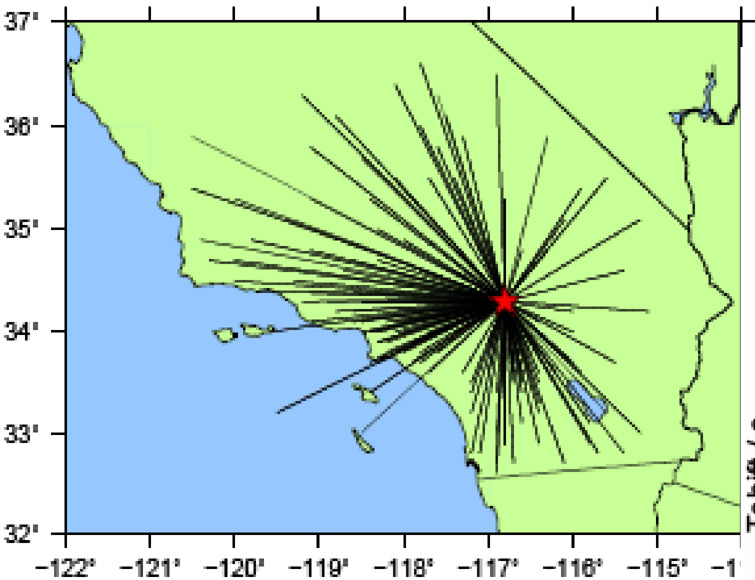
Preliminary Processing:

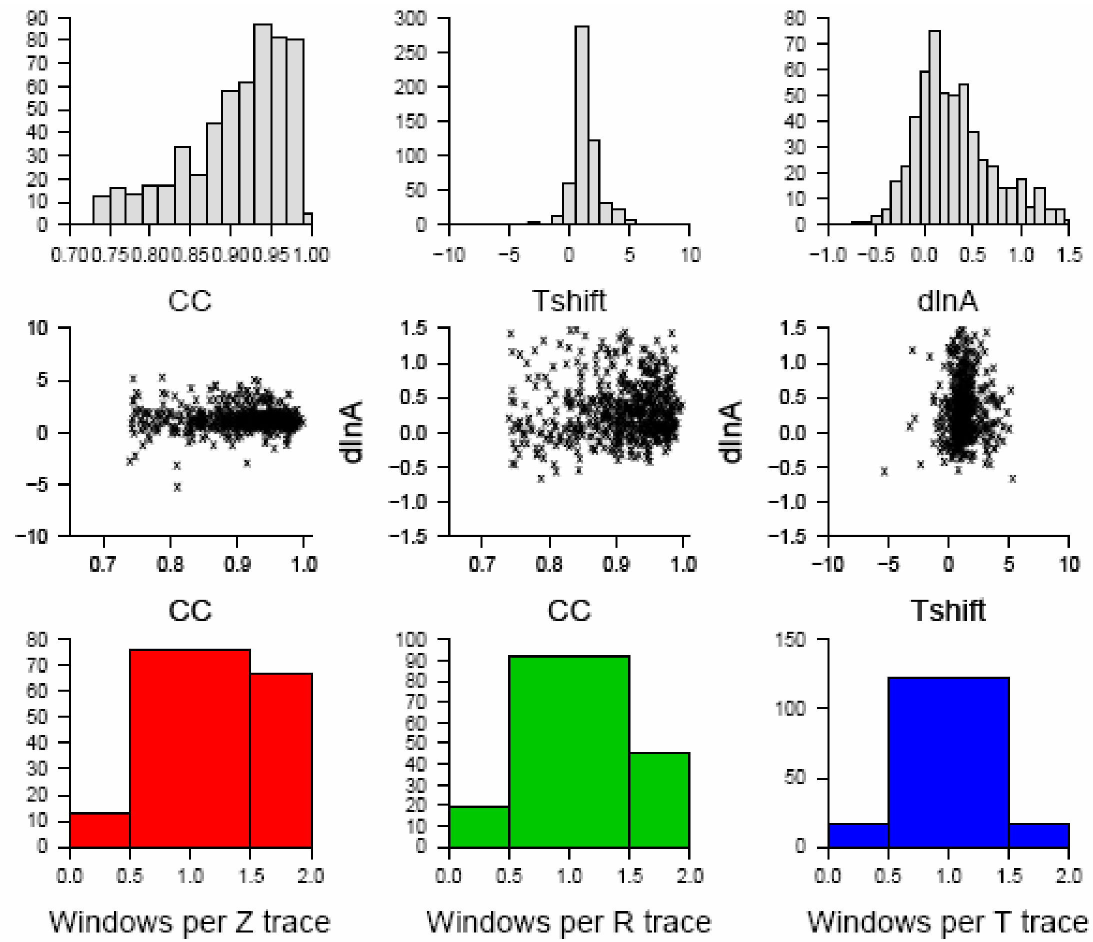
203 events

400+ stations

20056 picks!

Automatic Windowing Scheme - Example





MEASURE

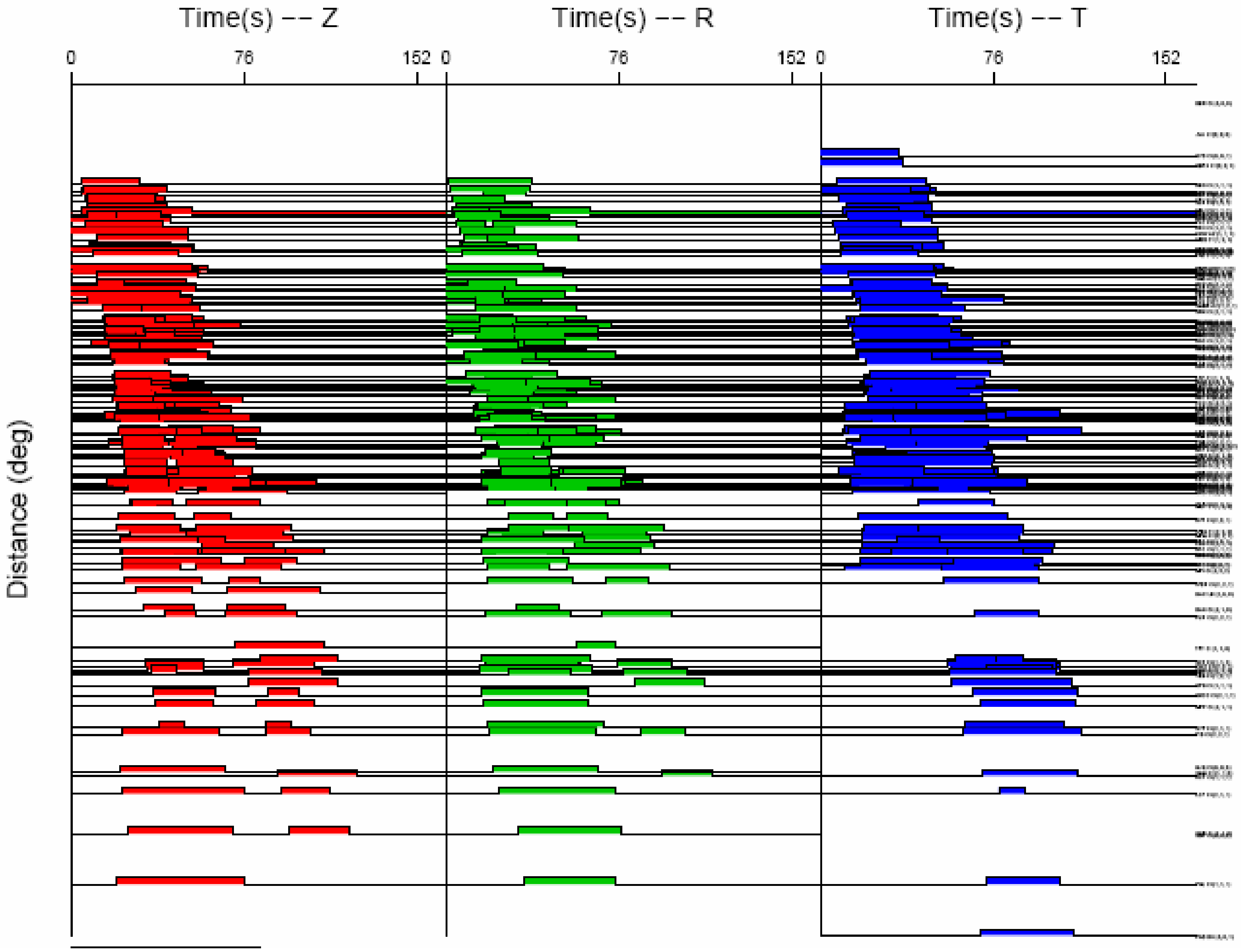
Epicenter: (34.3, −116.8).

Depth: 3.6 km.

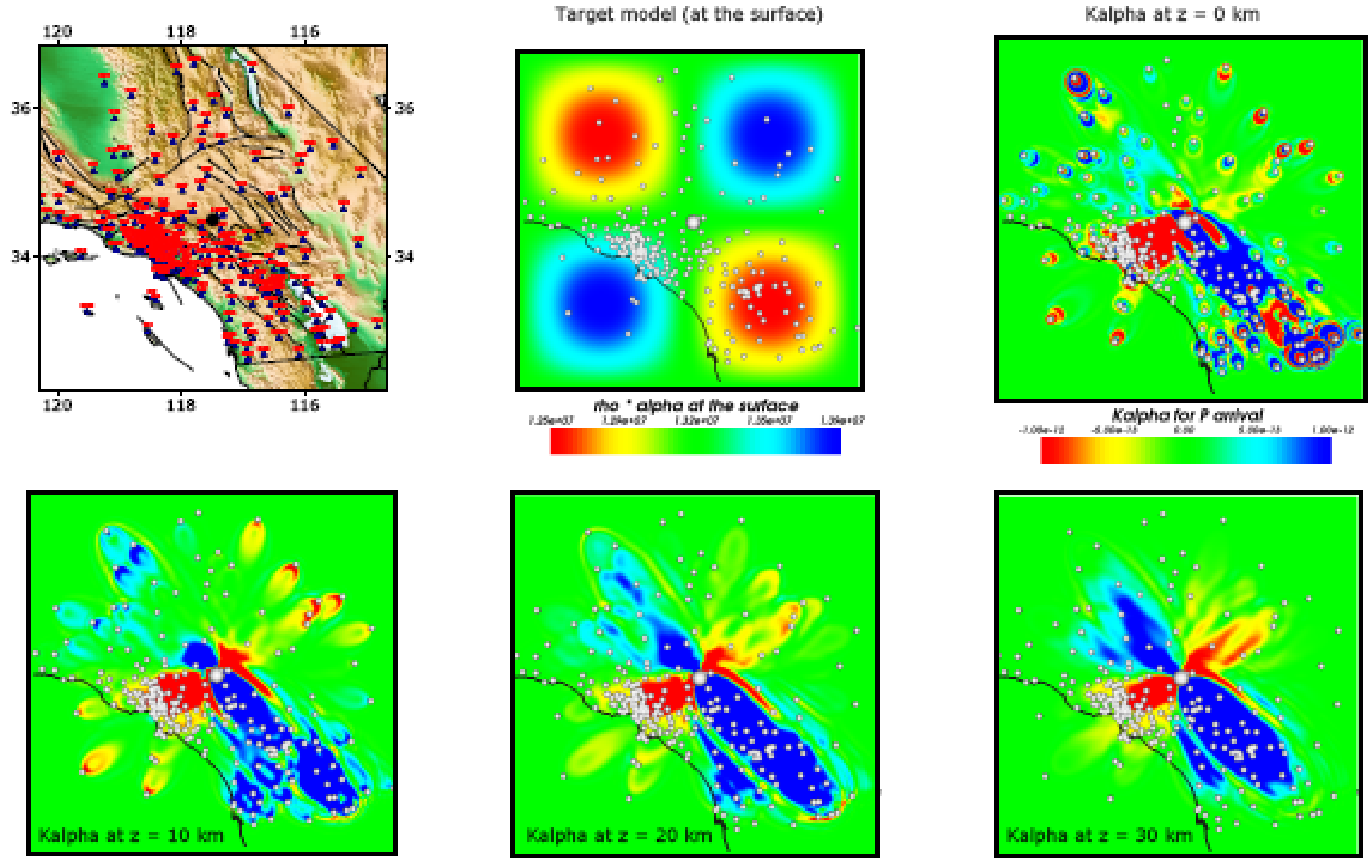
Records with measurement windows: 419.

Measurement windows: 548.

(Maggi et al 2007, Tape et al 2007)


Time(s) -- Z
Time(s) -- R
Time(s) -- T
0
76
152
Distance (deg)

Event Kernels



(Tape et al 2007)

Conclusions

✓ We demonstrate the application of spectral-element method to simulation seismic wave propagation at both global and regional scales, and its great improvement in predicting the recorded waveforms.

✓ We compute 3D sensitivity kernels using the adjoint methods.

✓ We use the Pdiff kernels computed by the adjoint method in the traditional tomographic inversion.

✓ We approach the seismic tomography problem by combining the adjoint method and the conjugate gradient optimization scheme, which allows us to use a 3-D reference model, compute accurate misfit gradients, and assimilate as many phases as possible in the seismograms.

www.shakemovie.caltech.edu/global