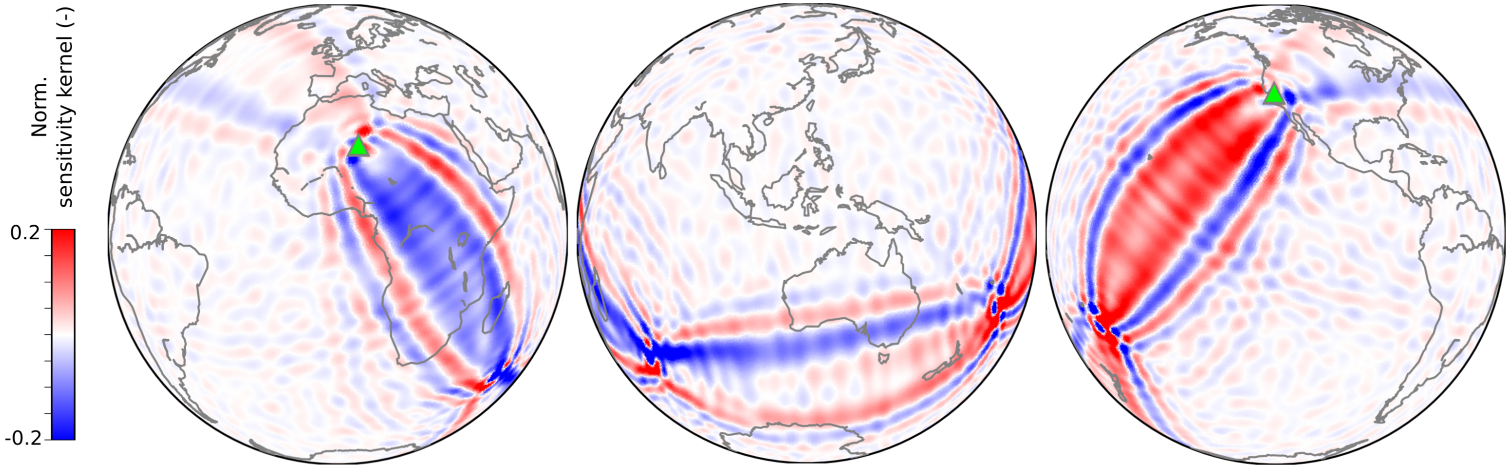
Lecture 3

Adjoint methods and sensitivity kernels

Andreas Fichtner and Christian Boehm

and the ETH Seismology and Wave Physics Group



ETH Seismology & Wave Physics

OUTLINE

PART I: The full-waveform inversion concept

- Summary of a dream
- Formulation as an optimisation problem
- Gradient-based descent methods

PART II: The adjoint method

- Problem statement
- Discrete adjoint method
- Continuous adjoint method
- Sensitivity kernels

➢ Break. Time for questions and short discussion.

PART III: Advanced Topics

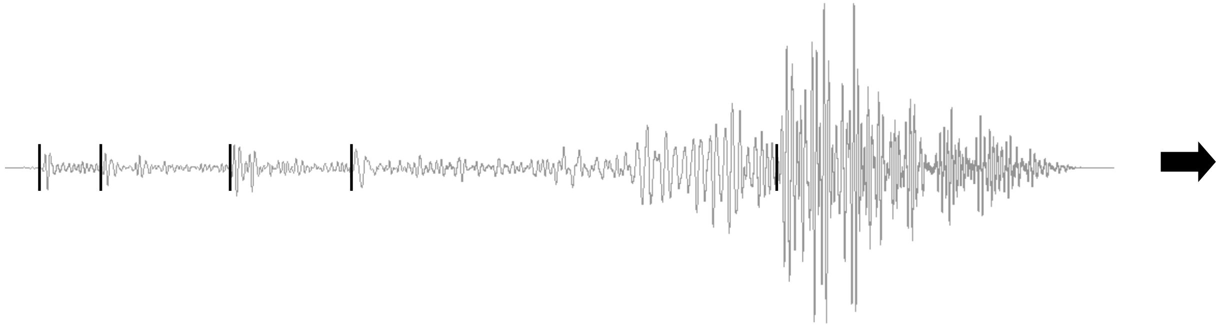
- Local minima and the multiscale approach
- Compressed wavefield storage
- Second derivatives

PART I

The full-waveform inversion concept

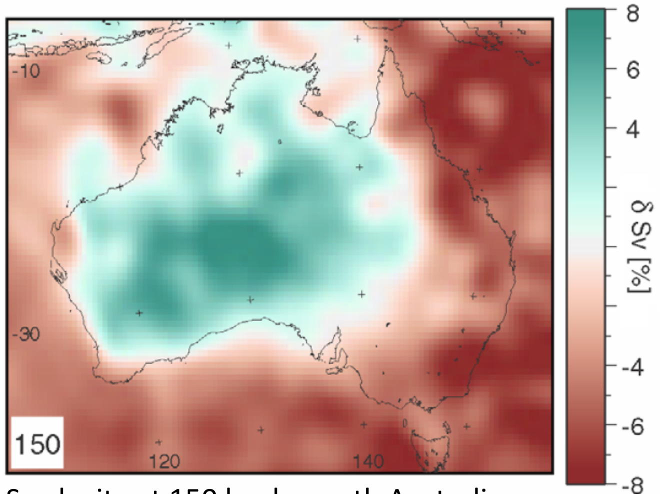
1. Summary of a dream

From Traveltimes To 'Full' Waveforms



'traditional' traveltime tomography

traveltime measurements



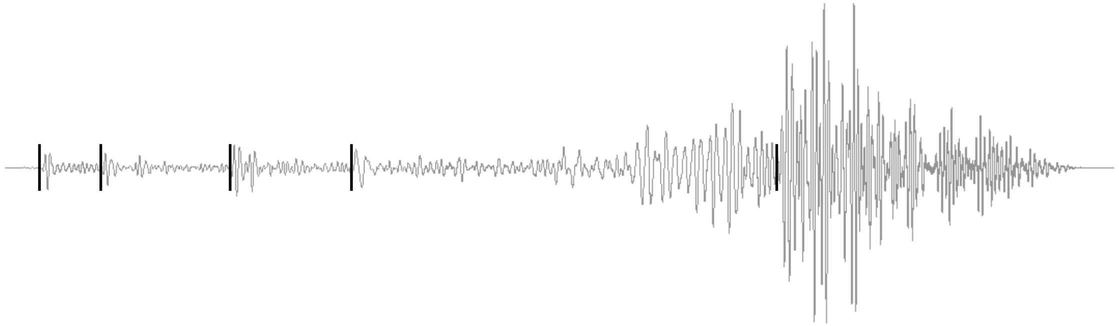
S velocity at 150 km beneath Australia

Fishwick et al., 2005

Extremely successful!

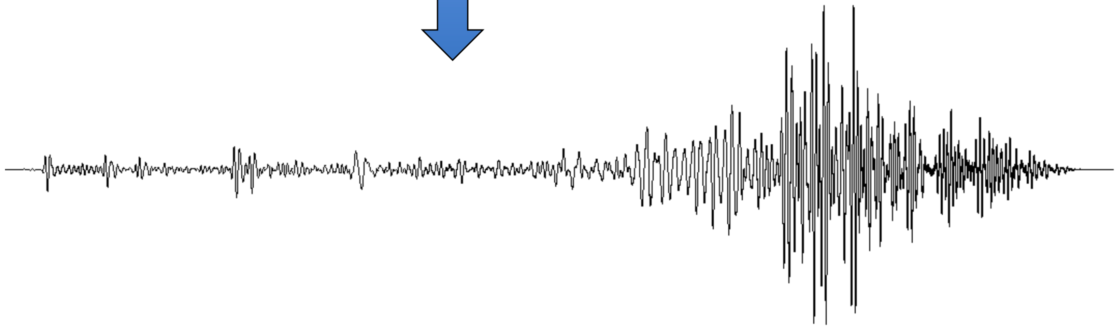
Can assimilate enormous quantities of data.

Still THE most widely used tomographic method.



'traditional' traveltime tomography

traveltime measurements



full-waveform inversion

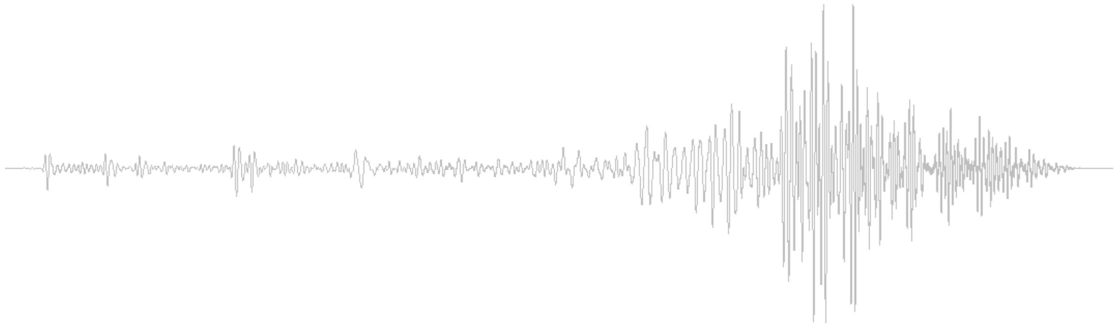
complete seismic recordings

GOALS

- **Explain broadband seismograms wiggle by wiggle ...**
- **... with hardly any human intervention** [Tarantolian black box]
- Better resolved tomographic images
  - thermochemical structure of the Earth
  - evolution and dynamics of the Earth
  - improved ground motion predictions
  - improved earthquake source inversion
    - emergency response, tsunami warning
    - tectonic interpretation
  - improved reservoir characterisation
  - ...

CHALLENGES

- Seismic wave propagation through complex media.
- Computational power.
- Nonlinear relation between waveforms and 3D Earth structure.
- Meaningful measurement of waveform differences.
- Algorithms to search for useful models [**all of them, ideally**].
- ...



full-waveform inversion
complete seismic recordings

GOALS

- **Explain broadband seismograms wiggle by wiggle** ...
- **... with hardly any human intervention** [Tarantolian black box]
- Better resolved tomographic images
  - thermochemical structure of the Earth
  - evolution and dynamics of the Earth
  - improved ground motion predictions
  - improved earthquake source inversion
    - emergency response, tsunami warning
    - tectonic interpretation
  - improved reservoir characterisation
  - ...

2. Formulation as an optimisation problem

- Find an Earth model **m** such that a suitably defined misfit χ is minimal.
- The number of model parameters and the numerical cost of the forward problem prevent the application of probabilistic methods.
- The minimisation proceeds iteratively:

1. Start from initial Earth model $\mathbf{m}_0$
2. Update according to $\mathrm{m}_{i+1} = \mathrm{m}_i + \gamma_i \, \mathrm{h}_i$ with $\chi(m_{i+1}) < \chi(m_i)$

step length → γ_i; descent direction → h_i

- Find an Earth model **m** such that a suitably defined misfit χ is minimal.
- The number of model parameters and the numerical cost of the forward problem prevent the application of probabilistic methods.
- The minimisation proceeds iteratively:

1. Start from initial Earth model $\mathbf{m}_0$
2. Update according to $\mathrm{m}_{i+1} = \mathrm{m}_i + \gamma_i \, \mathrm{h}_i$ with $\chi(m_{i+1}) < \chi(m_i)$

step length → γ_i; descent direction → h_i

Comment:

Minimal does not mean the smallest misfit possible!

The misfit should become about as small as the observational and forward modelling errors.

!!!

1. Start from initial Earth model $\mathbf{m}_0$

2. Update according to $\mathrm{m}_{i+1} = \mathrm{m}_i + \gamma_i \, \mathrm{h}_i$ with $\chi(m_{i+1}) < \chi(m_i)$

step length → γ_i; descent direction → h_i

$$\mathrm{h}_\mathrm{i} \propto -\frac{\partial \chi}{\partial \mathrm{m}_\mathrm{i}}$$

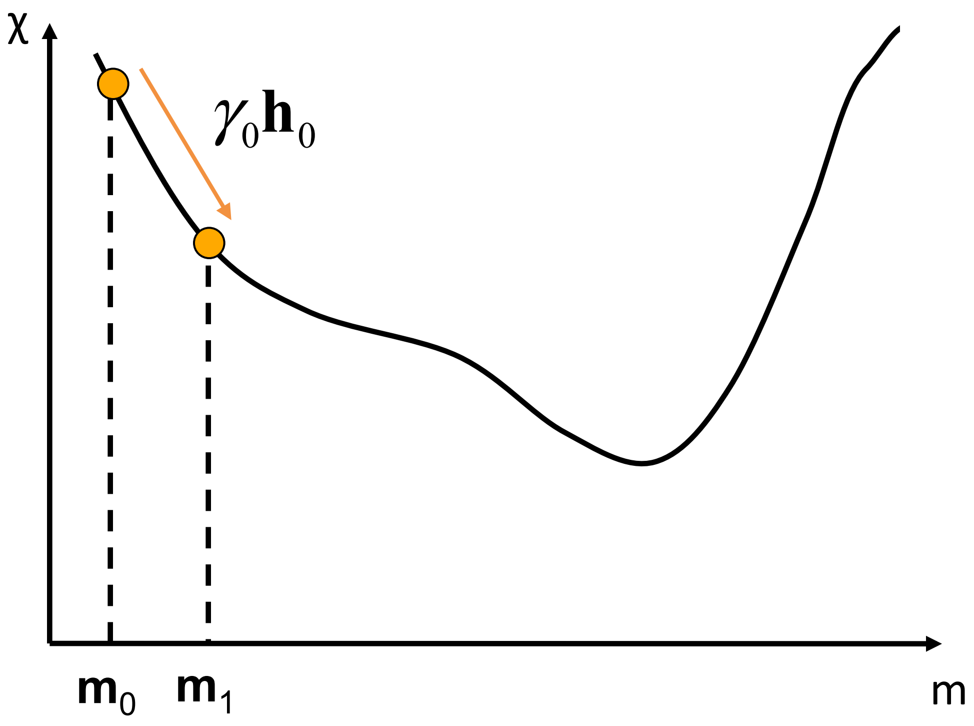
The family of gradient methods:

- method of steepst descent: $\mathrm{h}_\mathrm{i} = -\partial\chi/\partial\mathrm{m}$
- conjugate-gradient methods
- Newton and Newton-like methods
- BFGS and L-BFGS
- …

1. Start from initial Earth model $\mathbf{m}_0$

2. Update according to $\mathrm{m}_{i+1} = \mathrm{m}_i + \gamma_i \, \mathrm{h}_i$ with $\chi(m_{i+1}) < \chi(m_i)$

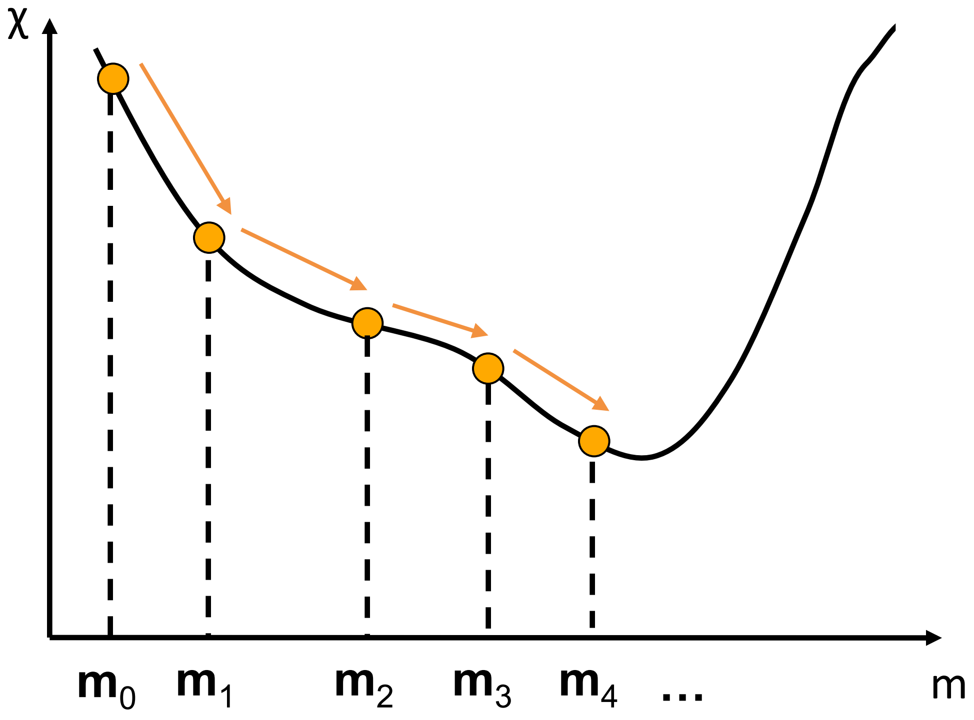
step length → γ_i; descent direction → h_i



1. Start from initial Earth model $\mathbf{m}_0$

2. Update according to $\mathbf{m}_{i+1} = \mathbf{m}_i + \gamma_i \mathbf{h}_i$ with $\chi(m_{i+1}) < \chi(m_i)$

step length (γ_i), descent direction ($\mathbf{h}_i$)



Iteratively approach the minimum misfit by following the local descent directions.

PART II

The adjoint method

1. Problem statement

So, Where Is The Problem?

- The full gradient – with all its components - is needed in each iteration.
- The most straightforward approach: approximate the gradient by finite-differences:

$$\frac{\partial \chi(m)}{\partial m_k} \approx \frac{\chi(\ldots, m_k + \delta m, \ldots) - \chi(\ldots, m_k, \ldots)}{\delta m}$$

- Example with 500,000 model parameters:

	500,001 forward simulations
×	0.5 h per simulation
×	126 compute cores
×	50 sources (earthquakes)
×	50 conjugate gradient iterations
	78e^{9} cpu hours ≈ 8,900,000 cpu years

2. The discrete adjoint method

Regular wave equation

$$\underline{\underline{L}}\,\underline{u} = \underline{f}$$

Adjoint wave equation

$$\underline{\underline{L}}^T \underline{v} = -\nabla\chi$$

Gradient equation

$$\frac{\partial \chi}{\partial m_i} = \underline{v}^T \frac{\partial \underline{\underline{L}}}{\partial m_i} \underline{u}$$

Regular wave equation

$$\underline{\underline{L}}\,\underline{u} = \underline{f}$$

Adjoint wave equation

$$\underline{\underline{L}}^T \underline{v} = -\nabla\chi$$

Gradient equation

$$\frac{\partial\chi}{\partial m_i} = \underline{v}^T \frac{\partial \underline{\underline{L}}}{\partial m_i} \underline{u}$$

Adjoint recipe

1. Solve forward problem [regular wave equation] to obtain $\underline{u}$.
2. Evaluate misfit χ.
3. Compute adjoint source, $-\nabla\chi$.
4. Solve adjoint equation to obtain adjoint field $\underline{v}$.
5. Plug $\underline{u}$ and $\underline{v}$ into the gradient equation.

Regular wave equation

$$\underline{\underline{L}}\,\underline{u} = \underline{f}$$

Adjoint wave equation

$$\underline{\underline{L}}^T \underline{v} = -\nabla\chi$$

Gradient equation

$$\frac{\partial\chi}{\partial m_i} = \underline{v}^T \frac{\partial \underline{\underline{L}}}{\partial m_i} \underline{u}$$

Adjoint recipe

1. Solve forward problem [regular wave equation] to obtain $\underline{u}$.
2. Evaluate misfit χ.
3. Compute adjoint source, $-\nabla\chi$.
4. Solve adjoint equation to obtain adjoint field $\underline{v}$.
5. Plug $\underline{u}$ and $\underline{v}$ into the gradient equation.

Comments

1. No need to explicitly compute the derivative of the wavefield $\underline{u}$ [by construction].
2. Gradient is entirely determined by the definition of the misfit [adjoint source is the only thing that explicitly depends on the misfit].
3. Computation of gradient requires storage of forward wavefield $\underline{u}$.

3. The continuous adjoint method

Discrete case [frequency domain]

$$\underline{\underline{L}}\,\underline{u} = (-\omega^2 \underline{\underline{M}} + \underline{\underline{K}})\underline{u}$$

$$\nabla\chi = \underline{v}^T \nabla \underline{\underline{L}}\,\underline{u}$$

Continuous case [time domain]

$$L(\underline{u}) = \rho\ddot{\underline{u}} - \nabla\cdot(C:\nabla\underline{u}) = \underline{f}$$

$$\nabla\chi' = \int \underline{v}^T \nabla L(\underline{u})\,dt$$

- The same formal derivation from the discrete case can be used in the continuous case.
  - Matrix $\underline{\underline{L}}$ becomes operator L.
  - Scalar product $\underline{a}^T\underline{b}$ becomes integral $\int a(x)b(x)\,dx$.
- In somewhat loose terms, $\nabla\chi$ is called a **sensitivity** or **Fréchet kernel** and symbolised by K.
- The only question: What is L^T in the continuous case? ... **See Russel Hewett's lecture!**

Regular wave equation

momentum balance

$$\rho(\mathbf{x})\ddot{\mathbf{u}}(\mathbf{x},t) - \nabla \cdot \sigma(\mathbf{x},t) = \mathbf{f}(\mathbf{x},t)$$

stress-strain relation

$$\sigma(\mathbf{x},t) = \int_{\tau=t_0}^{\infty} \dot{\mathbf{C}}(\mathbf{x},t-\tau) : \nabla\mathbf{u}(\mathbf{x},\tau)\,d\tau$$

initial conditions

$$\mathbf{u}|_{t\leq t_0} = \dot{\mathbf{u}}|_{t\leq t_0} = \mathbf{0}$$

boundary conditions

$$\mathbf{n}\cdot\sigma|_{\mathbf{x}\in\partial G} = \mathbf{0}$$

Regular wave equation

momentum balance

$$\rho(\mathbf{x})\ddot{\mathbf{u}}(\mathbf{x},t) - \nabla \cdot \boldsymbol{\sigma}(\mathbf{x},t) = \mathbf{f}(\mathbf{x},t)$$

stress-strain relation

$$\boldsymbol{\sigma}(\mathbf{x},t) = \int_{\tau=t_0}^{\infty} \dot{\mathbf{C}}(\mathbf{x},t-\tau) : \nabla\mathbf{u}(\mathbf{x},\tau)\, d\tau$$

initial conditions

$$\mathbf{u}|_{t\leq t_0} = \dot{\mathbf{u}}|_{t\leq t_0} = \mathbf{0}$$

boundary conditions

$$\mathbf{n}\cdot\boldsymbol{\sigma}|_{\mathbf{x}\in\partial G} = \mathbf{0}$$

Adjoint wave equation

adjoint momentum balance

$$\rho\ddot{\mathbf{u}}^{\dagger} - \nabla\cdot\boldsymbol{\sigma}^{\dagger} = -\nabla_u \chi$$

adjoint stress-strain relation

$$\boldsymbol{\sigma}^{\dagger}(t) = \int_{\tau=t}^{t_1} \dot{\mathbf{C}}(\tau-t) : \nabla\mathbf{u}^{\dagger}(\tau)\, d\tau$$

terminal conditions

$$\mathbf{u}^{\dagger}|_{t\geq t_1} = \dot{\mathbf{u}}^{\dagger}|_{t\geq t_1} = \mathbf{0}$$

boundary conditions

$$\mathbf{n}\cdot\boldsymbol{\sigma}^{\dagger}|_{\mathbf{x}\in\partial G} = \mathbf{0}$$

notation

$$\mathrm{v} = \mathrm{u}^{\dagger}$$

Regular wave equation

momentum balance

$$\rho(\mathbf{x})\ddot{\mathbf{u}}(\mathbf{x},t) - \nabla\cdot\boldsymbol{\sigma}(\mathbf{x},t) = \mathbf{f}(\mathbf{x},t)$$

stress-strain relation

$$\boldsymbol{\sigma}(\mathbf{x},t) = \int_{\tau=t_0}^{\infty} \dot{\mathbf{C}}(\mathbf{x},t-\tau) : \nabla\mathbf{u}(\mathbf{x},\tau)\,d\tau$$

initial conditions

$$\mathbf{u}|_{t\le t_0} = \dot{\mathbf{u}}|_{t\le t_0} = \mathbf{0}$$

boundary conditions

$$\mathbf{n}\cdot\boldsymbol{\sigma}|_{\mathbf{x}\in\partial G} = \mathbf{0}$$

Adjoint wave equation

adjoint momentum balance

$$\rho\ddot{\mathbf{u}}^{\dagger} - \nabla\cdot\boldsymbol{\sigma}^{\dagger} = -\nabla_u\chi$$

adjoint stress-strain relation

$$\boldsymbol{\sigma}^{\dagger}(t) = \int_{\tau=t}^{t_1} \dot{\mathbf{C}}(\tau-t) : \nabla\mathbf{u}^{\dagger}(\tau)\,d\tau$$

terminal conditions

$$\mathbf{u}^{\dagger}|_{t\ge t_1} = \dot{\mathbf{u}}^{\dagger}|_{t\ge t_1} = \mathbf{0}$$

boundary conditions

$$\mathbf{n}\cdot\boldsymbol{\sigma}^{\dagger}|_{\mathbf{x}\in\partial G} = \mathbf{0}$$

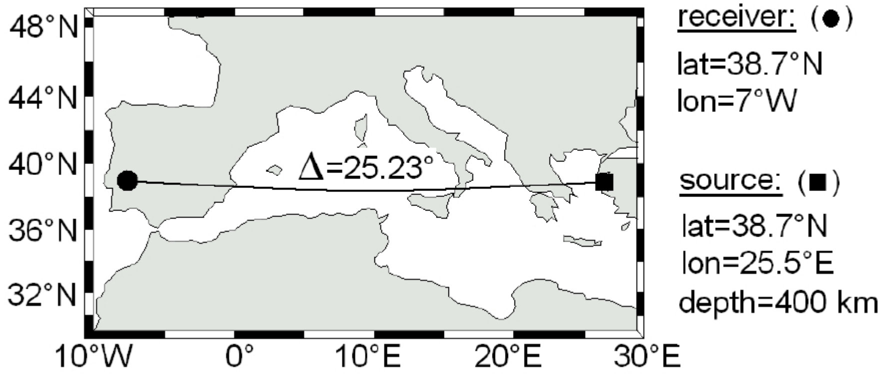
Comments

- Adjoint equation is a wave equation [same code can be used for its solution].
- Solving terminal conditions can be done by running code in reversed time.

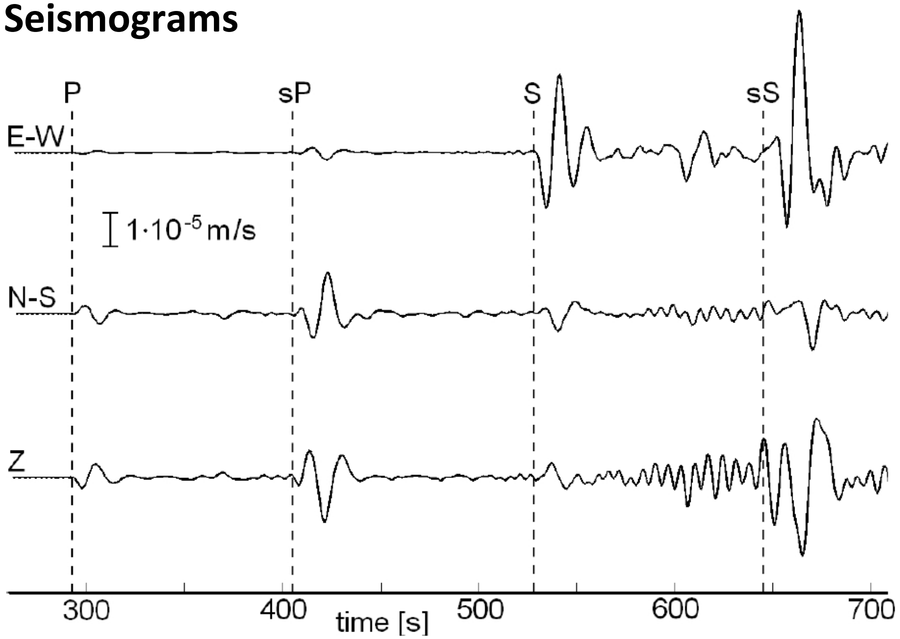
4. Sensitivity kernels

Traveltime Measurement On Specific Phases

Source-receiver geometry

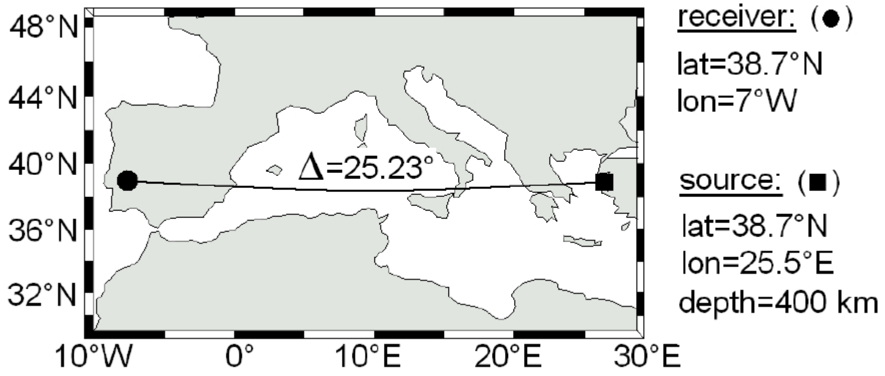


Seismograms

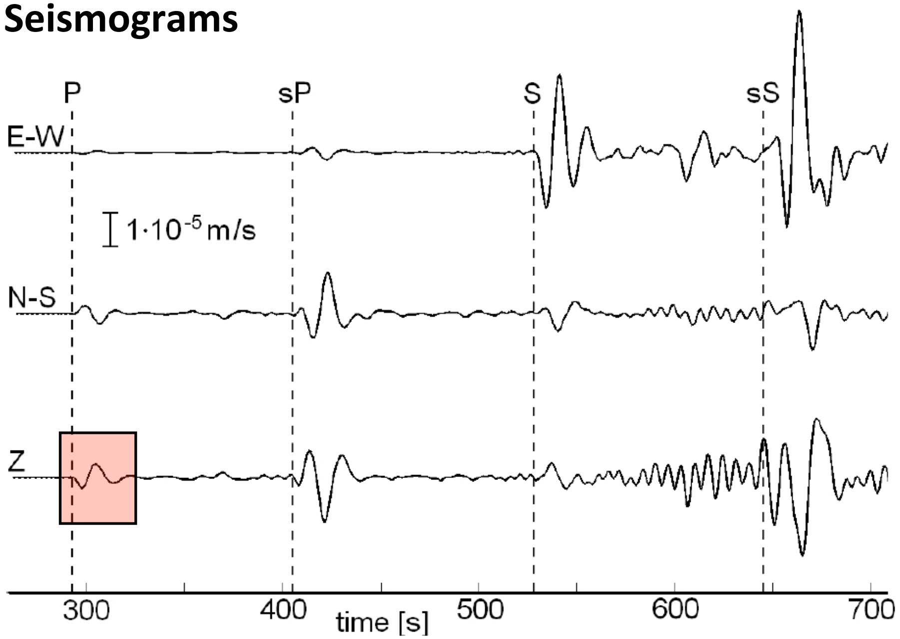


Traveltime Measurement On Specific Phases

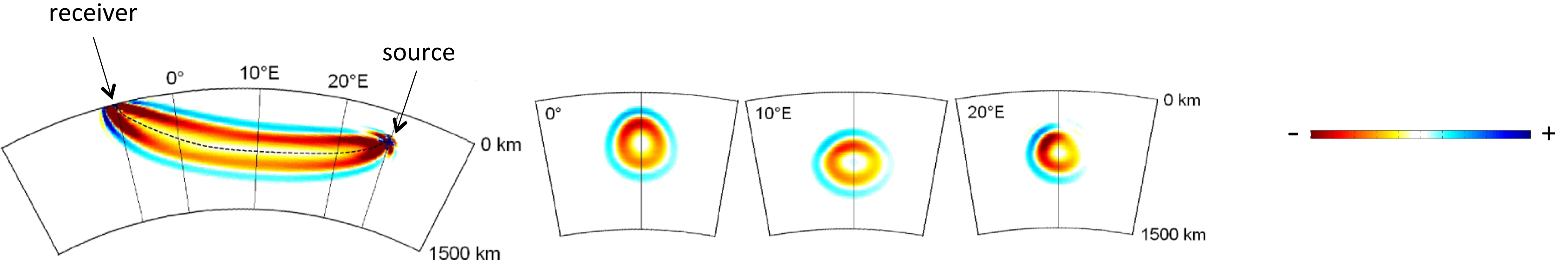
Source-receiver geometry



Seismograms

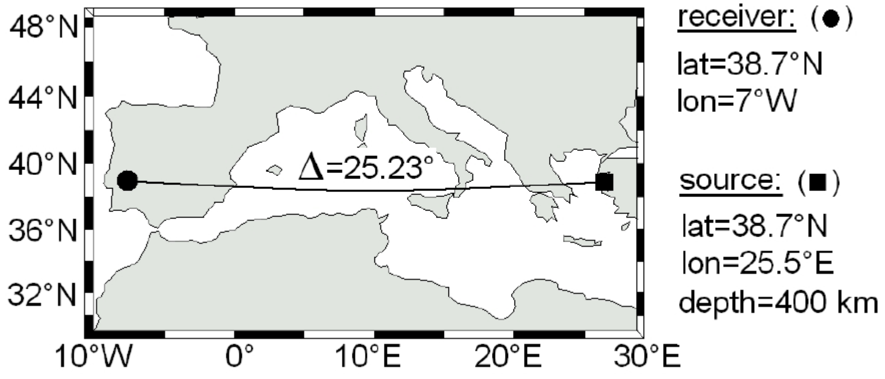


Sensitivity kernel for P wave velocity

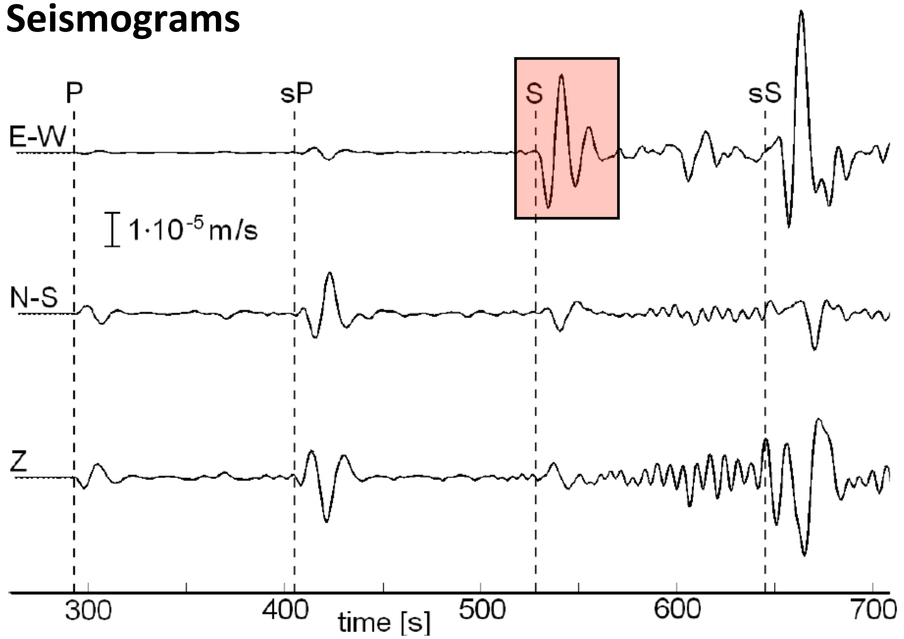


Traveltime Measurement On Specific Phases

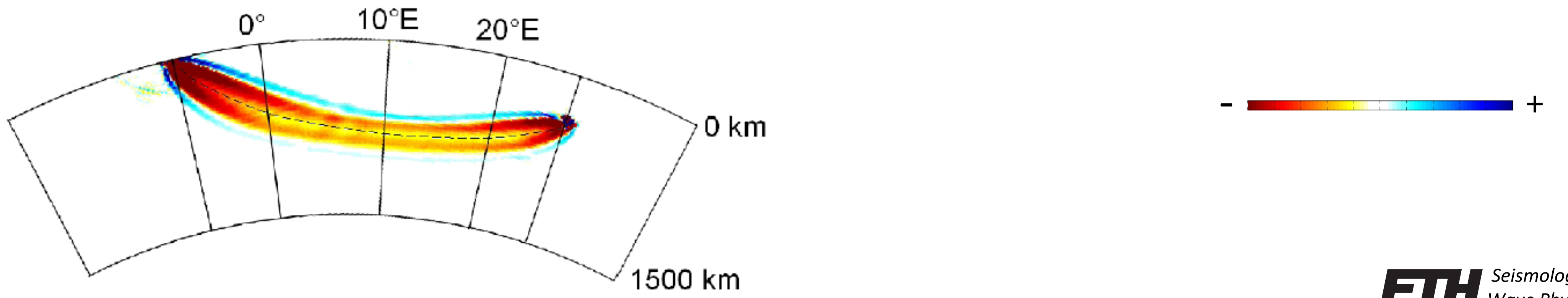
Source-receiver geometry



Seismograms

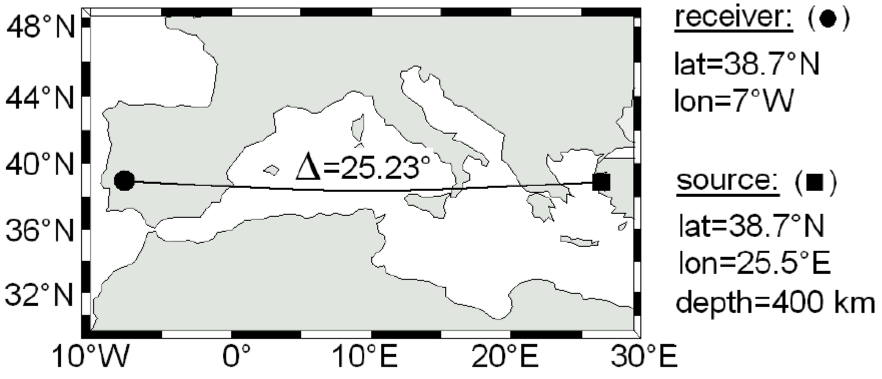


Sensitivity kernel for S wave velocity

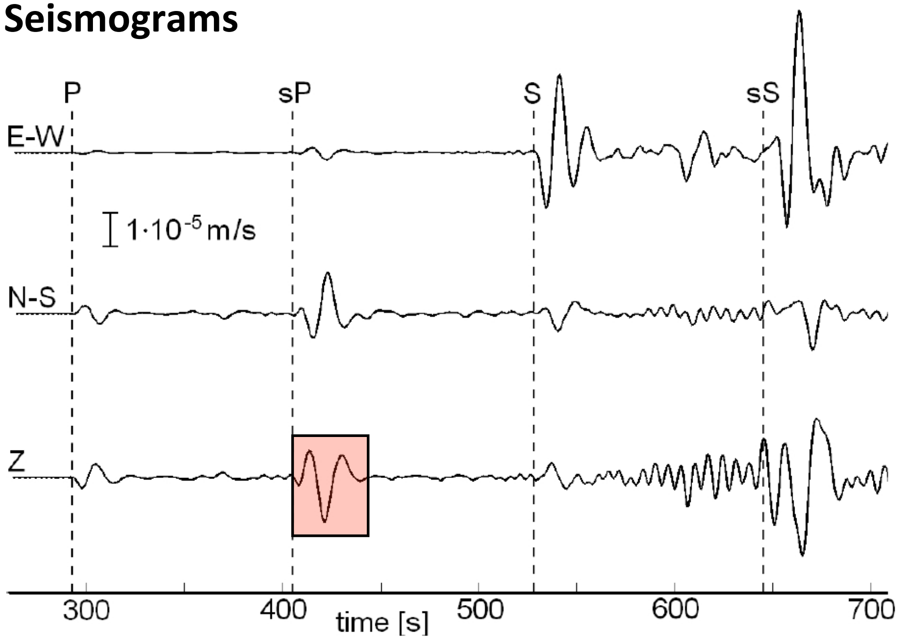


Traveltime Measurement On Specific Phases

Source-receiver geometry

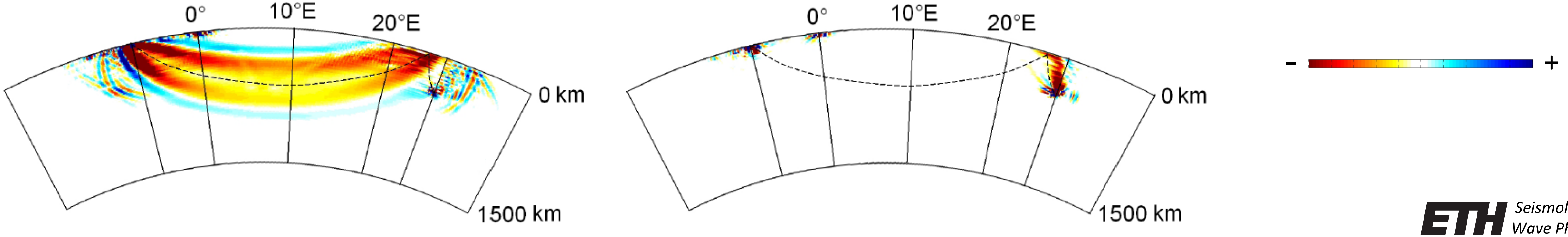


Seismograms



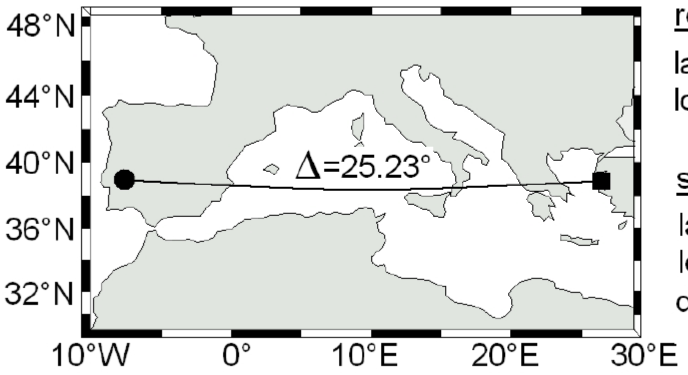
Sensitivity kernels for

P wave velocity and S wave velocity



Traveltime Measurement On Specific Phases

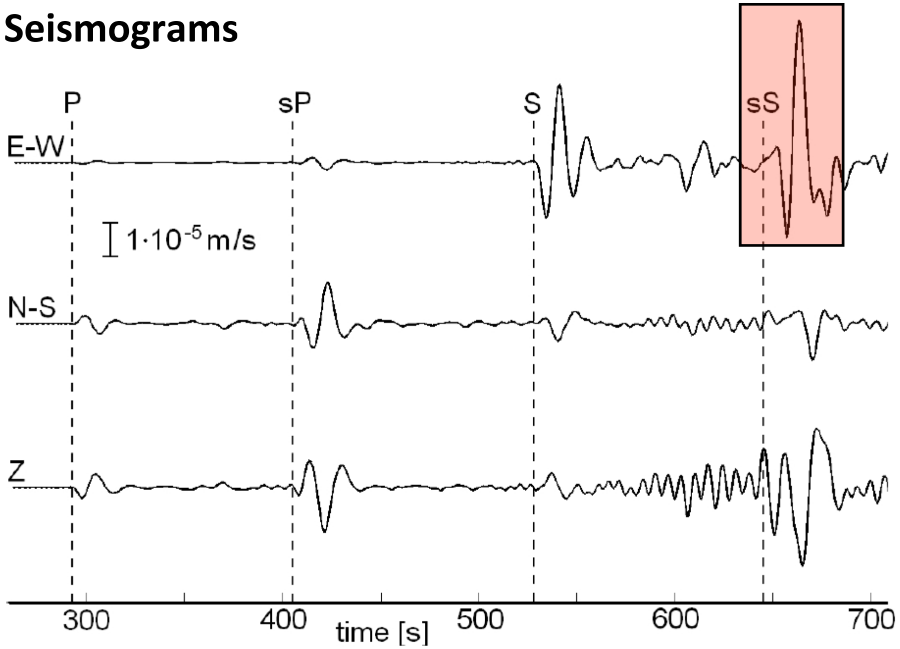
Source-receiver geometry



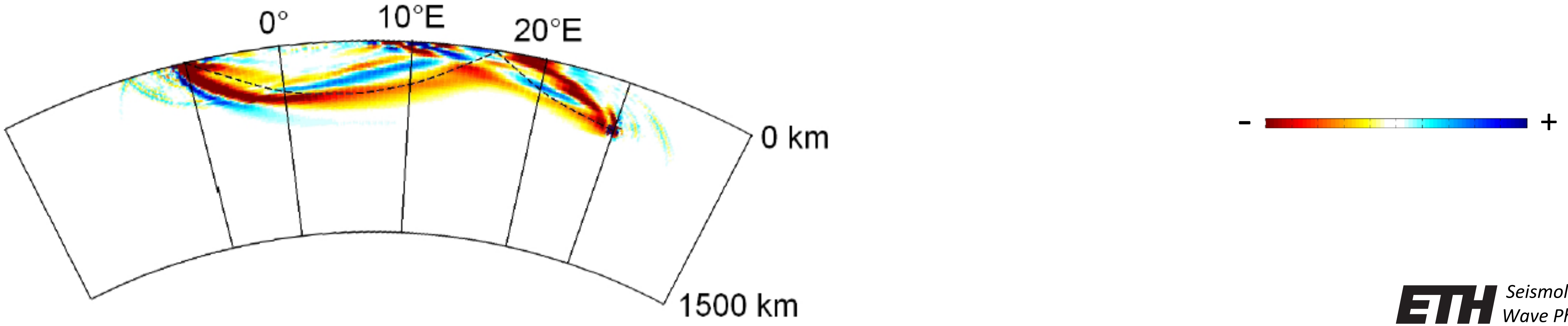
receiver: (●)
lat=38.7°N
lon=7°W

source: (■)
lat=38.7°N
lon=25.5°E
depth=400 km

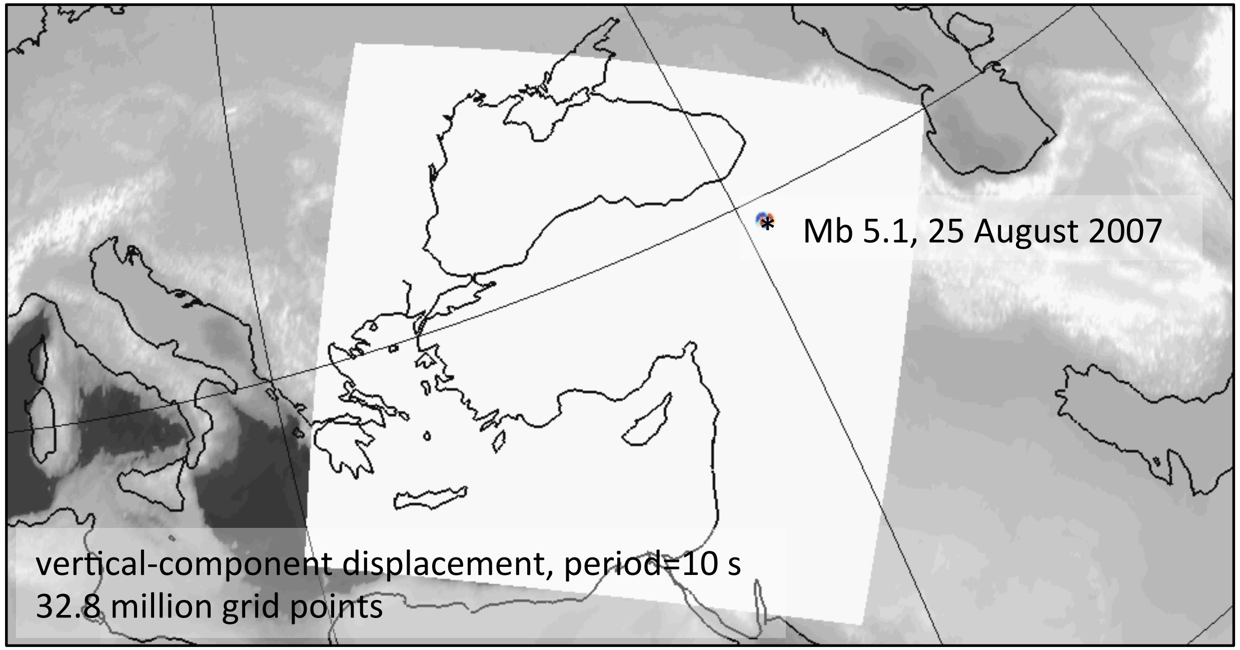
Seismograms

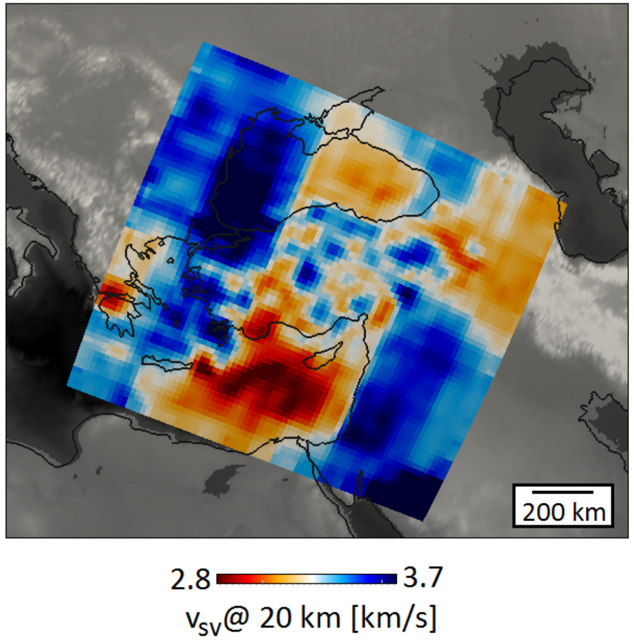


Sensitivity kernel for S wave velocity

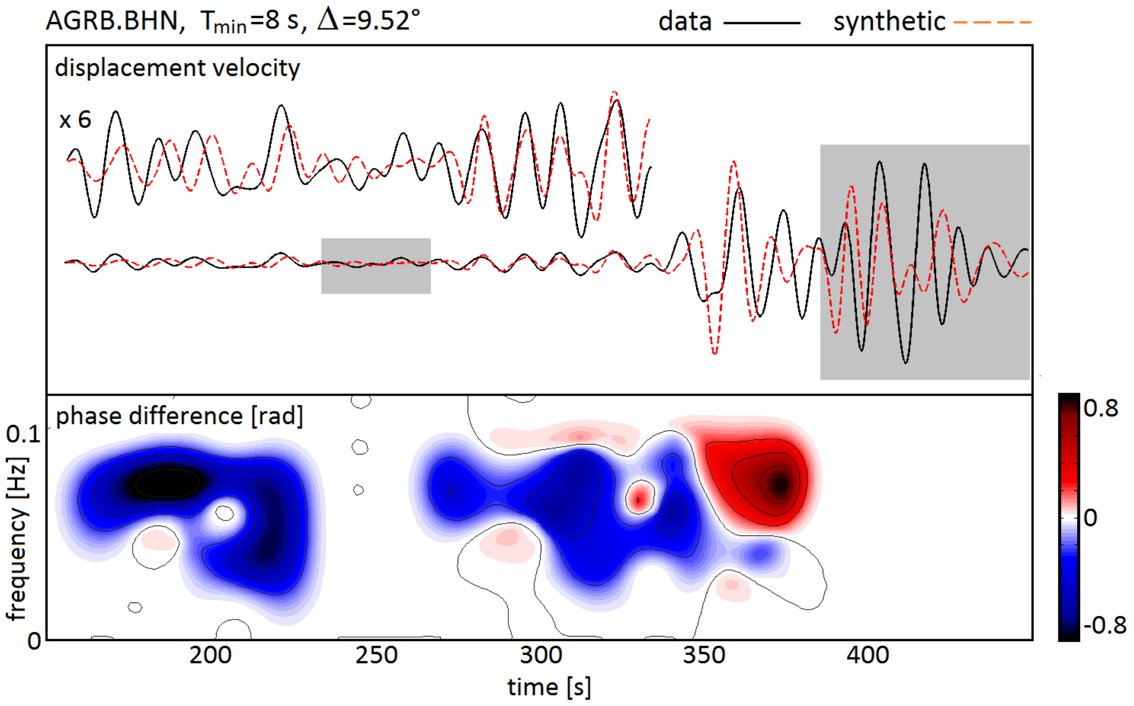


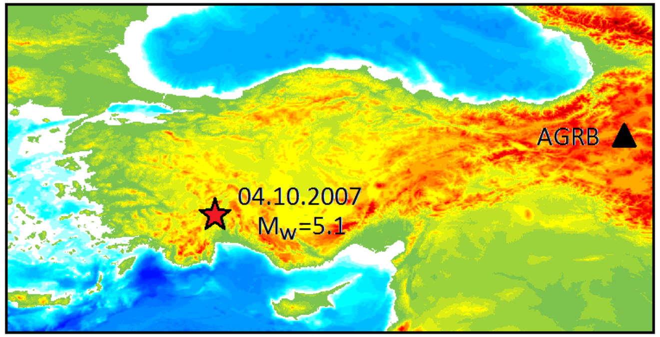
Seismology & Wave Physics


Mb 5.1, 25 August 2007
vertical-component displacement, period=10 s
32.8 million grid points


200 km
2.8
3.7
v_{sv}@ 20 km [km/s]

MEASURING TIME-FREQUENCY PHASE DIFFERENCES

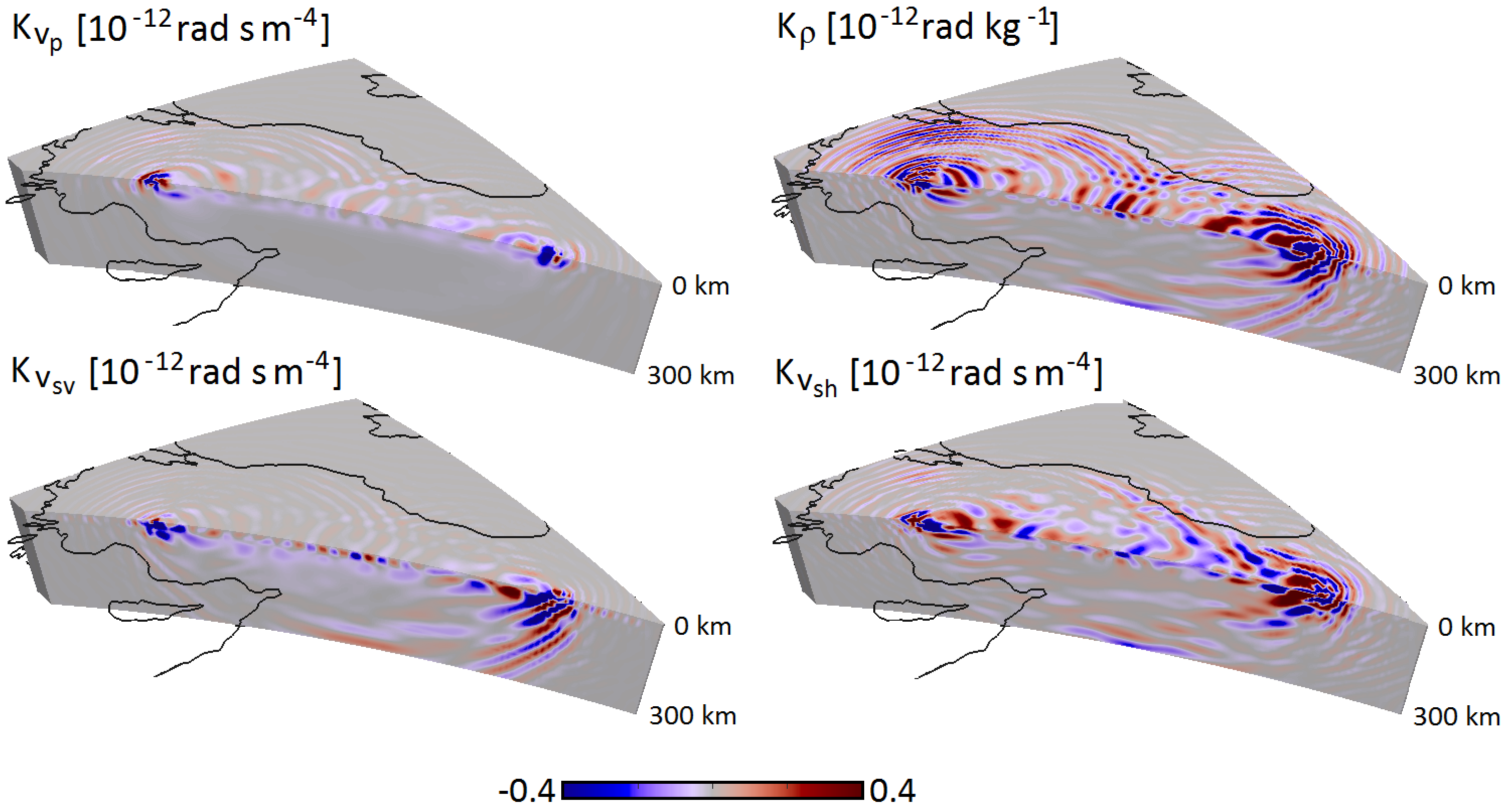




- Time- and frequency-dependent phase differences
- Based on selection of time windows where data and synthetics are similar
- Independent of absolute amplitudes

Measuring Time-Frequency Phase Differences

Sensitivity kernels



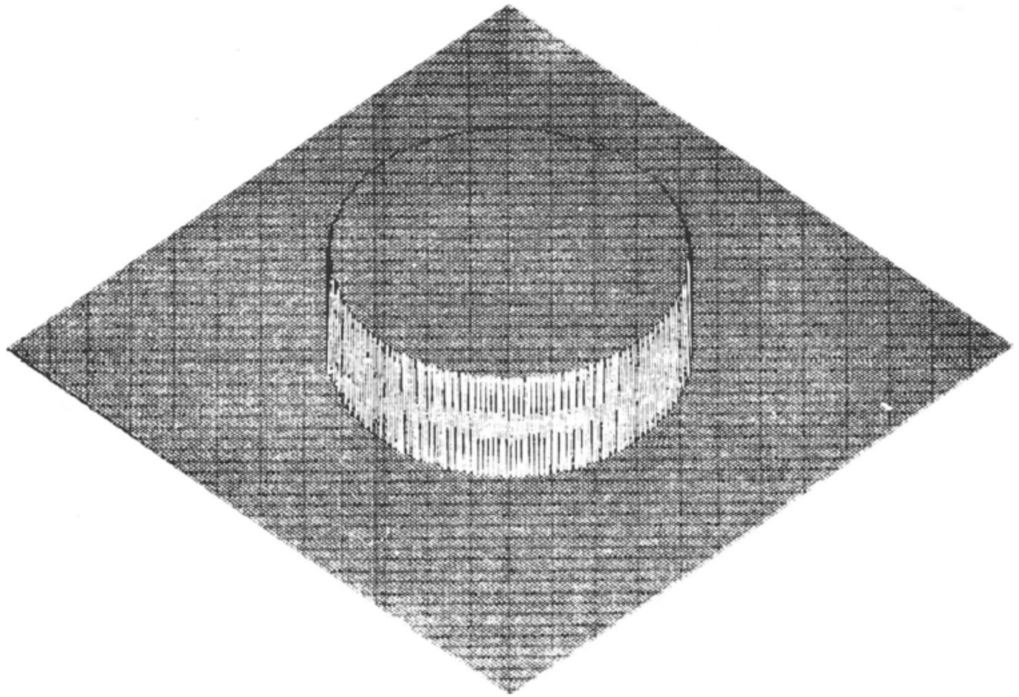
PART III

Advanced Topics

1. Local minima and the multiscale approach

The acoustic *Camembert Model*

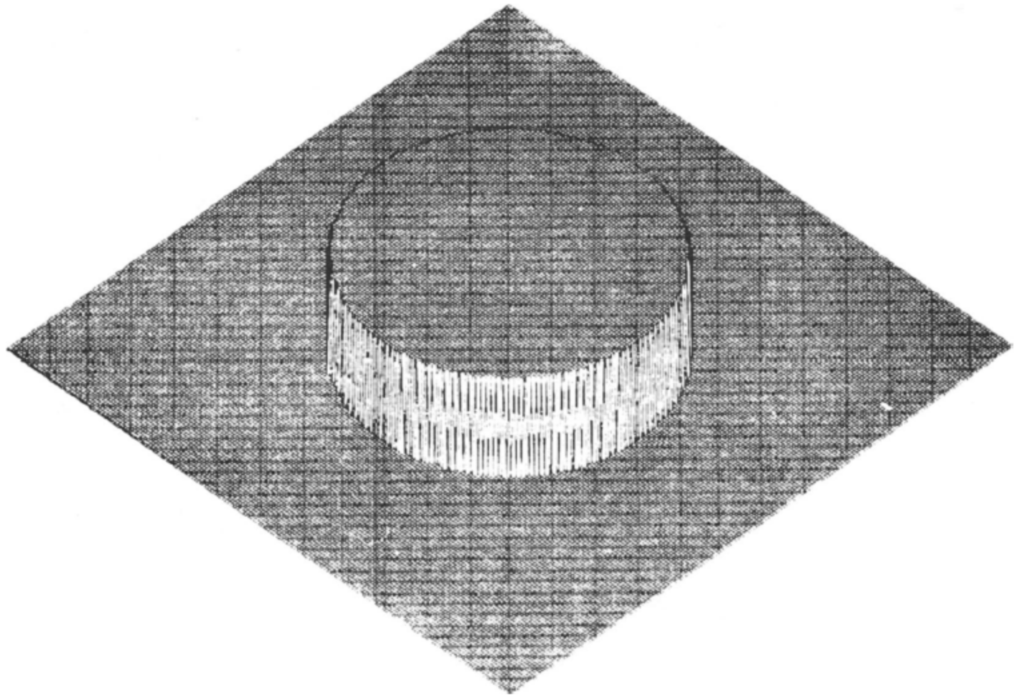
- 20 % velocity perturbation
- 8 sources and receivers around the model



Odile Gauthier, **Jean Virieux**, Albert Tarantola, Geophysics 1986.

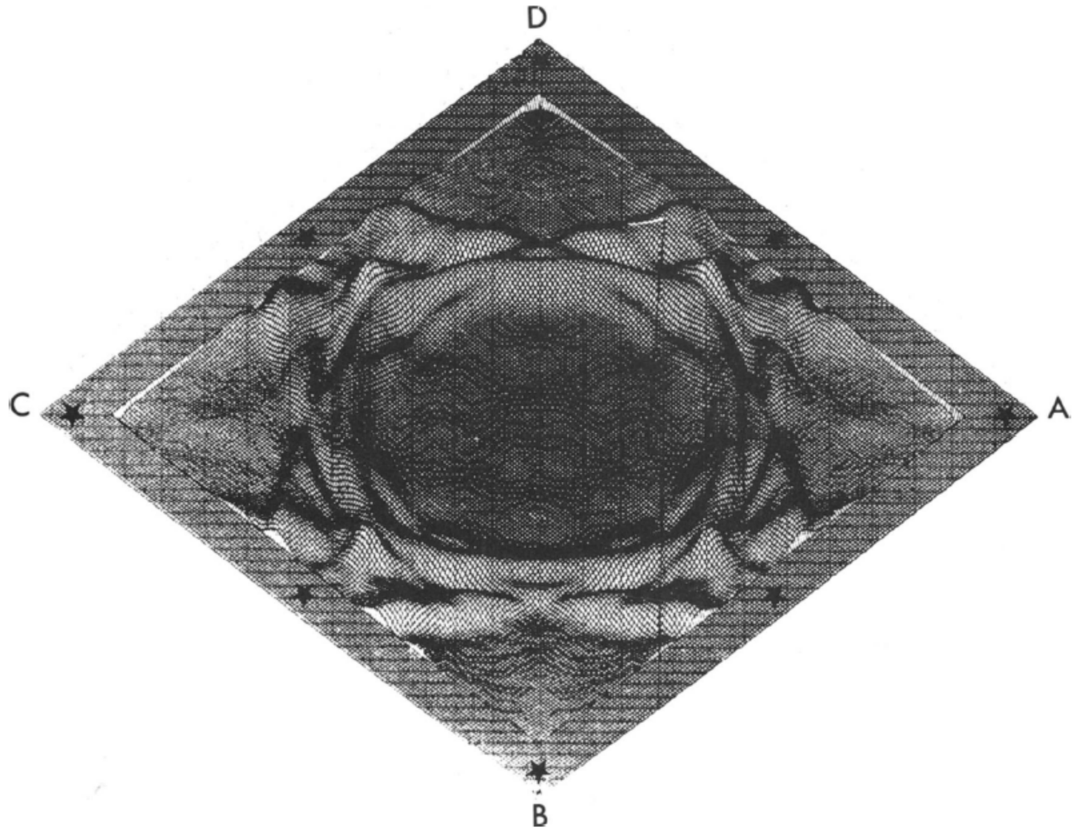
The acoustic *Camembert Model*

- 20 % velocity perturbation
- 8 sources and receivers around the model

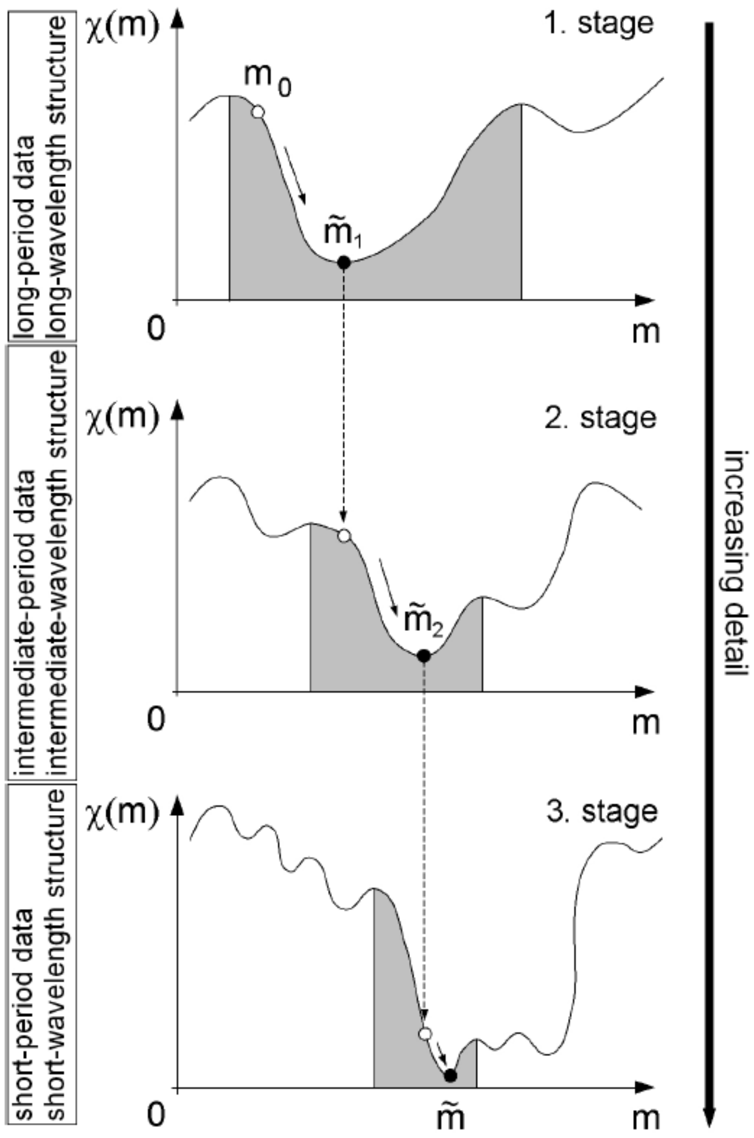


Inversion result after 5 iterations

- *Camembert* not recovered
- Stuck in a local minimum



Odile Gauthier, **Jean Virieux**, Albert Tarantola, Geophysics 1986.



- Identifies **cycle skipping** as main reason for nonlinearity.
- Misfit surface more complex the higher the frequency.
- Start with low frequencies.
- Work your way up to high frequencies.

- **Problem still**: Low frequencies may not always be available

2. Compressed wavefield storage

Sensitivity kernel examples

$$K_\rho = -\int_T \dot{\mathbf{u}}^\dagger \cdot \dot{\mathbf{u}}\, dt\,,$$

$$K_\lambda = \int_T (\nabla \cdot \mathbf{u})(\nabla \cdot \mathbf{u}^\dagger)\, dt\,,$$

$$K_\mu = \int_T [(\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u}) + (\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u})^T]\, dt$$

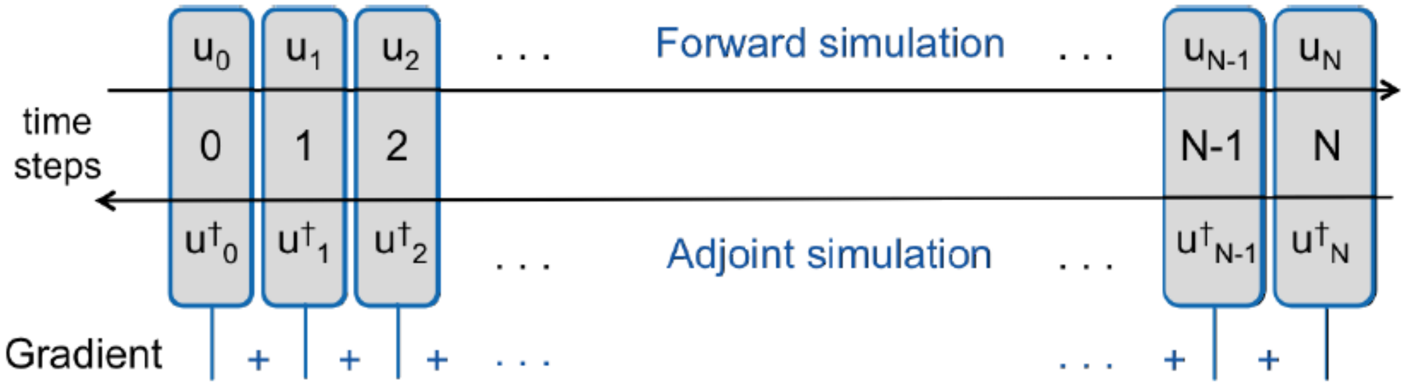
Sensitivity kernel examples

$$K_\rho = -\int_T \dot{\mathbf{u}}^\dagger \cdot \dot{\mathbf{u}}\, dt\,,$$

$$K_\lambda = \int_T (\nabla \cdot \mathbf{u})(\nabla \cdot \mathbf{u}^\dagger)\, dt\,,$$

$$K_\mu = \int_T [(\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u}) + (\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u})^T]\, dt$$

- Forward and adjoint fields must be **known at the same time**.
- This is not naturally the case.
- Forward wavefield needs to be **stored**.
- This is extremely **expensive**!



Sensitivity kernel examples

$$K_\rho = -\int_T \dot{\mathbf{u}}^\dagger \cdot \dot{\mathbf{u}}\, dt\,,$$

$$K_\lambda = \int_T (\nabla \cdot \mathbf{u})(\nabla \cdot \mathbf{u}^\dagger)\, dt\,,$$

$$K_\mu = \int_T [(\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u}) + (\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u})^T]\, dt$$

- Forward and adjoint fields must be **known at the same time**.
- This is not naturally the case.
- Forward wavefield needs to be **stored**.
- This is extremely **expensive**!

Can we somehow <u>compress the wavefield</u> *u* such that the kernel integrals are still sufficiently <u>accurate</u>?

1. Requantisation
   - adjust number of bits to represent field values
   - large number of bits in regions with large amplitude variations and vice versa

1. Requantisation
   - adjust number of bits to represent field values
   - large number of bits in regions with large amplitude variations and vice versa

2. p-coarsening
   - store wavefield with polynomial degree p as a new polynomial of degree $p_{new} < p$
   - re-interpolate to approximate the kernel integral

1. Requantisation
   - adjust number of bits to represent field values
   - large number of bits in regions with large amplitude variations and vice versa

2. p-coarsening
   - store wavefield with polynomial degree p as a new polynomial of degree $p_{new}<p$
   - re-interpolate to approximate the kernel integral

3. Temporal interpolation
   - Store wavefield only every n^{th} time step
   - Spline interpolation to fill missing time steps for kernel integral

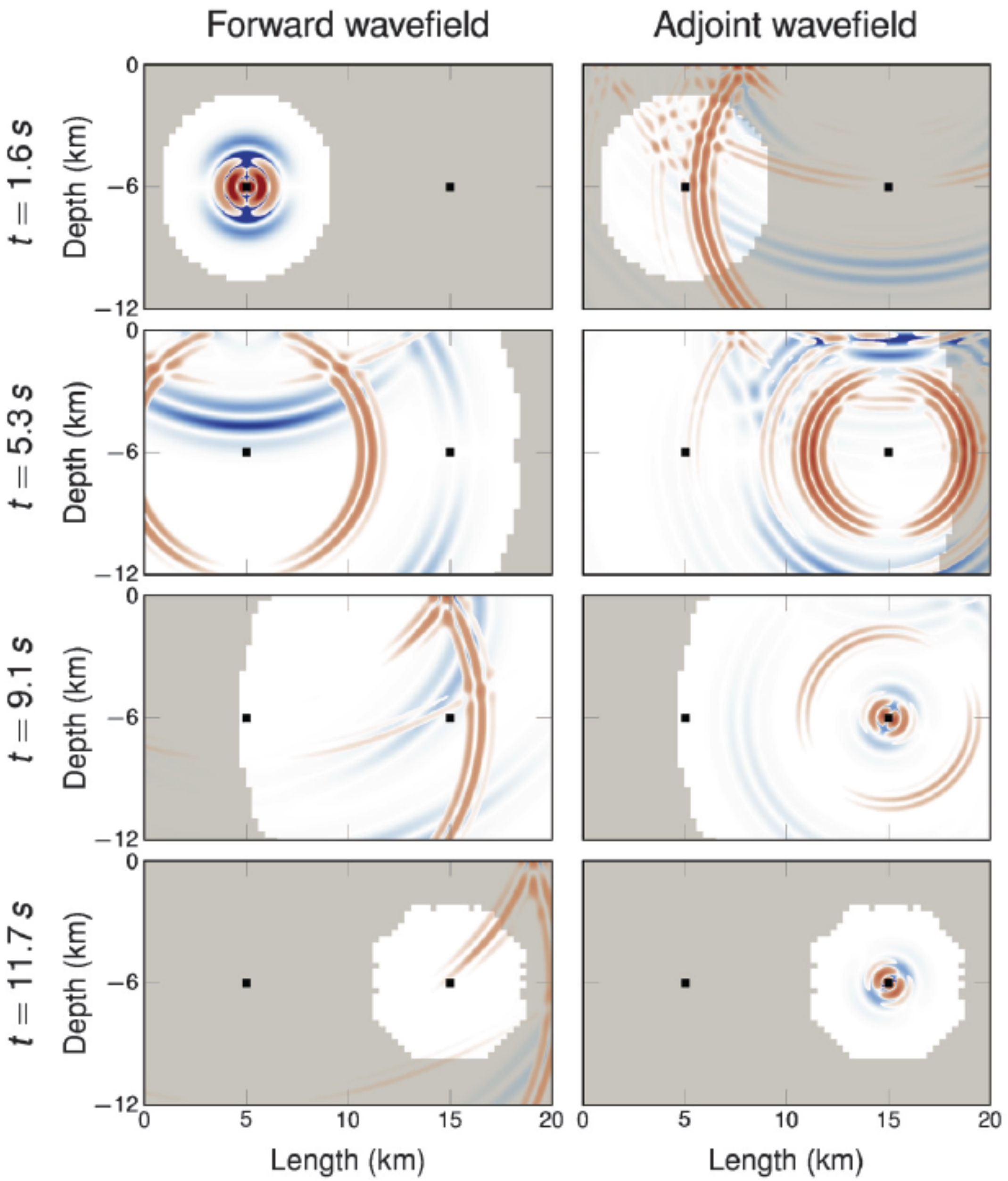
1. Requantisation
   - adjust number of bits to represent field values
   - large number of bits in regions with large amplitude variations and vice versa

2. p-coarsening
   - store wavefield with polynomial degree p as a new polynomial of degree $p_{new}<p$
   - re-interpolate to approximate the kernel integral

3. Temporal interpolation
   - Store wavefield only every n^{th} time step
   - Spline interpolation to fill missing time steps for kernel integral

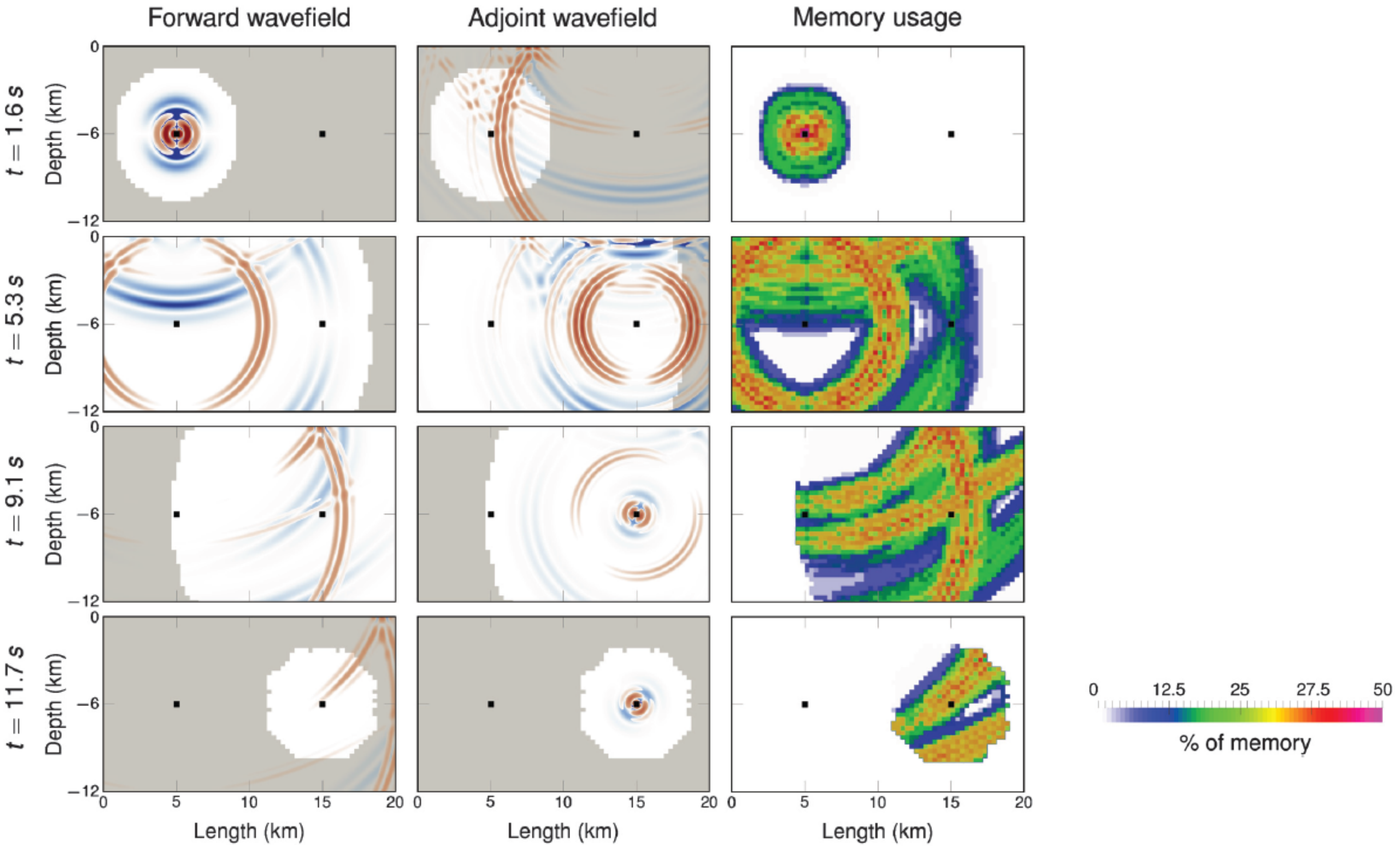
4. Lazy forward and adjoint simulations
   - Store forward and adjoint field only in regions where they overlap

EXAMPLE [COMBINING THESE STRATEGIES]



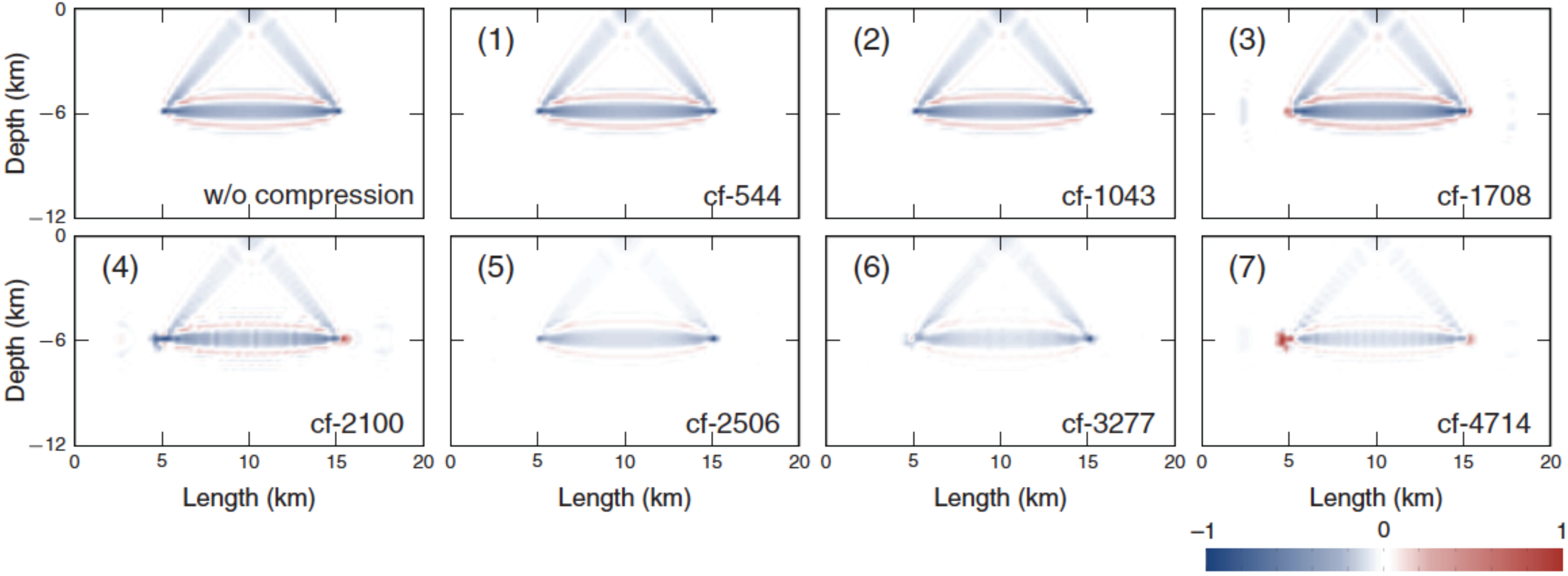
Boehm & Fichtner, Geophysics 2016

EXAMPLE [COMBINING THESE STRATEGIES]



Boehm & Fichtner, Geophysics 2016

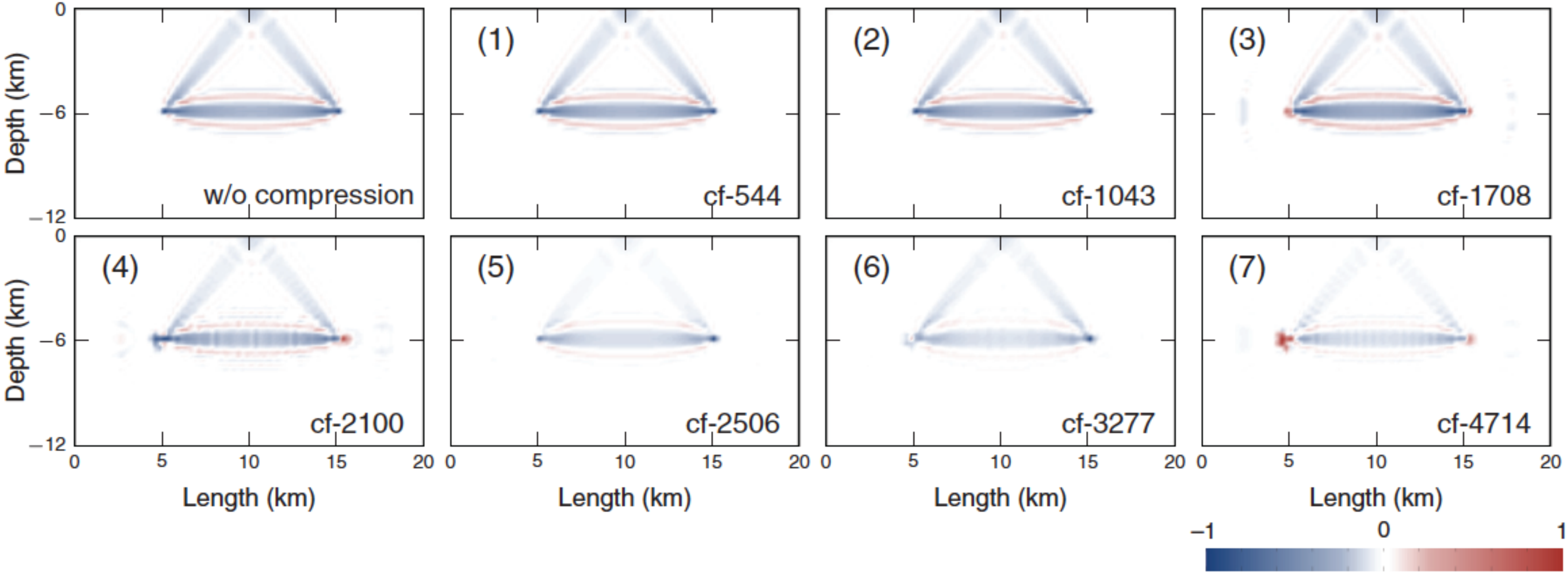
EXAMPLE [COMBINING THESE STRATEGIES]



cf = compression factor

Boehm & Fichtner, Geophysics 2016

EXAMPLE [COMBINING THESE STRATEGIES]



cf = compression factor

A compression factor of $\mathcal{O}(1000)$ is often feasible.

Boehm & Fichtner, Geophysics 2016

ETH Seismology & Wave Physics

3. Second derivatives

WHY ARE WE INTERESTED IN SECOND DERIVATIVES?

- Quadratic approximation of the misfit functional near the optimal model [approximately vanishing first derivative].

$$\chi(\mathbf{m}_{opt} + \delta\mathbf{m}) \approx \chi(\mathbf{m}_{opt}) + \delta\mathbf{m}^T \mathbf{H}\, \delta\mathbf{m}$$

misfit functional

optimal Earth model

Hessian at m_{opt}

- The Hessian **H** [second-derivative matrix]:
  - Local geometry of the misfit surface
  - resolution and trade-offs
  - H = inverse posterior covariance
  - ➢ **H contains information on uncertainties!**

- **H** cannot be computed explicity, and if we could, we would not be able to store it!
- But we can compute **H** d**m** for any arbitrary d**m**:

Second Derivatives

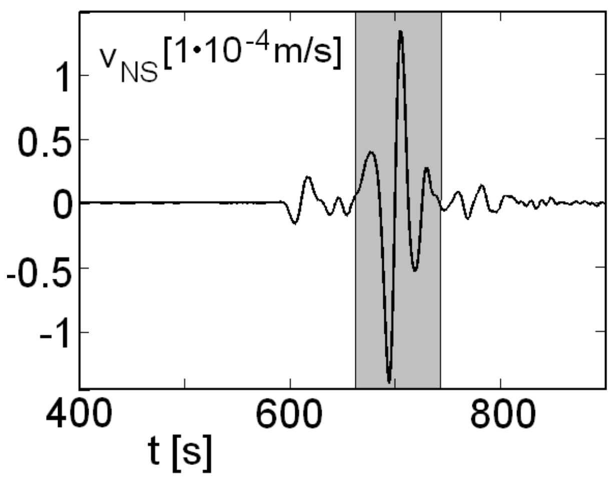
- **H** cannot be computed explicity, and if we could, we would not be able to store it!
- But we can compute **H** d**m** for any arbitrary d**m**:

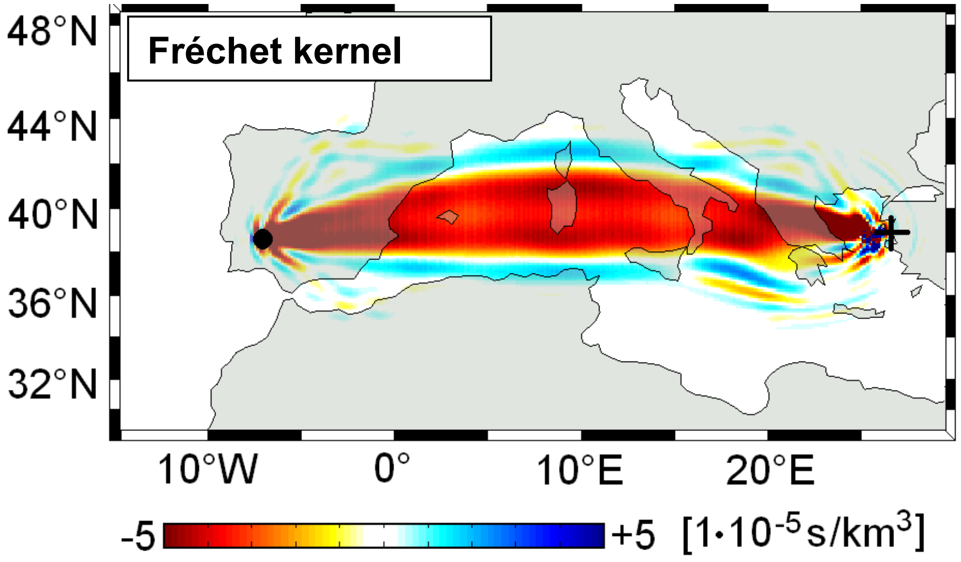
- Second derivative = first derivative (first derivative)
- Finite-difference approximation of second derivative = difference of first derivatives:

$$H(m)dm \propto K(m+dm)-K(m)$$

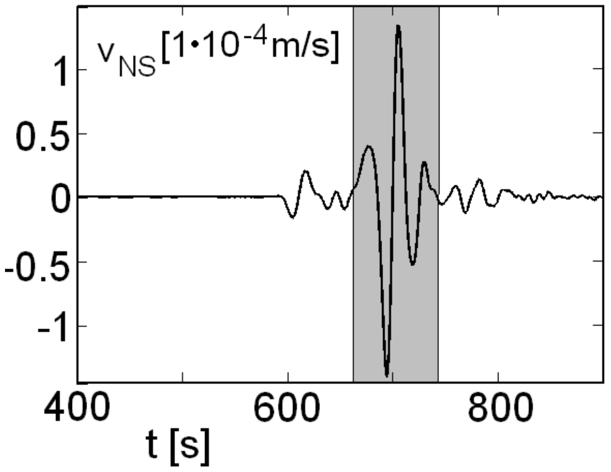
- **H** d**m** can trivially be approximated by subtracting two sensitivity kernels.
- Also possible without approximation [beyond scope of this lecture, details: Fichtner & Trampert, GJI 2011].

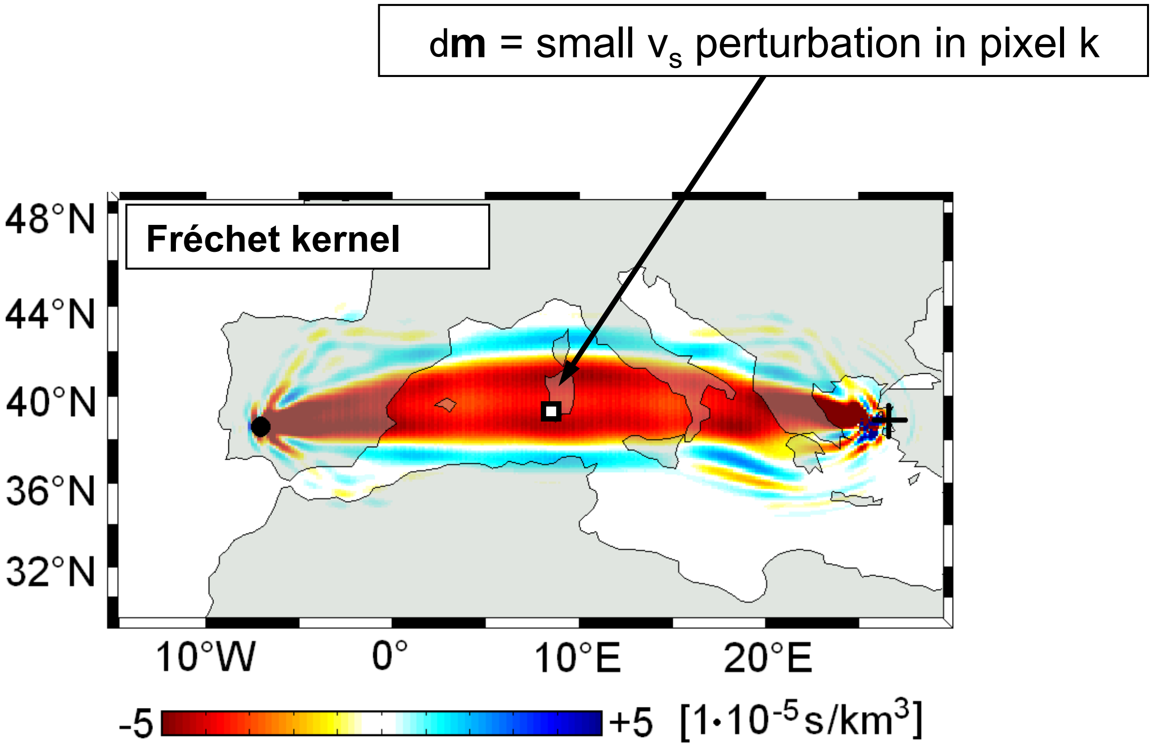
- 25 s Love wave
- finite-frequency traveltime





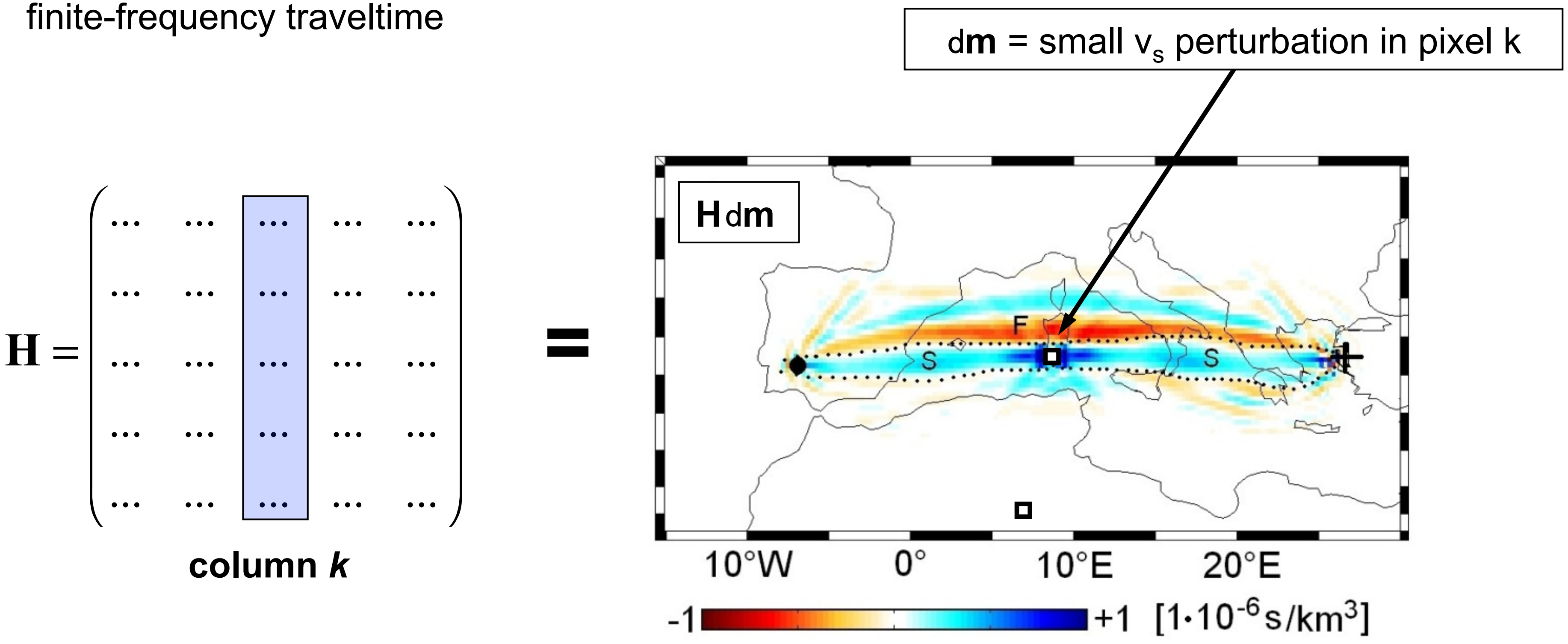
- 25 s Love wave
- finite-frequency traveltime



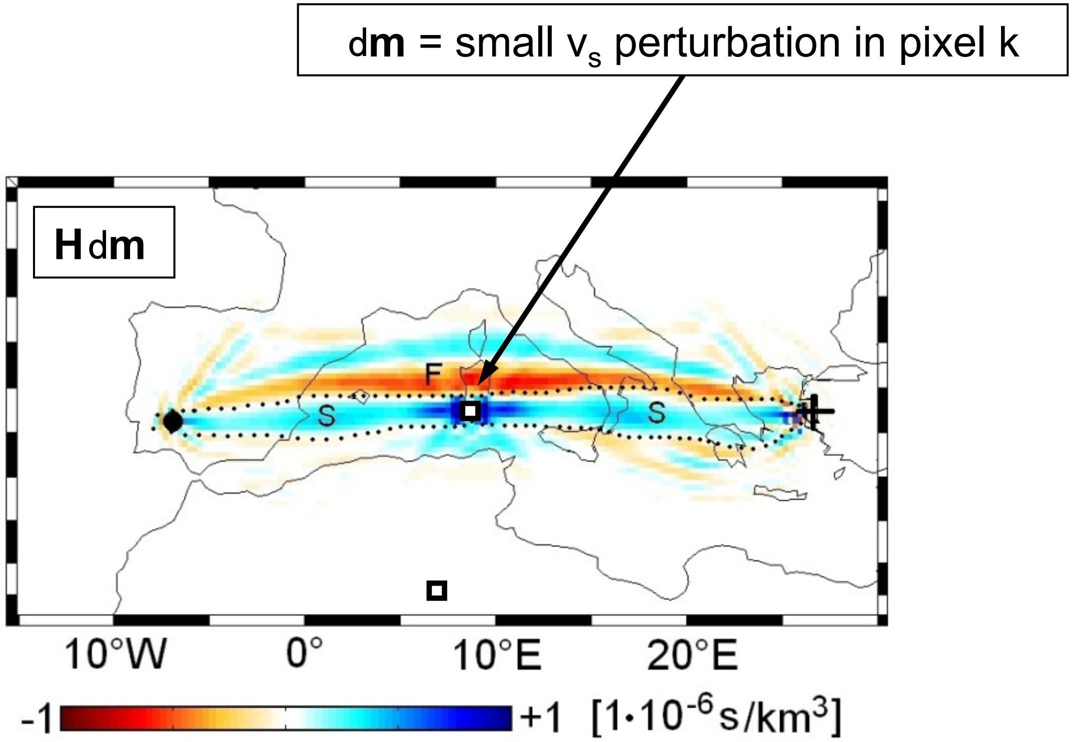


EXAMPLE

- 25 s Love wave
- finite-frequency traveltime



- 25 s Love wave
- finite-frequency traveltime



Two contributions:

F: First-order scattering
S: Second-order scattering

Thanks for your attention!