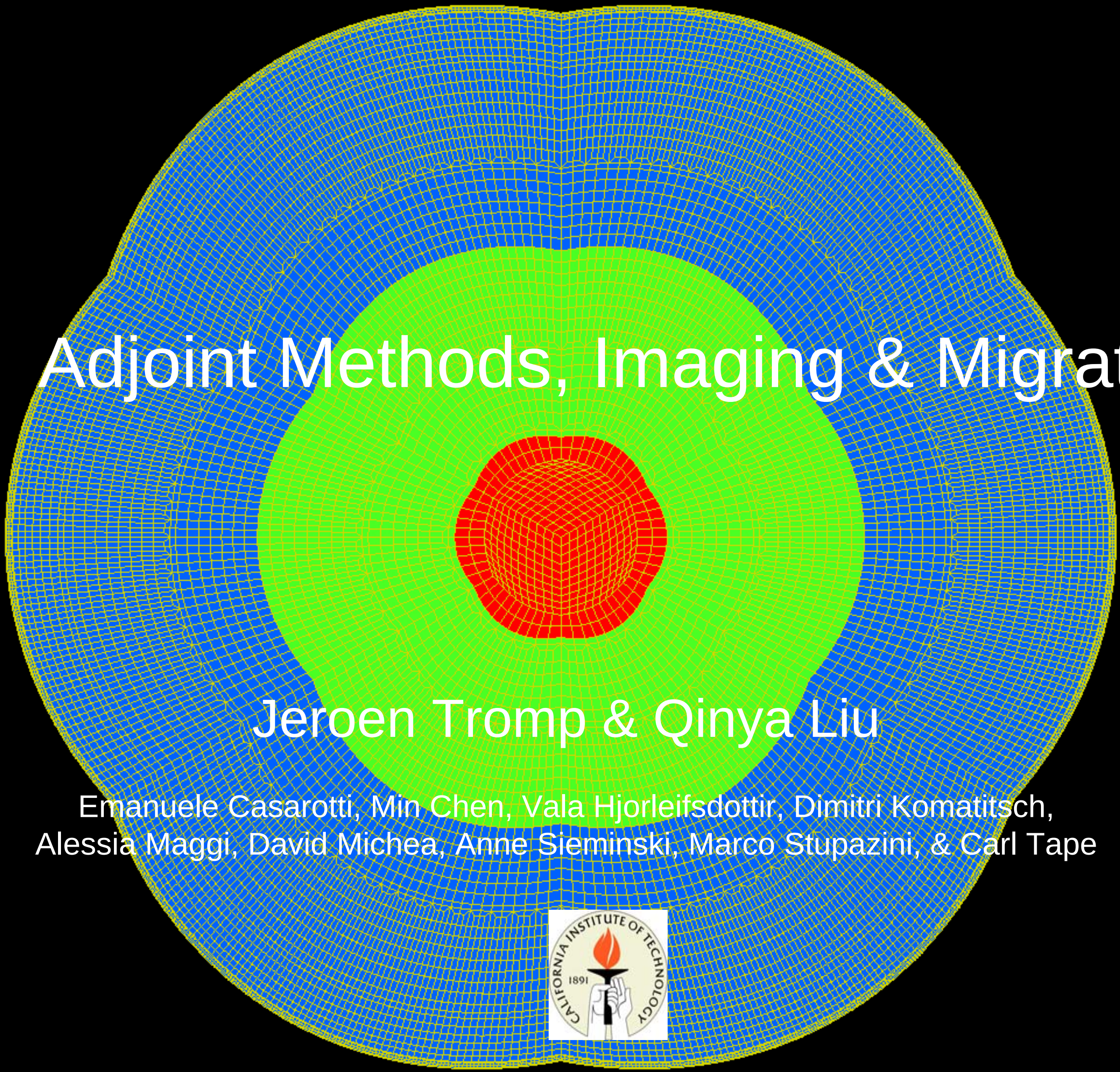
Adjoint Methods, Imaging & Migration



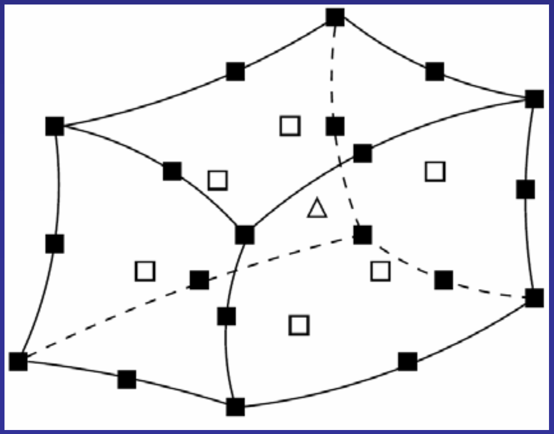
Jeroen Tromp & Qinya Liu

Emanuele Casarotti, Min Chen, Vala Hjorleifsdottir, Dimitri Komatitsch,
Alessia Maggi, David Michea, Anne Sieminski, Marco Stupazini, & Carl Tape

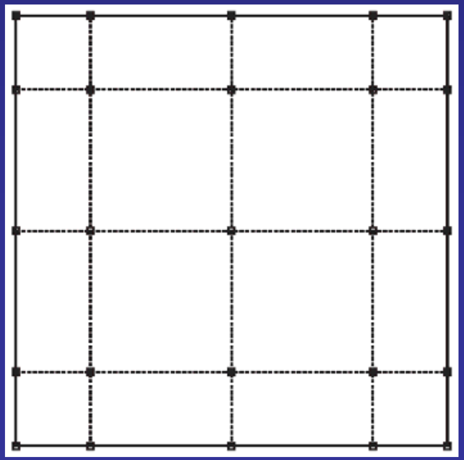
Spectral-Element Method

- Flexibility of the finite-element method
- Accuracy of a pseudospectral method
- Gauss-Lobatto-Legendre (GLL) quadrature
- Lagrange interpolants
- Diagonal mass matrix
- Explicit time integration

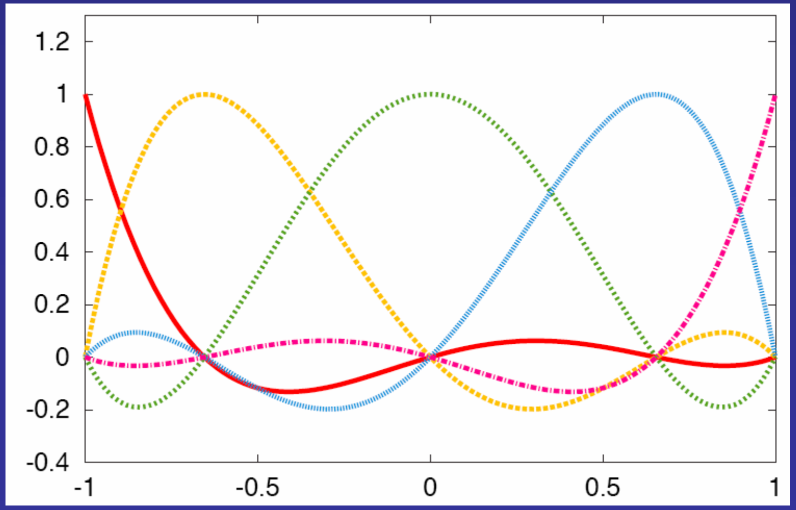
Hexahedral element



GLL integration points

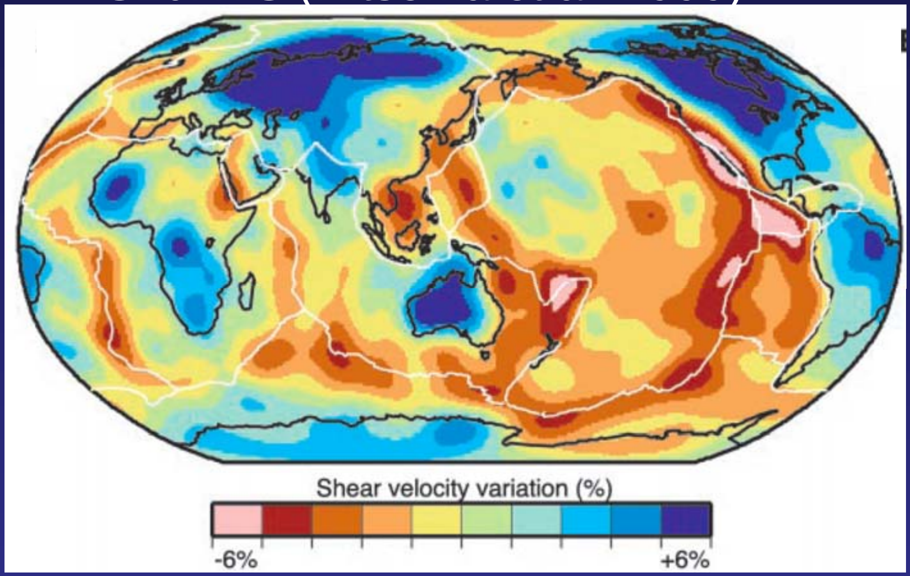


Lagrange polynomials

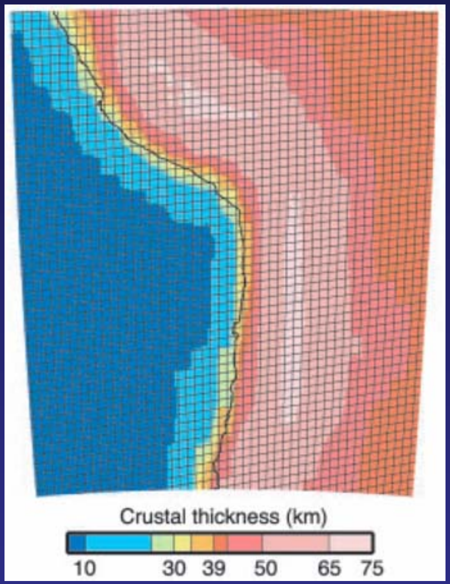


Forward Simulations

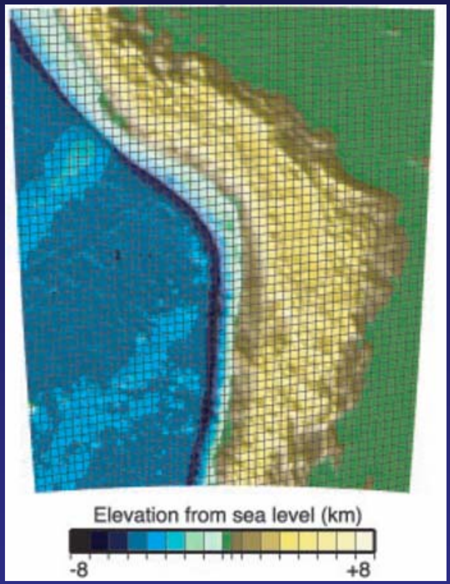
S20RTS (Ritsema et al. 1999)



Crust 2.0

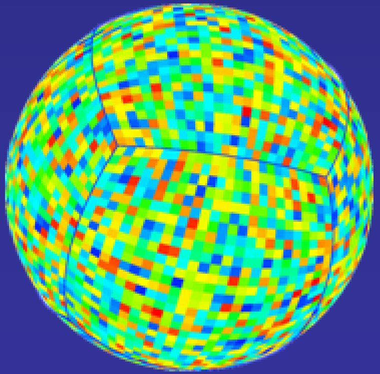


ETOPO5



- Topography & bathymetry
- Ellipticity
- 3D crust & mantle model
- 3D anisotropy
- 3D attenuation
- Rotation & self-gravitation
- Oceans

Cubed sphere mesh



6 n^2 mesh slices

Komatitsch & Tromp 2002

SPECFEM3D_GLOBE: Future Plans

On-demand TeraGrid applications:

- Automated, near real-time simulations of all M>6 earthquakes
- Analysis of past events (more than 20,000 events)
- Seismology Web Portal

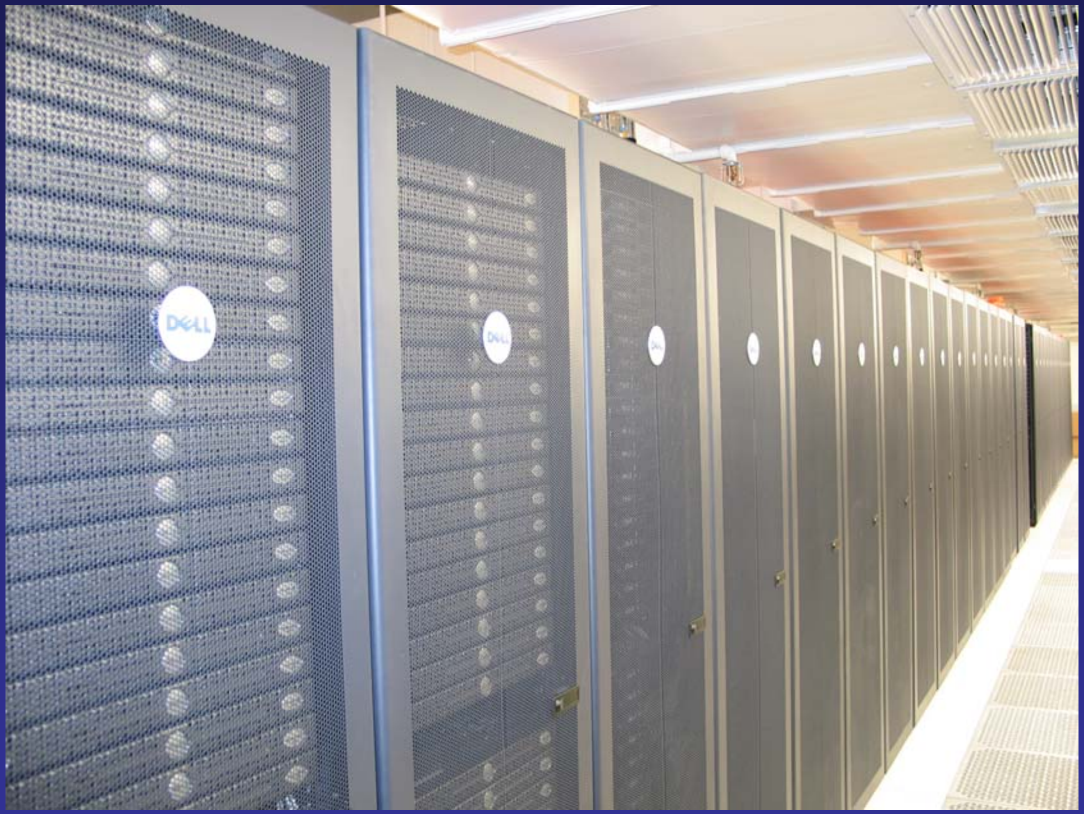
Petascale simulations:

- Global simulations at 1-2 Hz
  - New doubling brick (perfect load-balancing)
  - Cuthill-McKee sorting to reduce Cache misses
- Global tomography

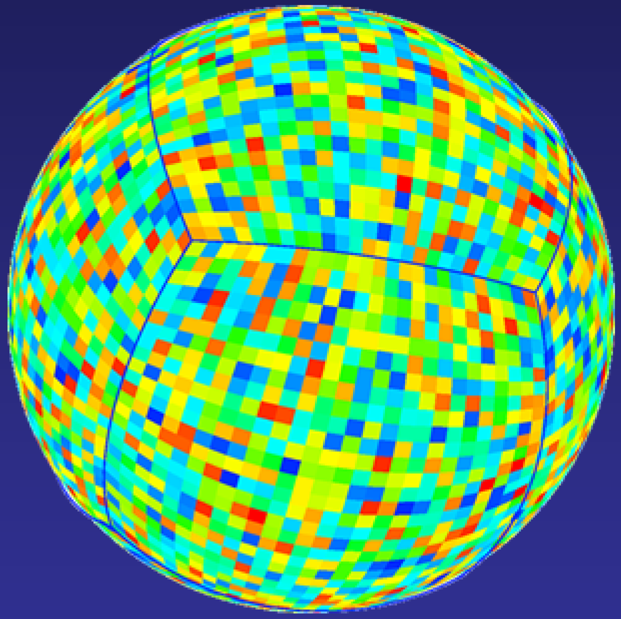
Opensource downloads:
www.geodynamics.org

High-Performance Computing

GPS Dell cluster



6 n^2 mesh slices



"Old" machine:

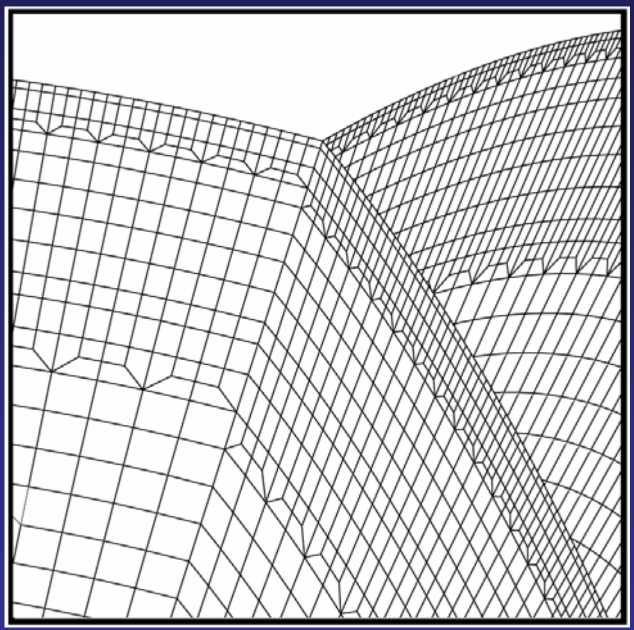
- 1024 nodes/2048 processors
- 3 TB of distributed memory
- 13.1 TFLOPS

New machine:

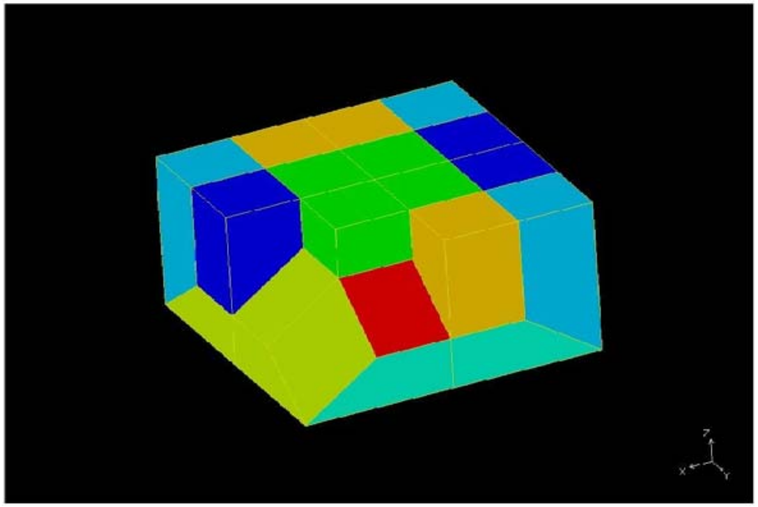
- 512 dual-processor quad-core nodes (4096 cores)
- 6 TB of distributed memory
- 22.6 TFLOPS
- Half the cooling & power

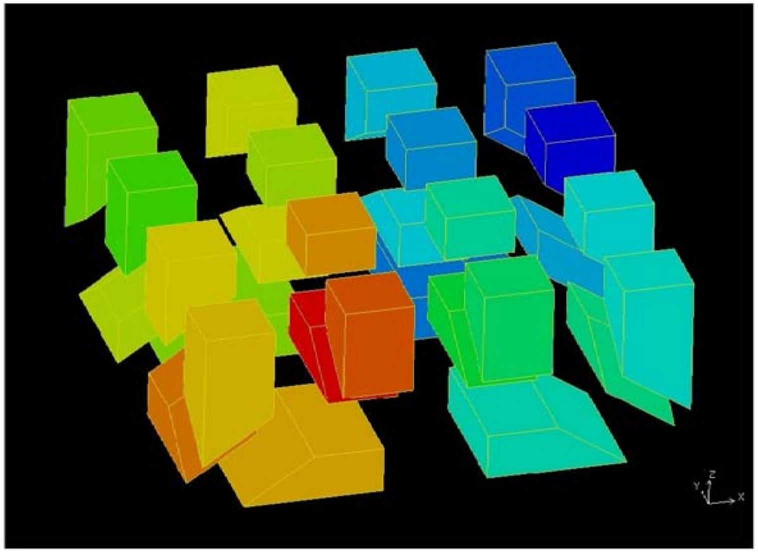
New Doubling Brick

V3.6



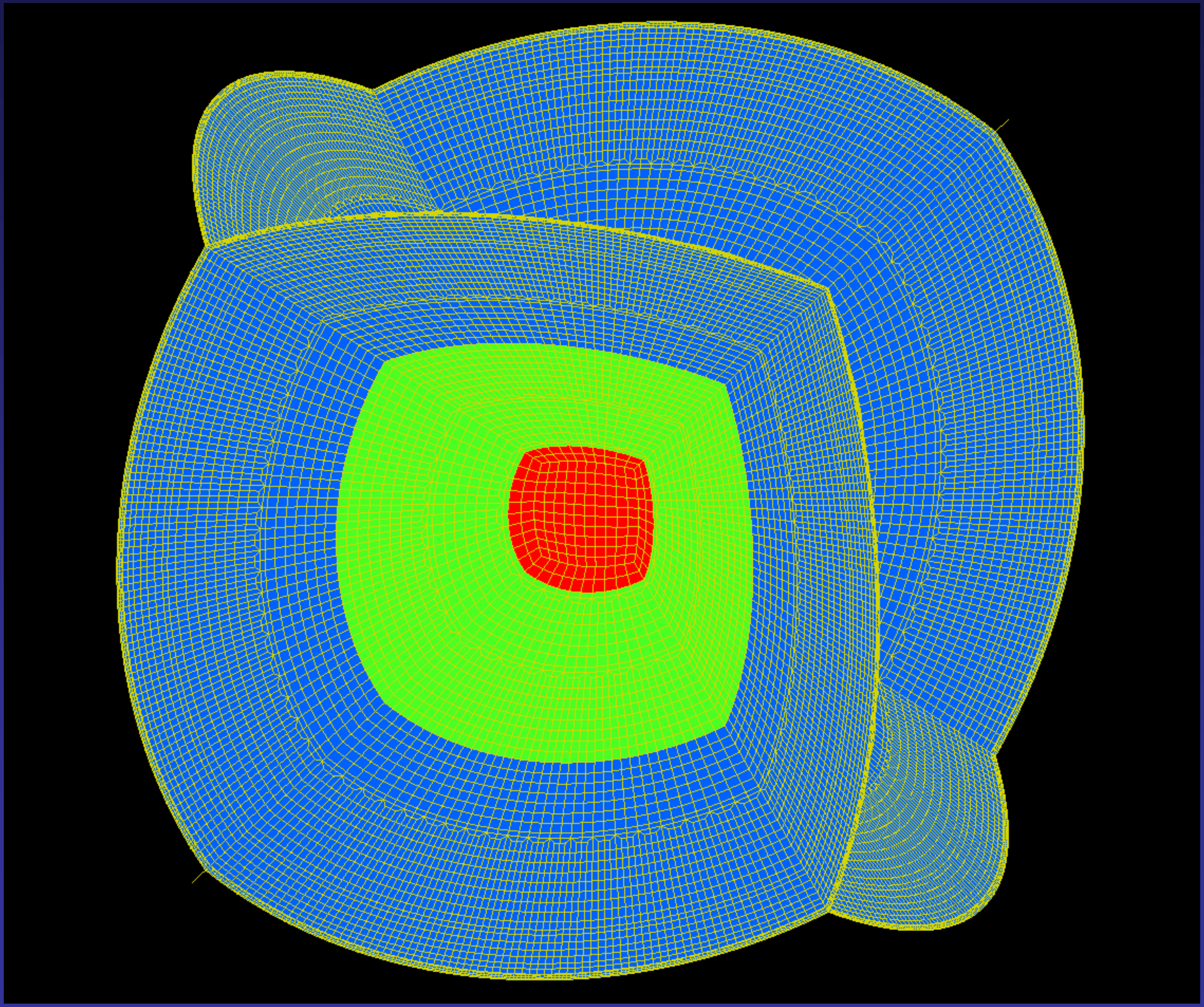
V4.0





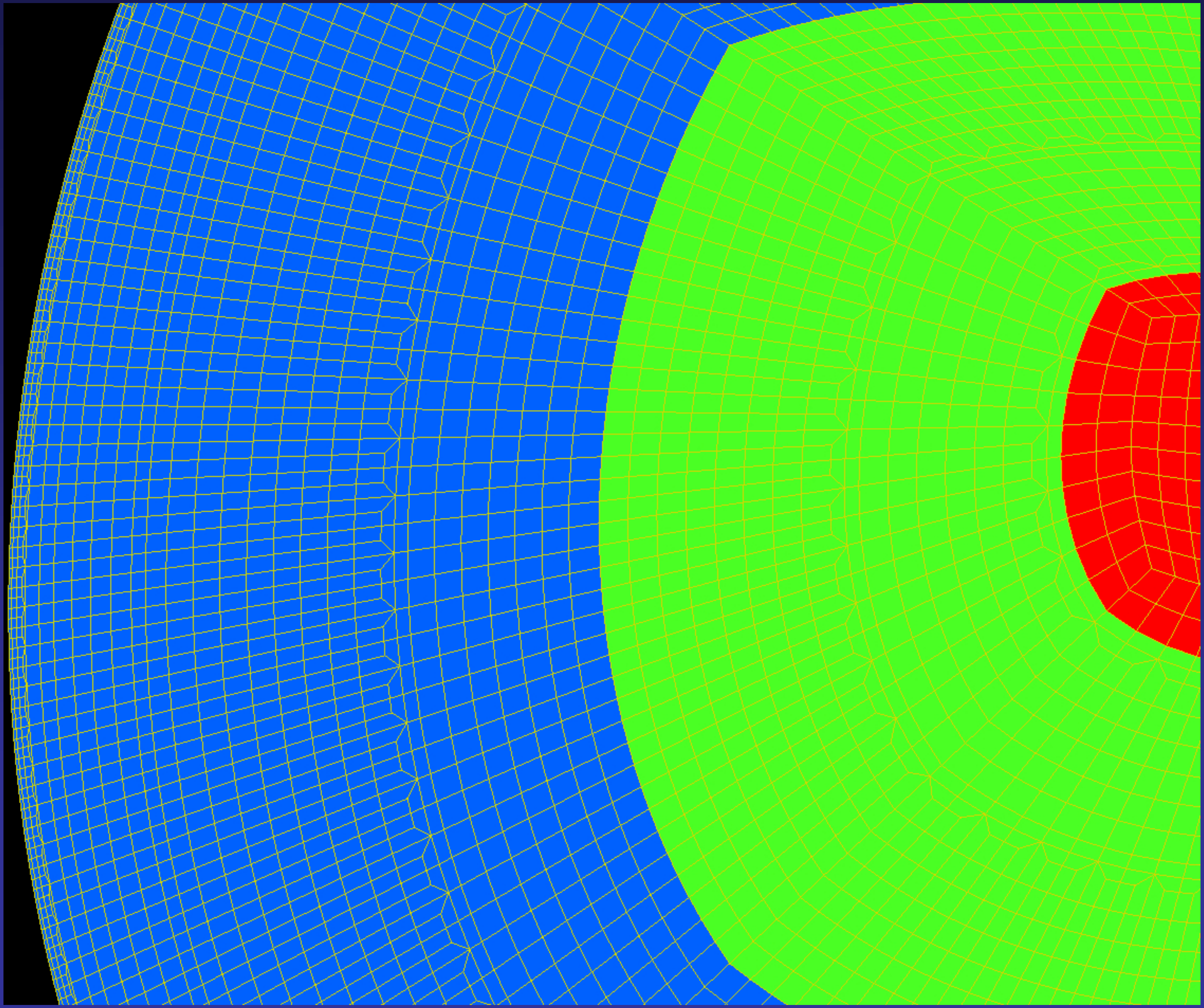
Michea & Komatitsch

New V4.0 Mesh



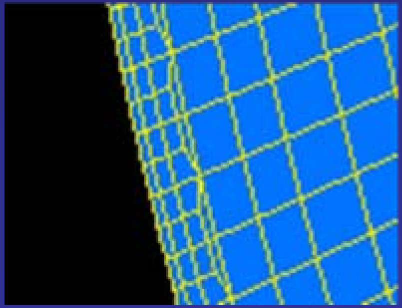
Michea & Komatitsch

New V4.0 Mesh



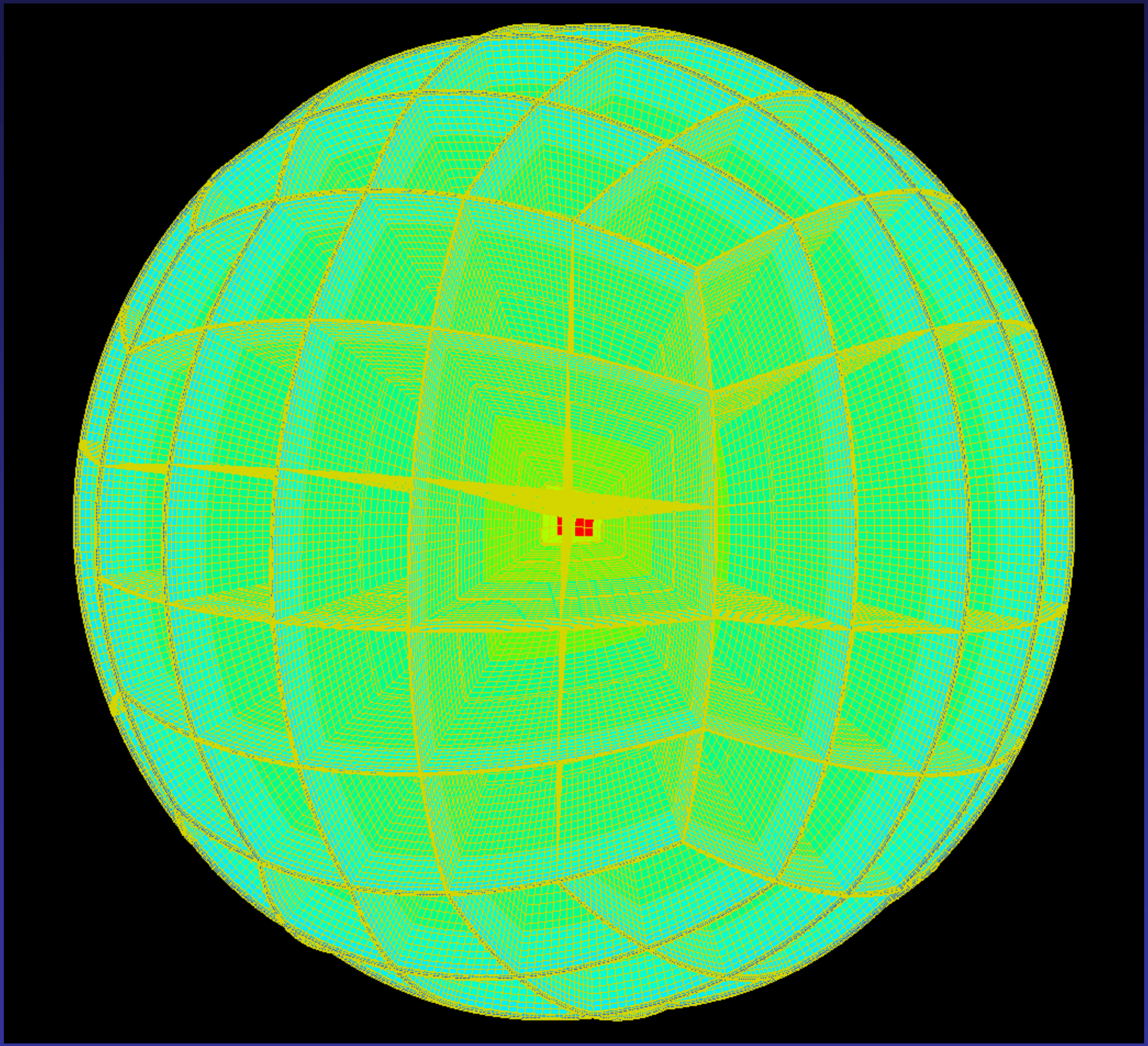
Four doublings:
- below the crust
- in the mid mantle
- two in the outer core

Note: two-layer crust



Michea & Komatitsch

New V4.0 Mesh



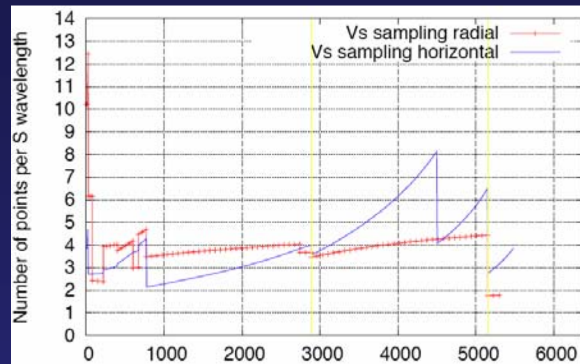
Example of 6 x 5 x 5 MPI mesh slices:

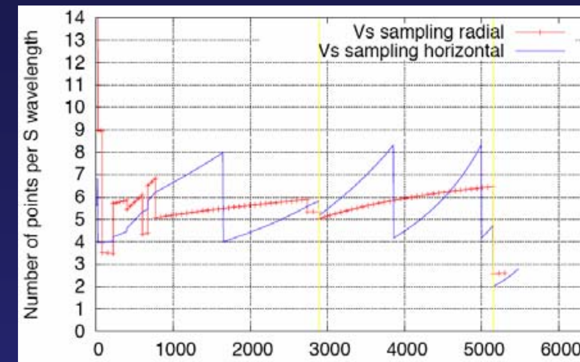
Perfect load balancing!

Michea & Komatitsch

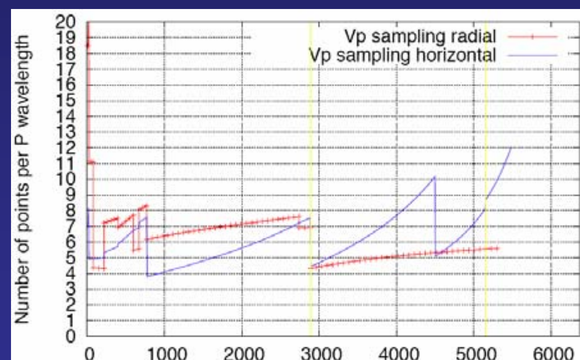
V3.6 versus V4.0

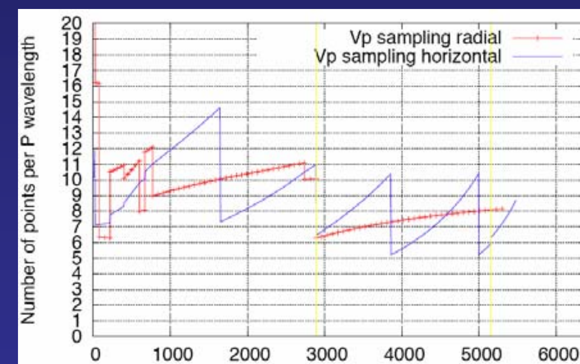
Grid points per S wavelength



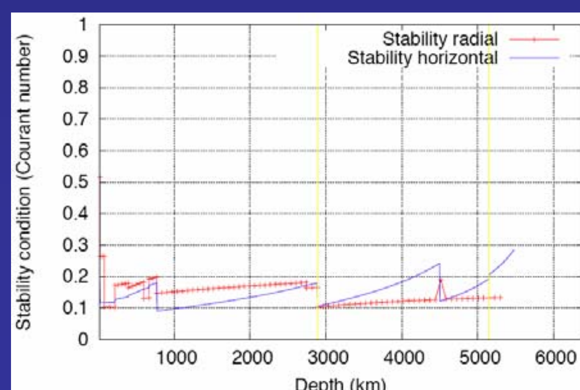


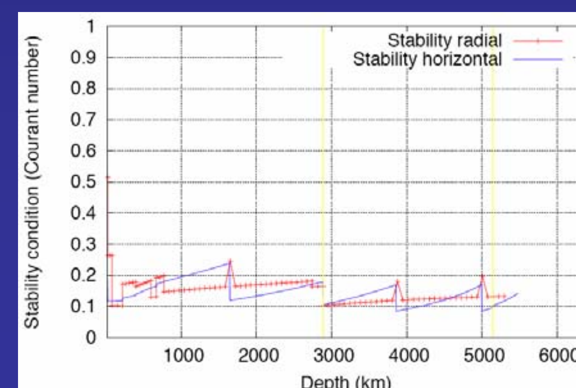
Grid points per ⚐ wavelength





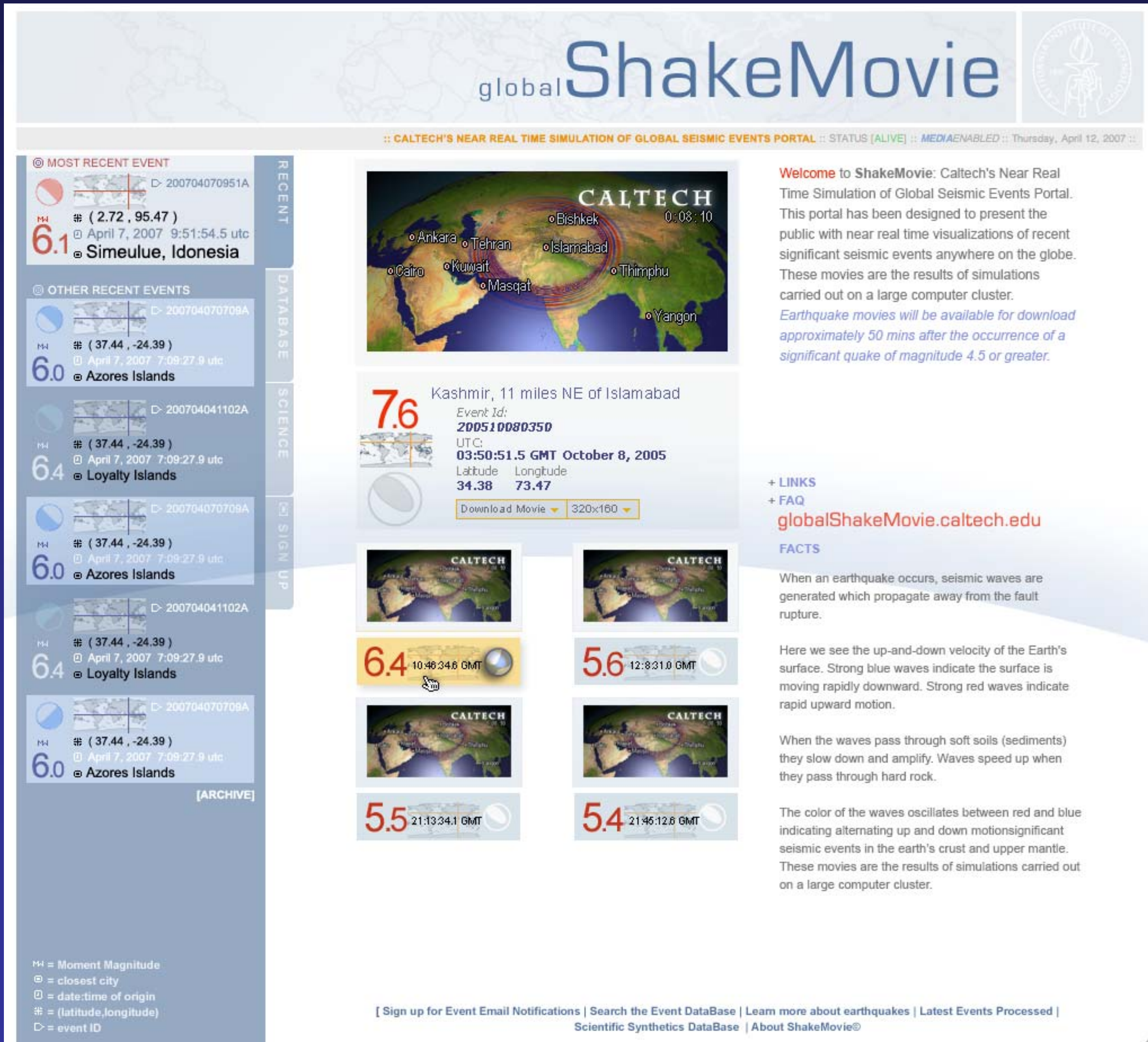
Courant stability condition





Michea & Komatitsch

Near Real-Time Global Simulations



Global ShakeMovies:

- All CMT events
  - Near real-time current events
  - Past events
- Synthetic seismograms:
  - Modes
  - SEM (need 3D REM)

2005 M 7.6 Kashmir Earthquake

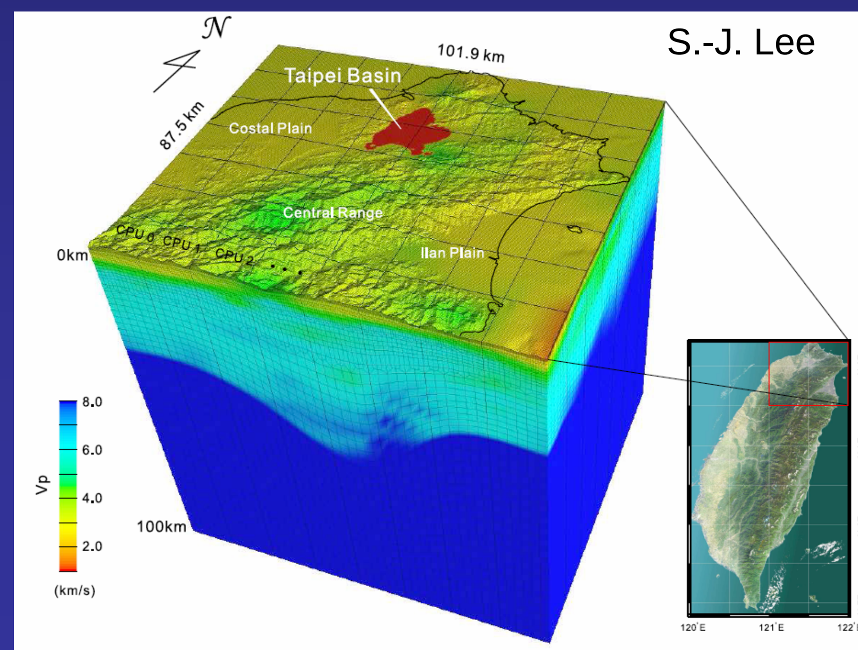
QuickTime™ and a
YUV420 codec decompressor
are needed to see this picture.

- 74,500 dead (Pakistan, India and Afghanistan)
- 106,000 injured
- 3,3 million homeless
- estimated damages: US$5 Billion

SPECFEM3D_BASIN: Future Plans

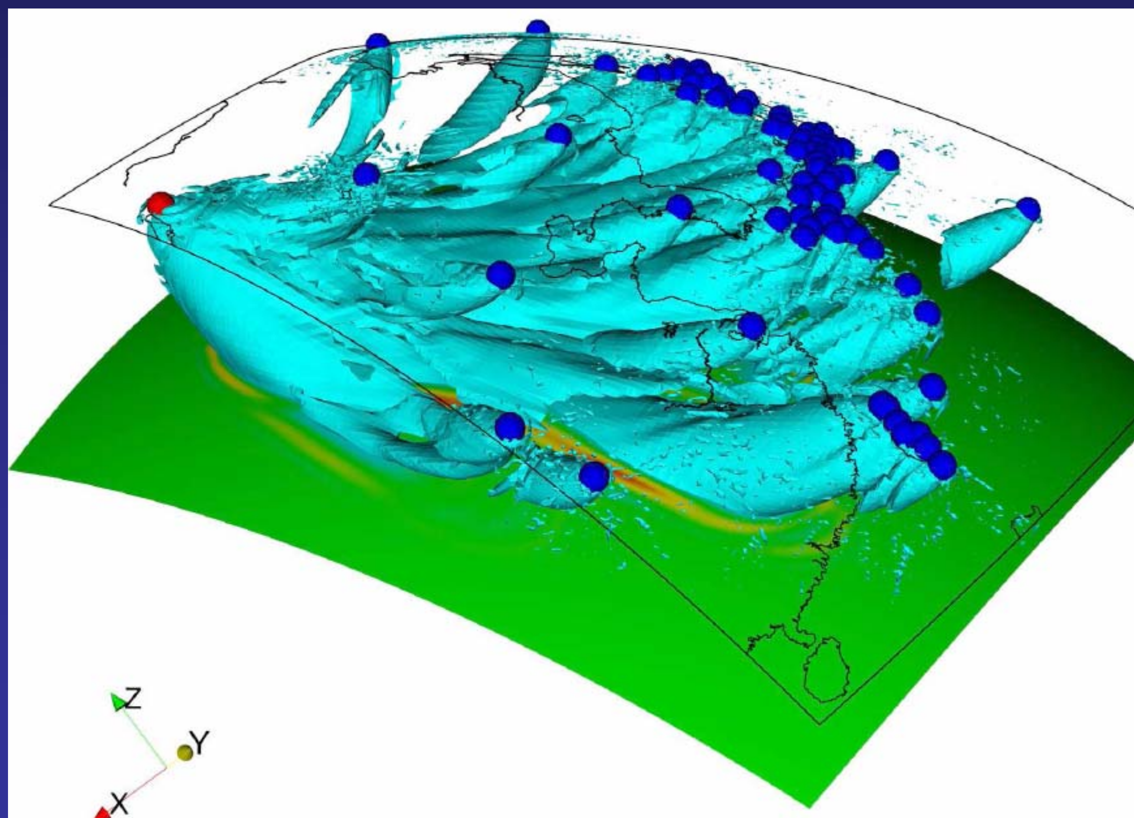
- GEOCUBIT: a (parallel) hexahedral finite-element mesher (Casarotti, Stupazini & Lee)
  - Topography & bathymetry
  - Major geological interfaces
  - Basins
  - Water layer
  - Fault surfaces
- Use ParMETIS or SCOTCH for mesh partitioning & load-balancing
- Retain the SPECFEM3D_BASIN solver (takes ParMETIS meshes; Komatitsch)
- Add dynamic rupture capabilities (Ampuero, Lapusta, Kaneko)

Stupazini



S.-J. Lee

Inverse Problems



Adjoint Tomography

PDE-constrained waveform tomography:

$$\chi = \tfrac{1}{2}\sum_r \int_0^T [\mathbf{s}(\mathbf{x}_r,t) - \mathbf{d}(\mathbf{x}_r,t)]^2\,\mathrm{d}t - \int_0^T \int_\Omega \boldsymbol{\lambda}\cdot(\rho\,\partial_t^2\mathbf{s} - \boldsymbol{\nabla}\cdot\mathbf{T} - \mathbf{f})\,\mathrm{d}^3\mathbf{x}\,\mathrm{d}t$$

Change in the waveform misfit function:

$$\begin{aligned}
\delta\chi = & \int_0^T \int_\Omega \sum_r [\mathbf{s}(\mathbf{x}_r,t) - \mathbf{d}(\mathbf{x}_r,t)]\delta(\mathbf{x}-\mathbf{x}_r)\cdot\delta\mathbf{s}(\mathbf{x},t)\,\mathrm{d}^3\mathbf{x}\,\mathrm{d}t \\
& - \int_0^T \int_\Omega (\delta\rho\boldsymbol{\lambda}\cdot\partial_t^2\mathbf{s} + \boldsymbol{\nabla}\boldsymbol{\lambda}:\delta\mathbf{c}:\boldsymbol{\nabla}\mathbf{s} - \boldsymbol{\lambda}\cdot\delta\mathbf{f})\,\mathrm{d}^3\mathbf{x}\,\mathrm{d}t - \int_0^T \int_\Omega [\rho\partial_t^2\boldsymbol{\lambda} - \boldsymbol{\nabla}\cdot(\mathbf{c}:\boldsymbol{\nabla}\boldsymbol{\lambda})]\cdot\delta\mathbf{s}\,\mathrm{d}^3\mathbf{x}\,\mathrm{d}t \\
& - \int_\Omega [\rho(\boldsymbol{\lambda}\cdot\partial_t\delta\mathbf{s} - \partial_t\boldsymbol{\lambda}\cdot\delta\mathbf{s})]_T\,\mathrm{d}^3\mathbf{x} - \int_0^T \int_{\partial\Omega} \hat{\mathbf{n}}\cdot(\mathbf{c}:\boldsymbol{\nabla}\boldsymbol{\lambda})\cdot\delta\mathbf{s}\,\mathrm{d}^2\mathbf{x}\,\mathrm{d}t,
\end{aligned}$$

Adjoint Equations

Define the adjoint wavefield: $\mathbf{s}^{\dagger}(\mathbf{x},t) \equiv \boldsymbol{\lambda}(\mathbf{x},T-t)$

Equation of motion: $\rho\,\partial_t^2\mathbf{s}^{\dagger} = \boldsymbol{\nabla}\cdot\mathbf{T}^{\dagger} + \mathbf{f}^{\dagger}$

Boundary conditions: $\hat{\mathbf{n}}\cdot\mathbf{T}^{\dagger} = \mathbf{0}$

Initial conditions: $\mathbf{s}^{\dagger}(\mathbf{x},0) = \mathbf{0}, \qquad \partial_t\mathbf{s}^{\dagger}(\mathbf{x},0) = \mathbf{0}$

The adjoint source is determined by the type of measurement and the misfit criterion, e.g., for l_2-norm waveform tomography:

$$\mathbf{f}^{\dagger}(\mathbf{x},t) = \sum_{r=1}^{N}[\mathbf{s}(\mathbf{x}_r,T-t)-\mathbf{d}(\mathbf{x}_r,T-t)]\delta(\mathbf{x}-\mathbf{x}_r)$$

whereas for l_2-norm cross-correlation traveltime tomography:

$$\mathbf{f}^{\dagger} = \frac{1}{N}w(T-t)\partial_t s(\mathbf{x}_r,T-t)\delta(\mathbf{x}-\mathbf{x}_r)$$

Frechet derivative

The Frechet derivative may be expressed as:

$$\delta\chi = \int_\Omega (\delta\rho K_\rho + \delta\mathbf{c} :: \mathbf{K_c})\,\mathrm{d}^3\mathbf{x} + \int_0^T \int_\Omega \mathbf{s}^\dagger \cdot \delta\mathbf{f}\,\mathrm{d}^3\mathbf{x}\,\mathrm{d}t$$

Density and elastic tensor kernels:

$$K_\rho(\mathbf{x}) = -\int_0^T \mathbf{s}^\dagger(\mathbf{x}, T-t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x}, t)\,\mathrm{d}t$$

$$\mathbf{K_c}(\mathbf{x}) = -\int_0^T \boldsymbol{\nabla}\mathbf{s}^\dagger(\mathbf{x}, T-t)\,\boldsymbol{\nabla}\mathbf{s}(\mathbf{x}, t)\,\mathrm{d}t$$

Traveltime Frechet Derivatives

Traveltime tomography:

$$\chi(m) = \frac{1}{2}\sum_{r=1}^{N}[T_r(m) - T_r^{\mathrm{obs}}]^2$$

Change in the misfit function:

$$\delta\chi = \sum_{r=1}^{N}[T_r(m) - T_r^{\mathrm{obs}}]\delta T_r = \int K(\mathbf{x})\,\delta\ln m(\mathbf{x})\,\mathrm{d}^3\mathbf{x}$$

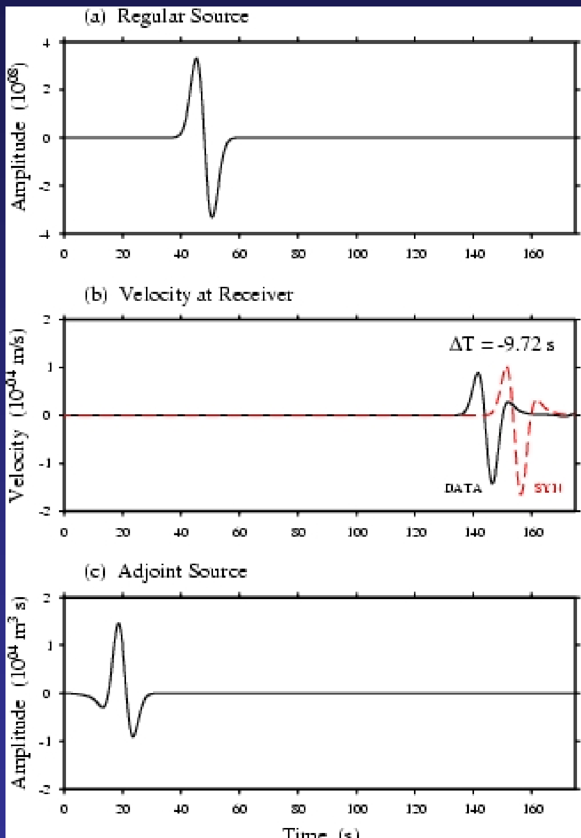
Traveltime anomaly in terms of *banana-donut* kernel (Dahlen):

$$\delta T_r = \int K_r(\mathbf{x})\,\delta\ln m(\mathbf{x})\,\mathrm{d}^3\mathbf{x}$$

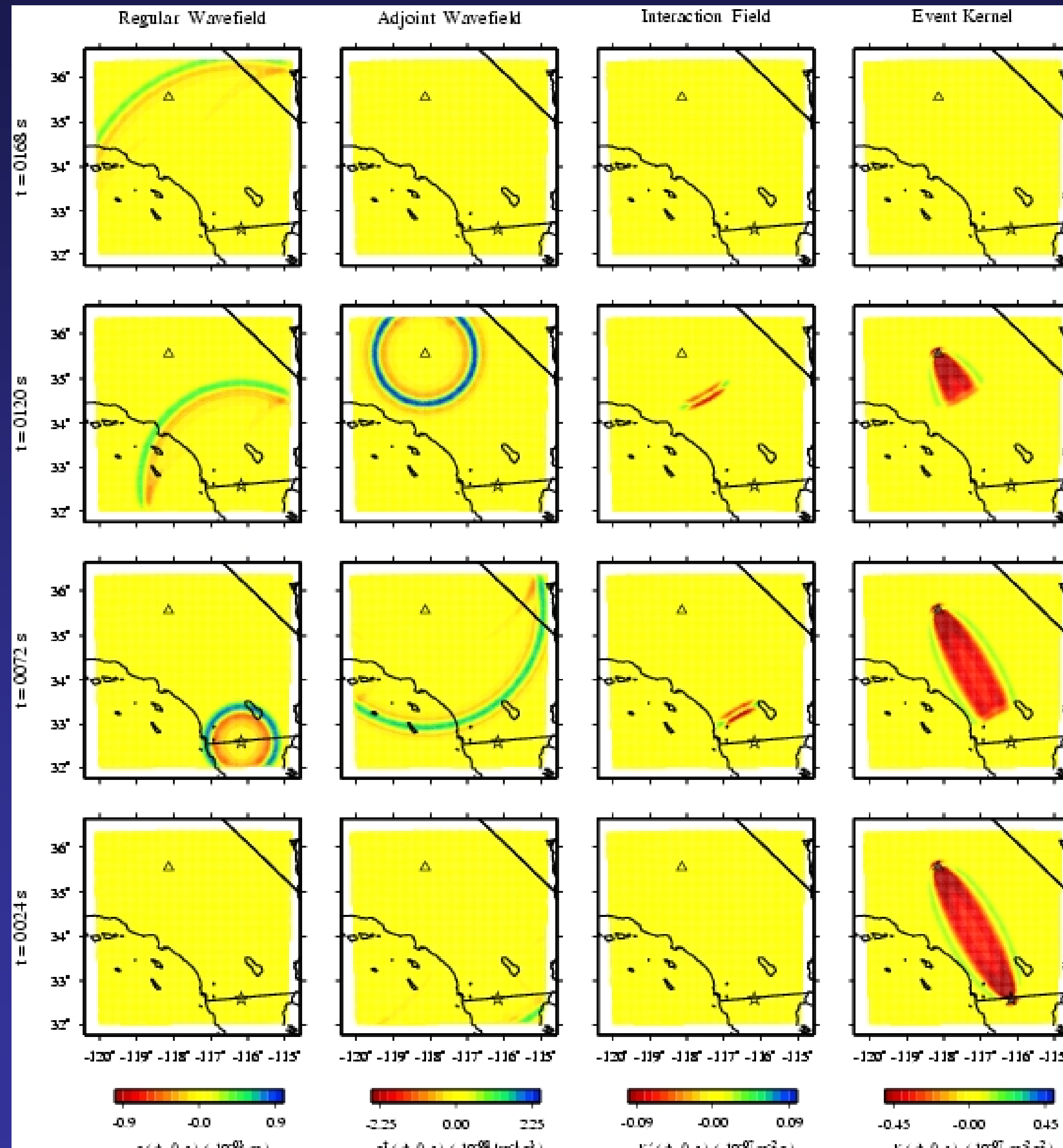
The kernel K is a weighted sum of banana-donut kernels K_r :

$$K(\mathbf{x}) = \sum_{r=1}^{N}[T_r(m) - T_r^{\mathrm{obs}}]\,K_r\,(\mathbf{x})$$

Construction of a **Banana-Donut** Kernel



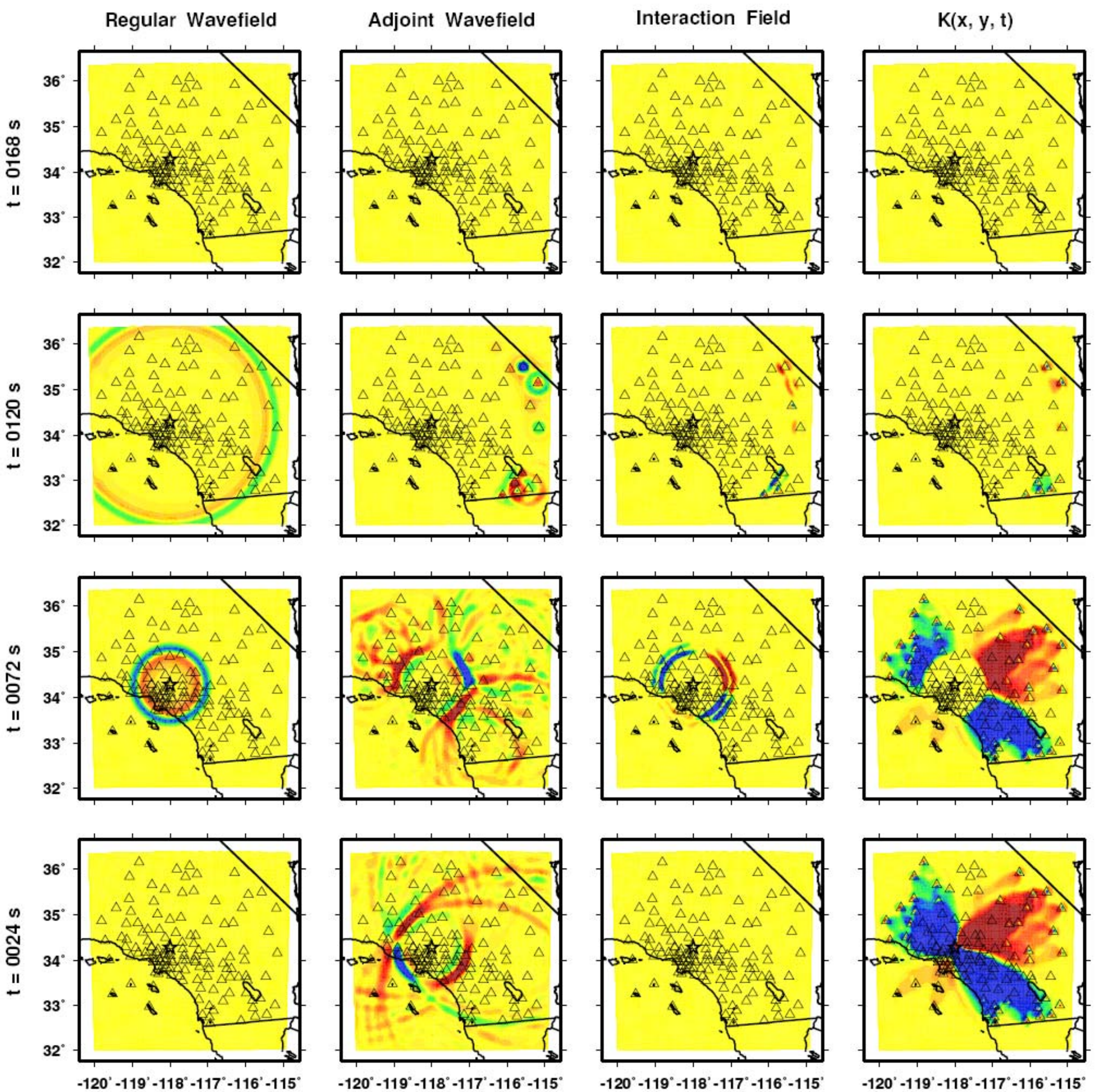
$$\delta T_r = \int K_r(\mathbf{x})\, \delta \ln m(\mathbf{x})\, \mathrm{d}^3\mathbf{x}$$



Tape et al. 2007

Construction of an **Event** Kernel

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



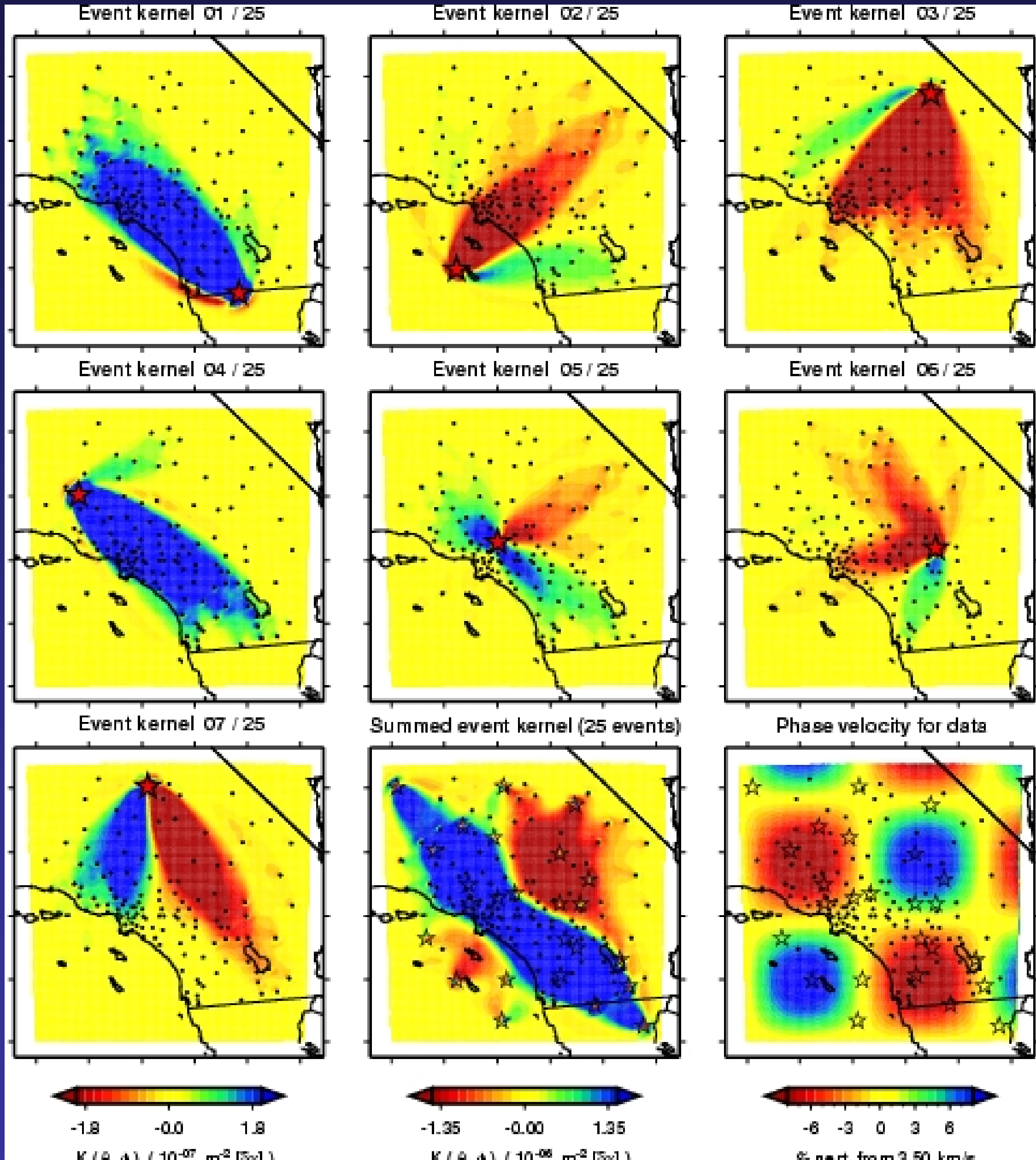
Event Kernel:

- Sum of weighted banana-donut kernels
- Two simulations per event

$$K(\mathbf{x}) = \sum_{r=1}^{N} [T_r(m) - T_r^{\mathrm{obs}}]\, K_r(\mathbf{x})$$

Tape et al. 2007

Construction of the **Misfit** Kernel



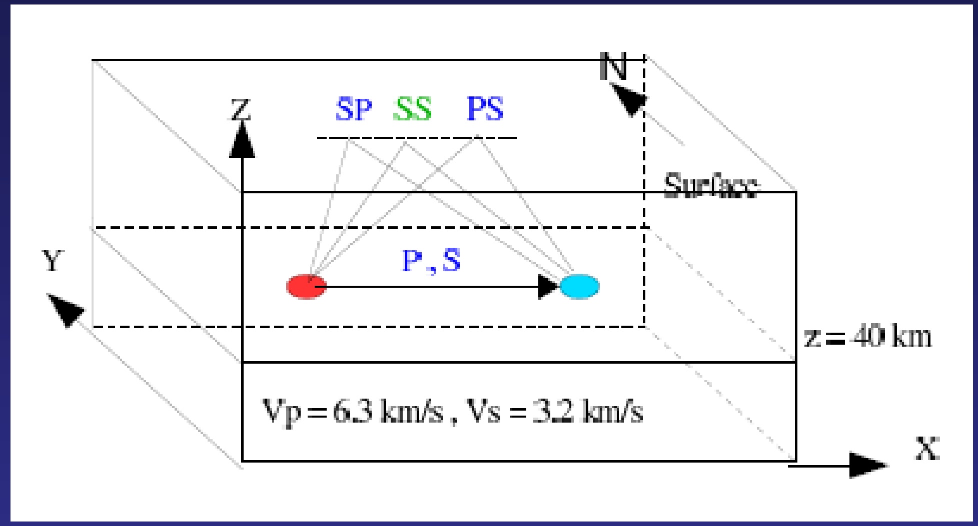
Misfit kernel:

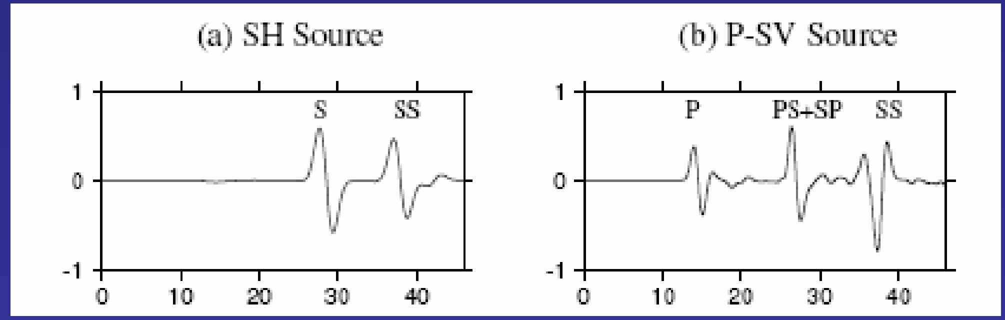
- Sum of all the event kernels
- Two simulations per event
- Gradient of misfit function

Tape et al. 2007

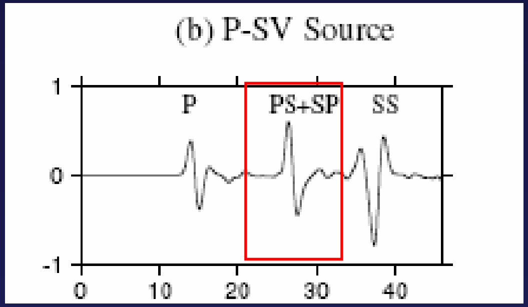
3D Examples of Banana-Donut Kernels

- Homogenous Model
- Point source

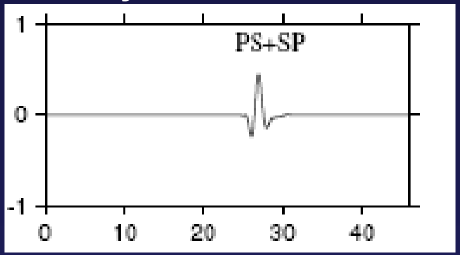


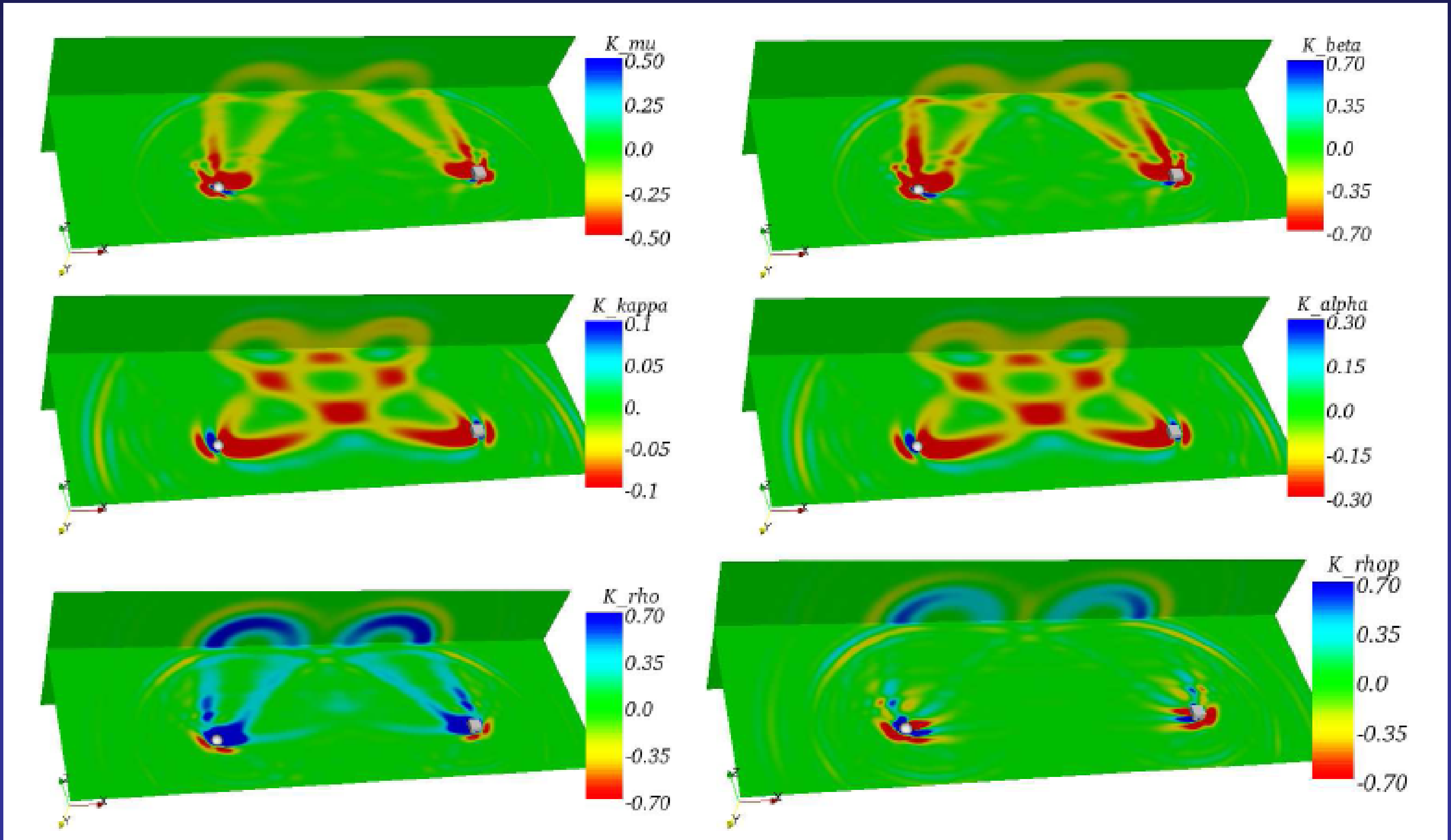


Liu & Tromp 2006



Adjoint source





$$\delta \ln \rho \, K_\rho + \delta \ln \mu \, K_\mu + \delta \ln \kappa \, K_\kappa = \delta \ln \rho \, K_{\rho'} + \delta \ln \beta \, K_\beta + \delta \ln \alpha \, K_\alpha$$

SENSITIVITY KERNEL

Distance: Δ = 67 km

Minimum period: 2 s

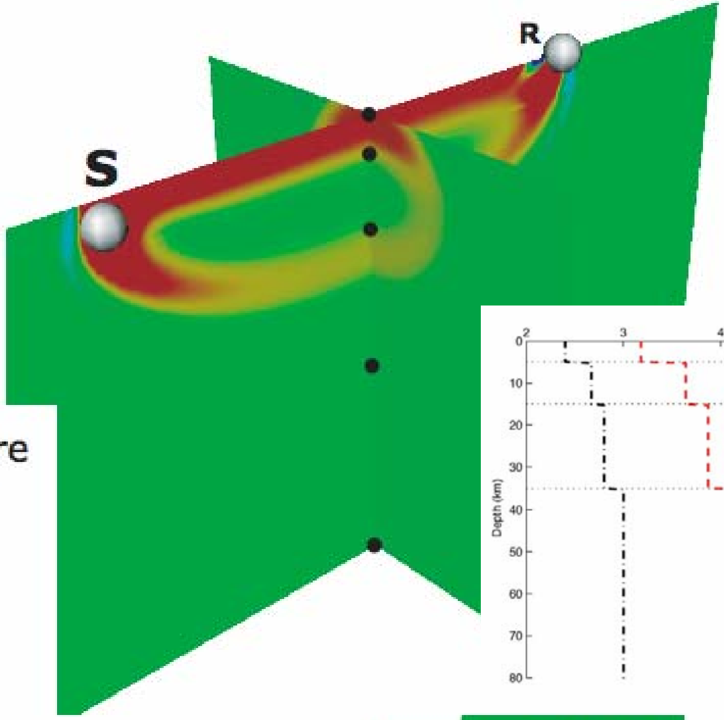
Reference model: SoCal-1D

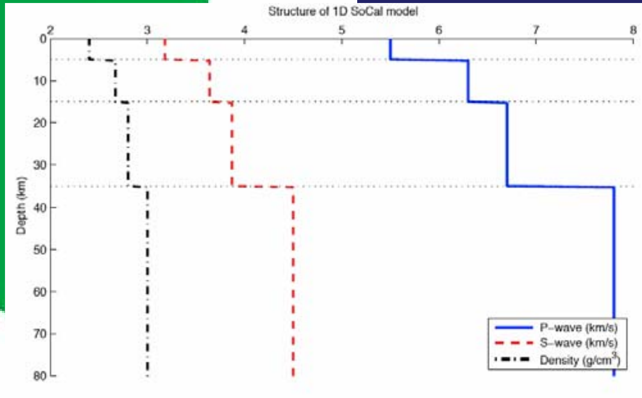
Source: Explosion (depth 2.75 km)

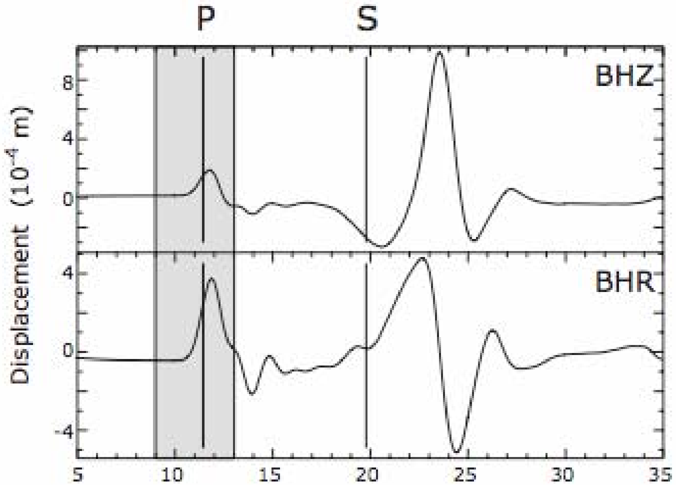
Time window: P arrival (Z, R)

Measurement: CC traveltime

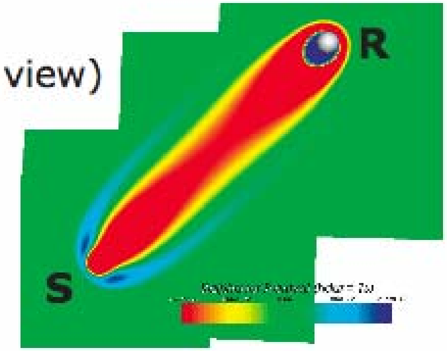
Model parameter: P-wave structure



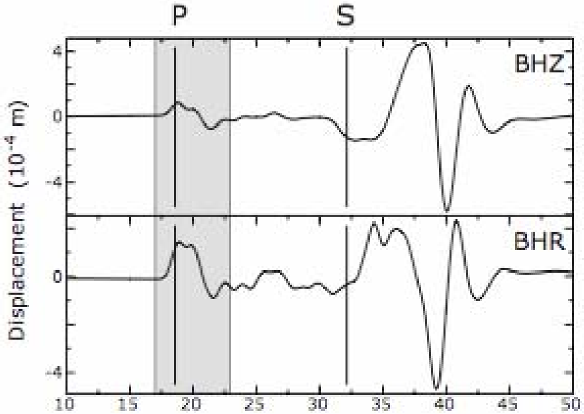




(Top view)



SEISMOLOGICAL LABORATORY
California Institute of Technology



SENSITIVITY KERNEL

Distance: Δ = 112 km

Minimum period: 2 s

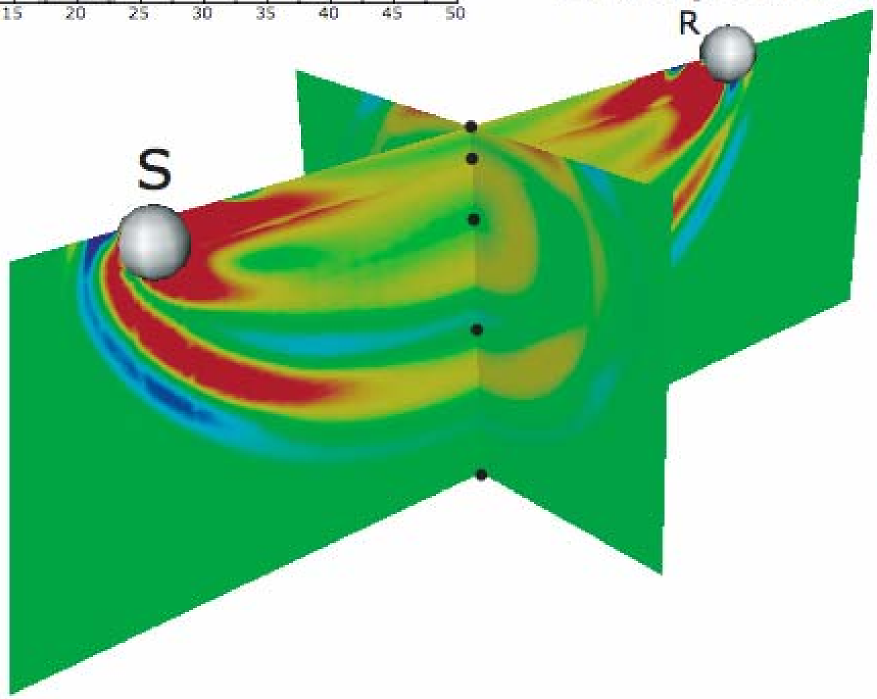
Reference model: SoCal-1D

Source: Explosion (depth 2.75 km)

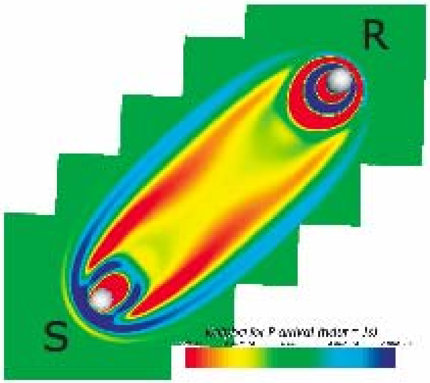
Time window: P arrival (Z, R)

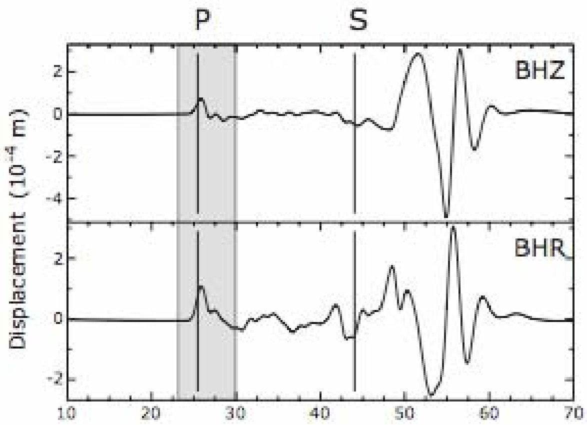
Measurement: CC traveltime

Model parameter: P-wave structure



(Top view)





SENSITIVITY KERNEL

Distance: Δ = 157 km

Minimum period: 2 s

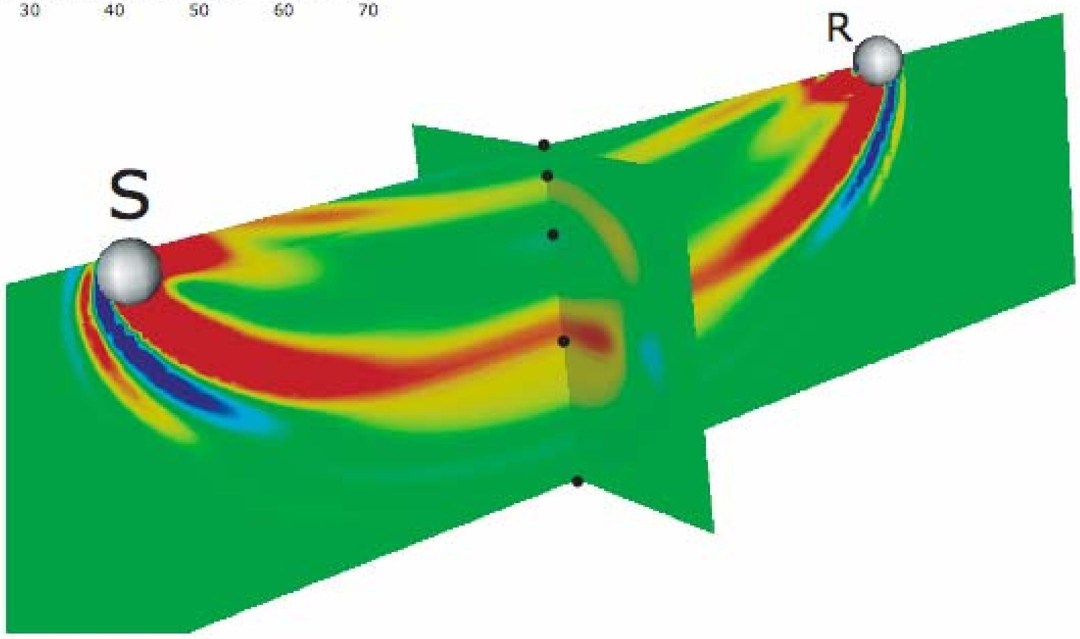
Reference model: SoCal-1D

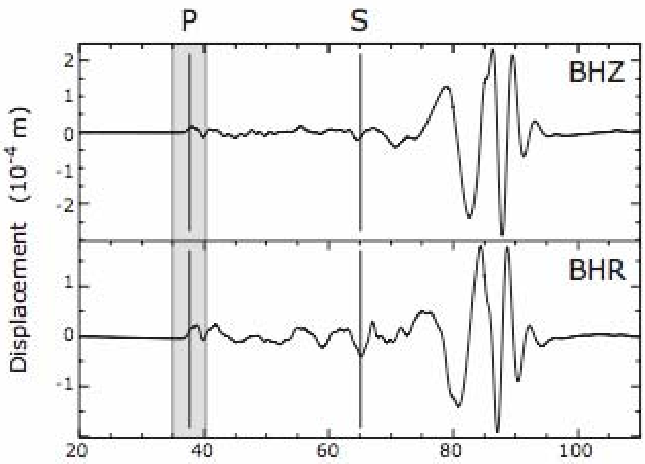
Source: Explosion (depth 2.75 km)

Time window: P arrival (Z, R)

Measurement: CC traveltime

Model parameter: P-wave structure





SENSITIVITY KERNEL

Distance: Δ = 247 km

Minimum period: 2 s

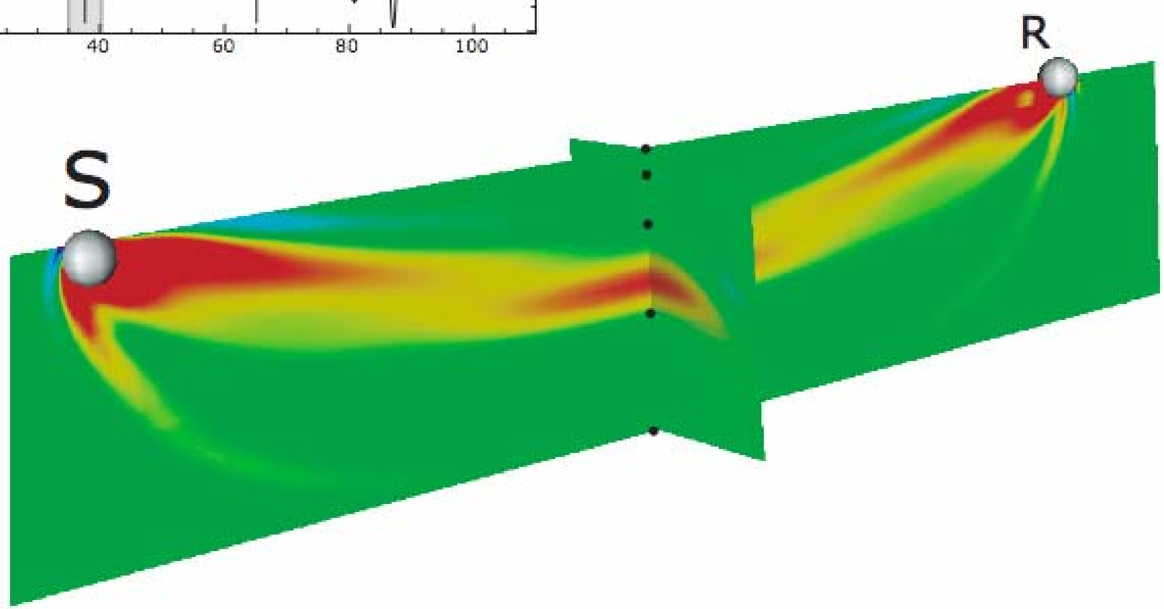
Reference model: SoCal-1D

Source: Explosion (depth 2.75 km)

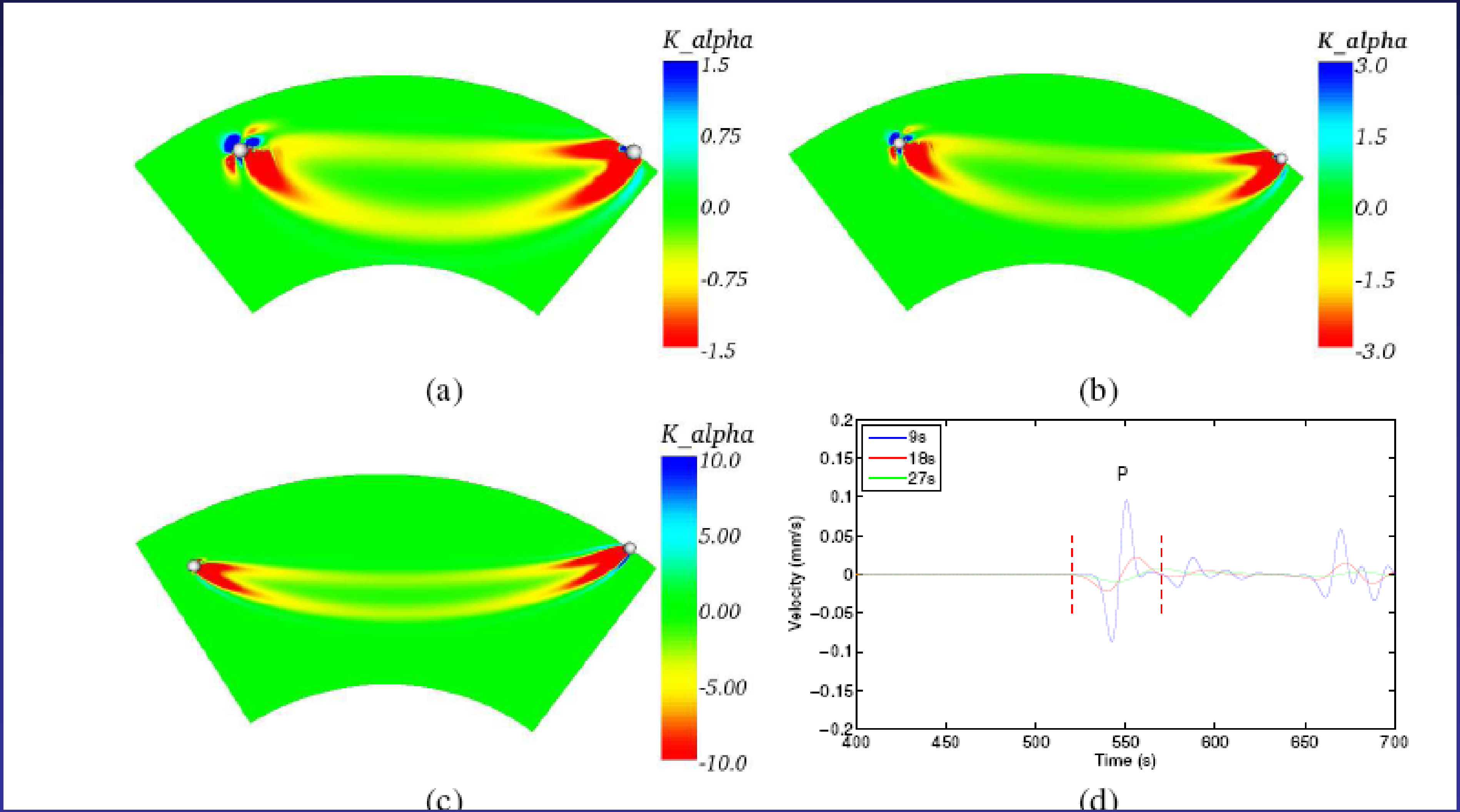
Time window: P arrival (Z, R)

Measurement: CC traveltime

Model parameter: P-wave structure

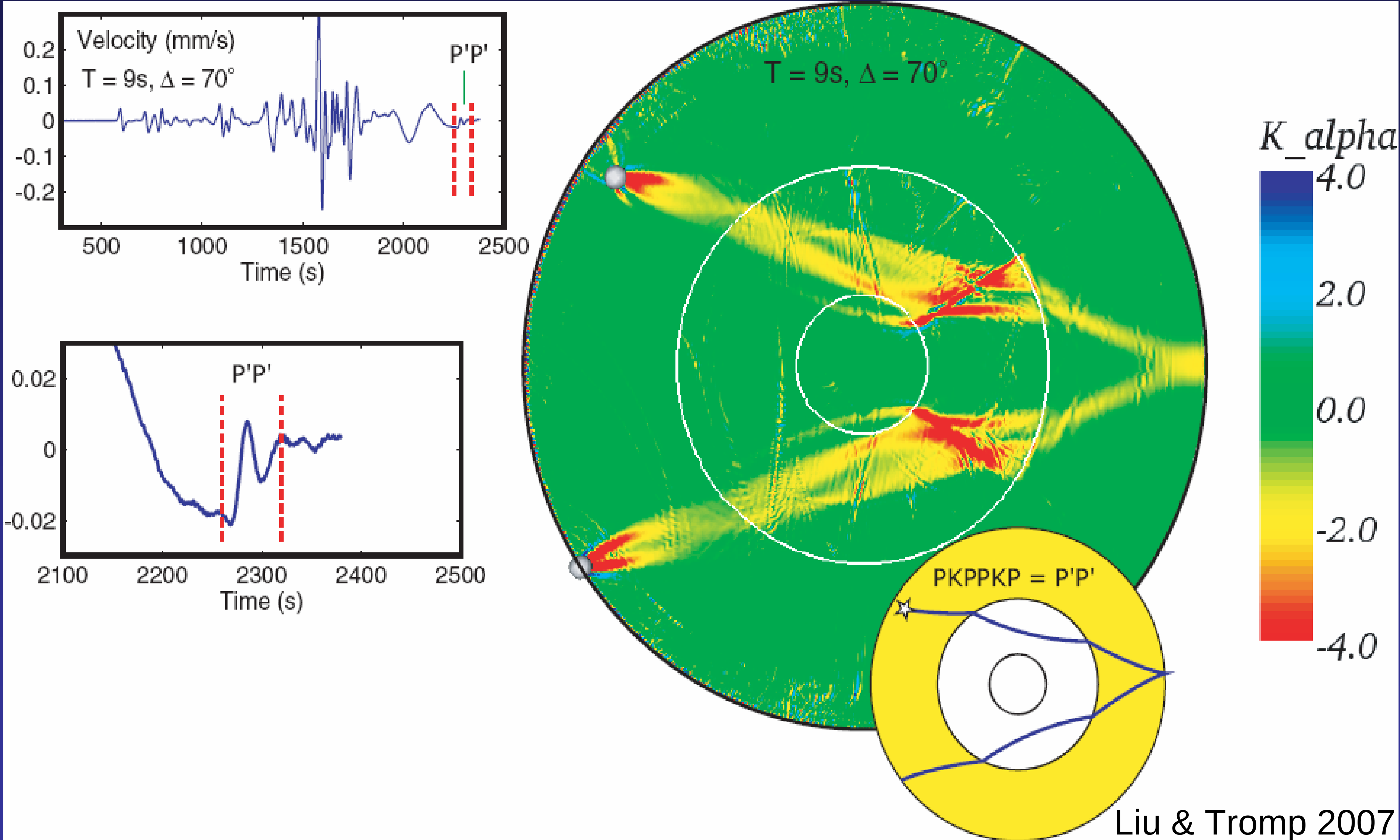


P waves: Finite-Frequency Effects



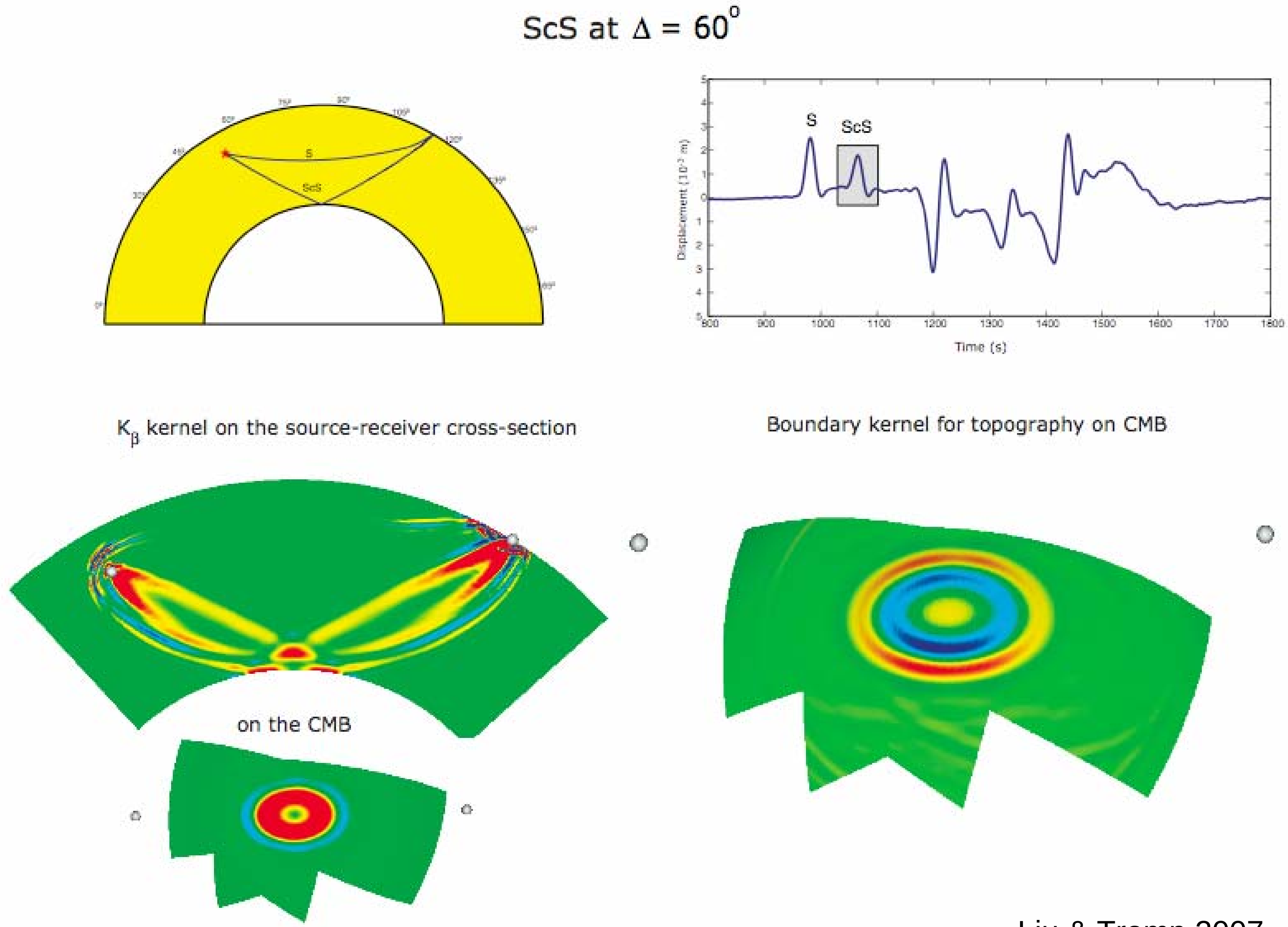
Liu & Tromp 2007

P'P' Kernel



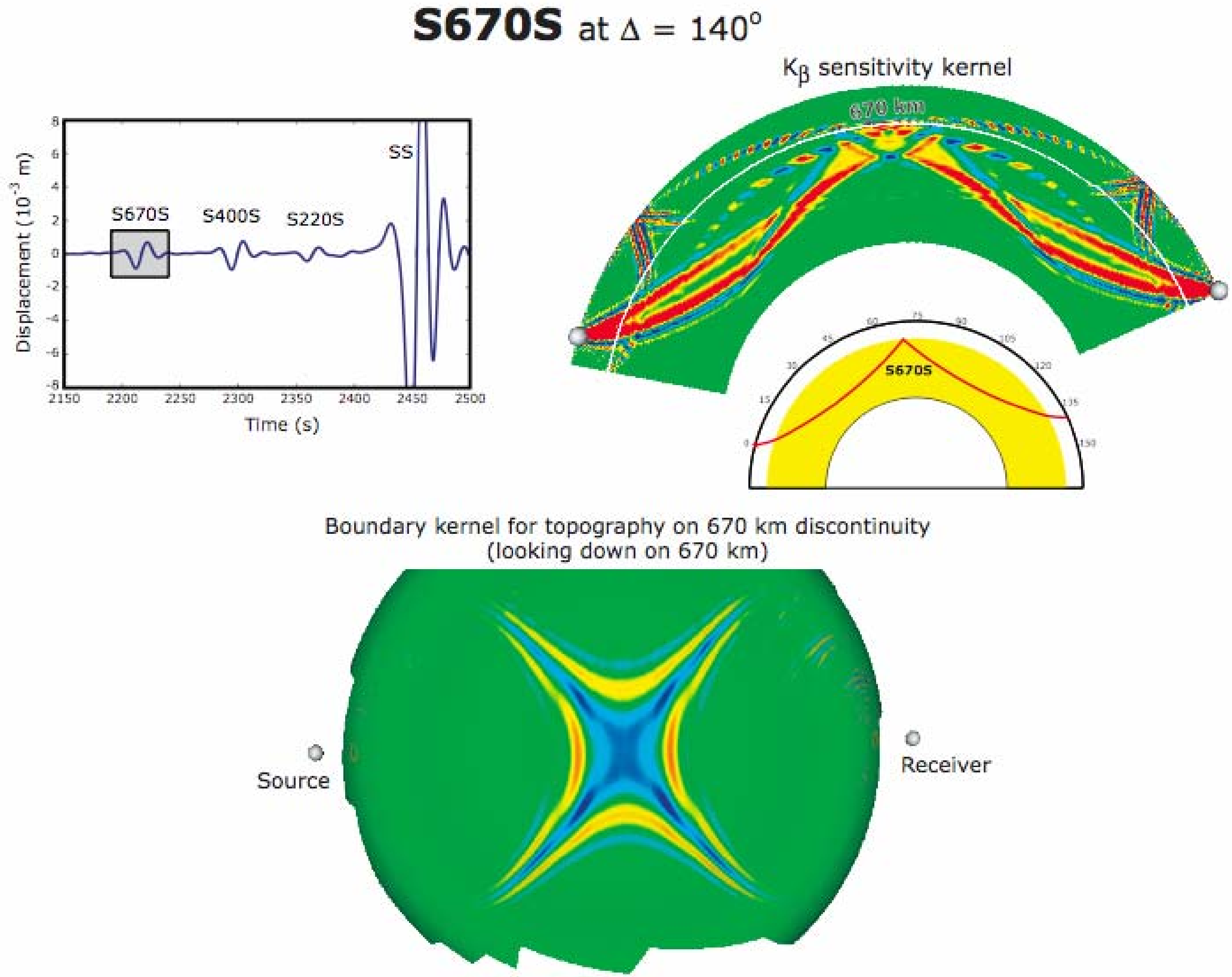
Liu & Tromp 2007

Boundary Kernels



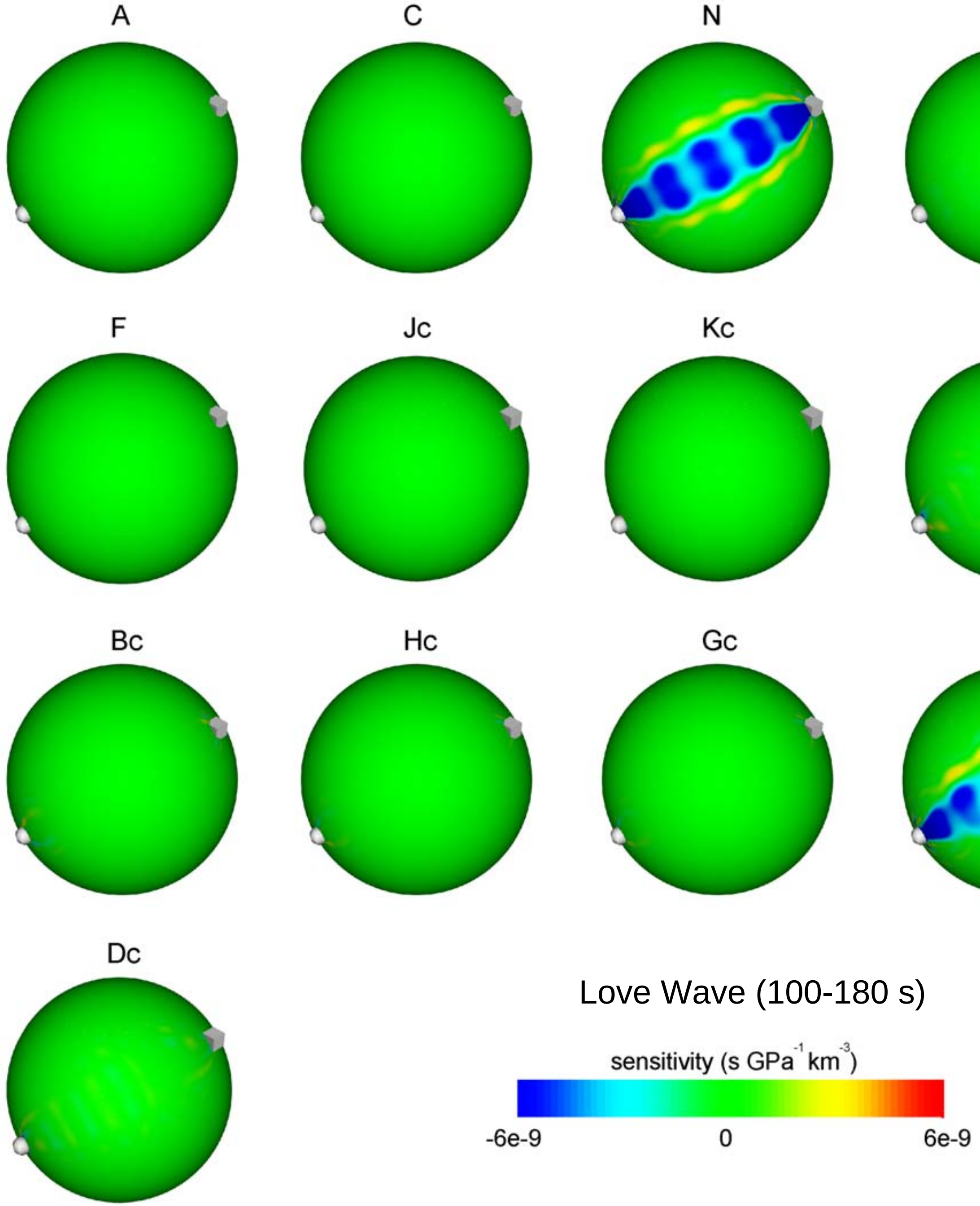
Liu & Tromp 2007

Boundary Kernels



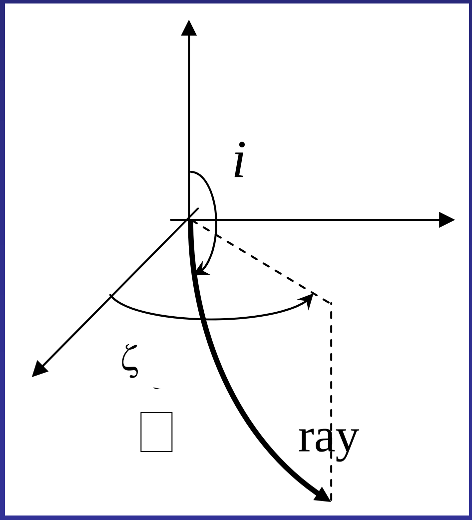
Liu & Tromp 2007

Anisotropic Surface-Wave Sensitivity Kernels



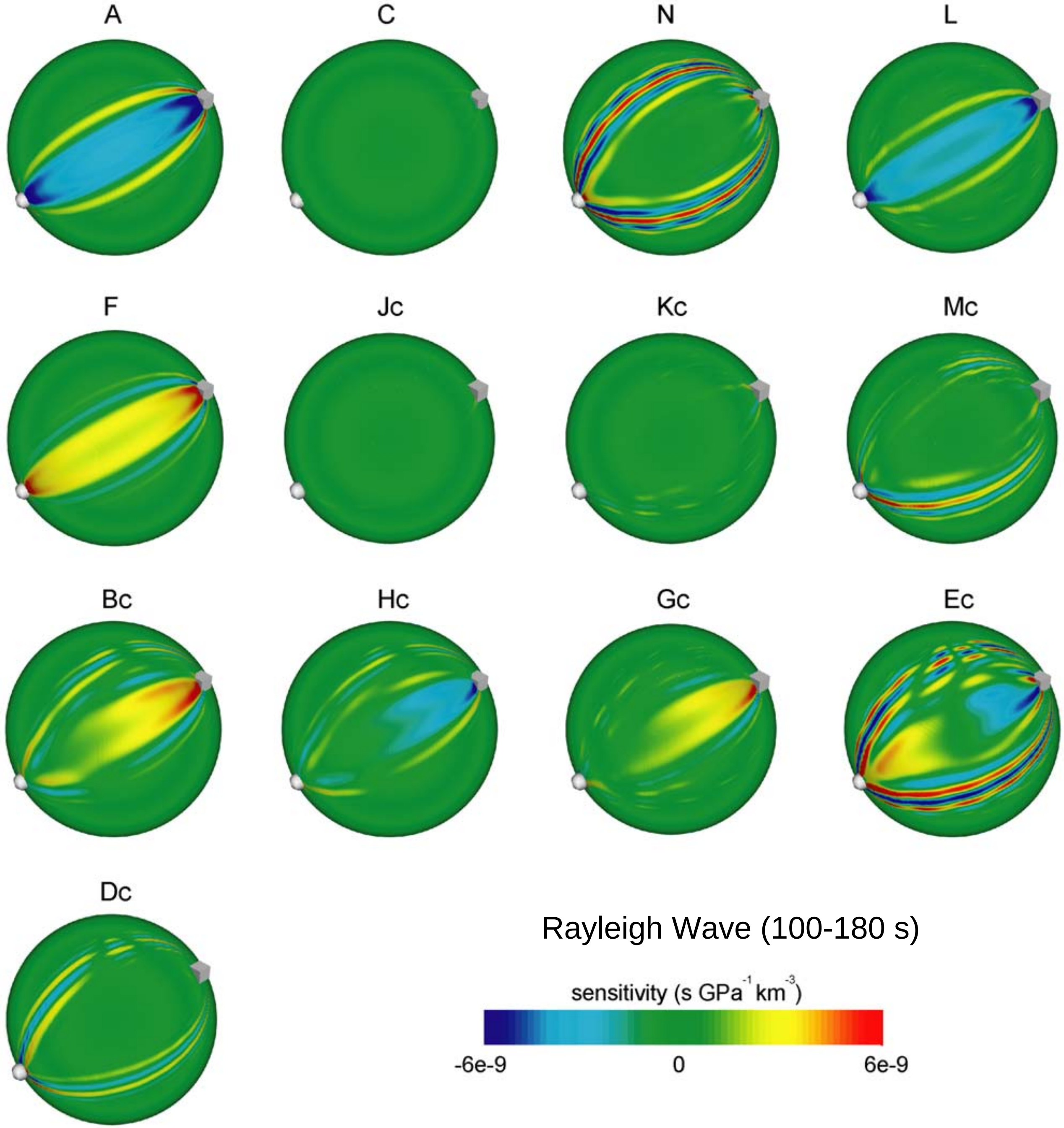
Smith & Dahlen, Montagner & Nataf: Surface waves are governed by 5 transversely isotropic plus 8 azimuthal parameters

Transversely Isotropic Parameters: A, C, L, N, F
1 ζ: J, K, M
2 ζ: G, B, H
3 ζ: D
4 ζ: E



Sieminski et al. 2006

Anisotropic Surface-Wave Sensitivity Kernels



Transversely Isotropic Parameters: A, C, L, N, F

1 ζ: J, K, M

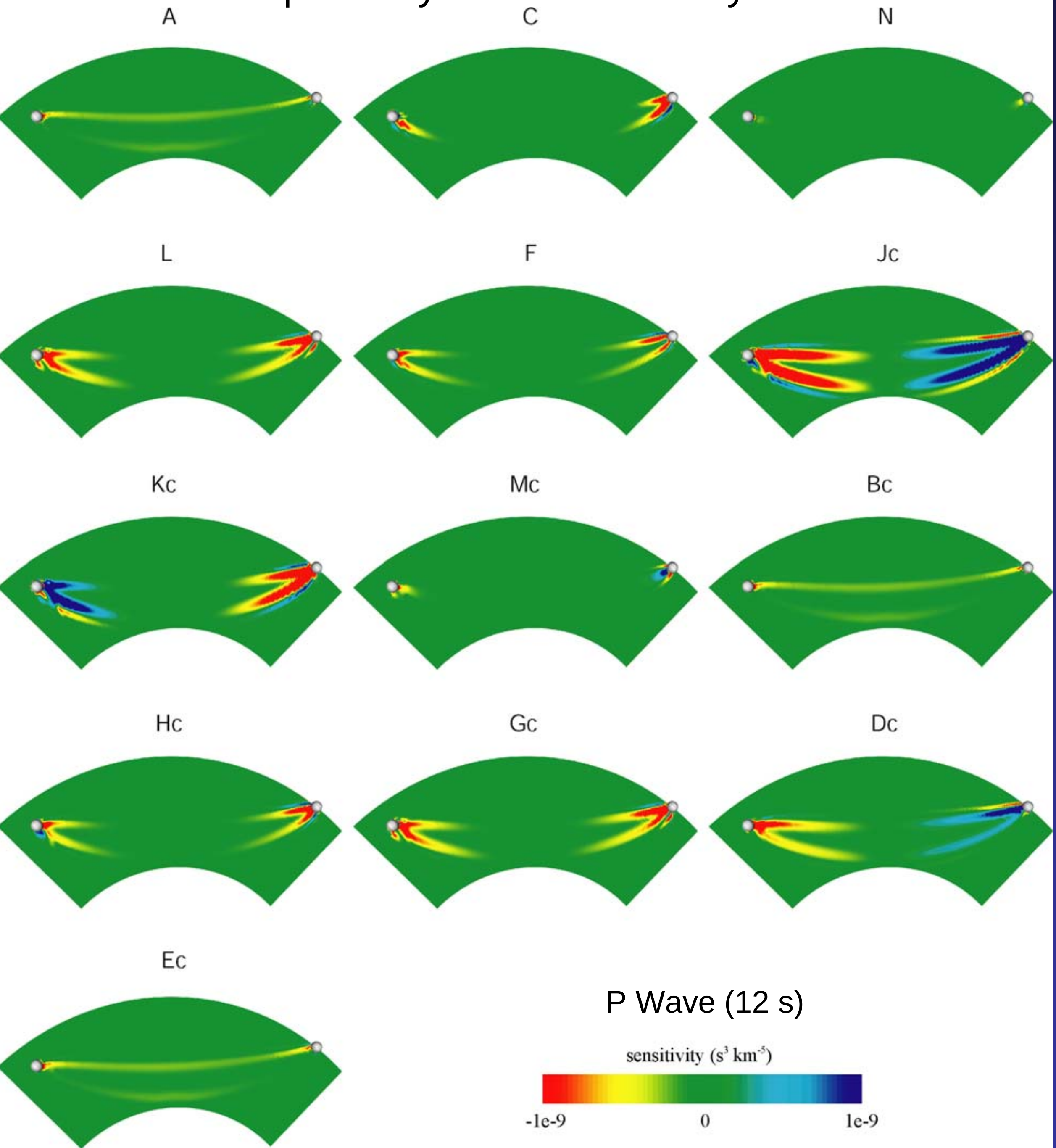
2 ζ: G, B, H

3 ζ: D

4 ζ: E

Sieminski et al. 2006

Anisotropic Body-Wave Sensitivity Kernels



Transversely Isotropic Parameters: A, C, L, N, F

1 ζ: J, K, M

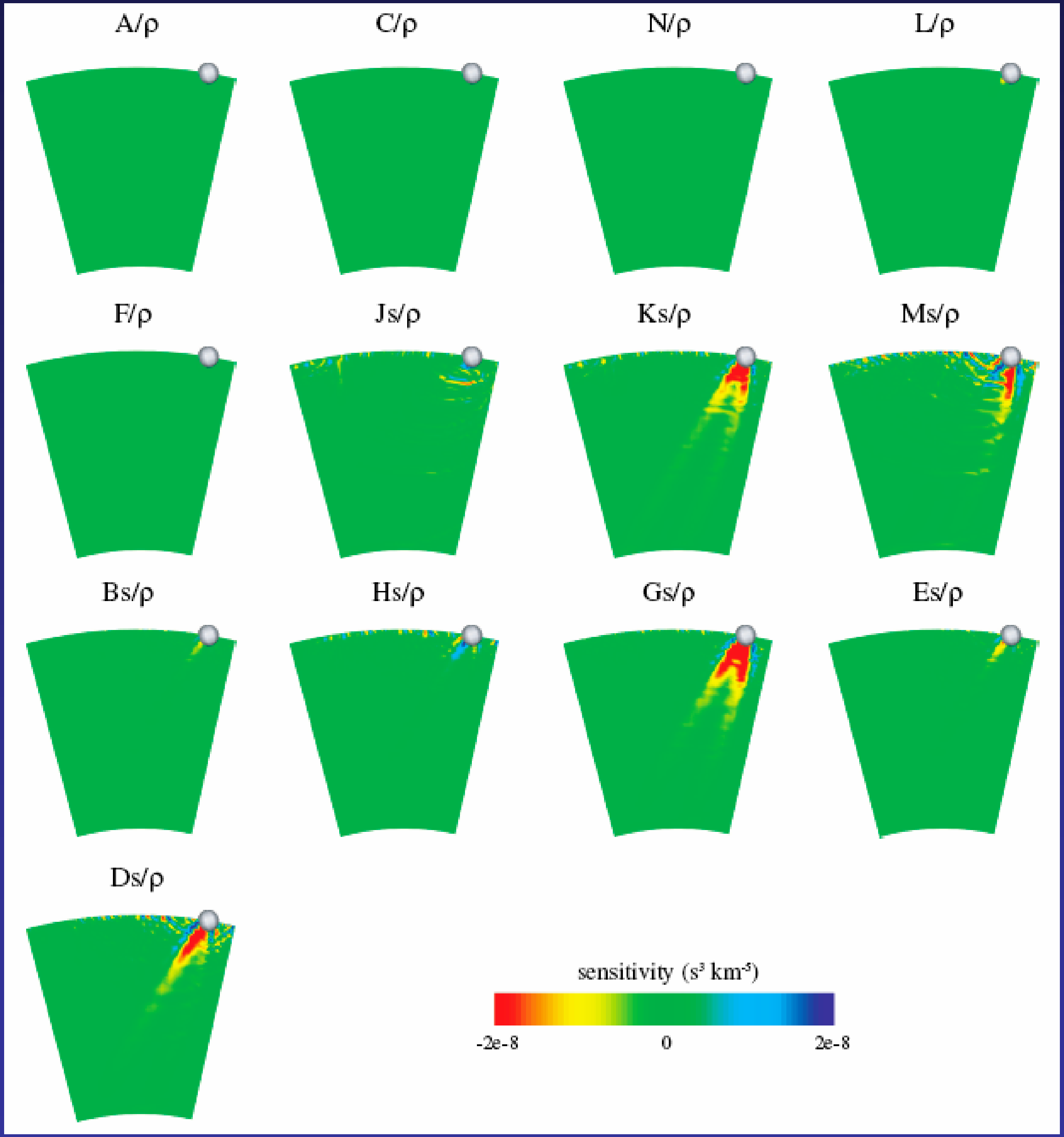
2 ζ: G, B, H

3 ζ: D

4 ζ: E

Sieminski et al. 2007

SKS Splitting Intensity



Transversely Isotropic Parameters: A, C, L, N, F
1 ζ: J, K, M
2 ζ: G, B, H
3 ζ: D
4 ζ: E

Not just Gc & Gs!

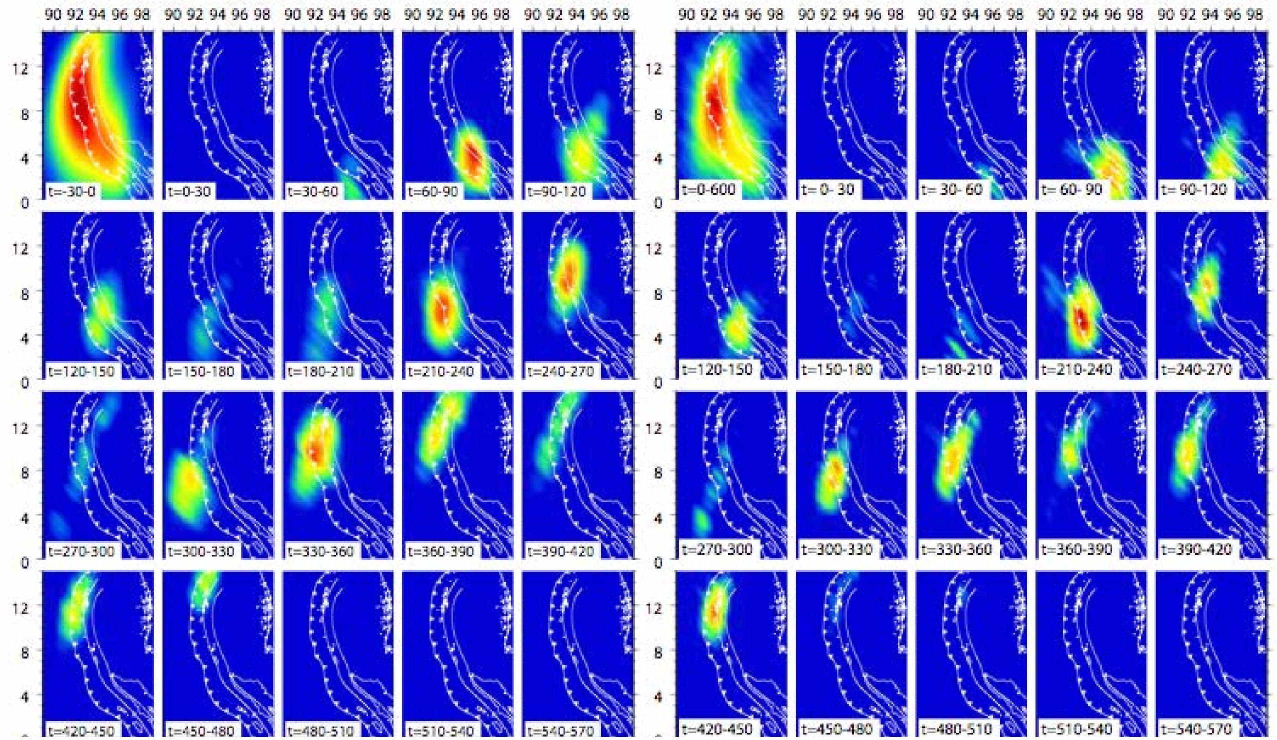
Sieminski et al. 2007

Adjoint Source Imaging

Ishii et al. (2005)

STACKS FILTERED @ 3 SEC

ADJOINT FIELD @ 3 SECONDS

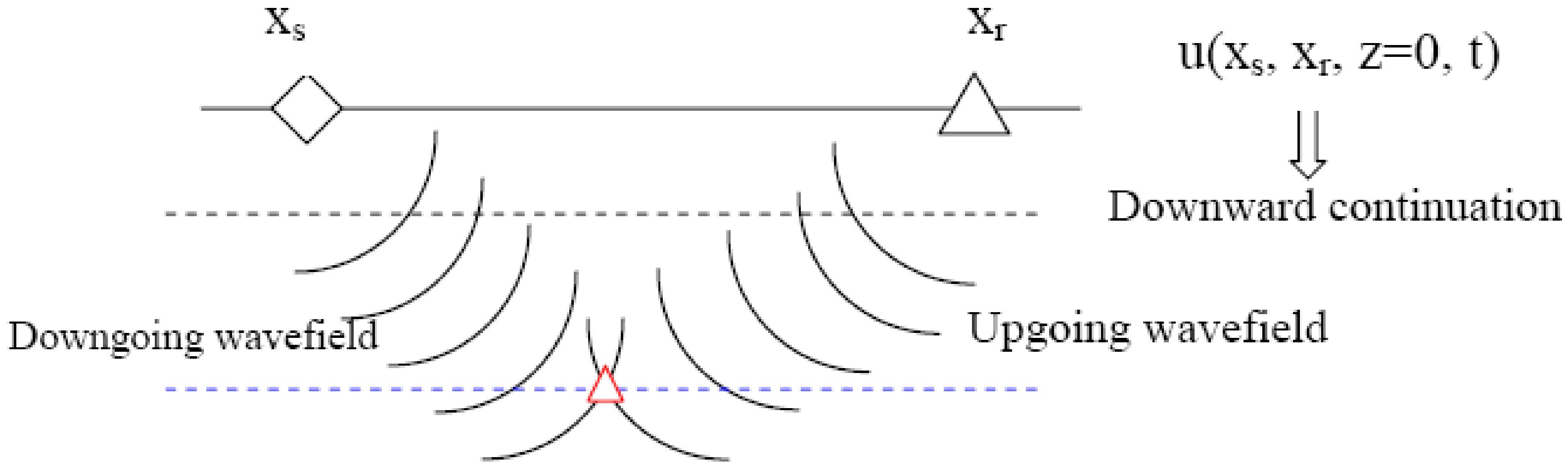


$$\delta\chi = \int_\Omega (\delta\rho K_\rho + \delta\mathbf{c} :: \mathbf{K_c})\, d^3\mathbf{x} + \int_0^T \delta\mathbf{M} : \epsilon^\dagger(\mathbf{x}_s, T-t) S(t)\, dt$$

$$+ \int_0^T \mathbf{M} : (\delta\mathbf{x}_s \cdot \nabla_s)\epsilon^\dagger(\mathbf{x}_s, T-t) S(t)\, dt + \int_0^T \mathbf{M} : \epsilon^\dagger(\mathbf{x}_s, T-t)\delta S(t)\, dt$$

Vala Hjorleifsdottir

Adjoint Methods and the "Imaging Principle"

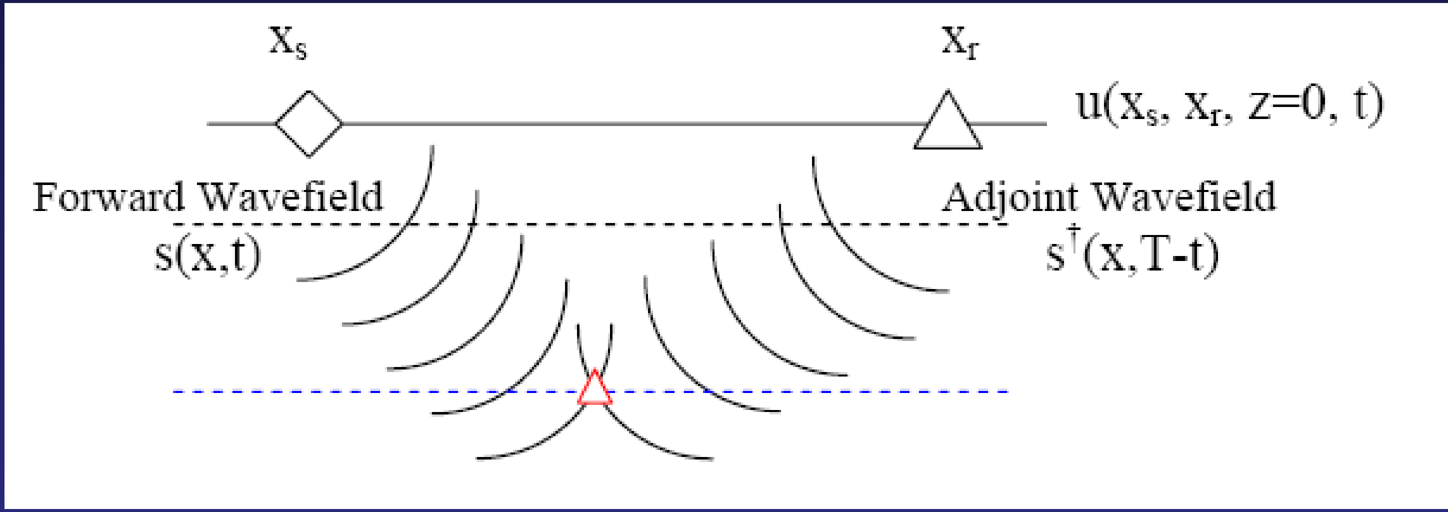


Imaging Principle:
Reflectors exist when up and down going waves coincide both in time and space

(Claerbout 1971)

$$I(\mathrm{x}_s, \mathrm{x}) = \frac{1}{P(\mathrm{x}_s, \mathrm{x})} \int u_i(\mathrm{x}_s, \mathrm{x}, \omega) u_b^*(\mathrm{x}_s, \mathrm{x}, \omega)\, d\omega$$

Adjoint Method (Hand-waving version)



$$\rho \ddot{\mathbf{s}}^{\dagger} = \nabla \cdot \mathbf{T}^{\dagger} + \mathbf{u}(T-t)\delta(\mathbf{x}-\mathbf{x}_r)$$

$$\text{B.C.} \quad \hat{z} \cdot \mathbf{T}^{\dagger} = 0$$

$$\text{I.C. at } t = 0$$

$$K(\mathbf{x}) \sim \int \mathbf{s}(\mathbf{x}, t)\, \mathbf{s}^{\dagger}(\mathbf{x}, T-t)\, dt$$

$$I(\mathbf{x}_s, \mathbf{x}) = \frac{1}{P(\mathbf{x}_s, \mathbf{x})} \int u_i(\mathbf{x}_s, \mathbf{x}, \omega) u_b^*(\mathbf{x}_s, \mathbf{x}, \omega)\, d\omega$$

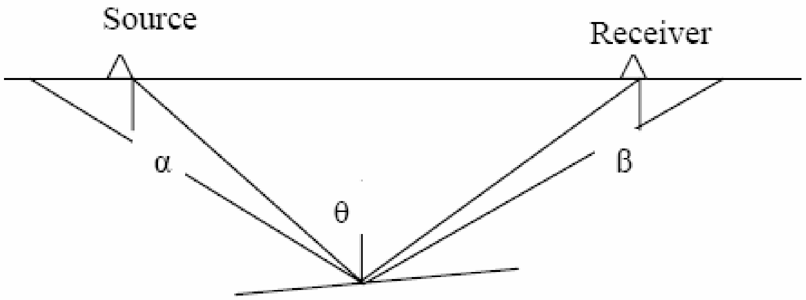
(Tarantola 1984)

Modified Imaging Principle

$$I(\mathbf{x}_s, \mathbf{x}) = \frac{1}{P(\mathbf{x}_s, \mathbf{x})} \int u_i(\mathbf{x}_s, \mathbf{x}, \omega) u_b^*(\mathbf{x}_s, \mathbf{x}, \omega) \frac{R(0)}{R(\theta)} d\omega \qquad \frac{R(0)}{R(\theta)} = \cos^2\theta$$

$$= \frac{1}{P(\mathbf{x}_s, \mathbf{x})} \int u_i(\mathbf{x}_s, \mathbf{x}, \omega) u_b^*(\mathbf{x}_s, \mathbf{x}, \omega) \boxed{\cos^2\theta}\, d\omega$$

$$\cos^2\theta \sim 1 + (\sin\alpha \sin\beta + \cos\alpha\cos\beta) \sim 1 + (\partial_x \partial_x^\dagger + \partial_z \partial_z^\dagger)$$



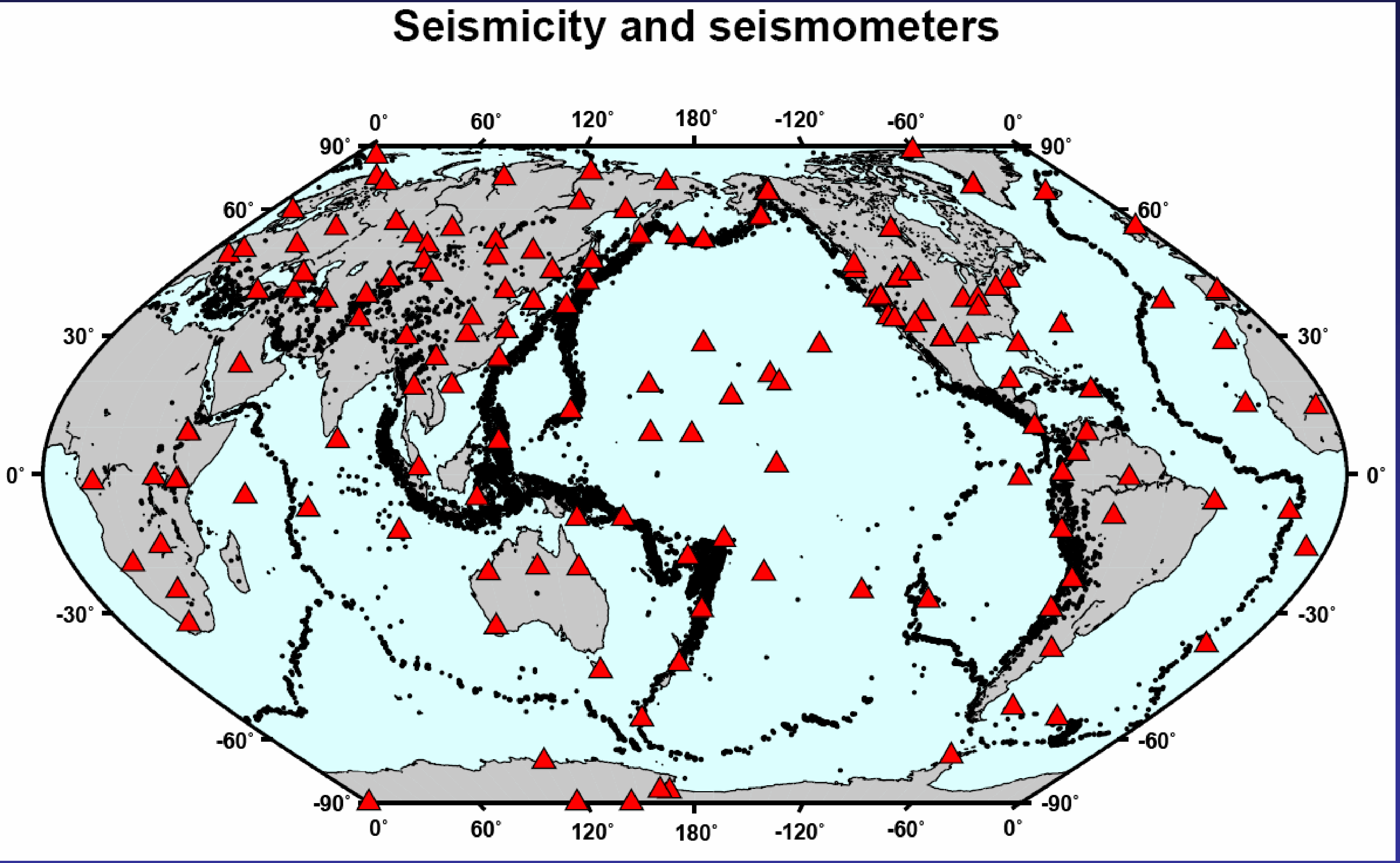
$$= \frac{1}{P(\mathbf{x}_s, \mathbf{x})} \int [u_i(\mathbf{x}_s, \mathbf{x}, \omega) u_b^*(\mathbf{x}_s, \mathbf{x}, \omega) - \frac{1}{k^2}(\partial_x u_i(\mathbf{x}_s, \mathbf{x}, \omega) \partial_x u_b^*(\mathbf{x}_s, \mathbf{x}, \omega) + \partial_z u_i(\mathbf{x}_s, \mathbf{x}, \omega) \partial_z u_b^*(\mathbf{x}_s, \mathbf{x}, \omega))] d\omega$$

If we identify $\dot{\mathrm{s}}^\dagger \sim u_b,\ \dot{s} \sim u_i,\ \rho\omega^2/k^2 \sim \kappa$

This looks like our $K_\rho + K_\kappa \sim -K_\mu$

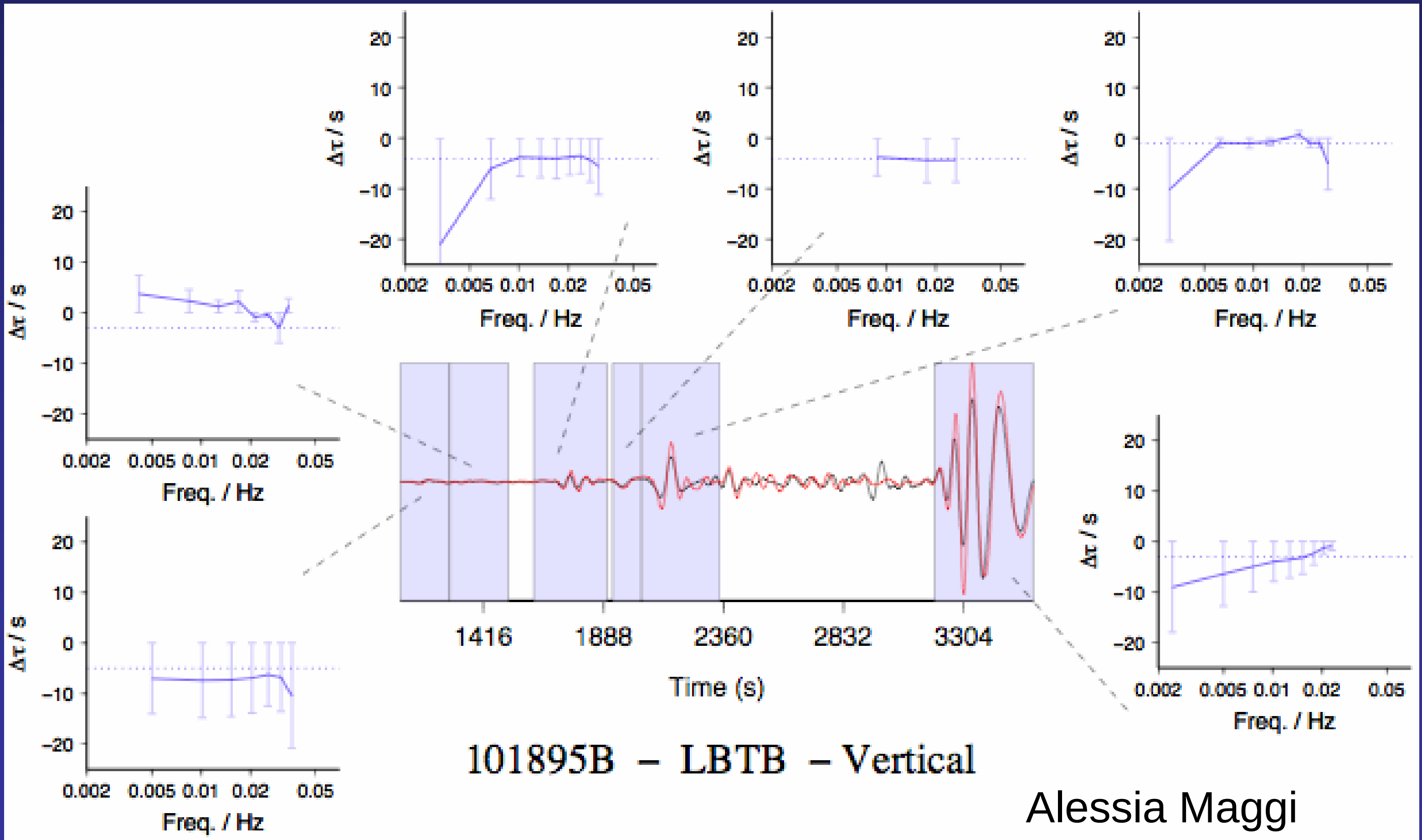
(Kiyashchenko et al. 2007)

Measure all suitable phases



Automated Measurements

- Window selection based upon short-term/long-term averages
- Window rejection based upon data/synthetics similarity criteria
- Measurement based upon either:
  - Waveform differences
  - Cross-correlation phase & amplitude anomalies
  - Multi-Taper phase & amplitude measurements

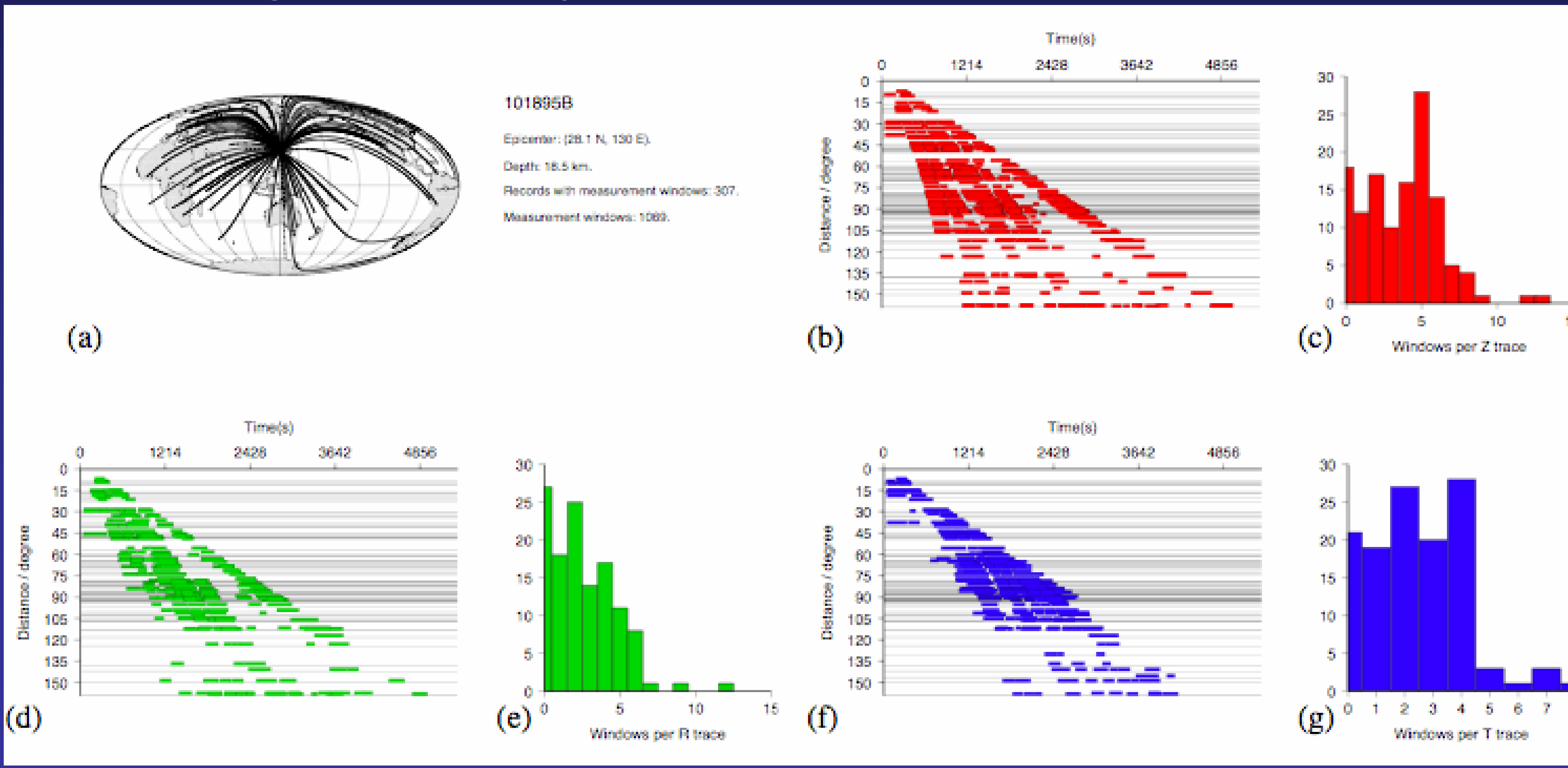


Alessia Maggi

Automated Measurements

1995 shallow event:

- 307 records
- 1069 windows
- average of 3 windows per record

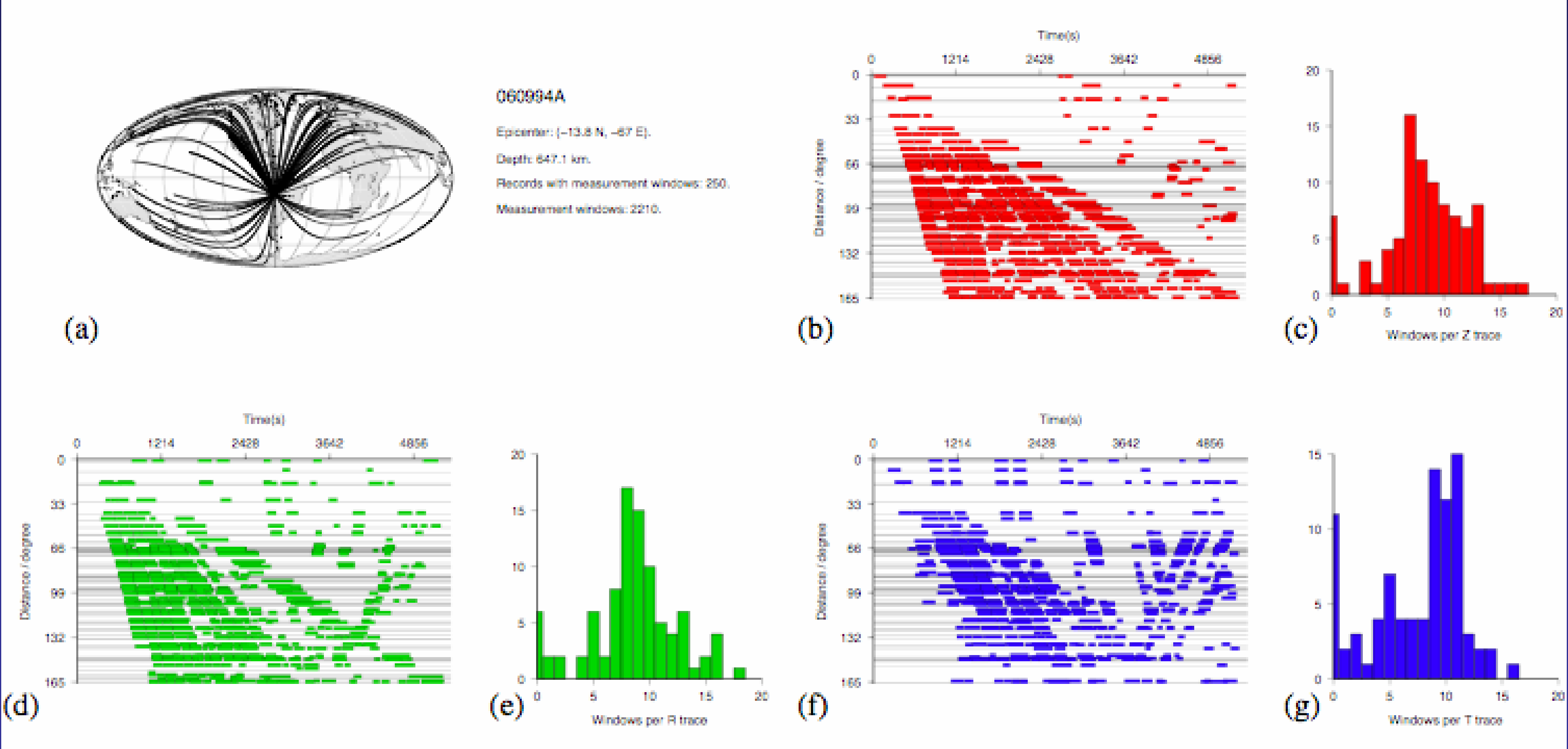


Alessia Maggi

Automated Measurements

1994 deep event:

- 250 records
- 2210 windows
- average of 9 windows per record



Alessia Maggi

Conclusions

Adjoint methods:

- Choose an observable, e.g., waveforms or cross-correlation traveltimes
- Choose a measure of misfit, e.g., least-squares
- Determine the appropriate adjoint source for this observable & measurement
- Use fully 3D reference models
- Any arrival suitable for measurement
- No dependence on the number of stations, components, or measurements
- 3D sensitivity kernels may be calculated based upon two forward simulations for each earthquake
- Number of simulations: 3 * (# earthquakes) * (# iterations)
- Full anisotropy for the same cost

Regional simulations:

- One 3 min forward simulation accurate to 1.5 s takes 45 min on a 75 node cluster
- 150 events and 3 iterations would require 1800 simulations, i.e., three weeks of dedicated CPU time on 75 nodes
- Regional tomography
- Near real-time simulations

Global simulations:

- One 1 h forward simulation accurate to 20 s takes 4 h on a 75 node cluster
- 500 events and 3 iterations requires 6,000 simulations, i.e., 100 days on a 750 node cluster
- Global tomography
- Near real-time simulations
- On-demand global seismology
- Petascale application