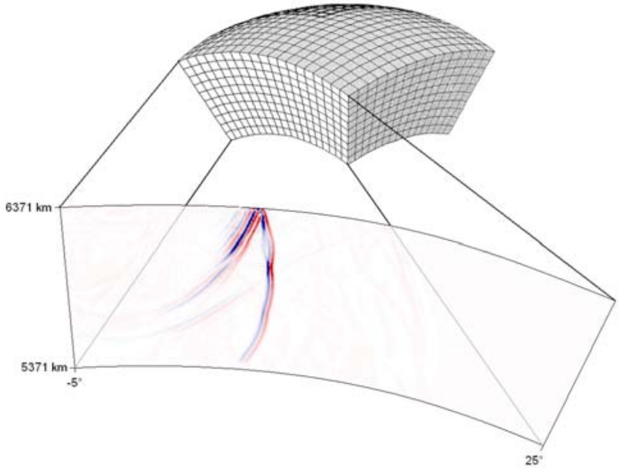
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
GEOPHYSICS
DEPARTMENT OF EARTH AND ENVIRONMENTAL SCIENCES

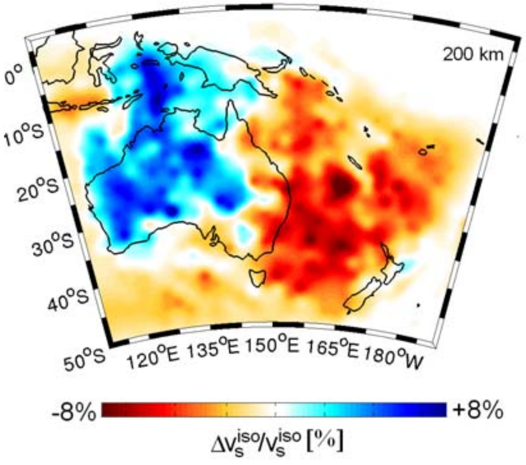
Short Course

Full Seismic Waveform Inversion

The Adjoint Method

Andreas Fichtner





OUTLOOK

1. Introduction

seismic tomography – purposes and difficulties

2. Waveform tomography

setup as optimisation problem

3. The Adjoint Method

motivation & general concept

sensitivity kernels

objective functionals: cross-correlation time shifts, rms amplitude differences, …

4. Solution of the optimisation problem

preparatory steps & initial model

conjugate-gradient algorithm

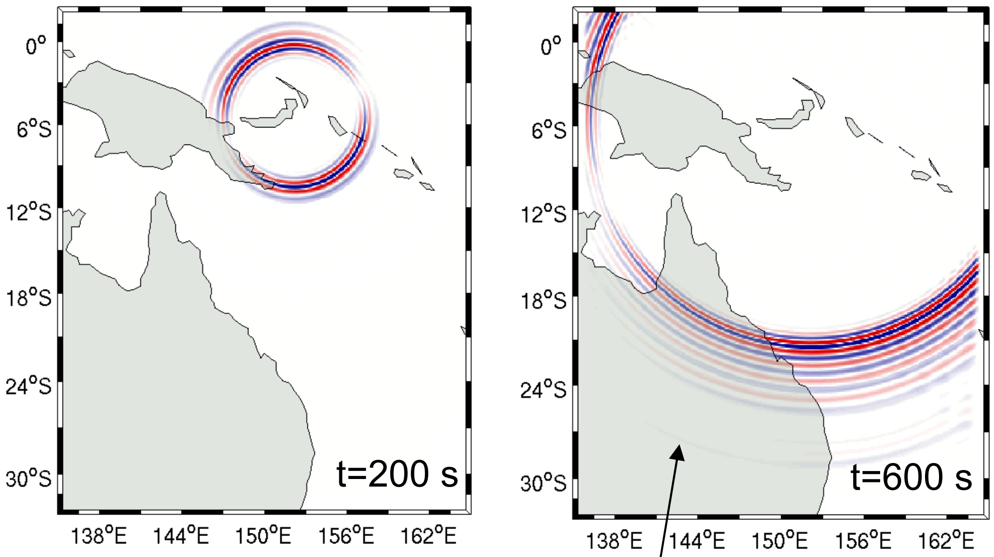
preconditioning

5. Results from our work

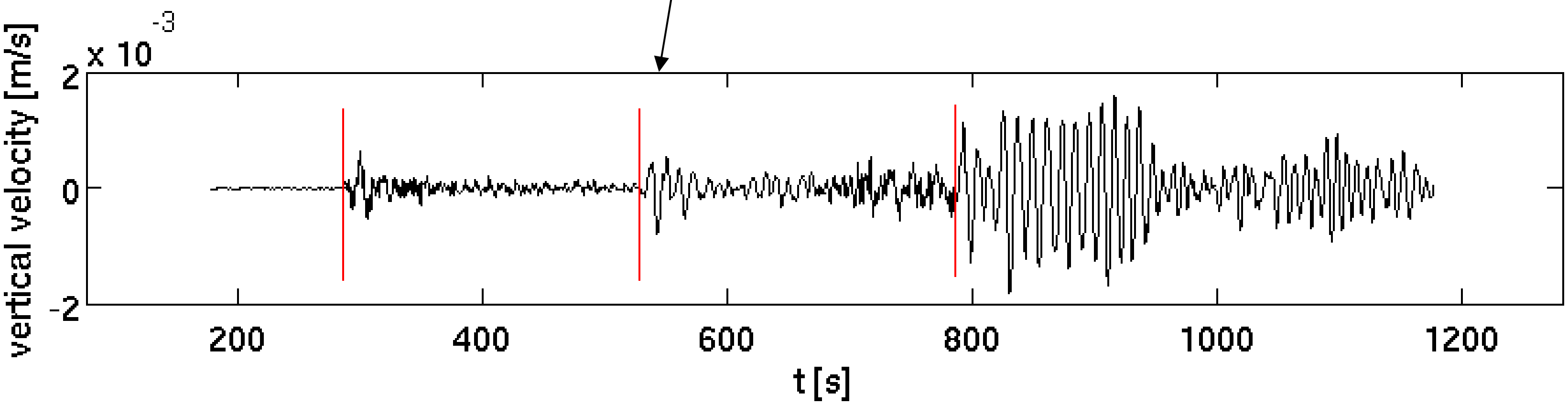
1. Introduction

INTRODUCTION: seismic tomography

wave field snapshots, horizontal slices @ 100 km depth



- **Earthquakes or active sources** excite waves that travel through the Earth.
- Recorded as **seismograms**.
- Waveforms depend on the Earth's **structure**.
- **Seismic tomography**: Infer Earth structure from seismograms.



- **High Performance Computing**: Simulation of elastic wave propagation through highly heterogeneous Earth models
- Exploit **more information** from seismograms → **More accurate Earth models.**

INTRODUCTION: objectives

Principal objectives of seismic tomography:

- infer the present **dynamics and evolution of the Earth**
- produce accurate Earth models for **reliable tsunami warnings** …
- … and the monitoring of the **Comprehensive Nuclear Test-Ban Treaty**
- search for **natural resources**
- assessment of **construction ground properties**
- monitoring of subsurface processes (**CO_2 storage, magma chambers**)
- …

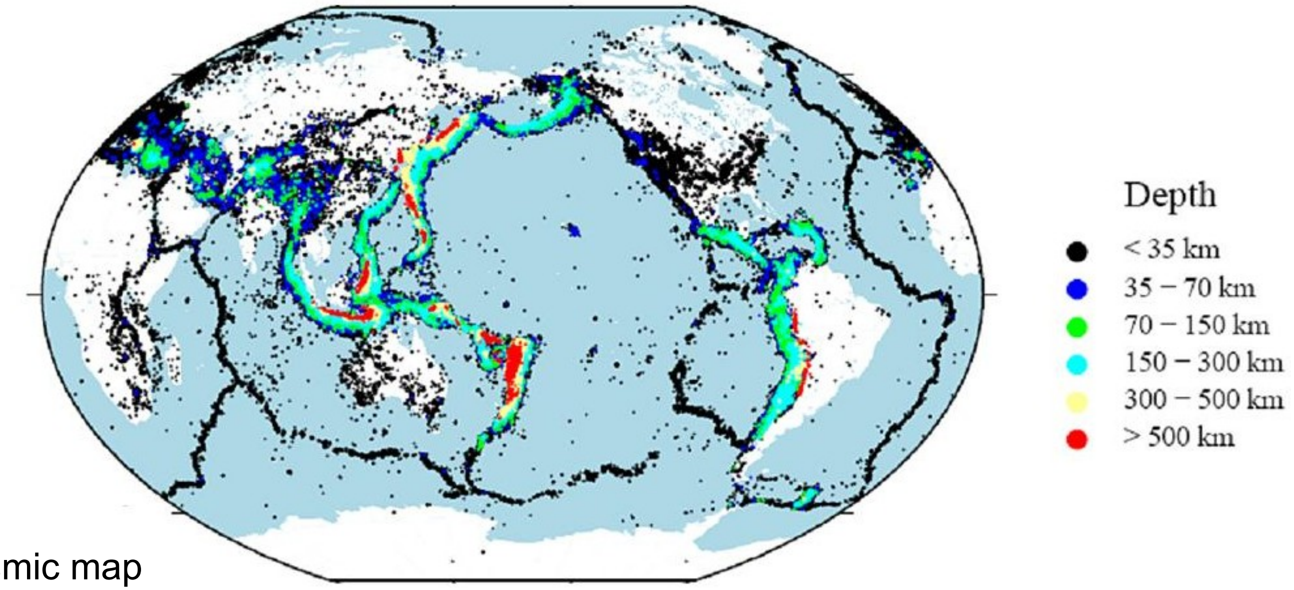
Principal objectives of seismic tomography:

- **non-destructive testing** in material sciences
- **medical imaging**
- …

INTRODUCTION: difficulties of seismic tomography

What makes seismic tomography difficult?

seismic tomography	medical tomography
• **receivers:** unevenly distributed	• **receivers:** well distributed
• **sources:** few and poorly distributed	• **sources:** well distributed
• **source mechanisms:** unknown	• **source mechanisms:** known
• **wave field:** highly complex	• **wave field:** rather simple
• **measurement:** very challenging	• **measurement:** rather simple
• **inversion:** non-linear and ill-posed	• **inversion:** rather well-posed, linear



World seismic map

2. Waveform tomography (general)

WAVEFORM TOMOGRAPHY: Setup as optimisation problem

Seismic tomography is usually set as a non-linear optimisation problem.

1 Solution of the forward problem

- Elastic wave equation
- No analytical solutions in realistically heterogeneous Earth model
- Numerical methods (FD, FEM, SEM, …)

2 Comparison of data and synthetics

- Physically meaningful measures of misfit that are applicable to imperfect data
- Physical intuition required

3 Gradient of misfit functional w/r model

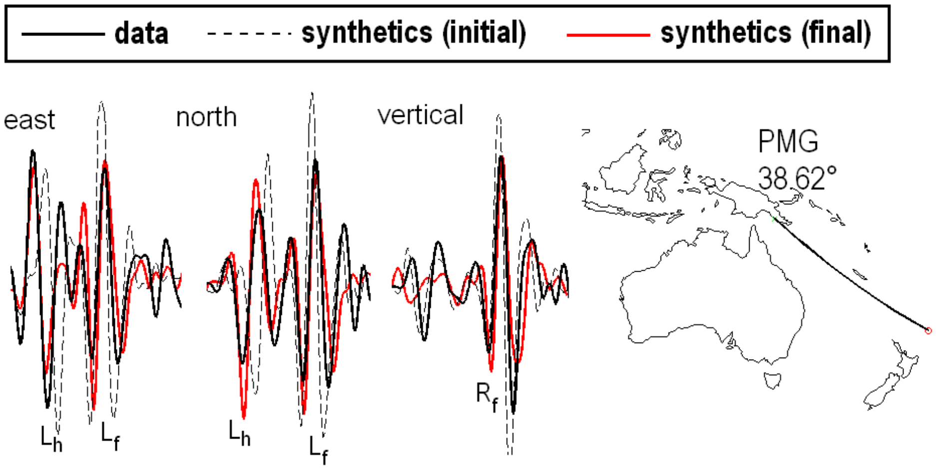
- Sensitivity densities (Fréchet kernels) via the adjoint method
- Choice of the model parameterisation

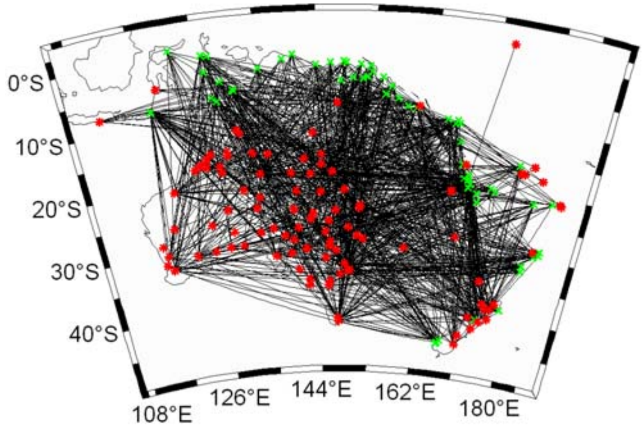
4 Iterative misfit minimisation

- steepest descent, conjugate gradients, …
- regularisation
- pre-conditioning

WAVEFORM TOMOGRAPHY: Character and size of the problem

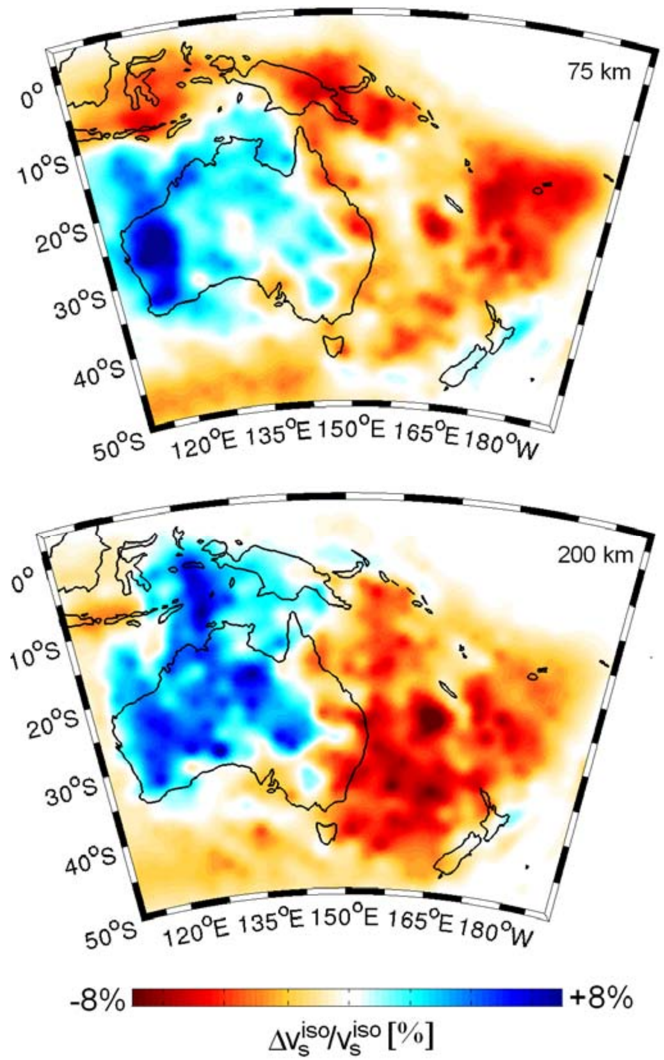
Data





≈ 3000 waveforms

Model



≈ 500 000 free parameters

WAVEFORM TOMOGRAPHY: Character and size of the problem

500 000 free parameters

+

gradient methods for misfit minimisation

⬇

500 000 partial derivatives of the misfit with respect to the model parameters

Different ways of computing partial derivatives:

- **Finite differencing:**

$$\frac{\partial \text{misfit(parameters)}}{\partial \text{parameter}_i} \approx \frac{\text{misfit(parameters} + \Delta\text{parameter}_i) - \text{misfit(parameters)}}{\Delta\text{parameter}_i}$$

	500,001 forward simulations
x	0.5 h per simulation
x	126 processors
x	50 earthquakes
x	4 simulations per conjugate gradient iteration
x	10 conjugate gradient iterations
	$6.3e^{10}$ cpu hours ≈ 720,000 cpu years

Inaccurate because we divide small numbers by small numbers.

- **Automatic differentiation:** knows nothing about physics → inefficient
- **Adjoint method**:

3. **The adjoint method**

3.1. General concept

ADJOINT METHOD

The adjoint method is a mathematical trick that allows us to compute the exact partial derivatives with 2 instead of 500,001 simulations.

Very general and widely used (meteorology, ground water modelling, optimal control problems, …)

The relevant equations can be derived in different ways:

1) Lagrange multiplier method (e.g. Liu & Tromp, GJI 2008)

2) Perturbation theory (e.g. Tarantola, PAGEOPH 1988)

3) Operator formulation (Fichtner et al., PEPI 2006)

- very compact
- applicable to any PDE, including the elastic wave equation

ADJOINT METHOD: Application to the elastic wave equation

Elastic wave equation:

$$\mathbf{L}(\mathbf{u},\rho,\mathbf{C}) = \mathbf{f}$$

$$\mathbf{L}(\mathbf{u},\rho,\mathbf{C}) = \rho(\mathbf{x})\partial_t^2\mathbf{u}(\mathbf{x},t) - \nabla\cdot\int_{-\infty}^{\infty}\dot{\mathbf{C}}(\mathbf{x},t-\tau):\nabla\mathbf{u}(\mathbf{x},t)\,d\tau$$

Subsidiary conditions:

$$\mathbf{u}(\mathbf{x},t)|_{t=t_0} = \mathbf{0} \qquad \partial_t\mathbf{u}(\mathbf{x},t)|_{t=t_0} = \mathbf{0} \qquad \mathbf{n}\cdot\int_{-\infty}^{t}\dot{\mathbf{C}}(\mathbf{x},t-\tau):\nabla\mathbf{u}(\mathbf{x},\tau)\,d\tau\,|_{\mathbf{x}\in\Gamma} = \mathbf{0}$$

Adjoint wave equation:

$$\mathbf{L}^t(\mathbf{u}^t,\mathbf{C},\rho) = \rho(\mathbf{x})\partial_t^2\mathbf{u}^t(\mathbf{x},t) - \nabla\cdot\int_{-\infty}^{\infty}\dot{\mathbf{C}}(\mathbf{x},\tau-t):\nabla\mathbf{u}^t(\mathbf{x},t)\,d\tau$$

$$\mathbf{L}^t(\mathbf{u}^t,\mathbf{C},\rho) = \mathbf{f}^t$$

← determined by the misfit measure

Adjoint subsidiary conditions:

$$\mathbf{u}^t(\mathbf{x},t)|_{t=t_1} = \mathbf{0} \qquad \partial_t\mathbf{u}^t(\mathbf{x},t)|_{t=t_1} = \mathbf{0} \qquad \mathbf{n}\cdot\int_{-\infty}^{t}\dot{\mathbf{C}}(\mathbf{x},\tau-t):\nabla\mathbf{u}^t(\mathbf{x},\tau)\,d\tau\,|_{\mathbf{x}\in\Gamma} = \mathbf{0}$$

3.2. Derivatives with respect to selected structural parameters

ADJOINT METHOD: Derivatives in an isotropic medium

Elastic tensor for isotropic media:

$$C_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu\delta_{ik}\delta_{jl} + \mu\delta_{il}\delta_{jk}$$

λ, μ: Lamé parameters

Partial derivatives:

$$D_\rho \mathfrak{E}(\mathbf{u})(\rho')|_{\lambda,\mu} = -\int_G \int_{t=t_0}^{t_1} \rho' \, \partial_t \mathbf{u}^\dagger \cdot \partial_t \mathbf{u} \, d^3\mathbf{x} \, dt \,,$$

$$D_\lambda \mathfrak{E}(\mathbf{u})(\lambda')|_{\mu,\rho} = \int_G \int_{t=t_0}^{t_1} \lambda' (\nabla \cdot \mathbf{u})(\nabla \cdot \mathbf{u}^\dagger) \, d^3\mathbf{x} \, dt \,,$$

$$D_\mu \mathfrak{E}(\mathbf{u})(\mu')|_{\lambda,\rho} = \int_G \int_{t=t_0}^{t_1} \mu' [(\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u}) + (\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u})^T] \, d^3\mathbf{x} \, dt$$

The terms under the volume integral are the *sensitivity kernels* or *Fréchet kernels* …

Fréchet kernels for an isotropic medium:

$$\delta_\rho \mathfrak{E}(\mathbf{u})|_{\lambda,\mu} = - \int_{t=t_0}^{t_1} \partial_t \mathbf{u}^\dagger \cdot \partial_t \mathbf{u} \, dt \, ,$$

$$\delta_\lambda \mathfrak{E}(\mathbf{u})|_{\mu,\rho} = \int_{t=t_0}^{t_1} (\nabla \cdot \mathbf{u})(\nabla \cdot \mathbf{u}^\dagger) \, dt \, ,$$

$$\delta_\mu \mathfrak{E}(\mathbf{u})|_{\lambda,\rho} = \int_{t=t_0}^{t_1} \left[(\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u}) + (\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u})^T \right] dt$$

ADJOINT METHOD: Derivatives in an isotropic medium

Partial derivatives with respect to different sets of parameters:

$$\kappa = \lambda + 2\mu/3$$

$$\delta_\rho \mathfrak{E}|_{\kappa,\mu} = \delta_\rho \mathfrak{E}\,,$$

$$\delta_\kappa \mathfrak{E}|_{\rho,\mu} = \delta_\lambda \mathfrak{E}\,,$$

$$\delta_\mu \mathfrak{E}|_{\rho,\kappa} = \delta_\mu \mathfrak{E} - \frac{2}{3}\delta_\lambda \mathfrak{E}$$

$$v_{\mathbf{S}} = \sqrt{\mu/\rho} \qquad v_{\mathbf{P}} = \sqrt{(\lambda + 2\mu)/\rho}$$

$$\delta_\rho \mathfrak{E}|_{\kappa,\mu} = \delta_\rho \mathfrak{E}\,,$$

$$\delta_\kappa \mathfrak{E}|_{\rho,\mu} = \delta_\lambda \mathfrak{E}\,,$$

$$\delta_\mu \mathfrak{E}|_{\rho,\kappa} = \delta_\mu \mathfrak{E} - \frac{2}{3}\delta_\lambda \mathfrak{E}$$

The elastic tensor in a medium with radial symmetry axis:

[→ Polarisation anisotropy]

$$\begin{pmatrix} C_{rrrr} & C_{rr\phi\phi} & C_{rr\theta\theta} & C_{rr\phi\theta} & C_{rrr\theta} & C_{rrr\phi} \\ C_{\phi\phi rr} & C_{\phi\phi\phi\phi} & C_{\phi\phi\theta\theta} & C_{\phi\phi\phi\theta} & C_{\phi\phi r\theta} & C_{\phi\phi r\phi} \\ C_{\theta\theta rr} & C_{\theta\theta\phi\phi} & C_{\theta\theta\theta\theta} & C_{\theta\theta\phi\theta} & C_{\theta\theta r\theta} & C_{\theta\theta r\phi} \\ C_{\phi\theta rr} & C_{\phi\theta\phi\phi} & C_{\phi\theta\theta\theta} & C_{\phi\theta\phi\theta} & C_{\phi\theta r\theta} & C_{\phi\theta r\phi} \\ C_{r\theta rr} & C_{r\theta\phi\phi} & C_{r\theta\theta\theta} & C_{r\theta\phi\theta} & C_{r\theta r\theta} & C_{r\theta r\phi} \\ C_{r\phi rr} & C_{r\phi\phi\phi} & C_{r\phi\theta\theta} & C_{r\phi\phi\theta} & C_{r\phi r\theta} & C_{r\phi r\phi} \end{pmatrix}$$

$$= \begin{pmatrix} \lambda+2\mu & \lambda+c & \lambda+c & 0 & 0 & 0 \\ \lambda+c & \lambda+2\mu+a & \lambda+a & 0 & 0 & 0 \\ \lambda+c & \lambda+a & \lambda+2\mu+a & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu+b & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu+b \end{pmatrix}$$

ADJOINT METHOD: Derivatives in a medium with radial symmetry

Fréchet kernels for a medium with radial anisotropy:

$$\delta_\rho \mathfrak{E}|_{\lambda,\mu,a,b,c} = -\int_{t=t_0}^{t_1} \partial_t \mathbf{u}^\dagger \cdot \partial_t \mathbf{u} \, dt \,,$$

$$\delta_\lambda \mathfrak{E}|_{\rho,\mu,a,b,c} = \int_{t=t_0}^{t_1} (\mathrm{tr}\ \boldsymbol{\epsilon}^\dagger)(\mathrm{tr}\ \boldsymbol{\epsilon}) \, dt \,,$$

$$\delta_\mu \mathfrak{E}|_{\rho,\lambda,a,b,c} = 2 \int_{t=t_0}^{t_1} \boldsymbol{\epsilon}^\dagger : \boldsymbol{\epsilon} \, dt \,,$$

$$\delta_a \mathfrak{E}|_{\rho,\lambda,\mu,b,c} = \int_{t=t_0}^{t_1} (\epsilon^\dagger_{\phi\phi} + \epsilon^\dagger_{\theta\theta})(\epsilon_{\phi\phi} + \epsilon_{\theta\theta}) \, dt \,,$$

$$\delta_b \mathfrak{E}|_{\rho,\lambda,\mu,a,c} = 4 \int_{t=t_0}^{t_1} (\epsilon^\dagger_{r\theta}\epsilon_{r\theta} + \epsilon^\dagger_{r\phi}\epsilon_{r\phi}) \, dt \,,$$

$$\delta_c \mathfrak{E}|_{\rho,\lambda,\mu,a,b} = \int_{t=t_0}^{t_1} [\epsilon^\dagger_{rr}(\epsilon_{\phi\phi} + \epsilon_{\theta\theta}) + \epsilon_{rr}(\epsilon^\dagger_{\phi\phi} + \epsilon^\dagger_{\theta\theta})] \, dt$$

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{dx_j} + \frac{\partial u_j}{dx_i}\right)$$ strain tensor

3.3. Derivatives with respect to the right-hand side (the source)

ADJOINT METHOD: Derivatives with respect to the right-hand side

- How does the misfit change as I change the source?
- Instead of L(u,m) = f, we consider L'(u,m,f) = L(u,m)-f = 0.
- Repeat the adjoint method recipe:

ADJOINT METHOD: Derivatives with respect to the right-hand side

- How does the misfit change as I change the source?
- Instead of L(u,m) = f, we consider L‘(u,m,f) = L(u,m)-f = 0.
- Repeat the adjoint method recipe:

→ The adjoint equations are the same as before, namely:

Adjoint wave equation:

$$\mathbf{L}^t(\mathbf{u}^t, \mathbf{C}, \rho) = \rho(\mathbf{x})\partial_t^2 \mathbf{u}^t(\mathbf{x}, t) - \nabla \cdot \int_{-\infty}^{\infty} \dot{\mathbf{C}}(\mathbf{x}, \tau - t) : \nabla \mathbf{u}^t(\mathbf{x}, t)\, d\tau$$

$$\mathbf{L}^t(\mathbf{u}^t, \mathbf{C}, \rho) = \mathbf{f}^t$$

← determined by the misfit measure

Adjoint subsidiary conditions:

$$\mathbf{u}^t(\mathbf{x}, t)|_{t=t_1} = \mathbf{0} \qquad \partial_t \mathbf{u}^t(\mathbf{x}, t)|_{t=t_1} = \mathbf{0} \qquad \mathbf{n} \cdot \int_{-\infty}^{t} \dot{\mathbf{C}}(\mathbf{x}, \tau - t) : \nabla \mathbf{u}^t(\mathbf{x}, \tau)\, d\tau\, |_{\mathbf{x} \in \Gamma} = \mathbf{0}$$

→ The derivative of E with respect to f is:

$$D_f E(u, m, f)(\delta f) = -\langle \delta f, u^t \rangle_U$$

ADJOINT METHOD: Derivatives with respect to the right-hand side

Example: Moment tensor point source

$$\mathbf{f}^{\Sigma}(\mathbf{x},t) = -\nabla \cdot [\mathbf{M}\,\delta(\mathbf{x}-\boldsymbol{\xi})]$$

ξ = point source location

M = moment tensor

Derivatives:

$$\frac{d}{dM_{ij}}\mathfrak{E}(\mathbf{u}) = -2 \int_{t=t_0}^{t_1} \epsilon^{\dagger}_{ij}(\boldsymbol{\xi},t)\,dt$$

… with respect to moment tensor comp.

$$\frac{d}{d\xi_i}\mathfrak{E}(\mathbf{u}) = -\mathbf{M} : \int_{t=t_0}^{t_1} \frac{d}{d\xi_i}\boldsymbol{\epsilon}^{\dagger}(\boldsymbol{\xi},t)\,dt$$

… with respect to source location

3.4. Objective functionals

OBJECTIVE FUNCTIONALS

The choice of physically meaningful objective functionals constitutes the *art of seismic tomography*.

Examples:

1. $E(u,u^0) = u_i(x^r,t^r)$

   of theoretical interest

2. $E(u,u^0) = ||u-u^0||$

   simple but not very useful

3. $E(u,u^0)$ = cross-correlation time shift

   frequently used, robust measurement

4. (u,u^0) = rms amplitude difference

   work in progress

5. Time frequency misfits

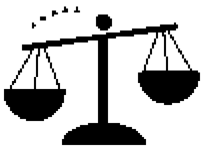
   that‘s what we use

OBJECTIVE FUNCTIONALS: Time-frequency misfits

We want to use full waveforms **but …**

phase information:

- can be measured reliably
- ± linearly related to Earth structure
- physically interpretable



amplitude information:

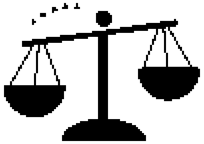
- hard to measure (earthquake magnitude often unknown)
- non-linearly related to structure

OBJECTIVE FUNCTIONALS: Time-frequency misfits

We want to use full waveforms **but …**

phase information:

- can be measured reliably
- ± linearly related to Earth structure
- physically interpretable



amplitude information:

- hard to measure (earthquake magnitude often unknown)
- non-linearly related to structure

$$\hat{u}(\omega, t) := G(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u(\tau)\, g(\tau - t)\, e^{-i\omega\tau} d\tau$$

[t-ω representation of synthetics, u(t)]

$$\hat{u}_0(\omega, t) := G(u_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u_0(\tau)\, g(\tau - t)\, e^{-i\omega\tau} d\tau$$

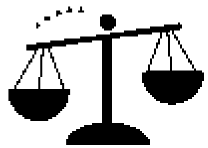
[t-ω representation of data, $u_0(t)$]

OBJECTIVE FUNCTIONALS: Time-frequency misfits

We want to use full waveforms **but …**

phase information:

- can be measured reliably
- ± linearly related to Earth structure
- physically interpretable



amplitude information:

- hard to measure (earthquake magnitude often unknown)
- non-linearly related to structure

phase misfit

$$E_{phase} = \iint (\varphi - \varphi_0)^2 \, d\omega \, dt$$

envelope misfit

$$E_{envelope} = \iint (|\hat{u}| - |\hat{u}_0|)^2 \, d\omega \, dt$$

adjoint method

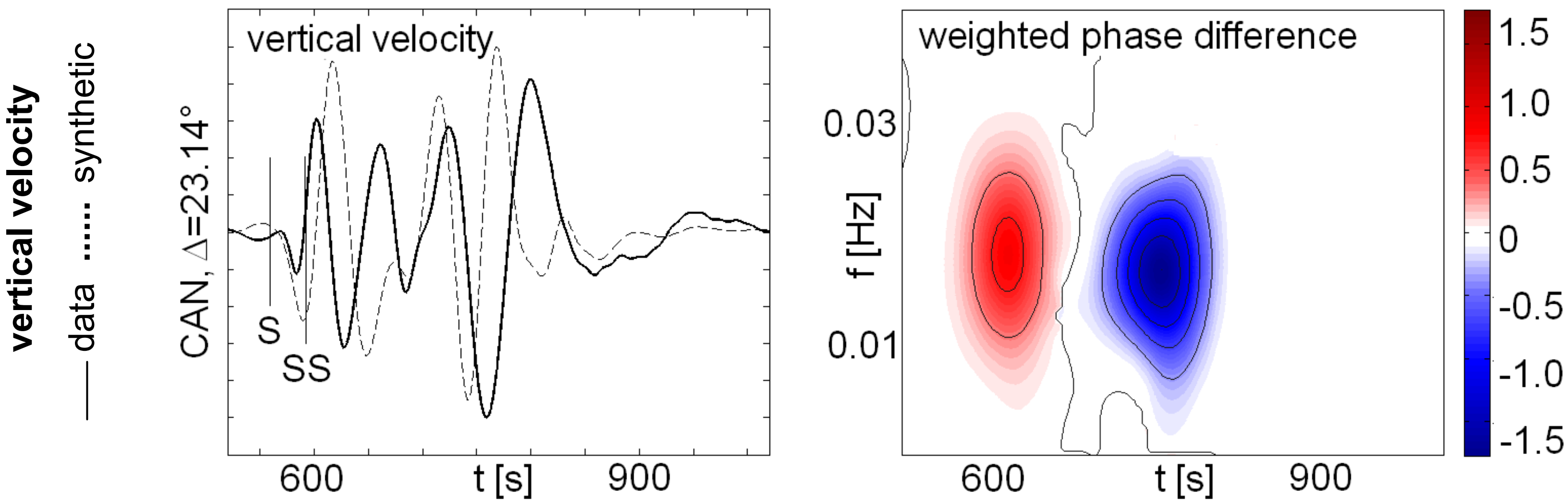
sensitivity densities

(Fréchet kernels)

OBJECTIVE FUNCTIONALS: Time-frequency misfits

• Concentrate on phase information.

• Compute time-frequency representations of data & synthetics.



• Quantify phase difference as a function of time and frequency.

• Phase advance = positive phase difference (red).

• Phase delay = negative phase difference (blue).

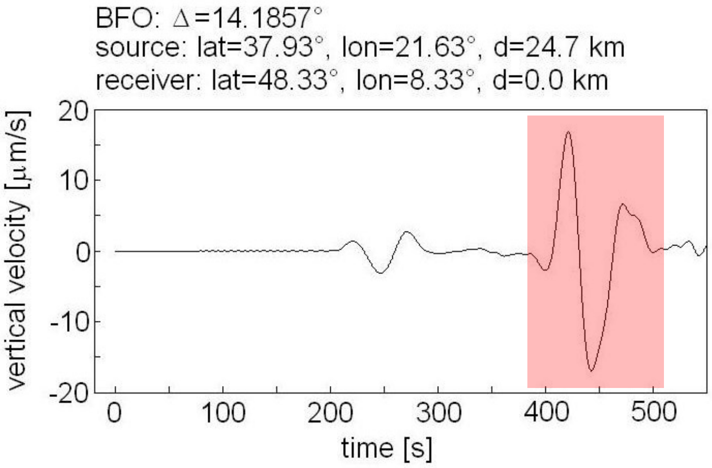
3.5. Kernel gallery

Regional-scale examples.

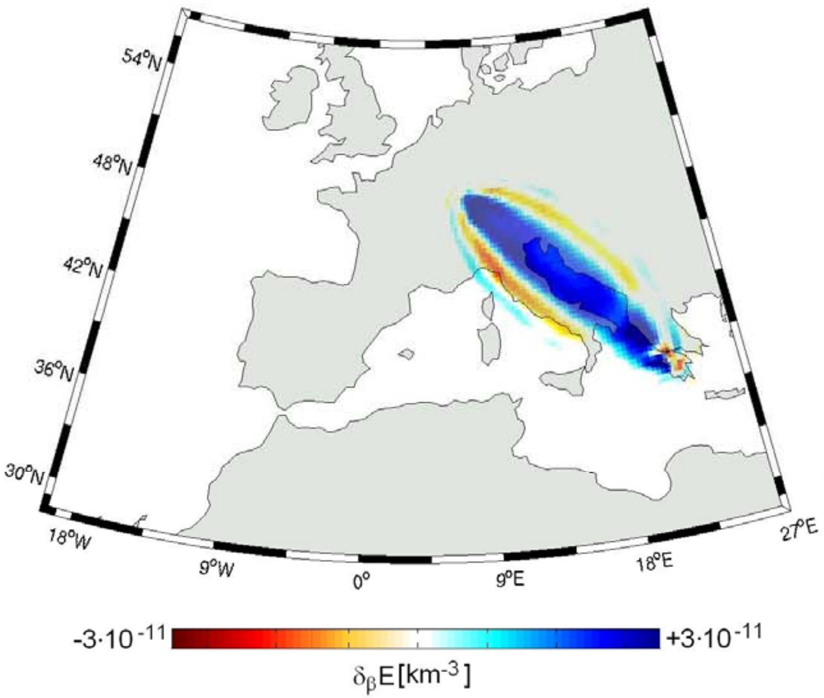
Global-scale examples.

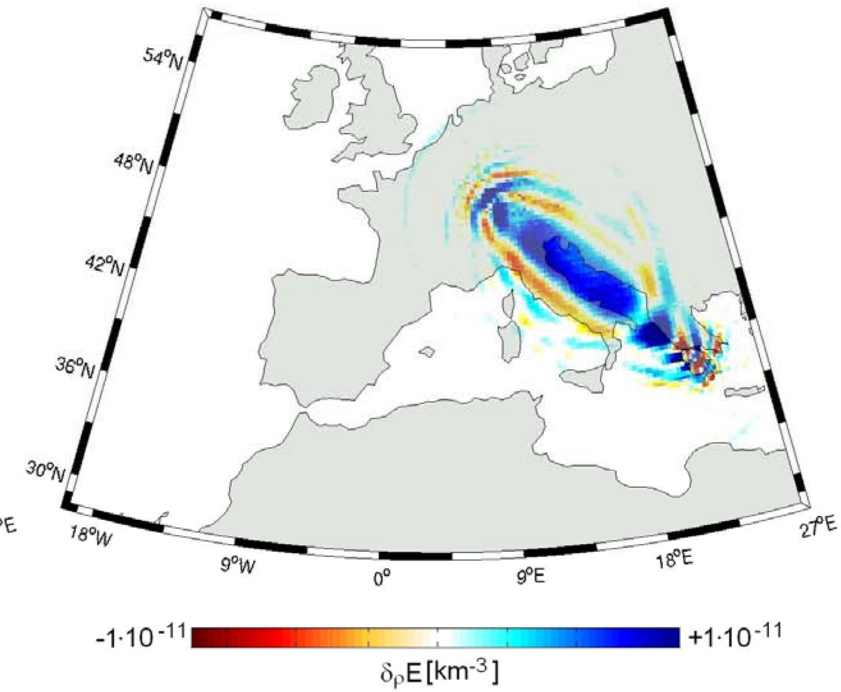
Anisotropy.

KERNEL GALLERY: Fundamental-mode Rayleigh wave (Monday's example)

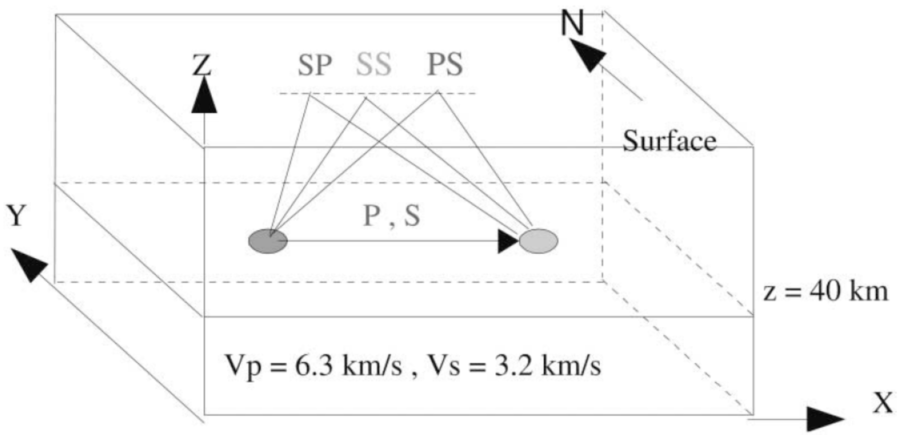


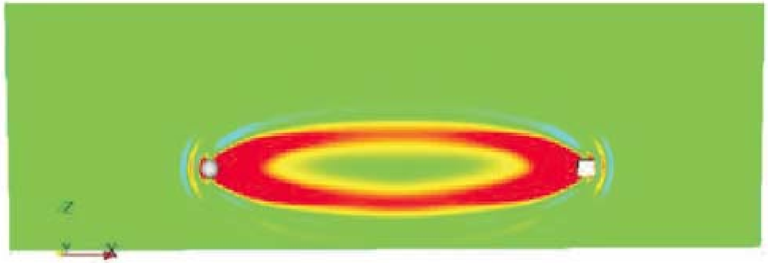
- measurement of cross-correlation time shift on a 50 s Rayleigh wave
- dominant first Fresnel zone
- increase of the S wave speed β leads to an increase of Δt → synthetic waveform arrives earlier



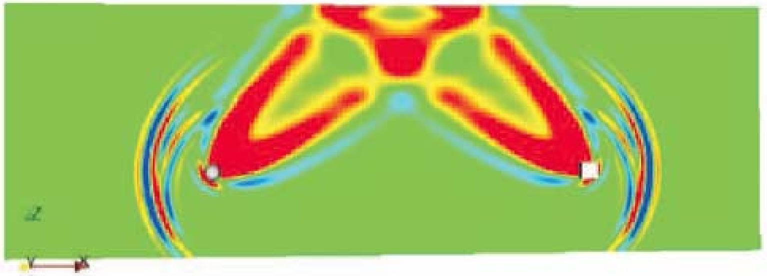


KERNEL GALLERY: Regional-scale homogeneous model (Liu & Tromp, BSSA 2006)

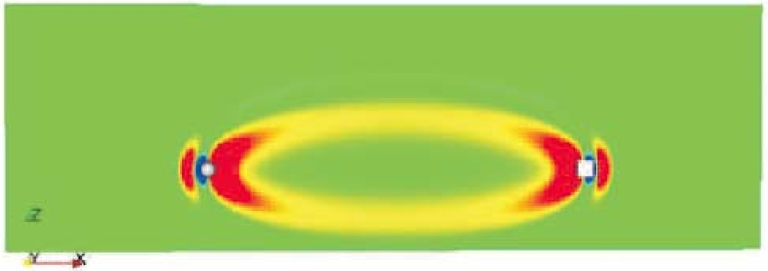




direct S wave



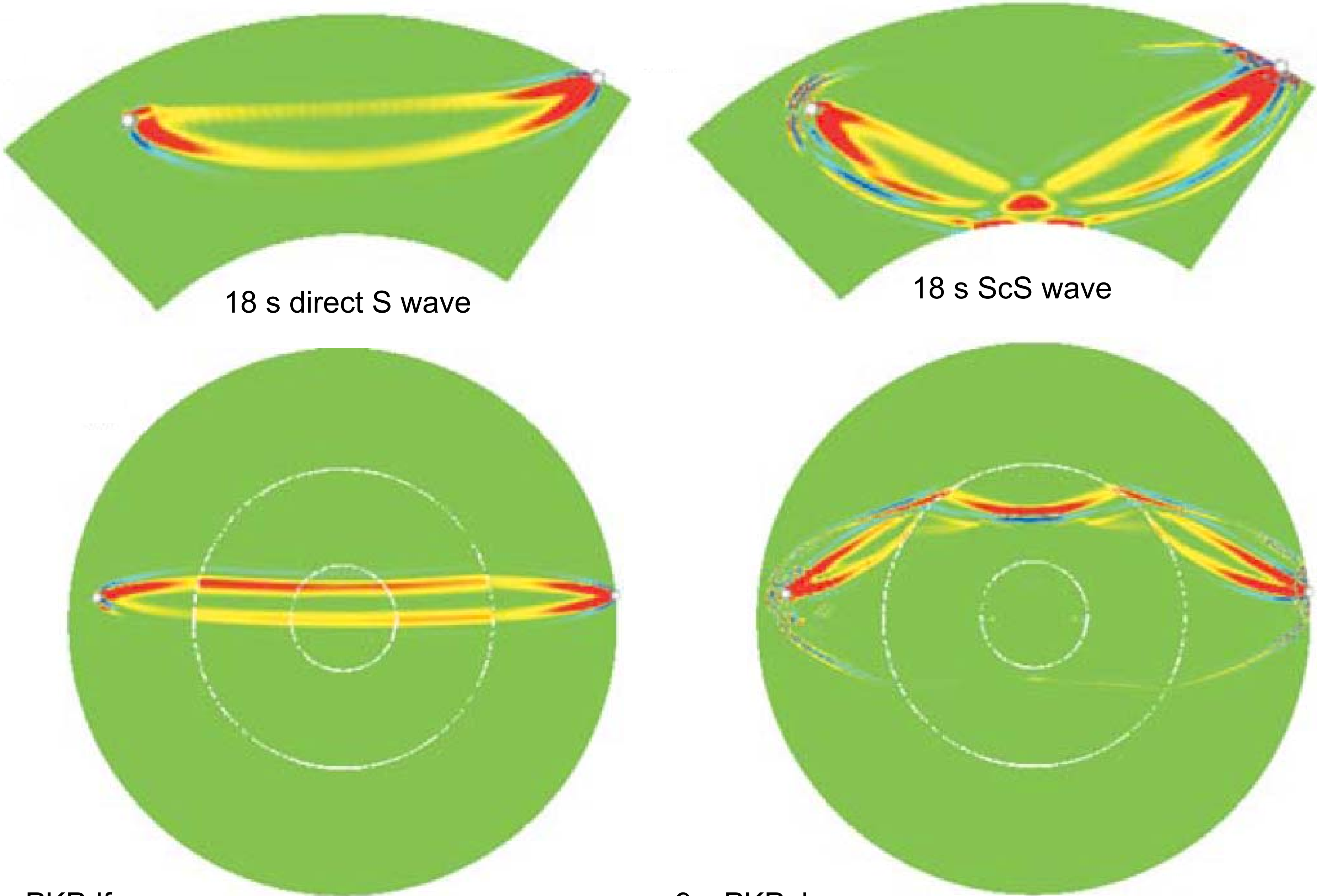
surface-reflected SS wave



direct P wave

- measurement of cross-correlation time shifts
- regional-scale layer over half space
- homogeneous

KERNEL GALLERY: Global-scale PREM model (Liu & Tromp, GJI 2008)



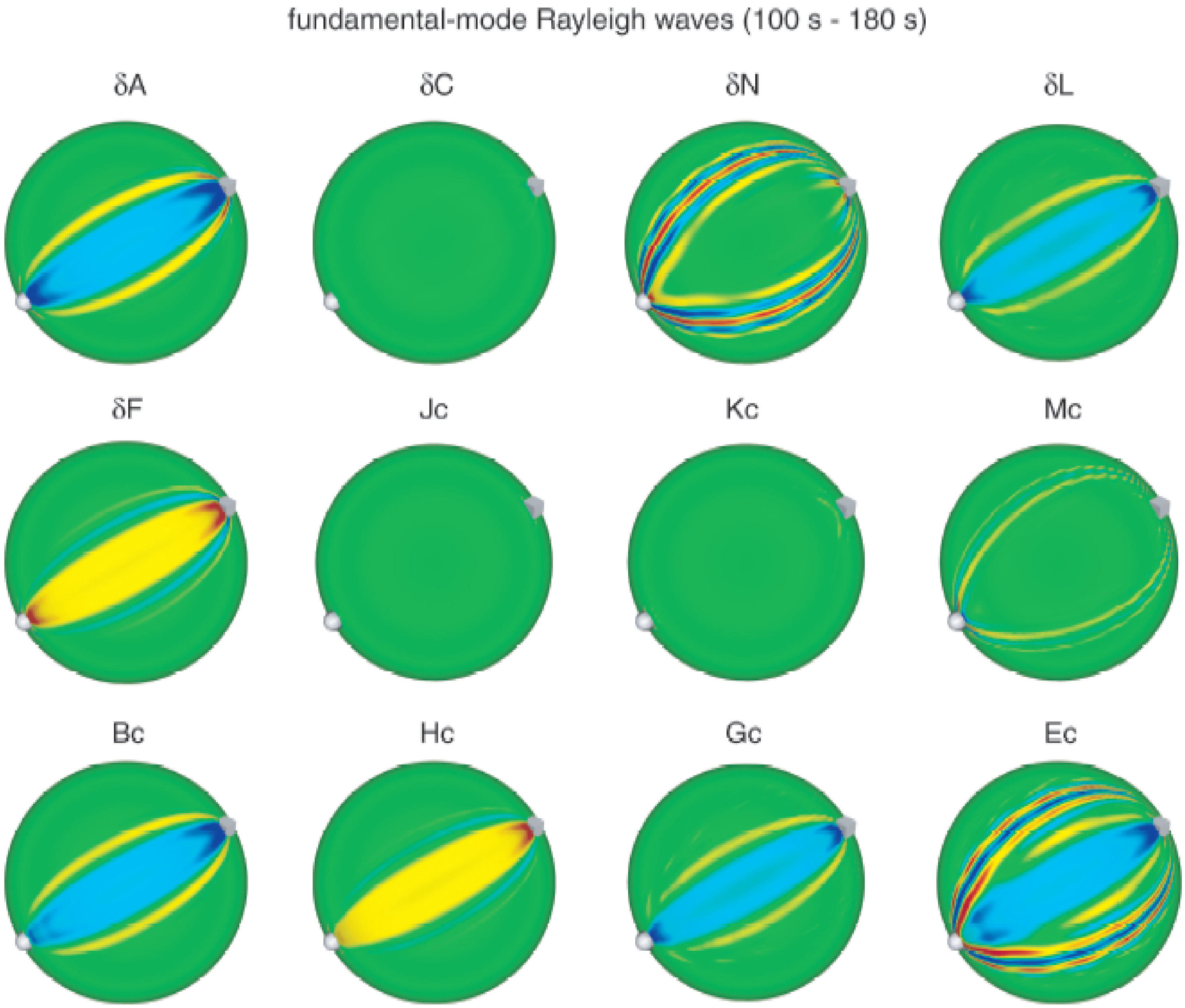
18 s direct S wave

18 s ScS wave

9 s PKPdf wave

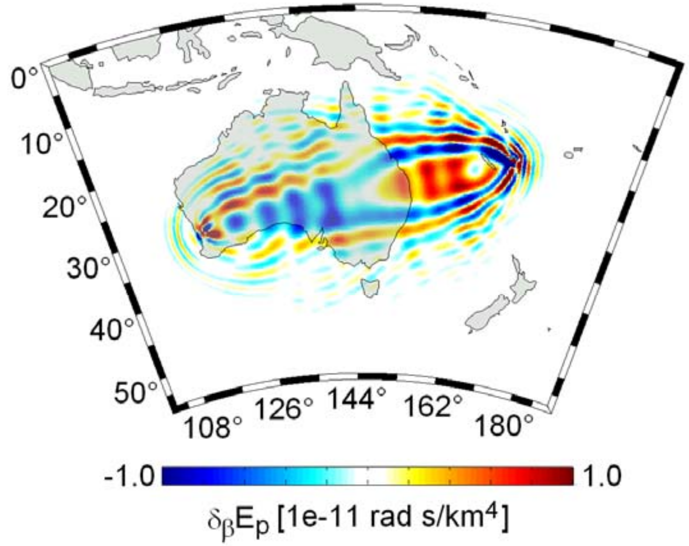
9 s PKPab wave

KERNEL GALLERY: Global-scale PREM model (Sieminski et al., GJI 2007)

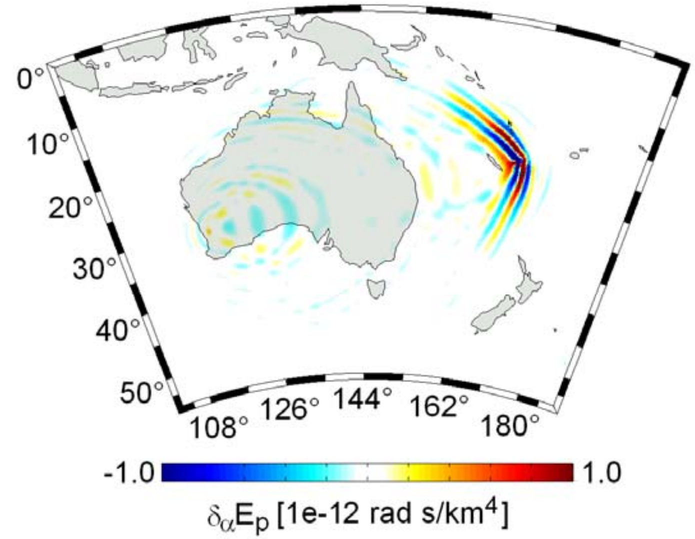


KERNEL GALLERY: Time-frequency phase misfit (Fichtner et al., GJI 2008)

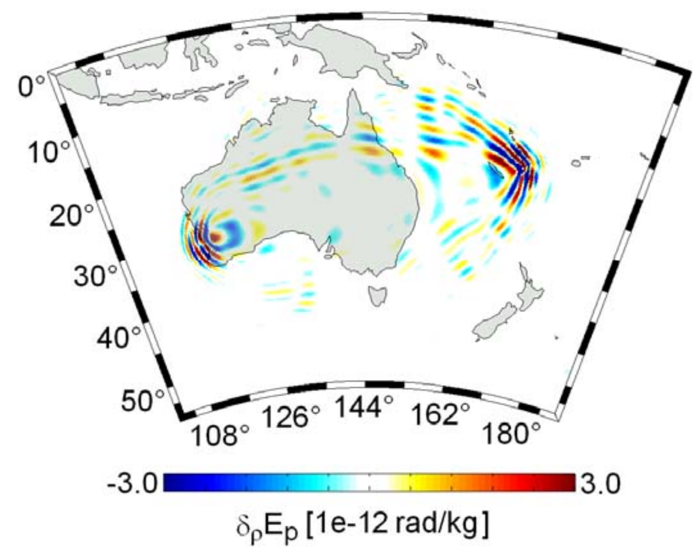
S wave speed (β) kernel



P wave speed (α) kernel



density (ρ) kernel



- **significant sensitivity off the ray path**
- **complex geometry**
- **little sensitivity w.r.t. P velocity & density**

3.6. The little important details

From kernels to gradients.

How do we solve the adjoint equations?

Accuracy-adaptive time integration.

IMPORTANT DETAILS: From Fréchet kernels to gradients

We have

$$D_m E(\delta m) = \int_G \delta m \cdot \delta_m E \, d^3x$$

For gradient-based misfit minimisation we need

$$\nabla_m E = \frac{dE}{dm_i}$$

where m_i is a model parameter.

IMPORTANT DETAILS: From Fréchet kernels to gradients

We have

$$D_m E(\delta m) = \int_G \delta m \cdot \delta_m E \, d^3x$$

For gradient-based misfit minimisation we need

$$\nabla_m E = \frac{dE}{dm_i}$$

where m_i is a model parameter.

$$\nabla_m E = \frac{dE}{dm_i} = \int_G b_i \cdot \delta_m E \, d^3x$$

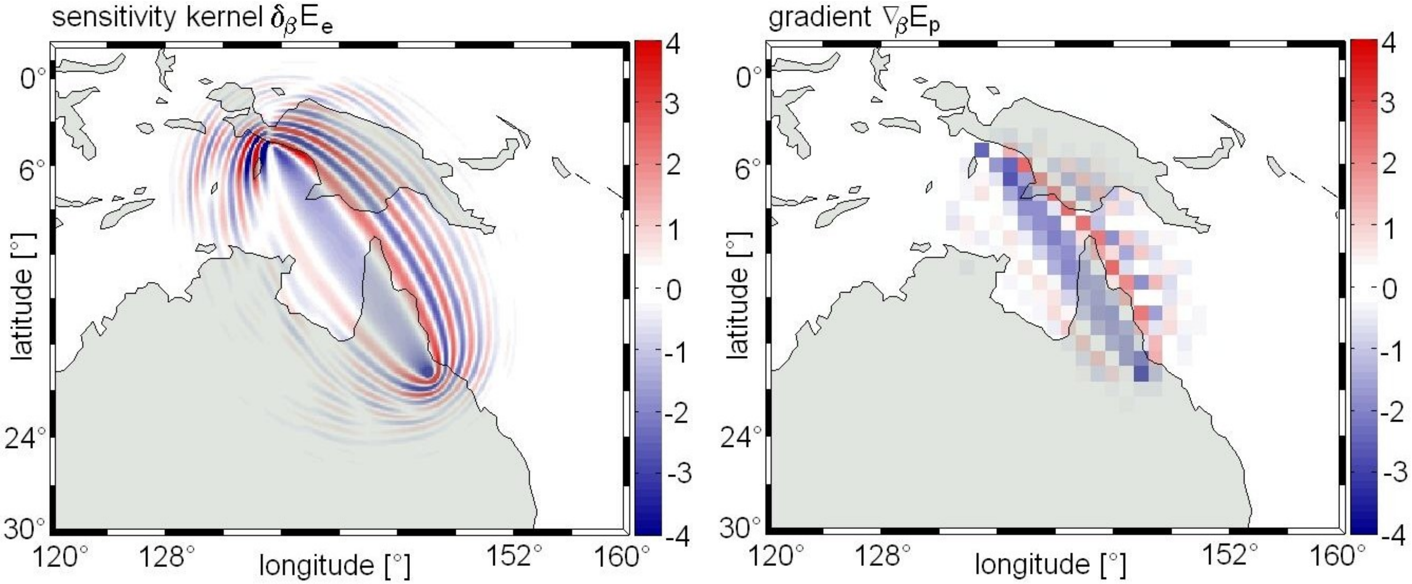
- The gradient is the projection of the sensitivity kernels onto the basis functions.
- The gradient depends on the basis functions!
- The basis functions are used for regularisation.

IMPORTANT DETAILS: From Fréchet kernels to gradients

Example:

- Sensitivity of 50 s Rayleigh wave amplitude w.r.t. S wave speed
- Basis functions: blocks (1° x 1° x 10 km)
- Small-scale features integrate out!
- The gradient resembles a smoothed ray.

→ Using 3D kernels instead of rays makes the largest difference in cases where ray theory breaks down (3D Earth structure, caustics, …)



IMPORTANT DETAILS: How do we solve the adjoint equations?

Elastic wave equation:

$$\mathbf{L}(\mathbf{u},\rho,\mathbf{C}) = \mathbf{f}$$

$$\mathbf{L}(\mathbf{u},\rho,\mathbf{C}) = \rho(\mathbf{x})\partial_t^2\mathbf{u}(\mathbf{x},t) - \nabla\cdot\int_{-\infty}^{\infty}\mathbf{C}(\mathbf{x},t-\tau):\nabla\mathbf{u}(\mathbf{x},t)\,d\tau$$

Subsidiary conditions:

$$\mathbf{u}(\mathbf{x},t)|_{t=t_0} = \mathbf{0} \qquad \partial_t\mathbf{u}(\mathbf{x},t)|_{t=t_0} = \mathbf{0} \qquad \mathbf{n}\cdot\int_{-\infty}^{t}\mathbf{C}(\mathbf{x},t-\tau):\nabla\mathbf{u}(\mathbf{x},\tau)\,d\tau\,|_{\mathbf{x}\in\Gamma} = \mathbf{0}$$

Adjoint wave equation:

$$\mathbf{L}^t(\mathbf{u}^t,\mathbf{C},\rho) = \rho(\mathbf{x})\partial_t^2\mathbf{u}^t(\mathbf{x},t) - \nabla\cdot\int_{-\infty}^{\infty}\mathbf{C}(\mathbf{x},\tau-t):\nabla\mathbf{u}^t(\mathbf{x},t)\,d\tau$$

$$\mathbf{L}^t(\mathbf{u}^t,\mathbf{C},\rho) = \mathbf{f}^t$$

← determined by the misfit measure

Adjoint subsidiary conditions:

$$\mathbf{u}^t(\mathbf{x},t)|_{t=t_1} = \mathbf{0} \qquad \partial_t\mathbf{u}^t(\mathbf{x},t)|_{t=t_1} = \mathbf{0} \qquad \mathbf{n}\cdot\int_{-\infty}^{t}\mathbf{C}(\mathbf{x},\tau-t):\nabla\mathbf{u}^t(\mathbf{x},\tau)\,d\tau\,|_{\mathbf{x}\in\Gamma} = \mathbf{0}$$

IMPORTANT DETAILS: How do we solve the adjoint equations?

- **regular wavefield: initial** conditions → simulated **forward** in time
- **adjoint wavefield: terminal** conditions → simulated **backwards** in time

- **regular** and **adjoint** wave fields must be known **simultaneously** at time t because:

Fréchet kernels of the misfit functional:

$$\delta_\rho E(\mathbf{x}) = \int_{t_0}^{t_1} \mathbf{u}^{\mathbf{t}}(\mathbf{x}, t) \cdot \partial_t^2 \mathbf{u}(\mathbf{x}, t)\, dt \qquad \delta_\Phi E = \int_{t_0}^{t_1} \int_{\tau=t_0+t}^{t_1} \nabla \mathbf{u}^{\mathbf{t}}(\mathbf{x}, \tau) \otimes \nabla \mathbf{u}(\mathbf{x}, \tau - t)\, d\tau\, dt$$

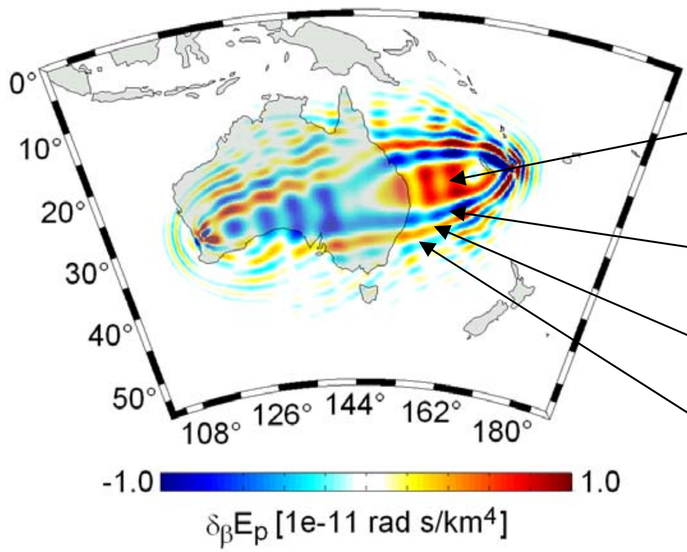
→ The regular wavefield must be stored.

→ At every time step (10.000): **7 TB per earthquake**

This can be reduced with physics:

IMPORTANT DETAILS: Accuracy-adaptive time integration

S wave speed (β) kernel



1st Fresnel zone

2nd Fresnel zone

3rd Fresnel zone

4th Fresnel zone

major feature

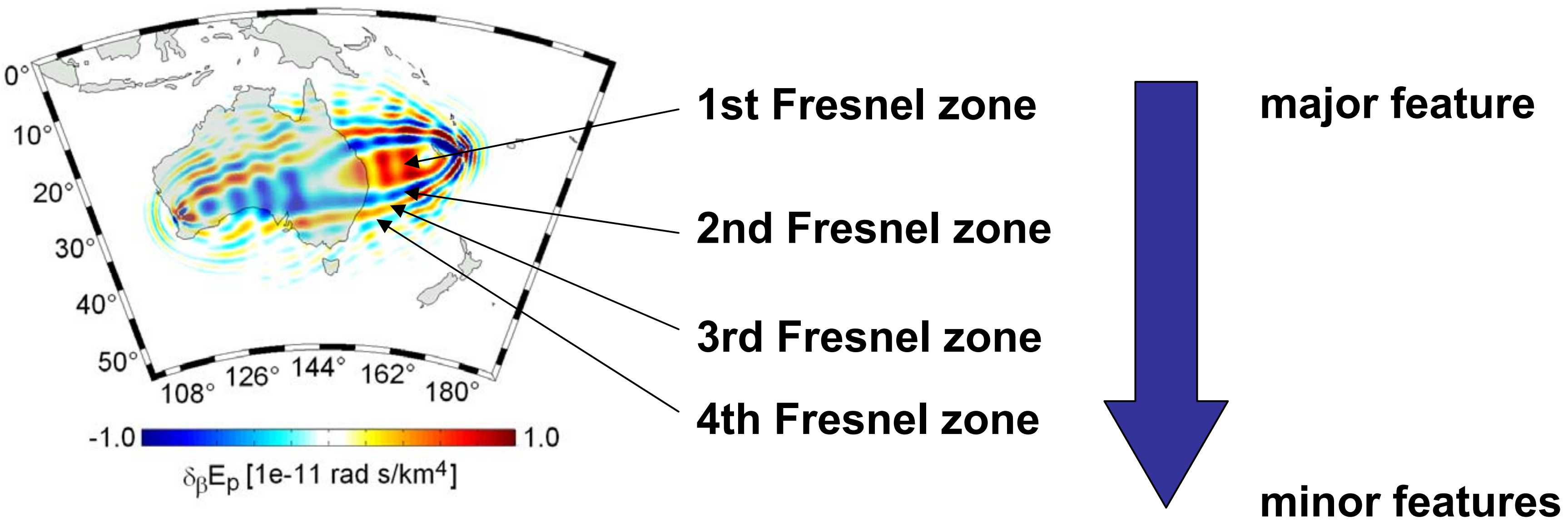
minor features

The higher Fresnel zones become increasingly unimportant for the tomography.

→ No need to model all of them very accurately.

IMPORTANT DETAILS: Accuracy-adaptive time integration

S wave speed (β) kernel



<u>Physical arguments:</u>

accurate mth Fresnel zone: save wavefield at least every

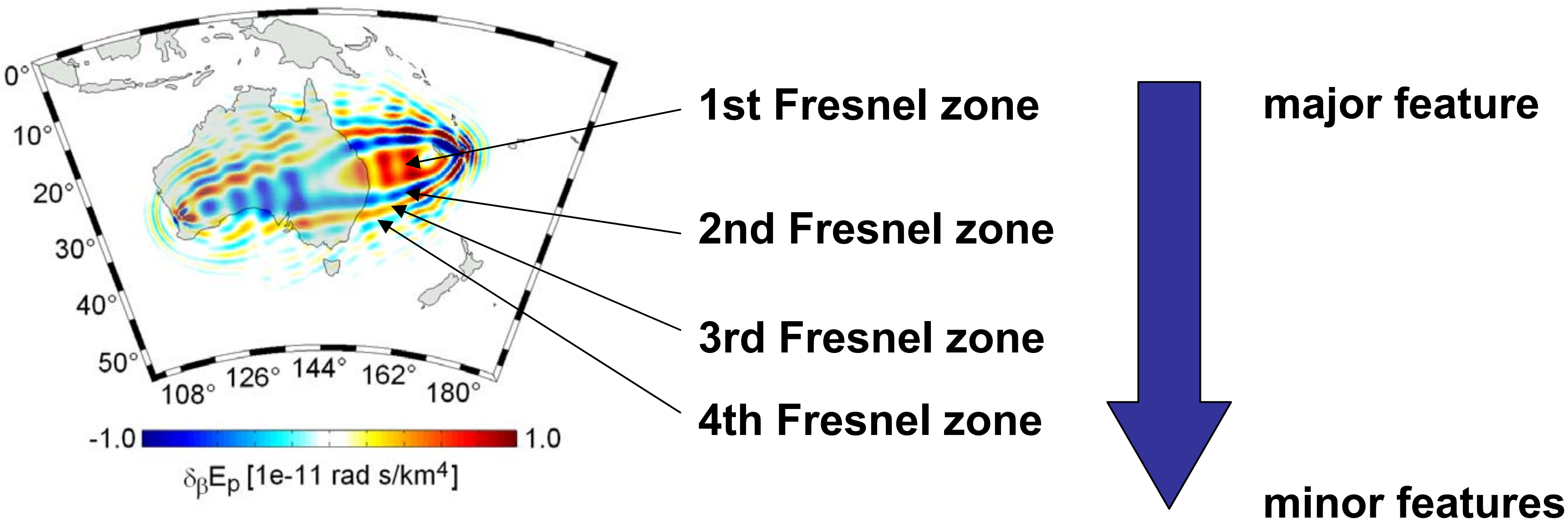
$$\frac{5}{2}\frac{\alpha_{max}}{C}\sqrt{\frac{\ell}{T\beta^3}}\left(\sqrt{m}-\sqrt{m-1}\right)$$

time steps .

$\alpha_{max} =$	max. P wave speed
$\beta =$	approx. S wave speed
$T =$	dominant period
$C =$	Courant number
$\ell =$	epicentral distance

IMPORTANT DETAILS: Accuracy-adaptive time integration

S wave speed (β) kernel



Physical arguments: save every 50 time steps → 1st - 3rd Fresnel zones correct.

→ 140 GB storage per earthquake (instead of 7 TB)

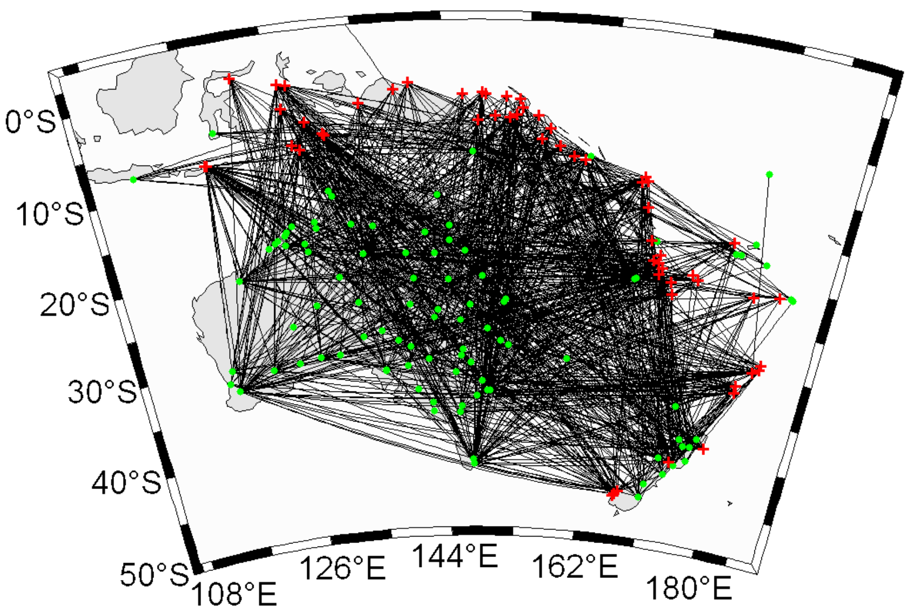
accuracy-adaptive storage interval

4. Iterative solution of the waveform tomographic problem

PREPARATORY STEPS: Data

1. Data selection

- Few high-quality data (thousands instead of millions)
- Earthquakes recorded by many stations (efficiency of the adjoint method)
- Good coverage of the target region

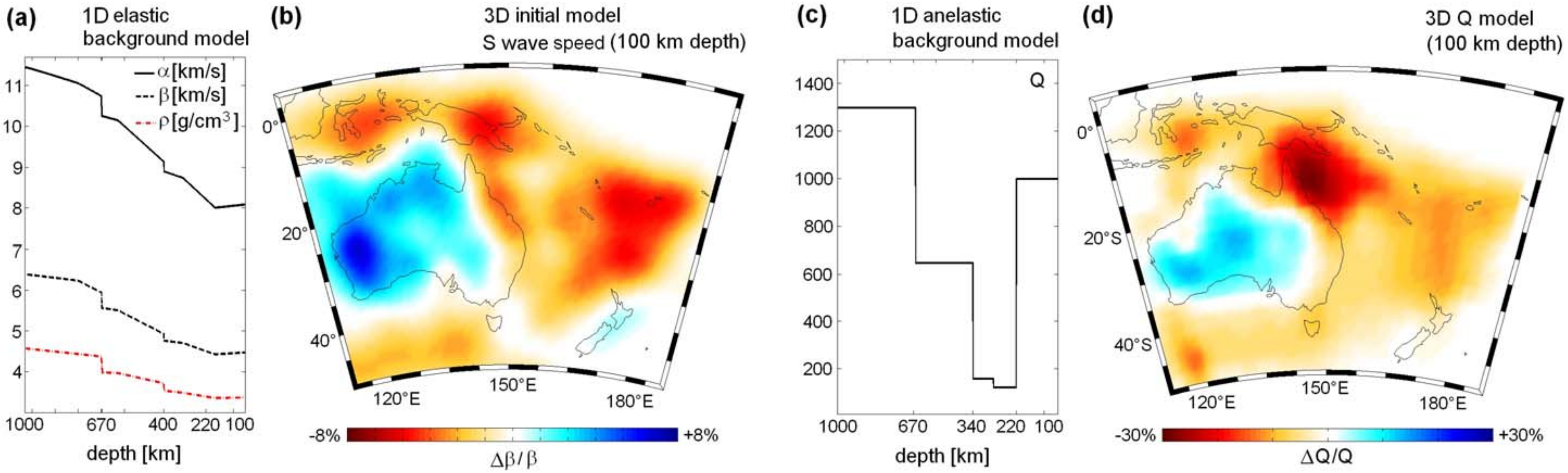


Our work:

- 57 earthquakes in the Australasian region
- approximately 3000 usable three-component recordings (mostly surface waves)
- periods down to 50 s

2. Initial Models

- ensure convergence to the global optimum



<u>elastic</u>: Smoothed surface wave tomography (*Fishwick et al., 2005*)

<u>anelastic</u>: 3D Q model from multi-frequency body wave amplitudes (Abdulah, 2007)

ITERATIVE OPTIMISATION: General strategy

Preconditioned conjugate-gradient method (Fletcher & Reeves)

(i) Initialisation: Choose an initial model $\mathbf{m}^{(0)}$. Set the iteration index k to 0 and the initial search direction, $\mathbf{d}^{(0)}$

$$\mathbf{d}^{(0)} = -P^{(0)}\,\nabla E_p^{(0)}(\mathbf{m}^{(0)})\,.$$

(ii) Update: Find an efficient step length, $\sigma^{(k)}$, (see section 3.3.2 for details) and set

$$\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \sigma^{(k)}\,P^{(k)}\,\mathbf{d}^{(k)}\,.$$

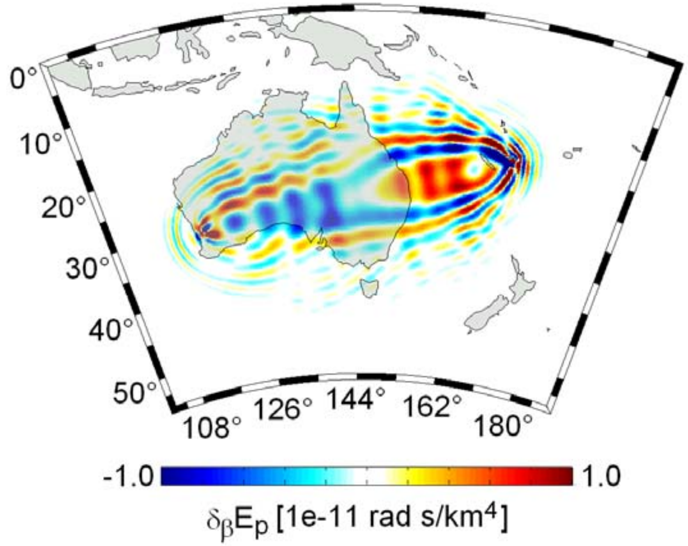
(iii) Search direction: Compute the next search direction, $\mathbf{d}^{(k+1)}$, according to

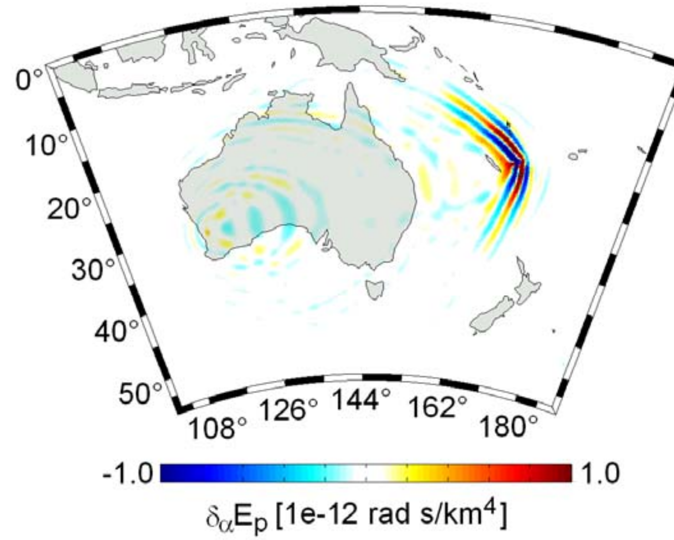
$$\beta^{(k)} = \frac{||P^{(k)}\,\nabla E_p^{(k+1)}(\mathbf{m}^{(k+1)})||^2}{||P^{(k)}\,\nabla E_p^{(k)}(\mathbf{m}^{(k)})||^2}\,, \qquad \mathbf{d}^{(k+1)} = -P^{(k)}\,\nabla E_p^{(k)}(\mathbf{m}^{(k+1)}) + \beta^{(k)}\mathbf{d}^{(k)}\,.$$

(iv) Iteration: Set $k := k+1$ and go to (ii).

- Start with long periods (≈ 100 s) and successively increase bandwidth
  - preserves quasi-linearity
  - requires intuition and experience
- Step length through quadratic line search
  - guess two reasonable step lengths (again intuition and experience …)
  - interpolate with quadratic polynomial and choose the minimum

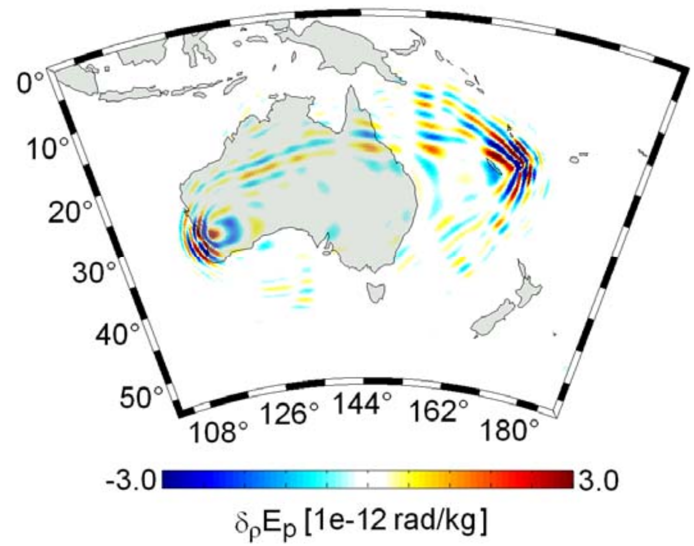
ITERATIVE OPTIMISATION: Preconditioning





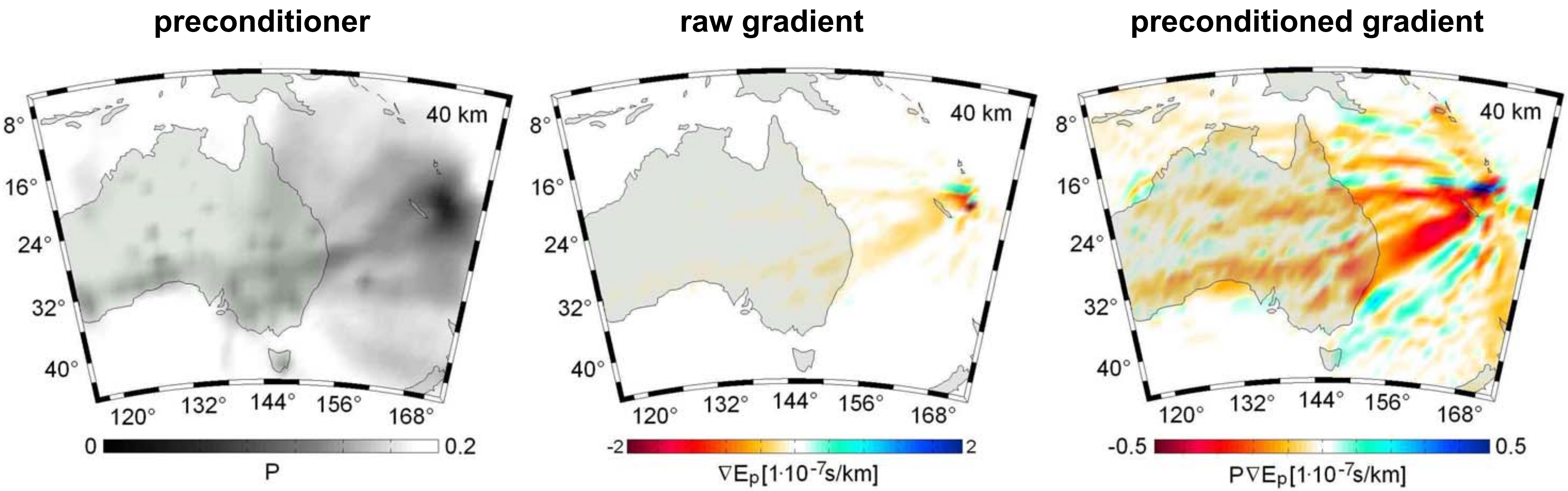
- Singularity at source &receiver.
- Changes of the Earth structure near the source and the receiver are not well represented by first derivatives.
- Using the kernels blindly, would result in nonsense tomographic images.
- Kernels must be modified in the source and receiver regions to ensure convergence to global optimum.

  → **Preconditioning.**



ITERATIVE OPTIMISATION: Preconditioning

- **Preconditioner is found empirically** (2D experiments).
- **Preconditioner is inversely proportional to the amplitudes of the regular and adjoint wavefields** (inverse geometrical spreading, Igel et al., 1996).
- **Preconditioner emphasises kernel contributions further away from sources and receiver.**



ITERATIVE OPTIMISATION: Remark

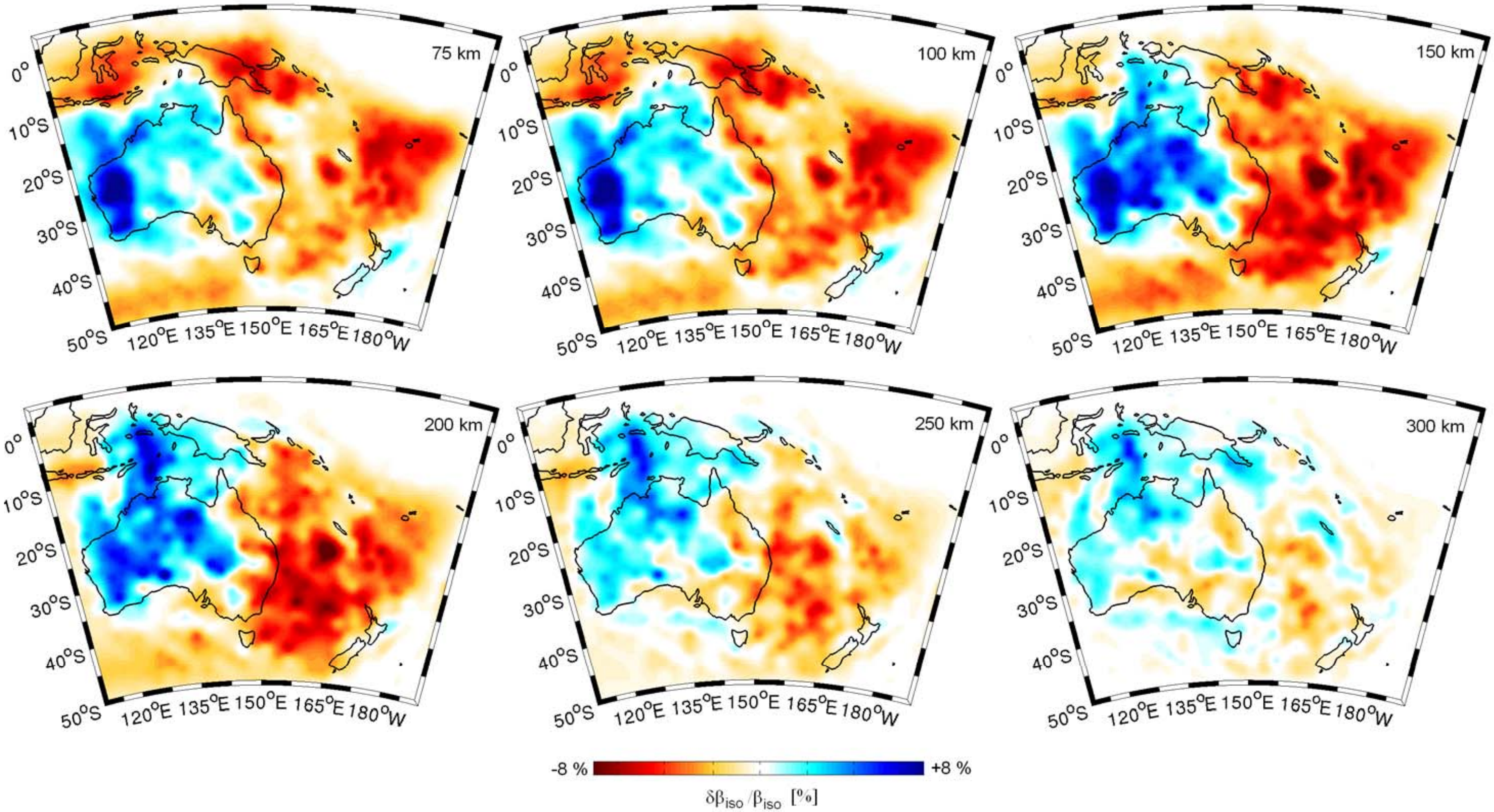
This optimisation scheme is not automatic even though it appears to be!

Seismic tomography requires human intervention in every single step!

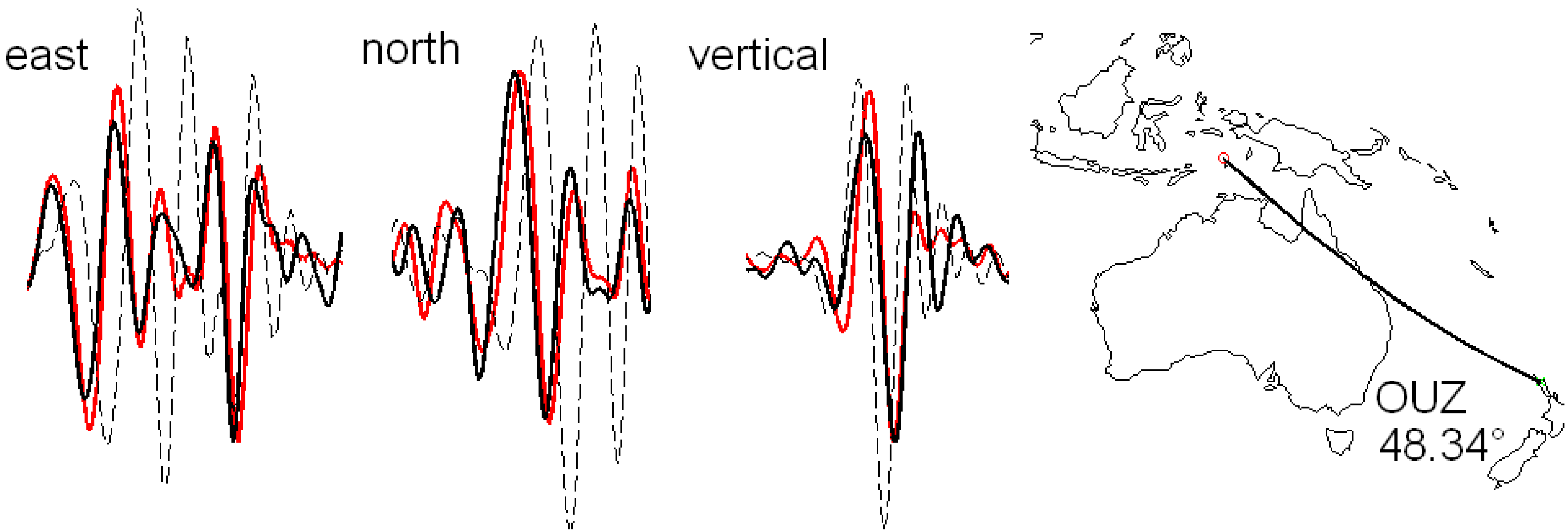
5. Some results of our work so far

RESULTS: S wave speed in the upper mantle beneath the Australian continent

S wave speed



RESULTS: Waveform fit

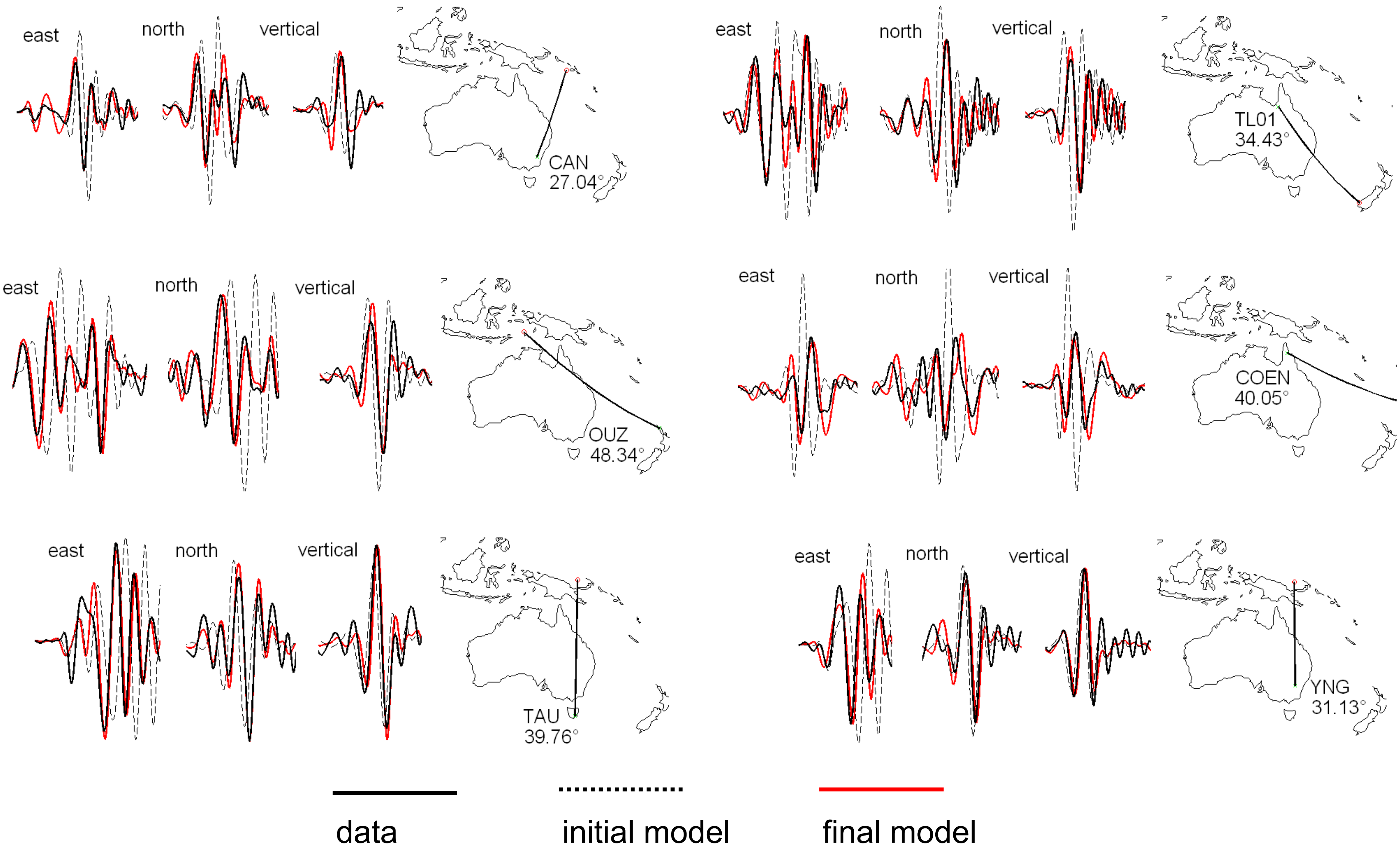


data

initial model

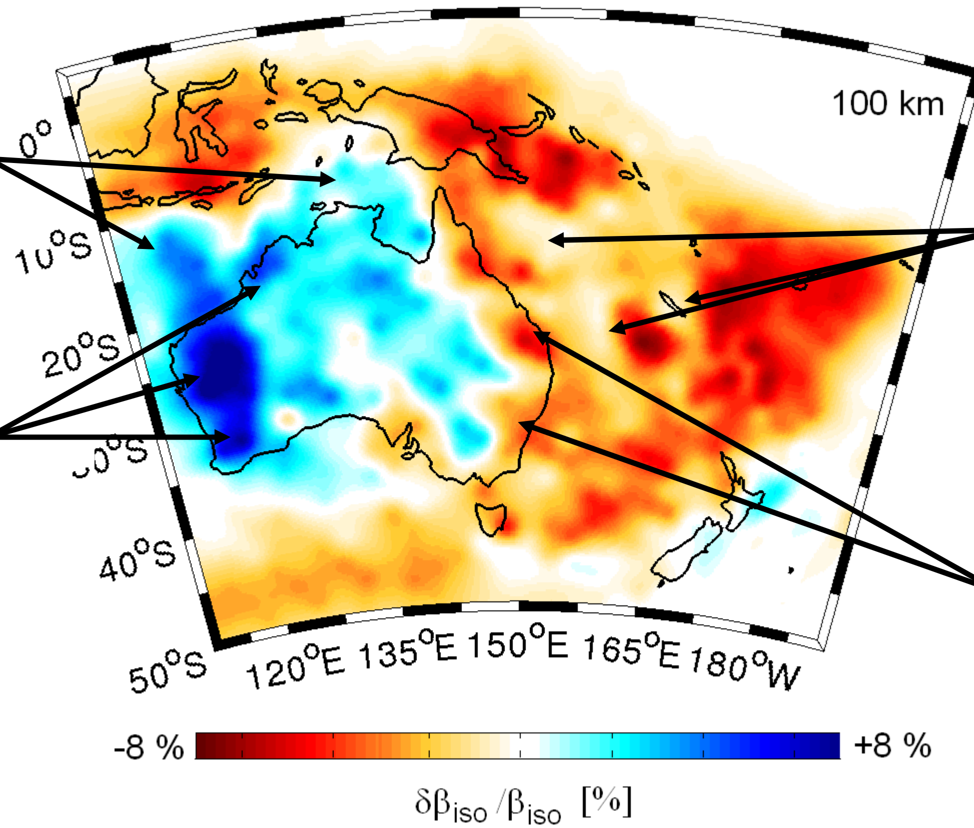
final model

RESULTS: Waveform fit



RESULTS: Interpretation

S wave speed - 100 km depth



northward continuation of the Australian continent

Precambrian units:

Pilbara Block,

Yilgarn Block,

Kimberley Craton

New Caledonia-Norfolk Ridge, Lord Howe Rise

micro-continental ribbons

Phanerozoic:

elevated surface heat flow,

recent volcanism,

seismicity

Fin …

… thanks for your attention!