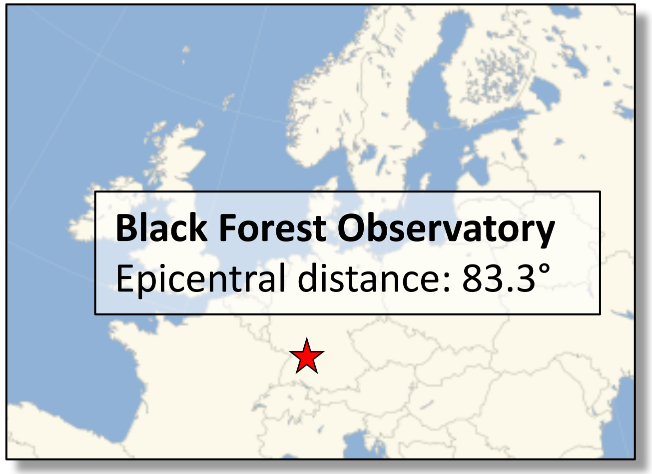
Full-waveform constraints on the lithosphere

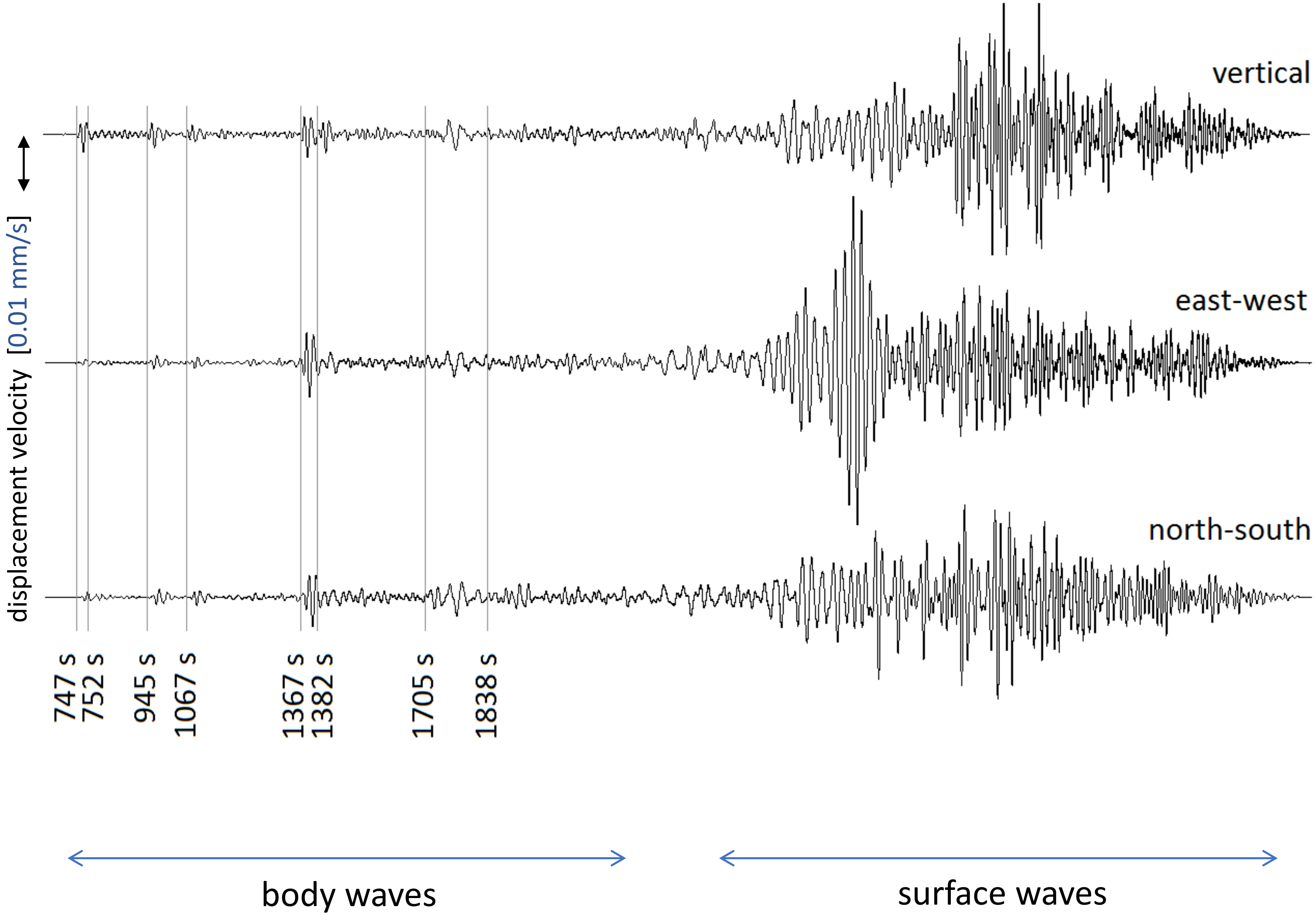
Scott Keating

Outline

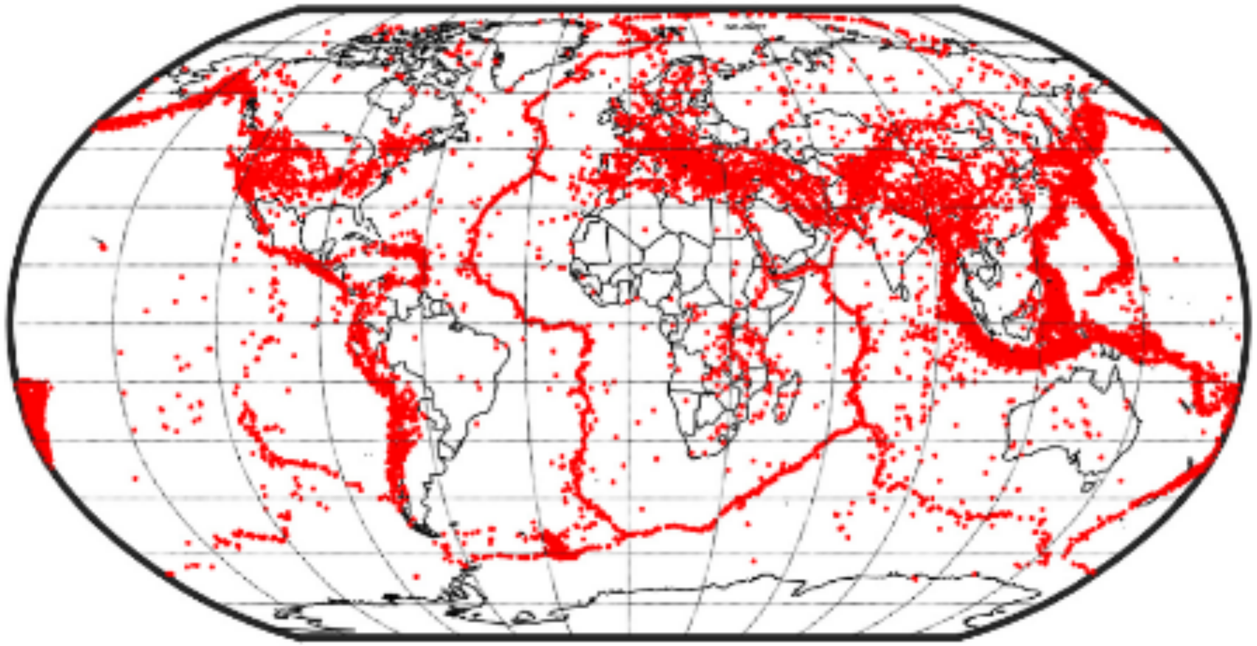
1. Overview of full-waveform inversion
2. Numerical modelling of seismic waves
3. Objective function
4. Optimization
5. Reducing computational cost
6. Uncertainty quantification

Seismic data





Data coverage

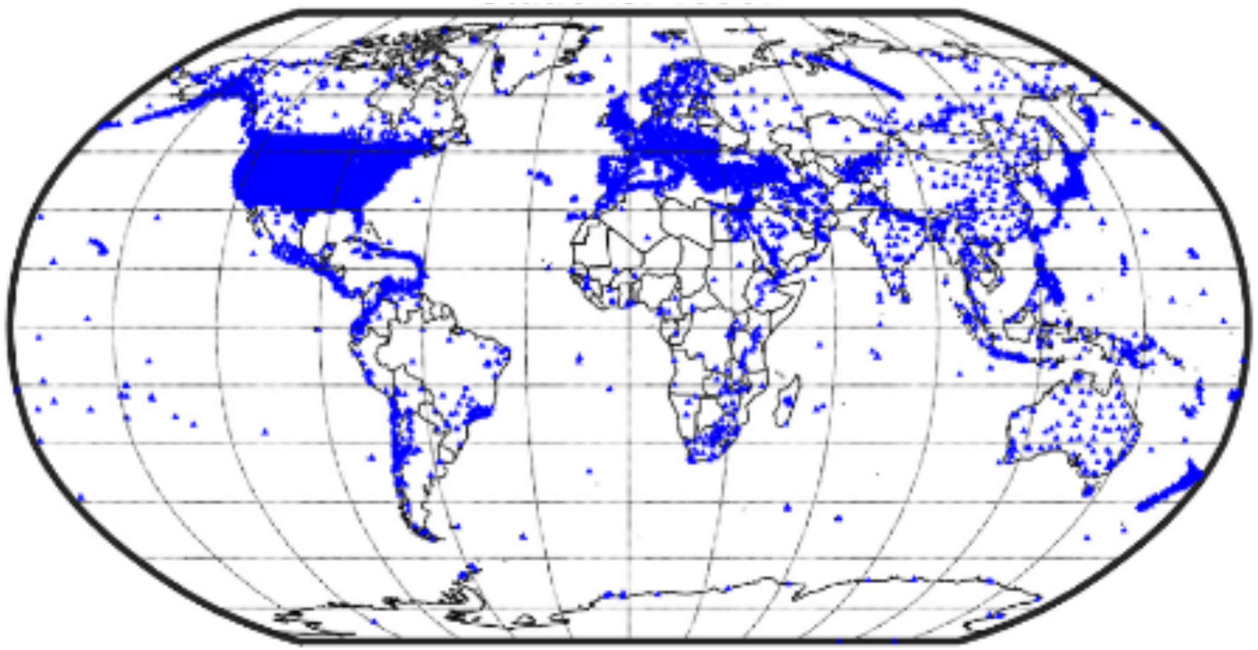


Earthquakes

- Distributed along plate boundaries
- Huge gaps in coverage

Simmons et al. 2021

Data coverage

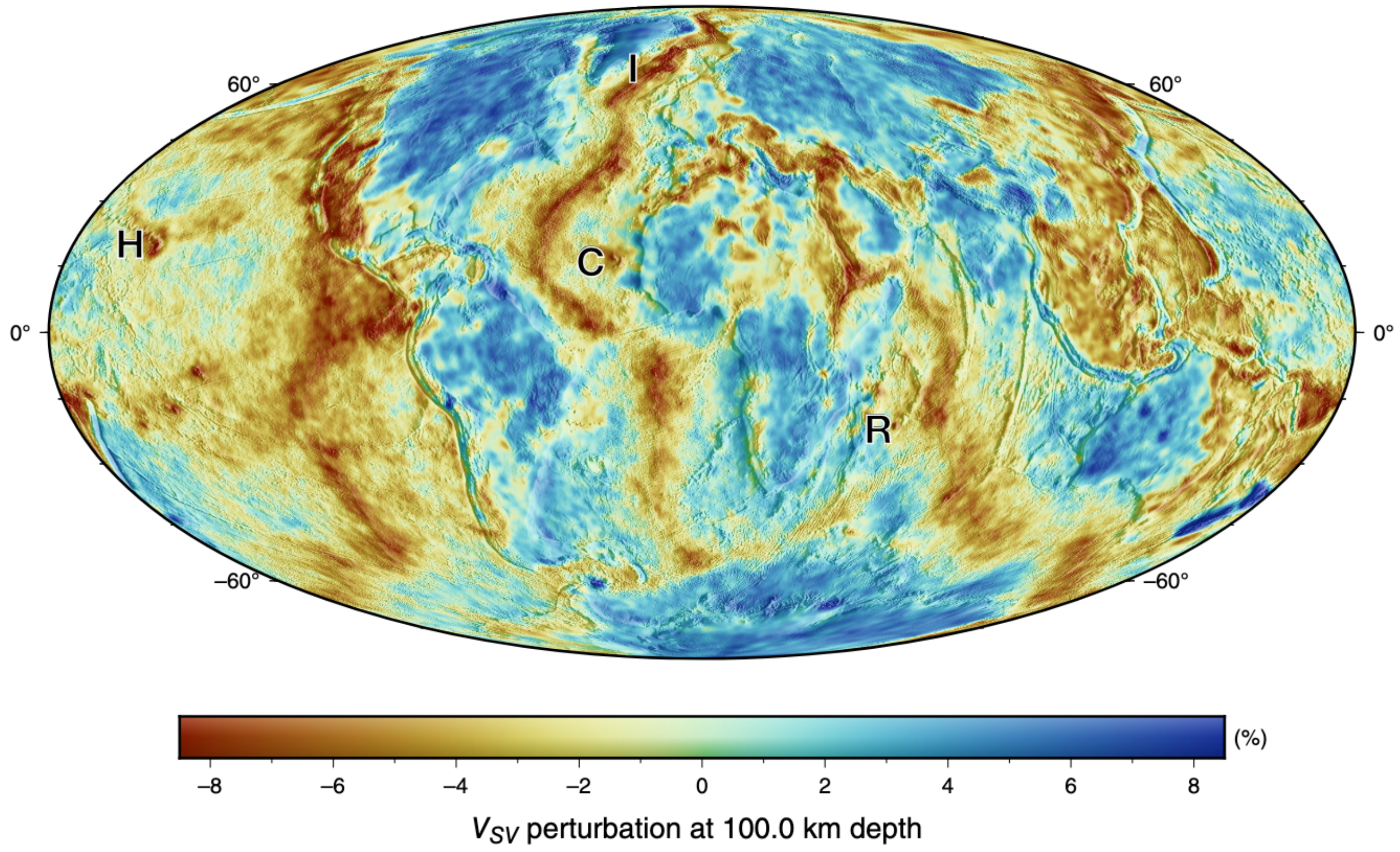


Receivers

- Dense in wealthy, populous areas
- Poor coverage of oceans, less active regions
- Huge gaps in coverage
- Typically measure ground velocity or acceleration

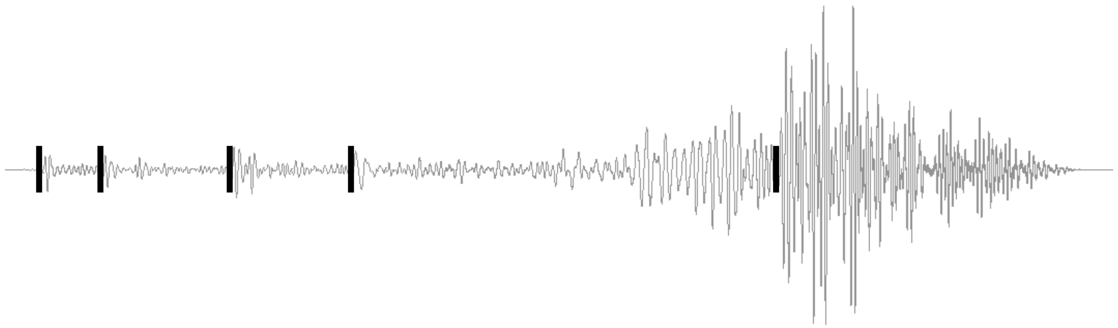
Simmons et al. 2021

Objective: high resolution Earth structure



Thrastarson et al. 2024

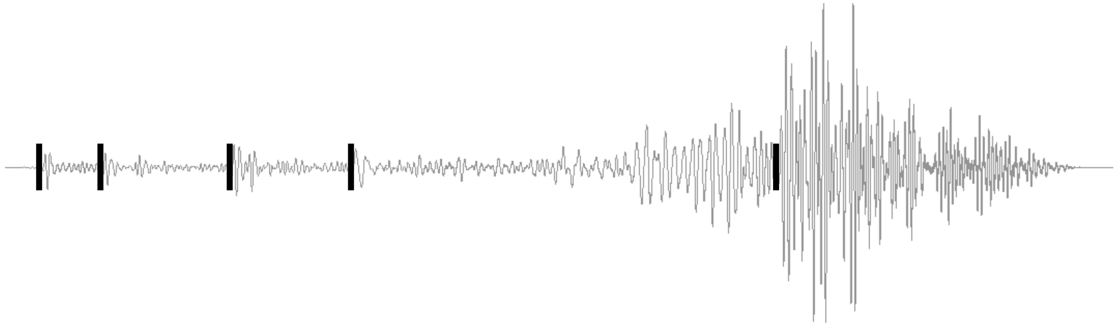
Using the full measurements



traditional traveltime tomography

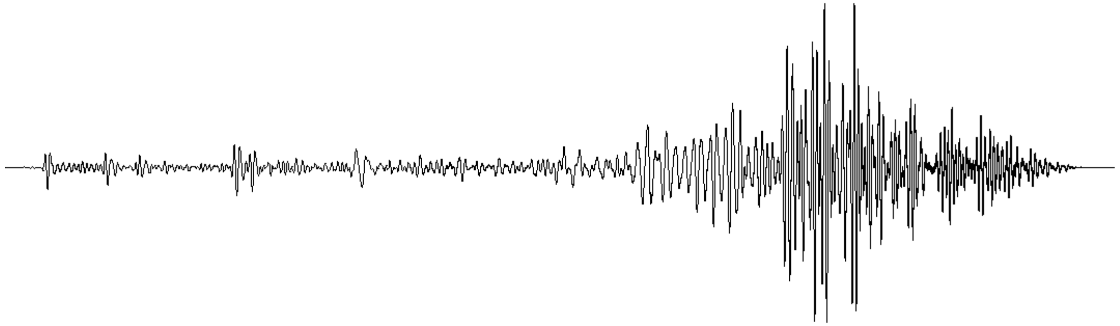
traveltime measurements

Using the full measurements



traditional traveltime tomography

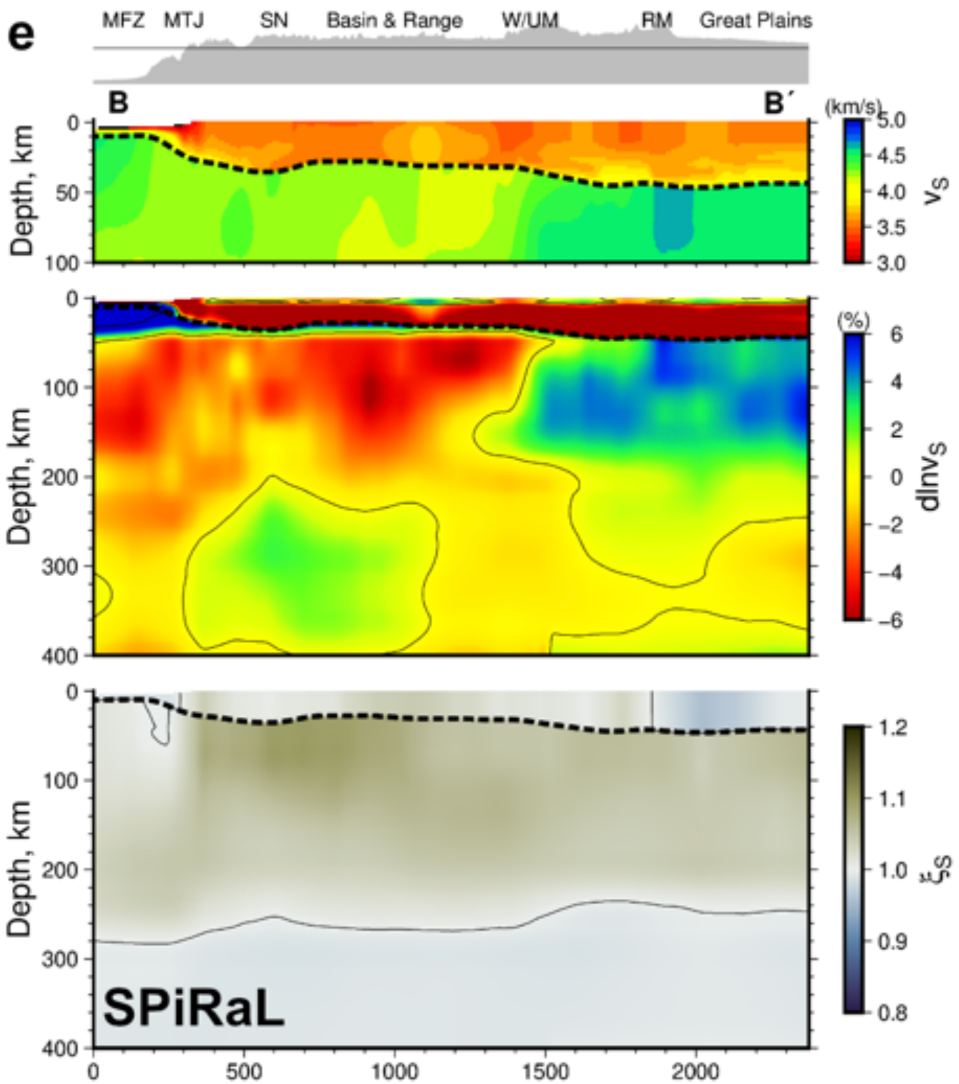
traveltime measurements

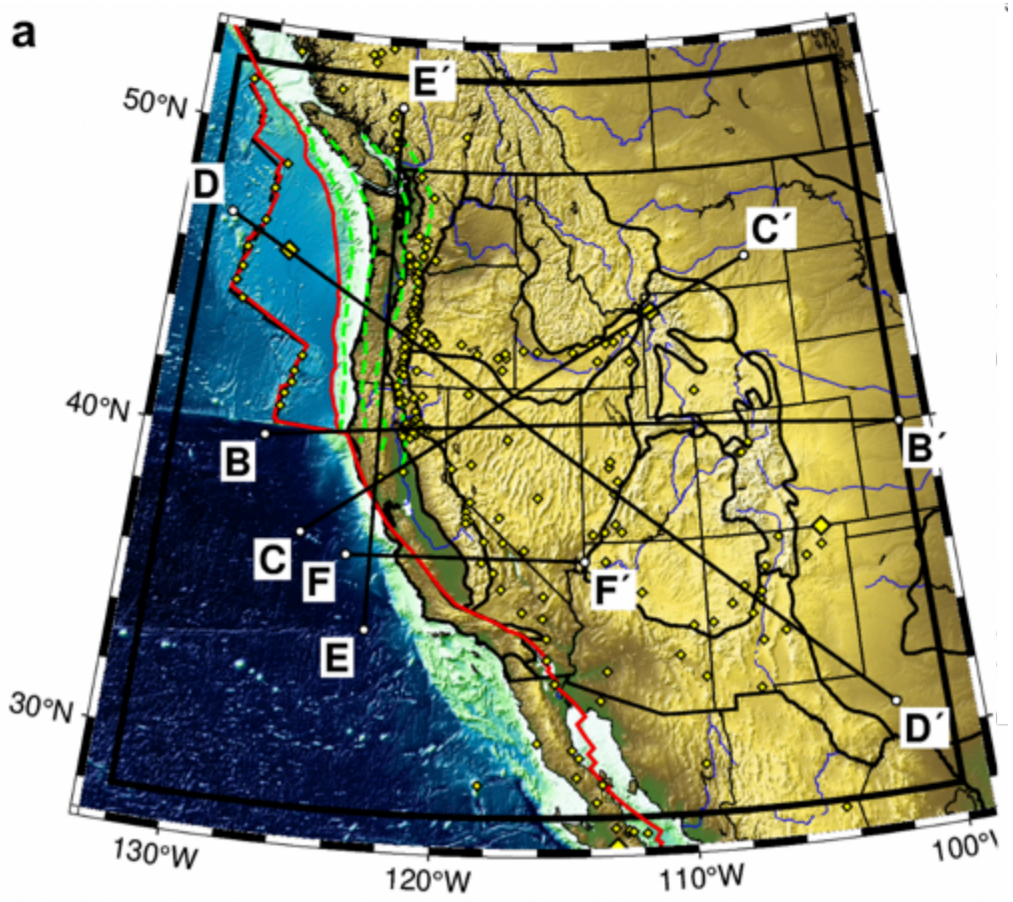


full-waveform inversion

complete seismic recordings

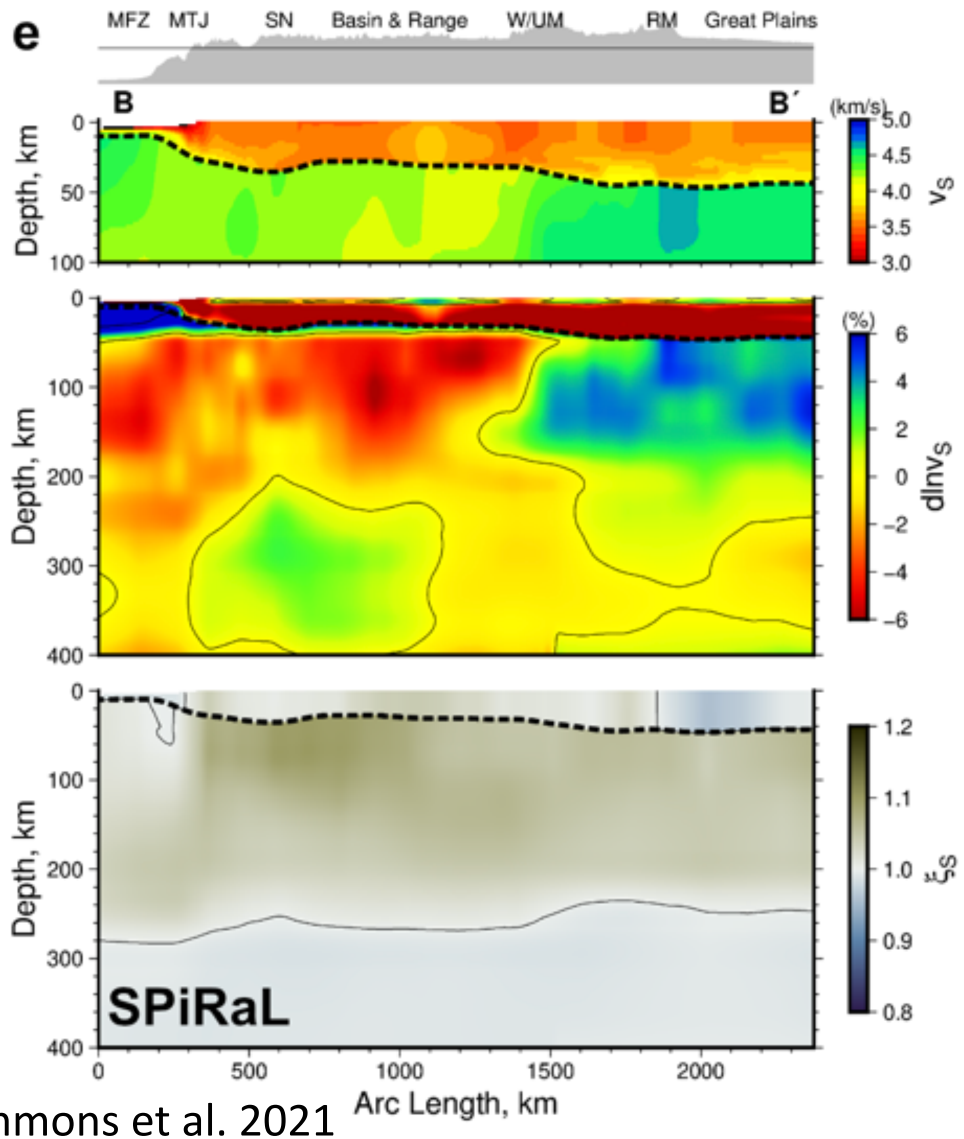
Traveltime





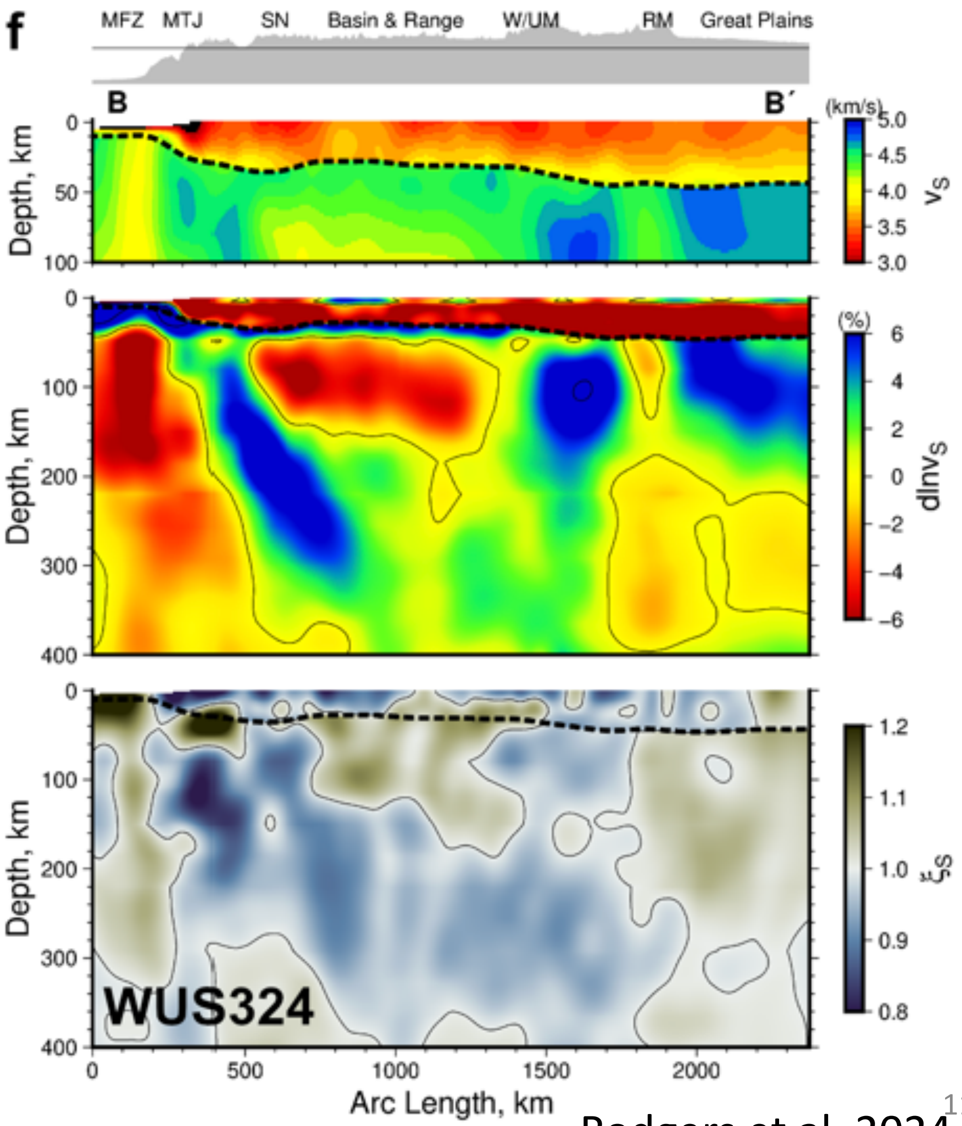
Simmons et al. 2021

Traveltime



Simmons et al. 2021

Full-waveform



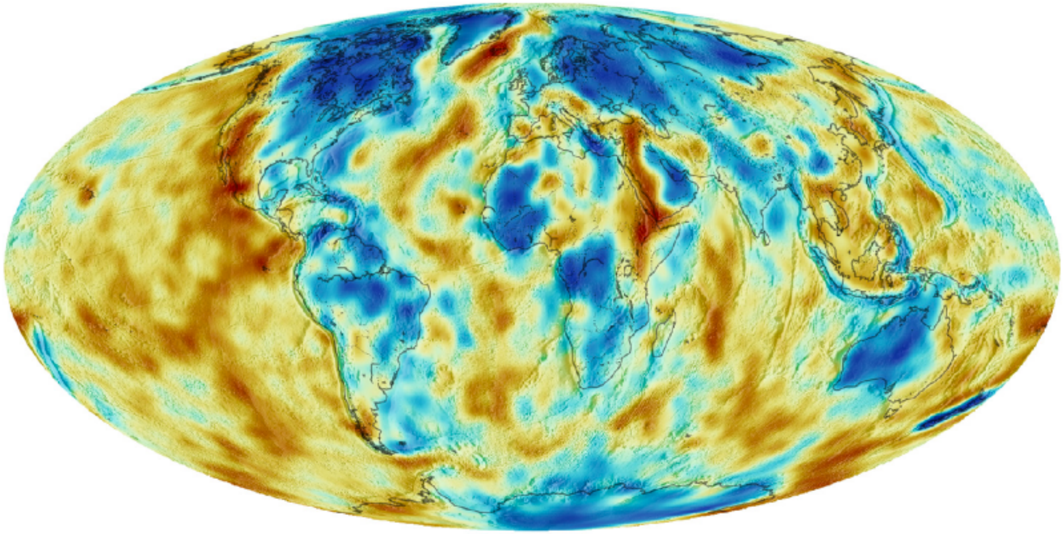
Rodgers et al. 2024

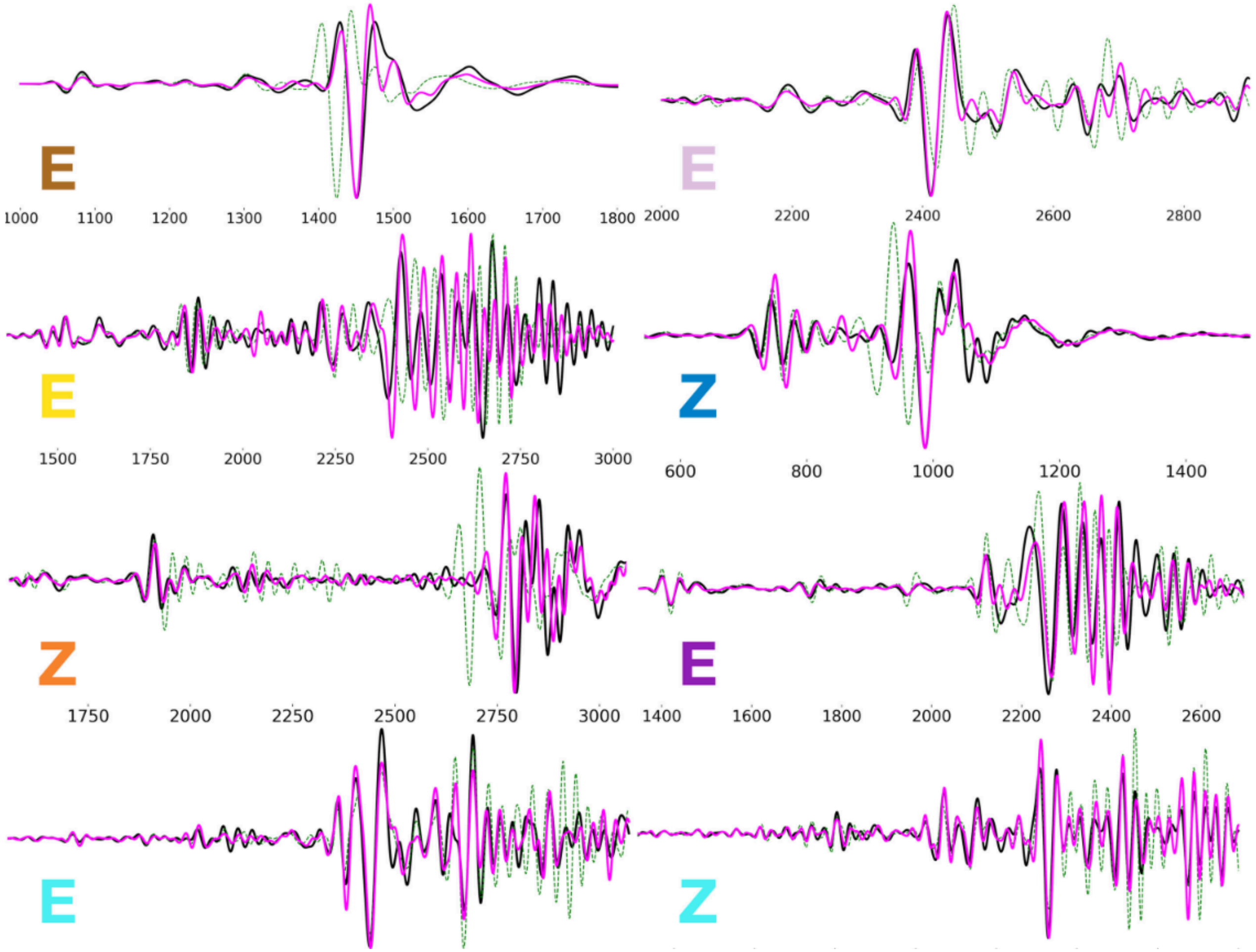
Basic concept

Find the Earth model that reproduces the data

Better data match -> Better models

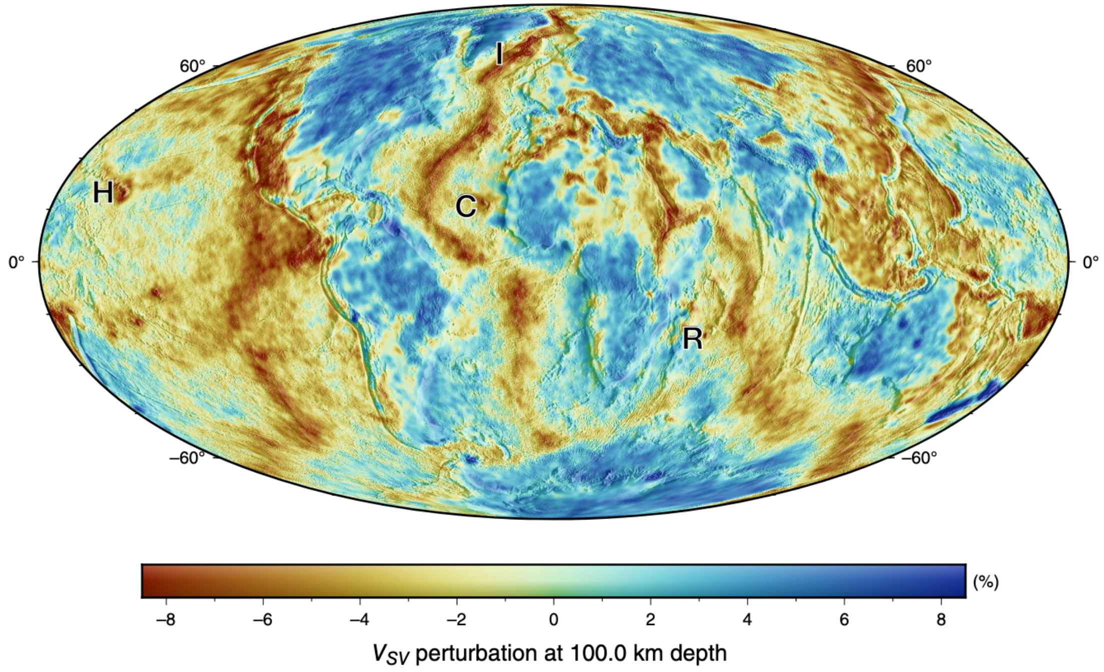
Basic concept

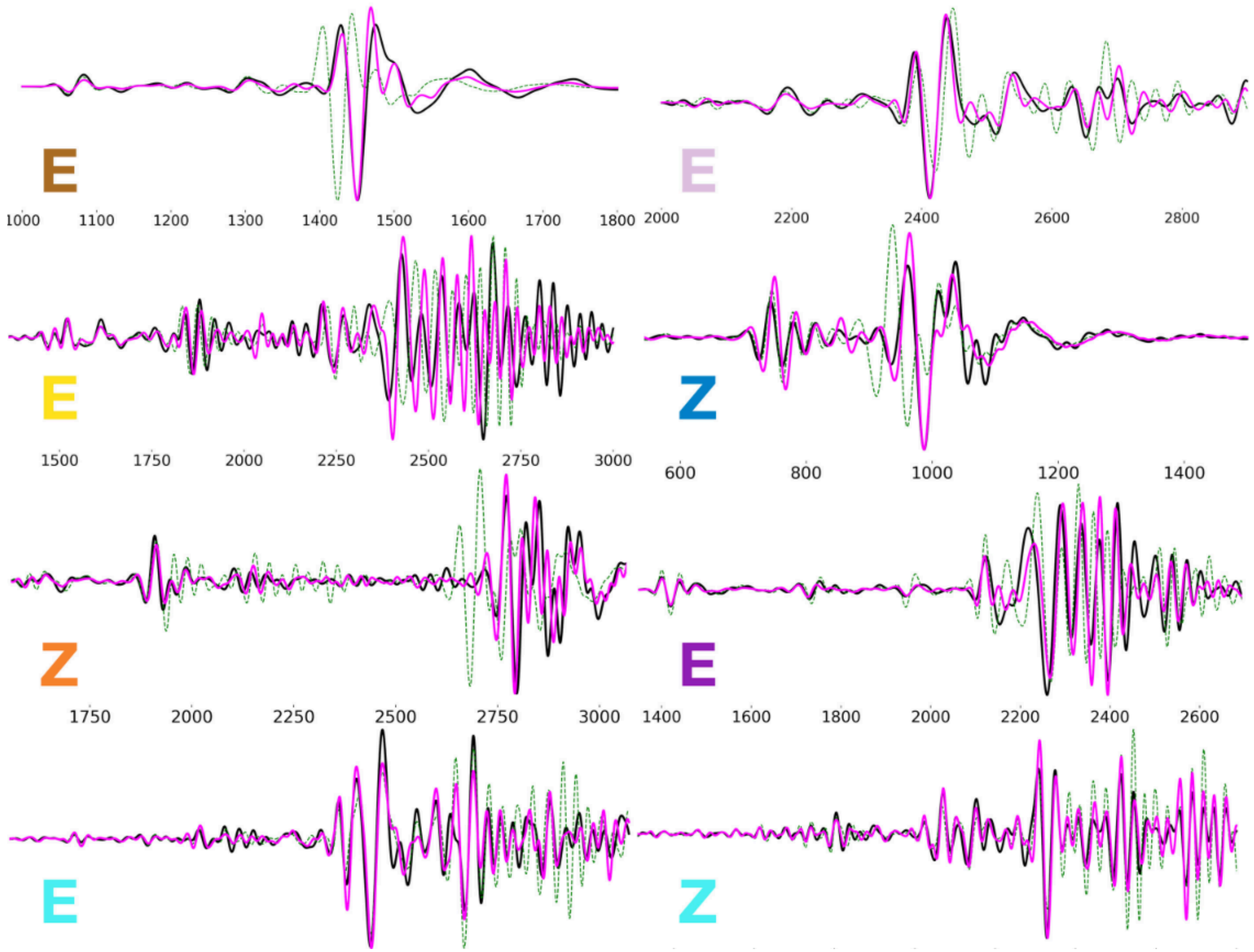




Thrastarson et al. 2022

Basic concept





Thrastarson et al. 2024

Full-waveform inversion: why is it difficult?

High computational costs – f^4

Full-waveform inversion: why is it difficult?

High computational costs – f^4

Large inversions running on supercomputers take weeks

Full-waveform inversion: why is it difficult?

High computational costs – f^4

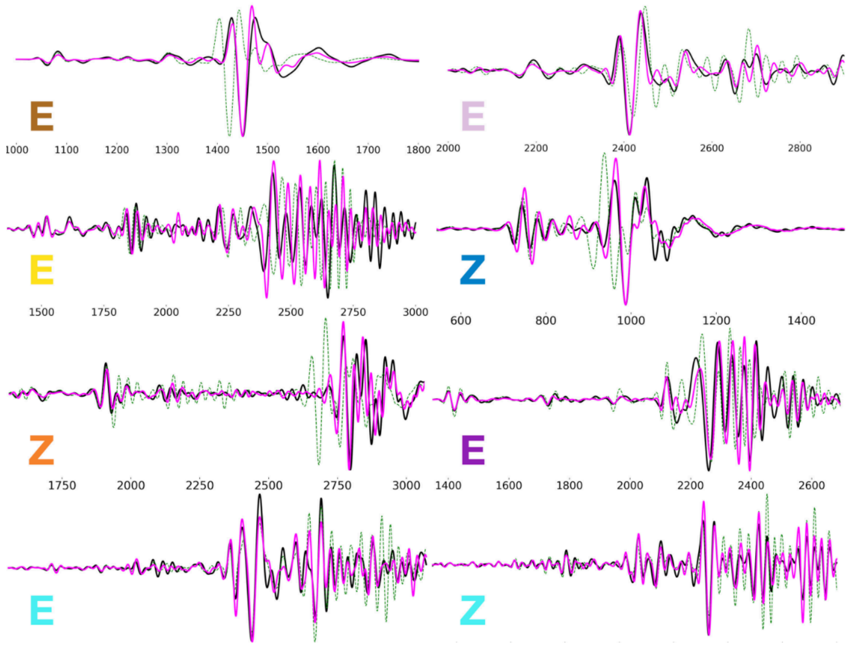
We cannot directly solve -> iterative methods

Full-waveform inversion: why is it difficult?

High computational costs – f^4

We cannot directly solve -> iterative methods

Difficult to define “better” and “worse” data matching



Full-waveform inversion: why is it difficult?

High computational costs – f^4

We cannot directly solve -> iterative methods

Difficult to define “better” and “worse” data matching

Uncertainty quantification is difficult

Full-waveform inversion: why is it difficult?

High computational costs – f^4

We cannot directly solve -> iterative methods

Difficult to define “better” and “worse” data matching

Uncertainty quantification is difficult

Data coverage and quality can be poor

Full-waveform inversion: why is it difficult?

High computational costs – f^4

We cannot directly solve -> iterative methods

Difficult to define “better” and “worse” data matching

Uncertainty quantification is difficult

Data coverage and quality can be poor

Physics assumptions can fail

Full-waveform inversion: why is it difficult?

High computational costs – f^4

We cannot directly solve -> iterative methods

Difficult to define “better” and “worse” data matching

Uncertainty quantification is difficult

Data coverage and quality can be poor

Physics assumptions can fail

High computational cost

Outline

1. Overview of full-waveform inversion
2. Numerical modelling of seismic waves
3. Objective function
4. Optimization
5. Reducing computational cost
6. Uncertainty quantification

Numerical modelling of seismic waves

1. Physics assumptions
2. Source assumptions
3. Finite-difference solvers
4. Finite-element solvers

Numerical modelling of seismic waves

To use the **full information content** of seismic waves, we must be able to reproduce the **full complexity** of seismic waves

Wave simulation - assumptions

A general elastic stiffness tensor has 21 independent components

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{bmatrix} = \begin{bmatrix} & & \\ & c_{nm} & \\ & & \end{bmatrix} \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ 2\varepsilon_{23} \\ 2\varepsilon_{13} \\ 2\varepsilon_{12} \end{bmatrix}$$

Wave simulation - assumptions

A general elastic stiffness tensor has 21 independent components

Real media are not perfectly elastic, so each of these components is also frequency-dependent

$$
\begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{bmatrix} = \begin{bmatrix} & & c_{nm}(\omega) & & \end{bmatrix} \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ 2\varepsilon_{23} \\ 2\varepsilon_{13} \\ 2\varepsilon_{12} \end{bmatrix}
$$

Wave simulation - assumptions

A general elastic stiffness tensor has 21 independent components

Real media are not perfectly elastic, so each of these components is also frequency-dependent

We do not have sufficient data to constrain all these parameters

Wave simulation - assumptions

Fortunately, simplified models do a good job of explaining most Earth materials

Wave simulation - assumptions

Fortunately, simplified models do a good job of explaining most Earth materials

Typically, we invert for v_P, v_S, density, and a small number of anisotropy parameters

Wave simulation - assumptions

Fortunately, simplified models do a good job of explaining most Earth materials

Typically, we invert for v_P, v_S, density, and a small number of anisotropy parameters

Attenuation is approximated in a scalar Q, and typically not updated

Wave simulation - assumptions

Fortunately, simplified models do a good job of explaining most Earth materials

Typically, we invert for v_P, v_S, density, and a small number of anisotropy parameters

Attenuation is approximated in a scalar Q, and typically not updated

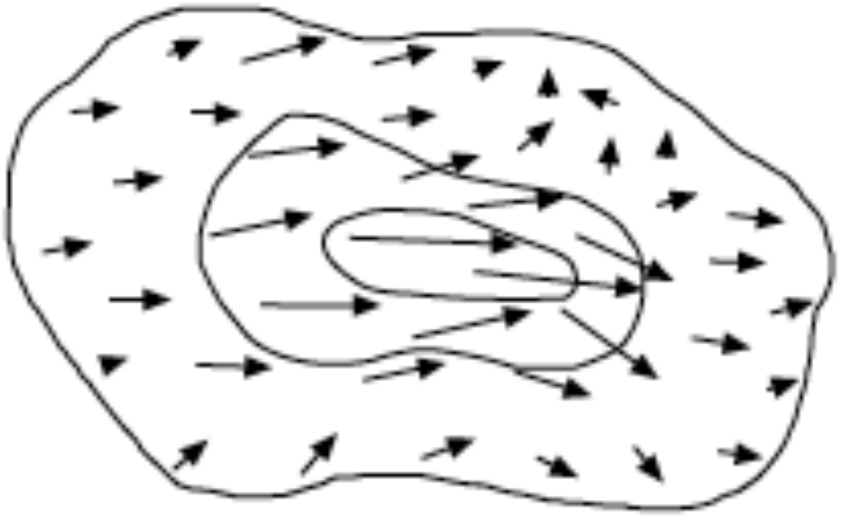
Even with this reduced set of parameters, we usually cannot constrain all parameters well

Sources - assumptions

Earthquakes are complex!

Different parts of the same fault move in different directions, at different speeds, with different start and end times

Estimating the source requires that we know the Earth model, and vice versa



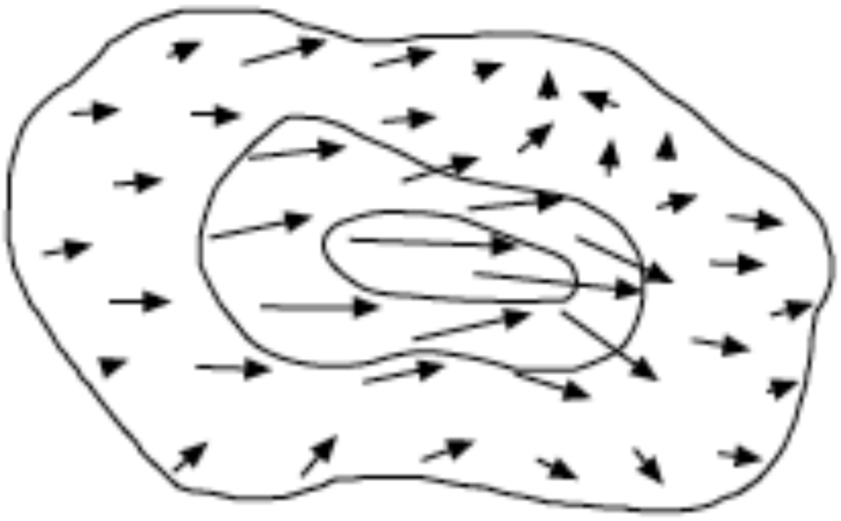
$$M = \int_A \mu D(A) dA$$

Sources - assumptions

We typically choose earthquakes small enough to treat as points in our inversions

We abstract the complex forces into a single moment tensor

We typically use non-full-waveform estimates and fix them during inversion



$$M = \int_A \mu D(A) dA$$

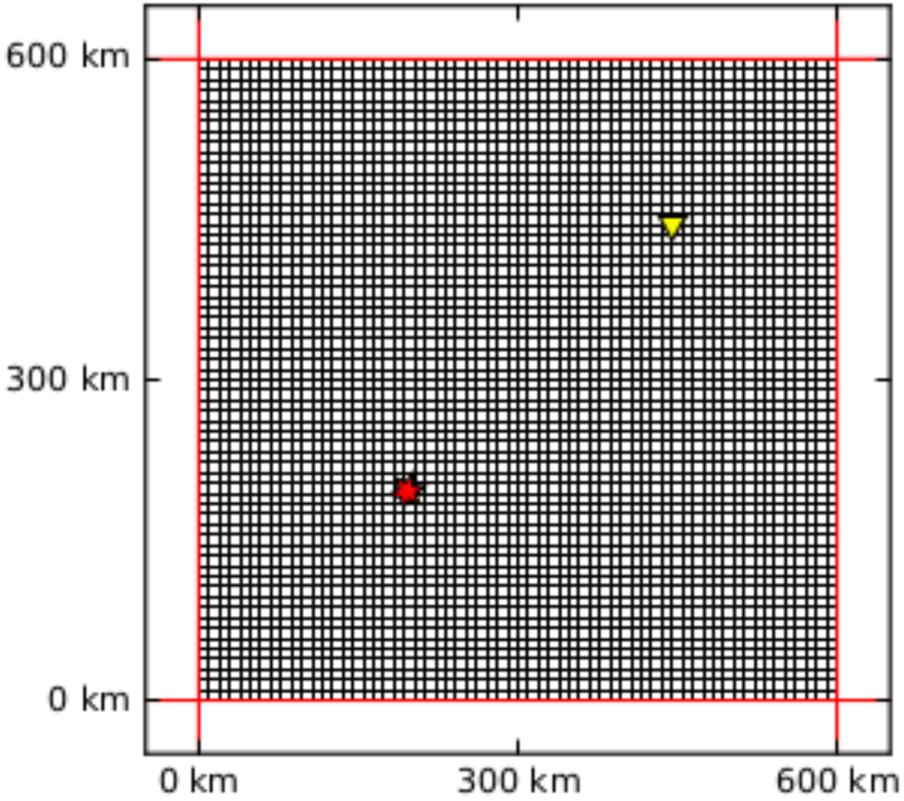
Modelling the entire wavefield

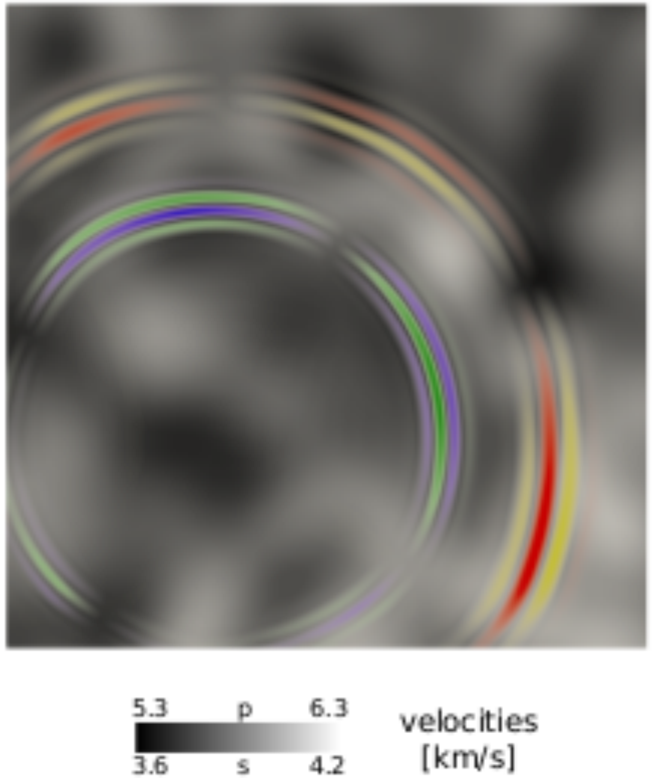
We want to solve equations like

$$\alpha^2\nabla(\nabla\cdot\mathbf{u})-\beta^2\nabla\times(\nabla\times\mathbf{u})+\mathbf{F}=\ddot{\mathbf{u}}$$

Analytic solutions are not available – we use numerical solutions of discretized approximations

Modelling the entire wavefield





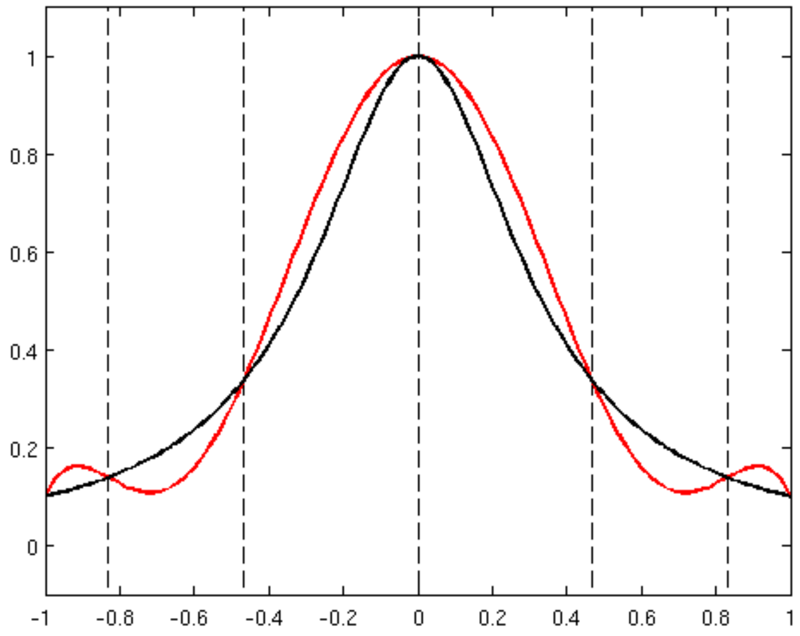
Numerical solvers

Finite differences

- Computationally efficient
- Simple to implement
- Cartesian grid
- Bad at handling irregular surfaces

$$\frac{d^2u}{dx^2} \approx \frac{u(x + \Delta x) - u(x - \Delta x)}{\Delta x}$$

Numerical solvers



Spectral elements

- Slower computation
- More challenging to implement
- Require meshing
- Naturally handle irregular surfaces

Numerical solvers

Finite differences

- Computationally efficient
- Simple to implement
- Cartesian grid
- Bad at handling irregular surfaces

- Often used at smaller scales

Spectral elements

- Slower computation
- More challenging to implement
- Require meshing
- Naturally handle irregular surfaces

- Often used at larger scales

Limitations

We invoke **approximations** in each of these approaches

Accurate approximation requires **both**

1. Enough time-samples per period
2. Enough space-samples per wavelength

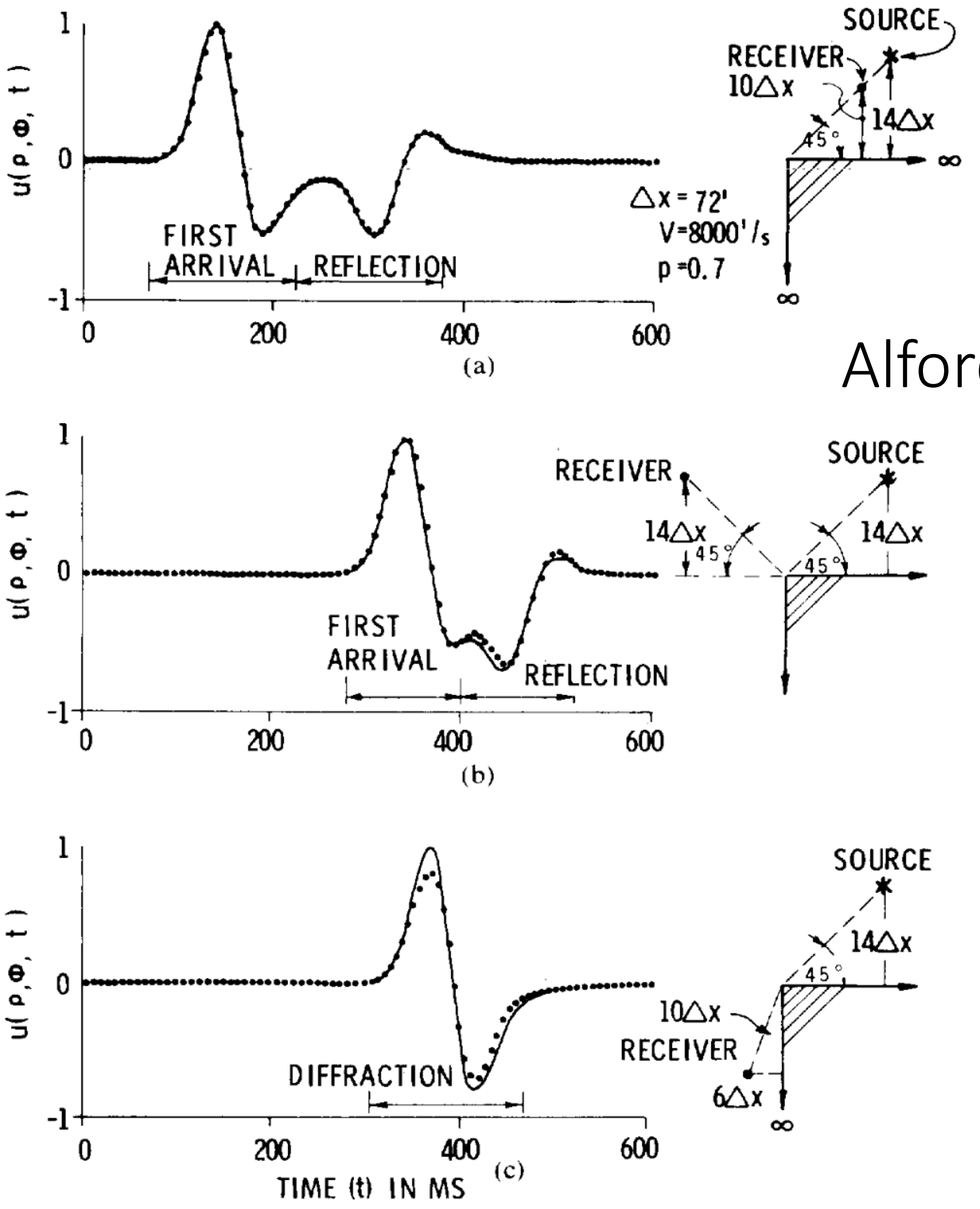


FIG. 4. Analytical solution (solid line) and fine grid ($G_0 \approx 11$) finite-difference solution (circles) for the second-order scheme.

Alford et al. 1974

Alford et al. 1974

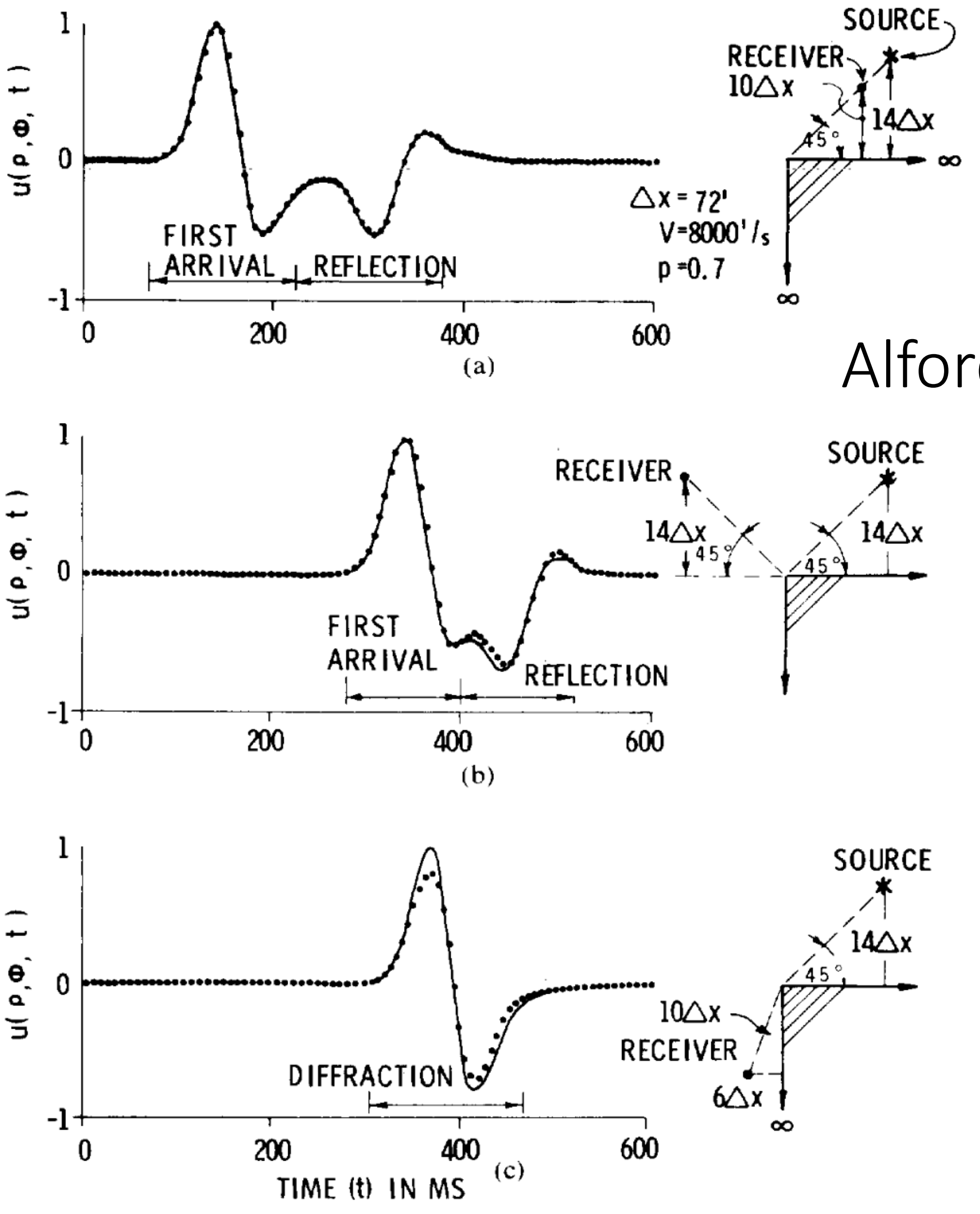


FIG. 4. Analytical solution (solid line) and fine grid ($G_0 \approx 11$) finite-difference solution (circles) for the second-order scheme.

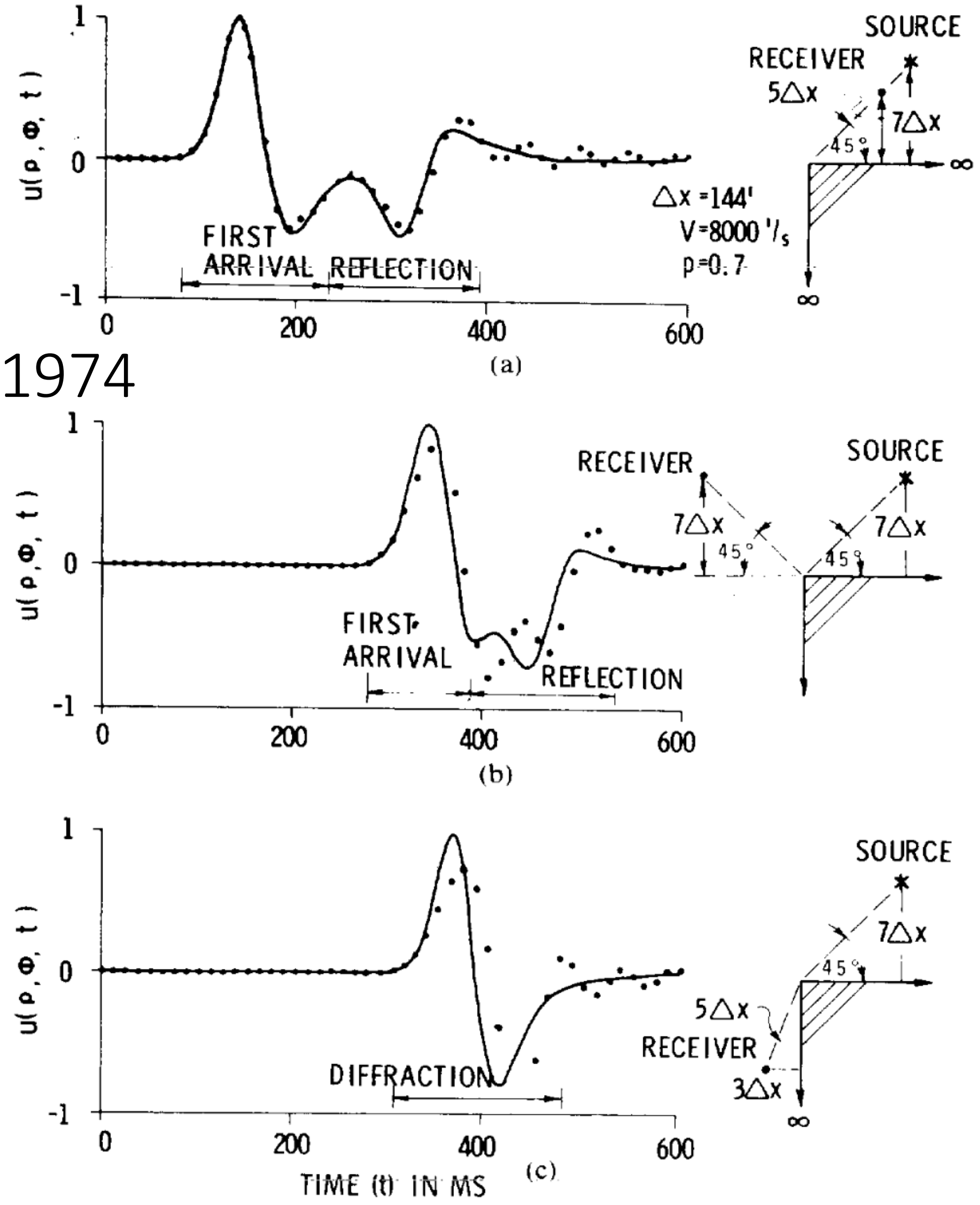
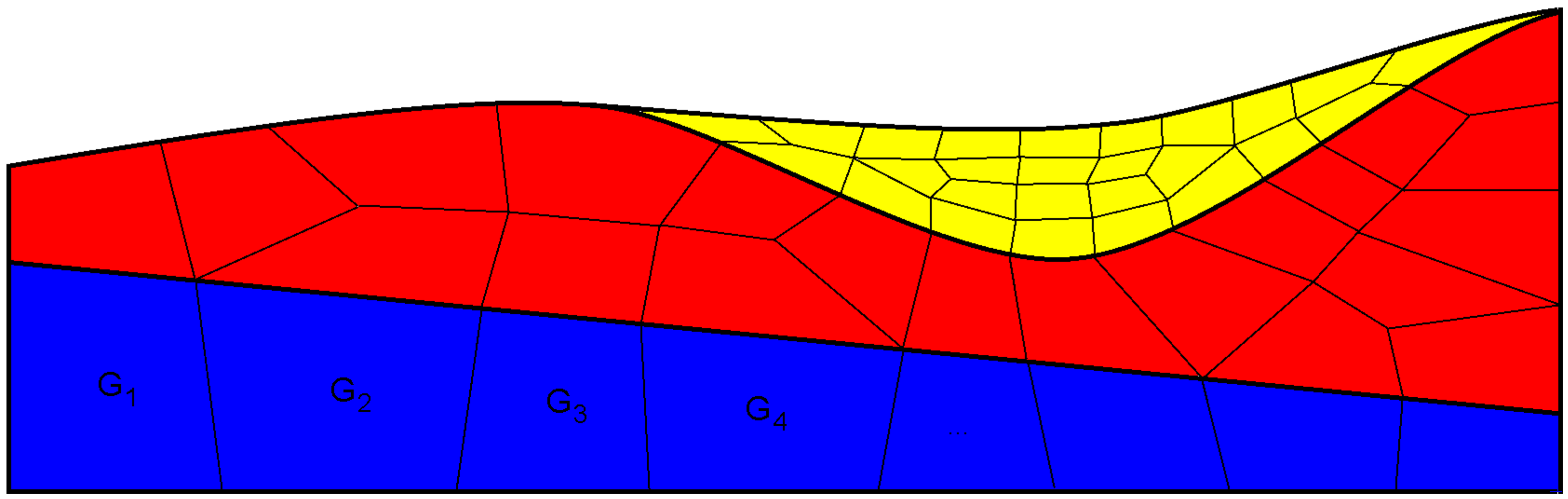


FIG. 5. Analytical solution (solid line) and coarse grid ($G_0 \approx 5.5$) finite-difference solution (circles) for the second-order scheme.



low velocities: short wavelength → small elements

high velocities: long wavelength → large elements

<u>accurate solutions</u>: discontinuities need to coincide with element boundaries

Limitations

We invoke **approximations** in each of these approaches

Accurate approximation requires **both**

1. Enough time-samples per period
2. Enough space-samples per wavelength

Slow speeds and **high frequencies** define the computational demand

Limitations

Accurate approximation requires **both**

1. Enough time-samples per period
2. Enough space-samples per wavelength

Wavelength is $\lambda = \frac{v}{f}$, so **increasing f lowers λ in three dimensions!**

Combined with the period decrease, this means cost scales with f^4**!**

Outline

1. Overview of full-waveform inversion
2. Numerical modelling of seismic waves
3. Objective function
4. Optimization
5. Reducing computational cost
6. Uncertainty quantification

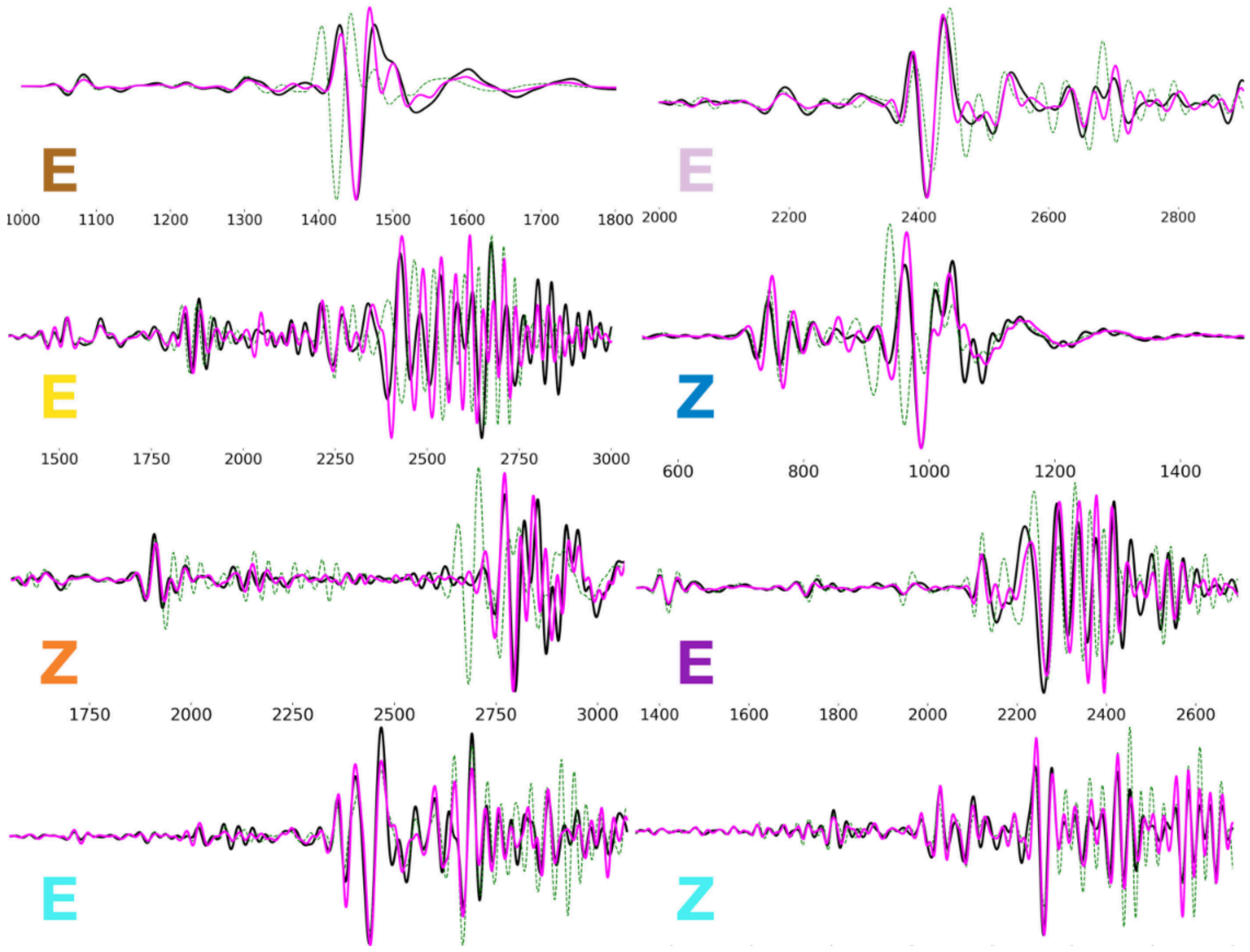
Objective functions

1. Basics - L^2
2. Cycle-skipping
3. Windowing
4. Multi-scaling
5. Time-frequency phase
6. Graph-space optimal transport

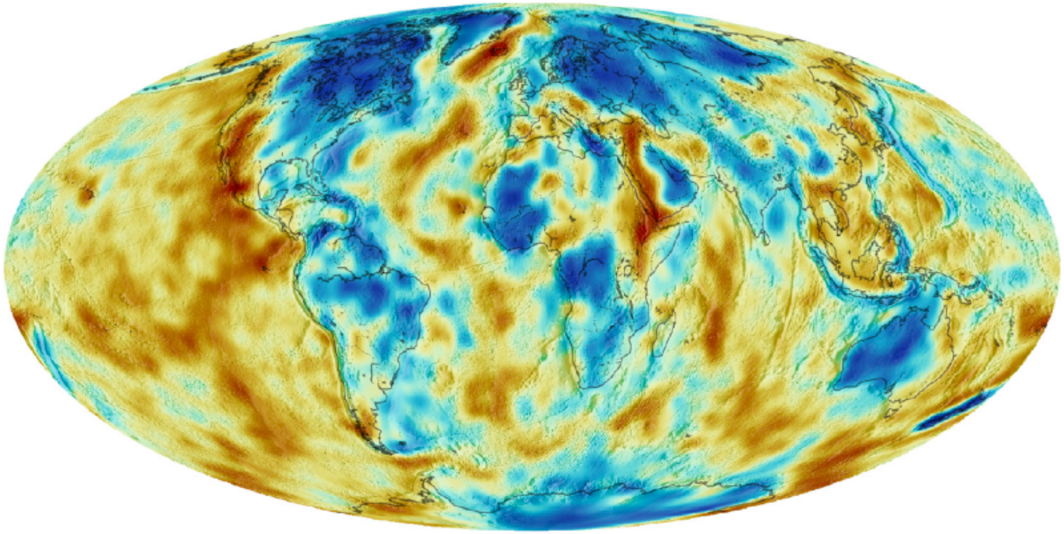
Objective functions

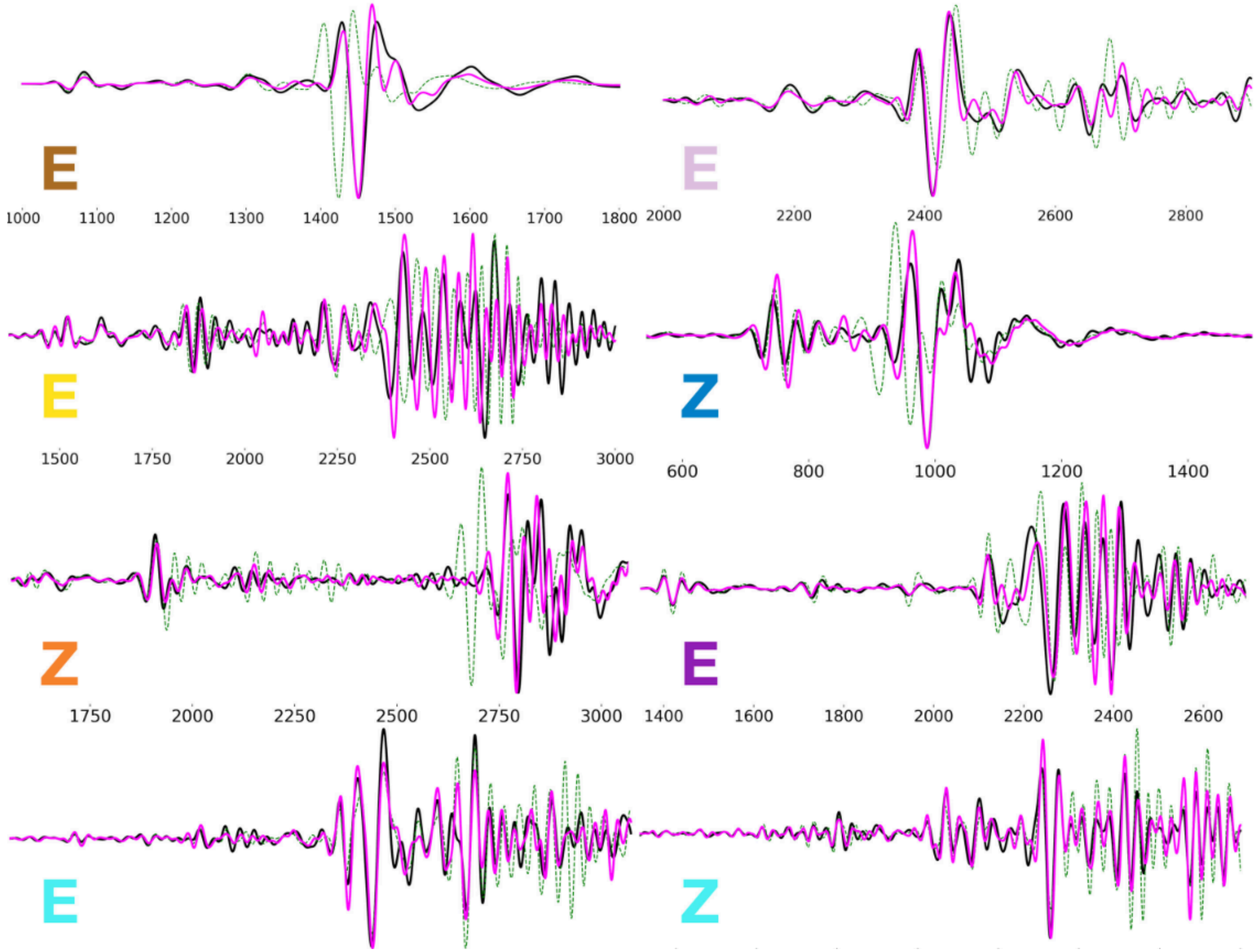
Iterative:
Worse models -> better models

How do we measure “worse” or “better”?



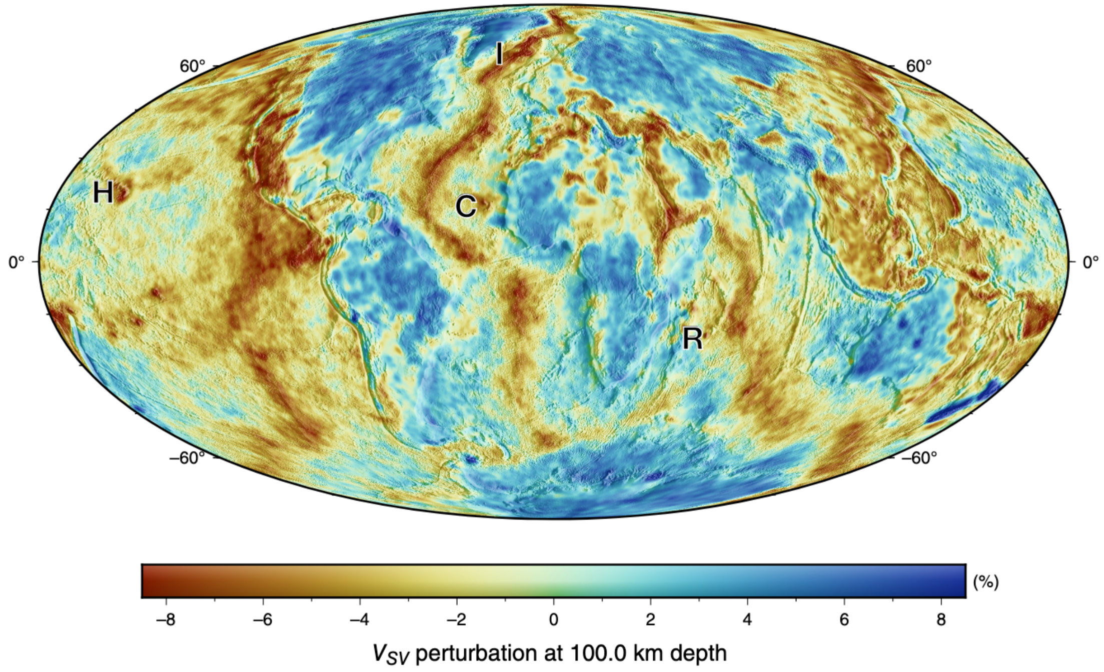
Objective functions

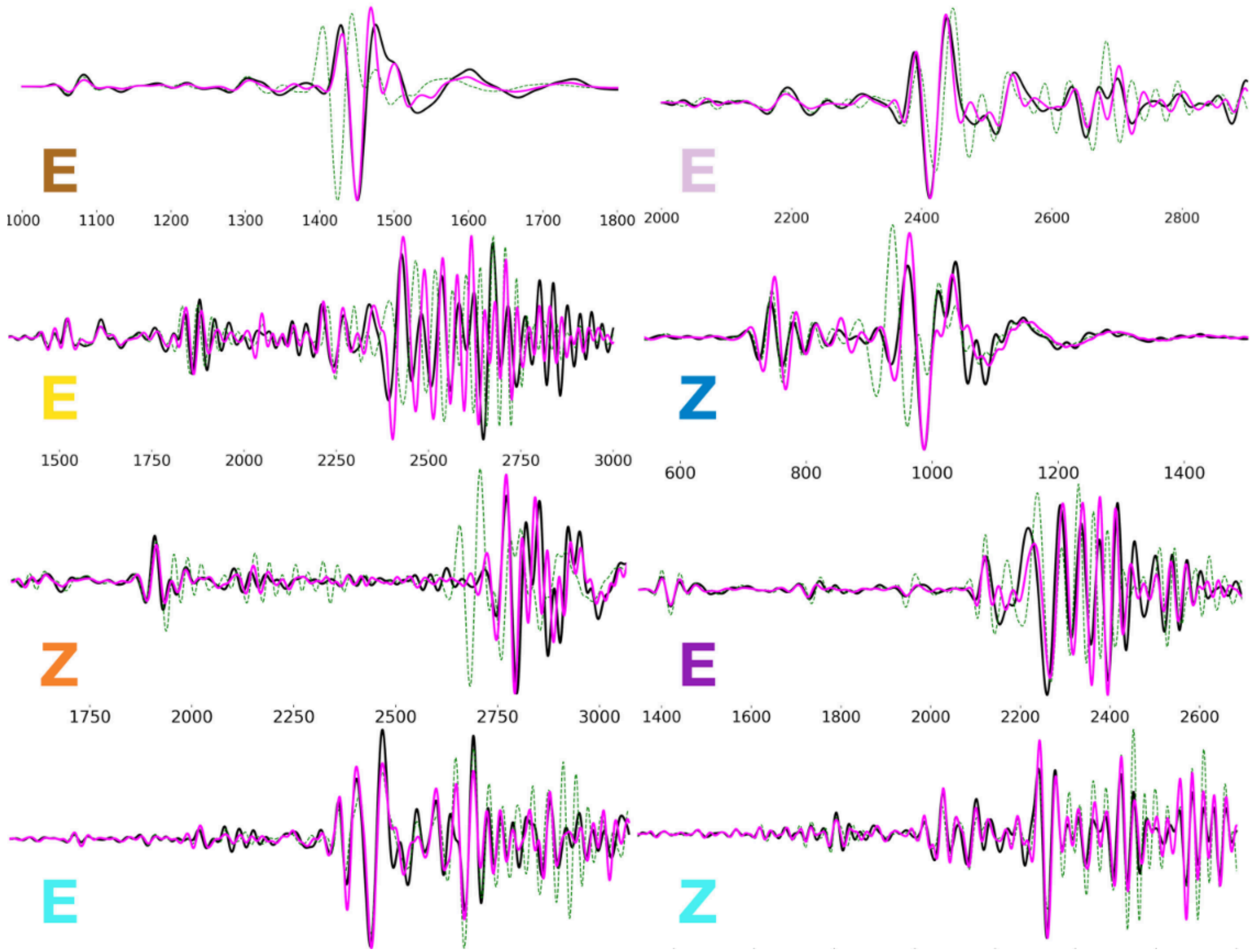




Thrastarson et al. 2022

Objective functions





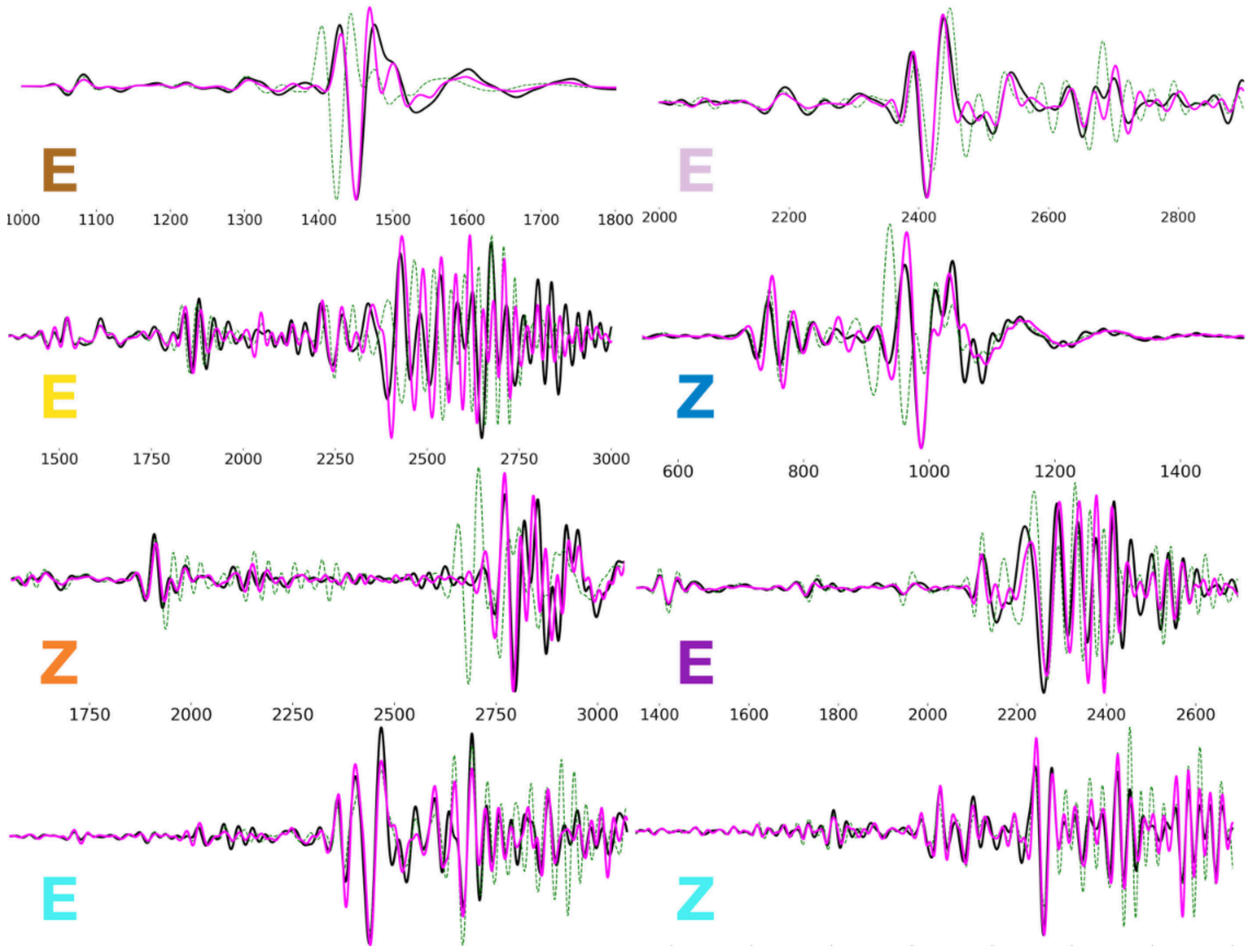
Thrastarson et al. 2024

Objective function

Scalar measure of data-fit

We seek local improvements

Inversion is driven by the gradient of this function

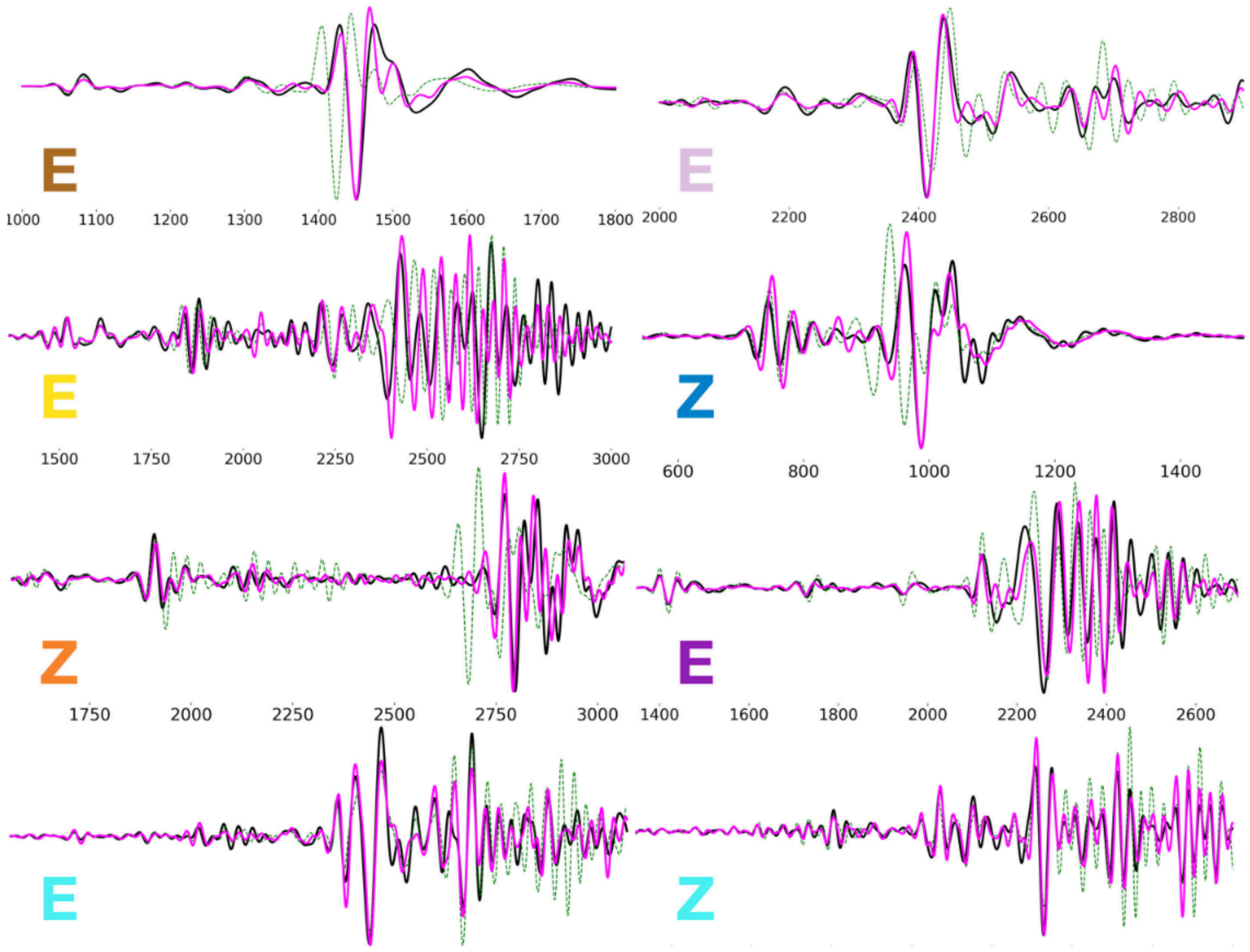


Objective function

Ideally:

1. Identify the importance of each piece of data

2. Assign a robust measure of the agreement with data



Objective function

In practice, we often use simpler formulations

$$\phi = \sum_{S} \frac{1}{2} ||R\boldsymbol{u} - \boldsymbol{d}||^2$$

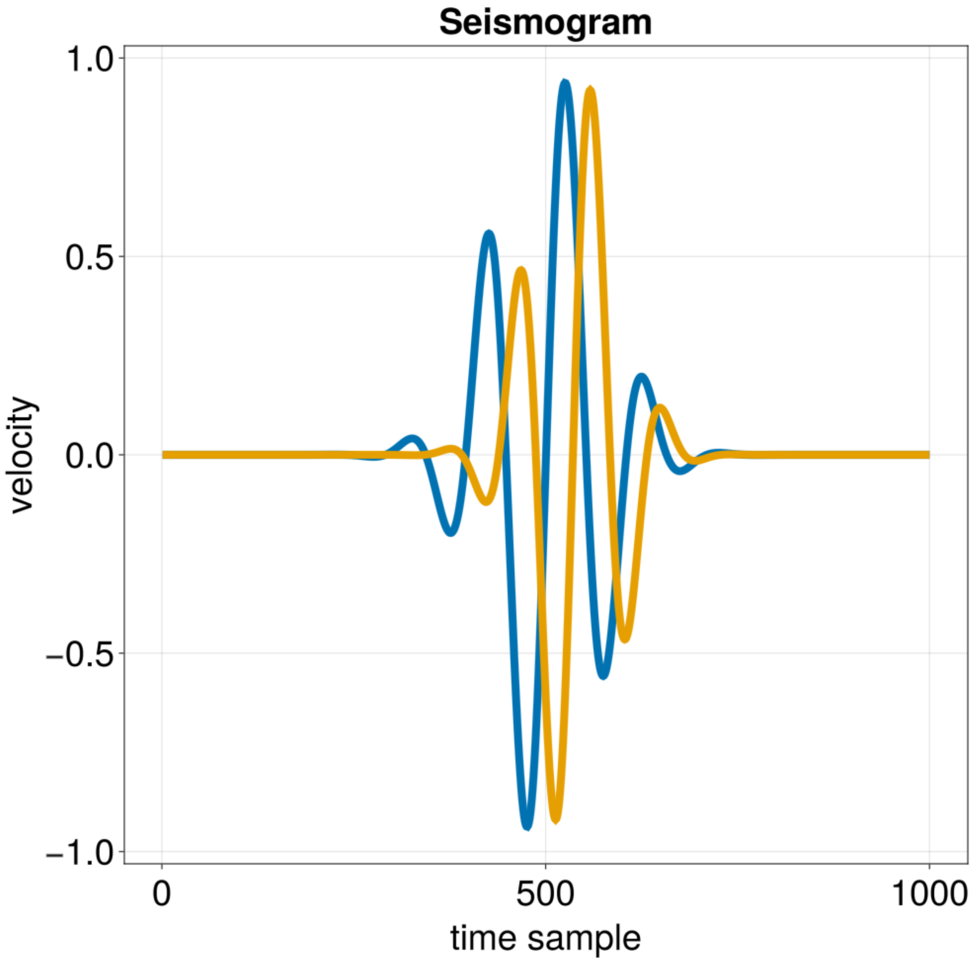
Objective function

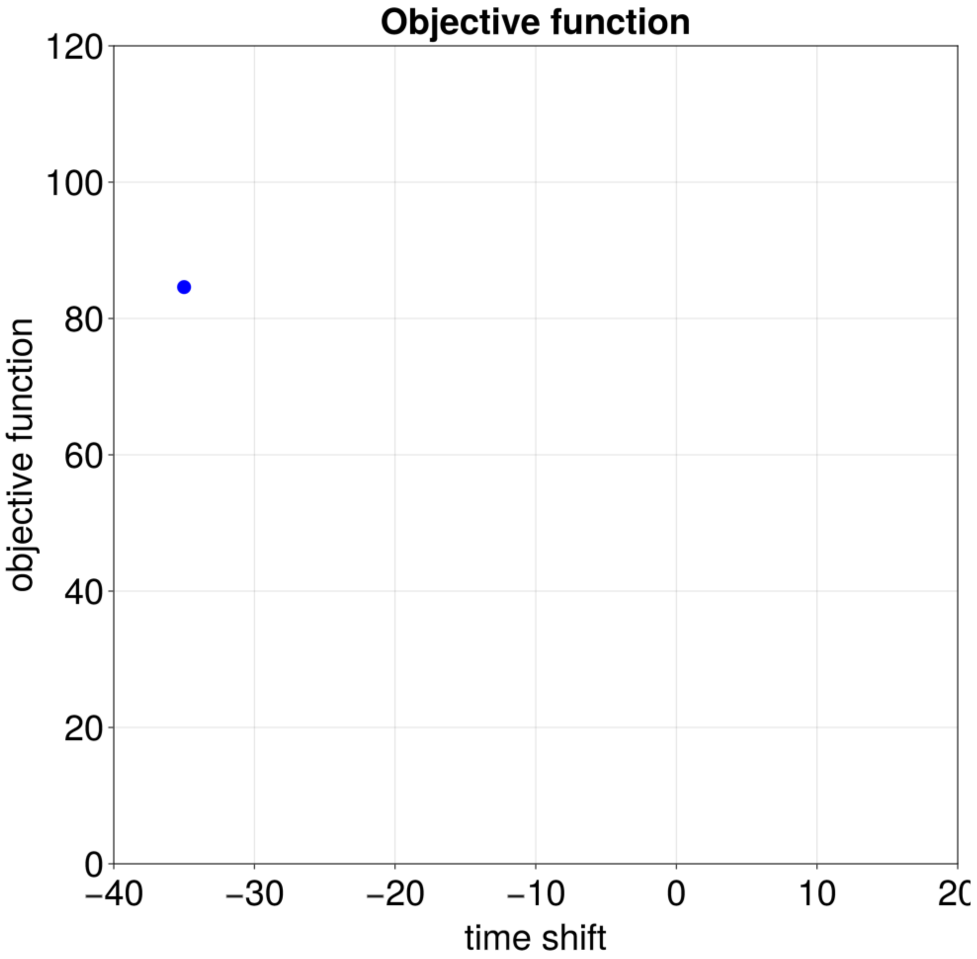
In practice, we often use simpler formulations

$$\phi = \sum_{S} \frac{1}{2} \left|\left| R\boldsymbol{u} - \boldsymbol{d} \right|\right|^2$$

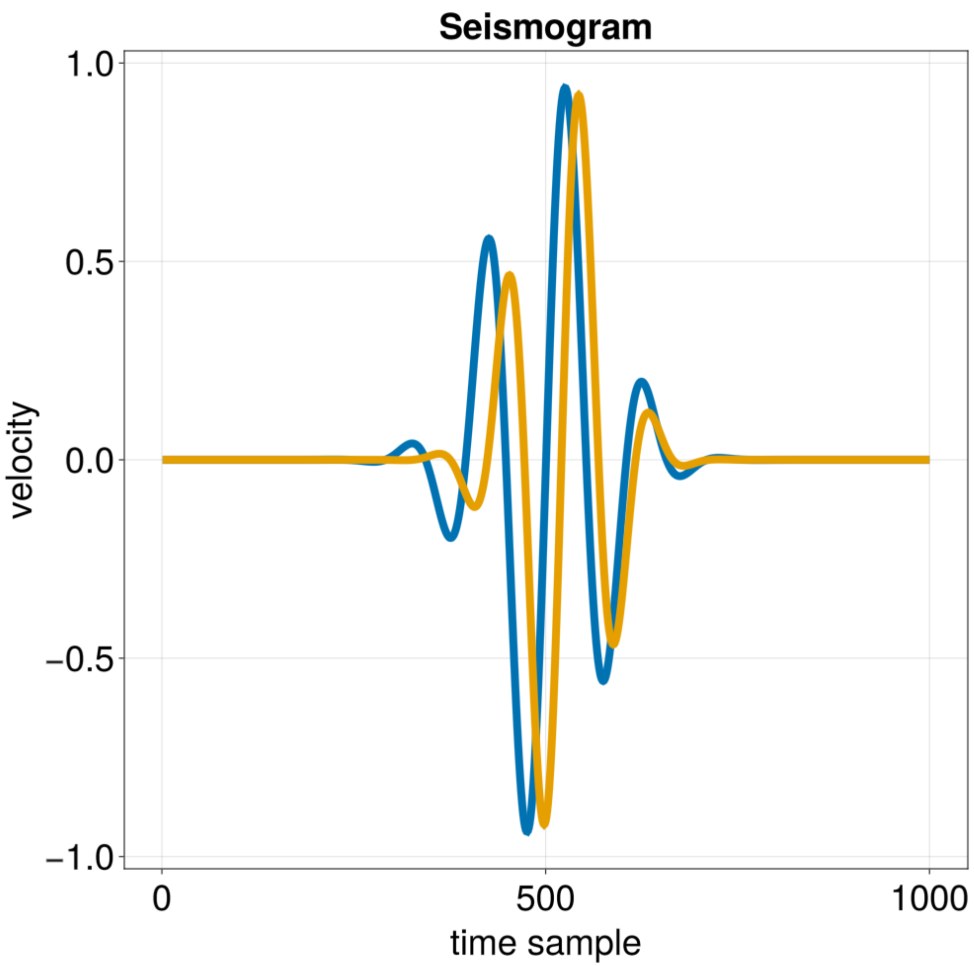
R – data sampling matrix
$\boldsymbol{u}$ – synthetic wavefield
$\boldsymbol{d}$ – measured data

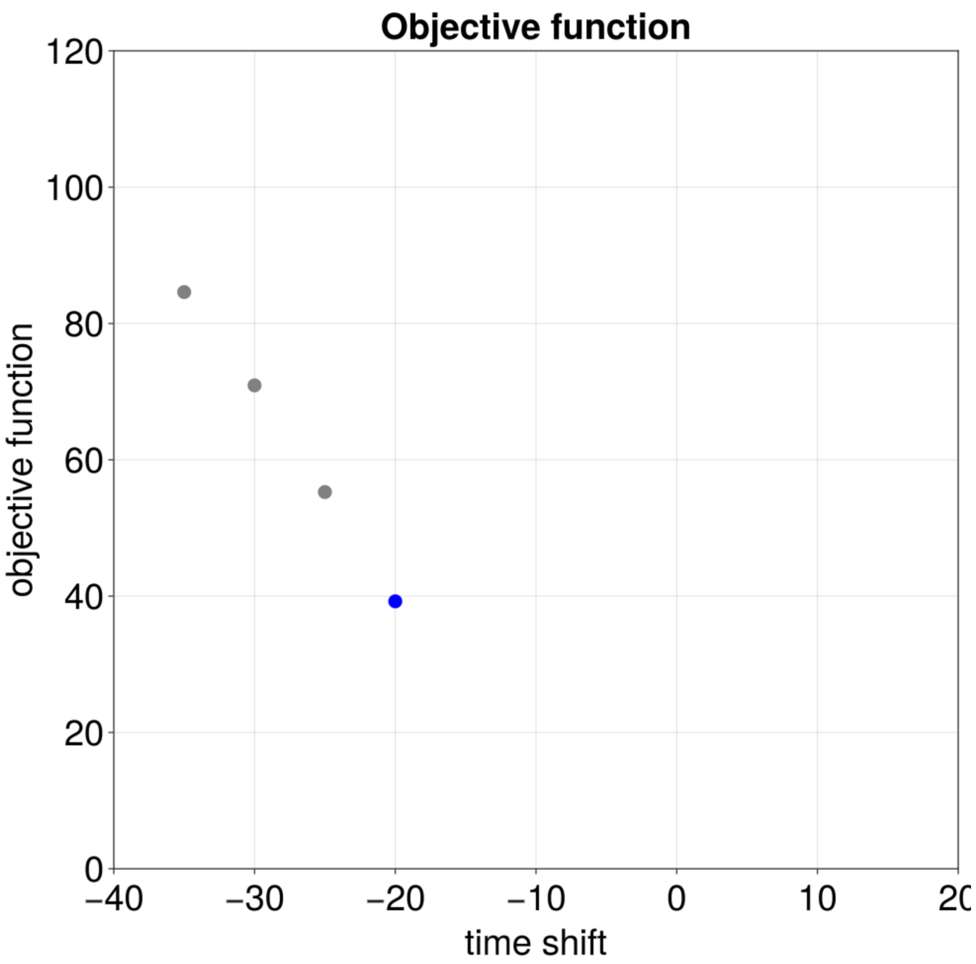
Objective functions



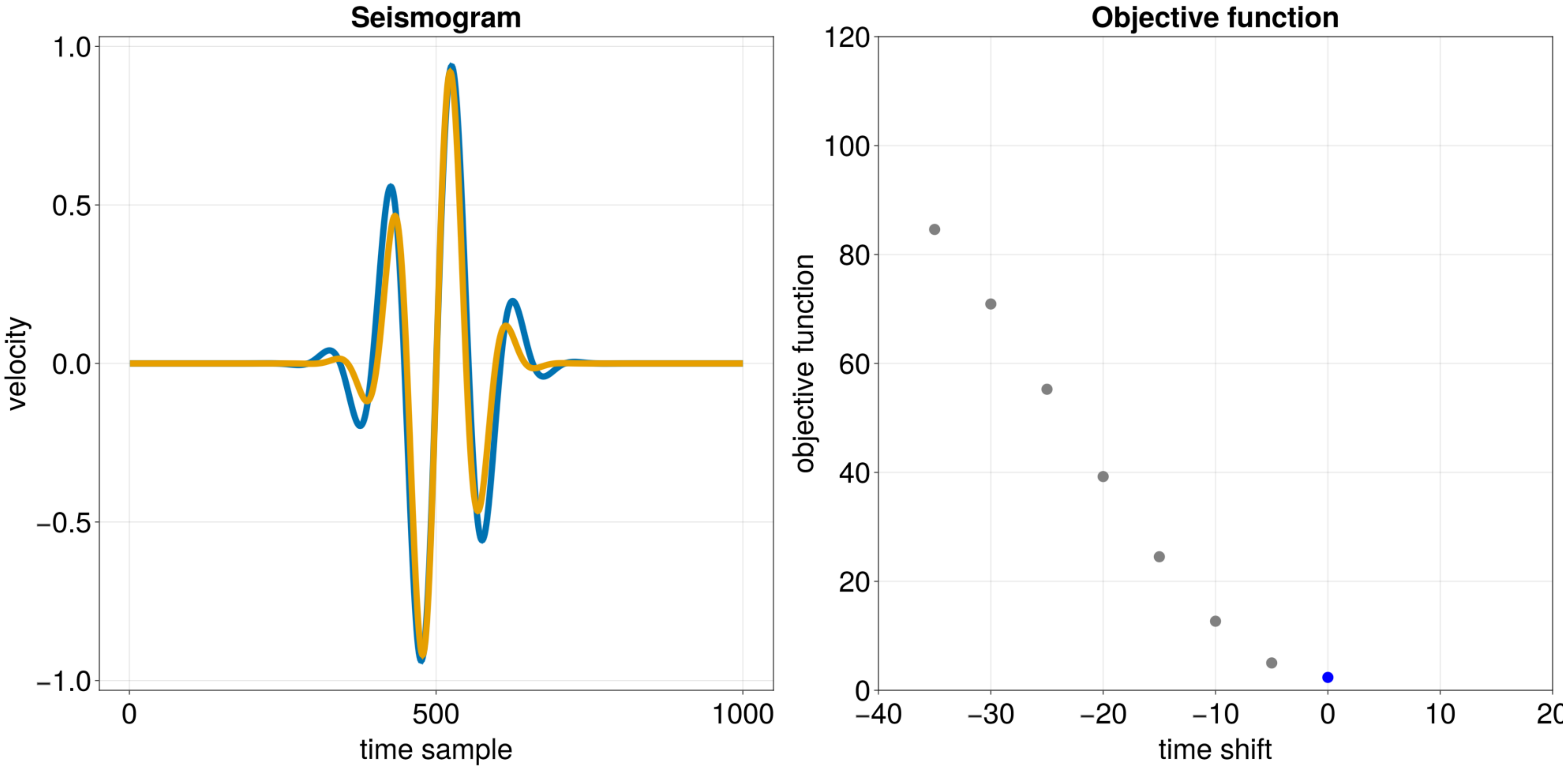


Objective functions

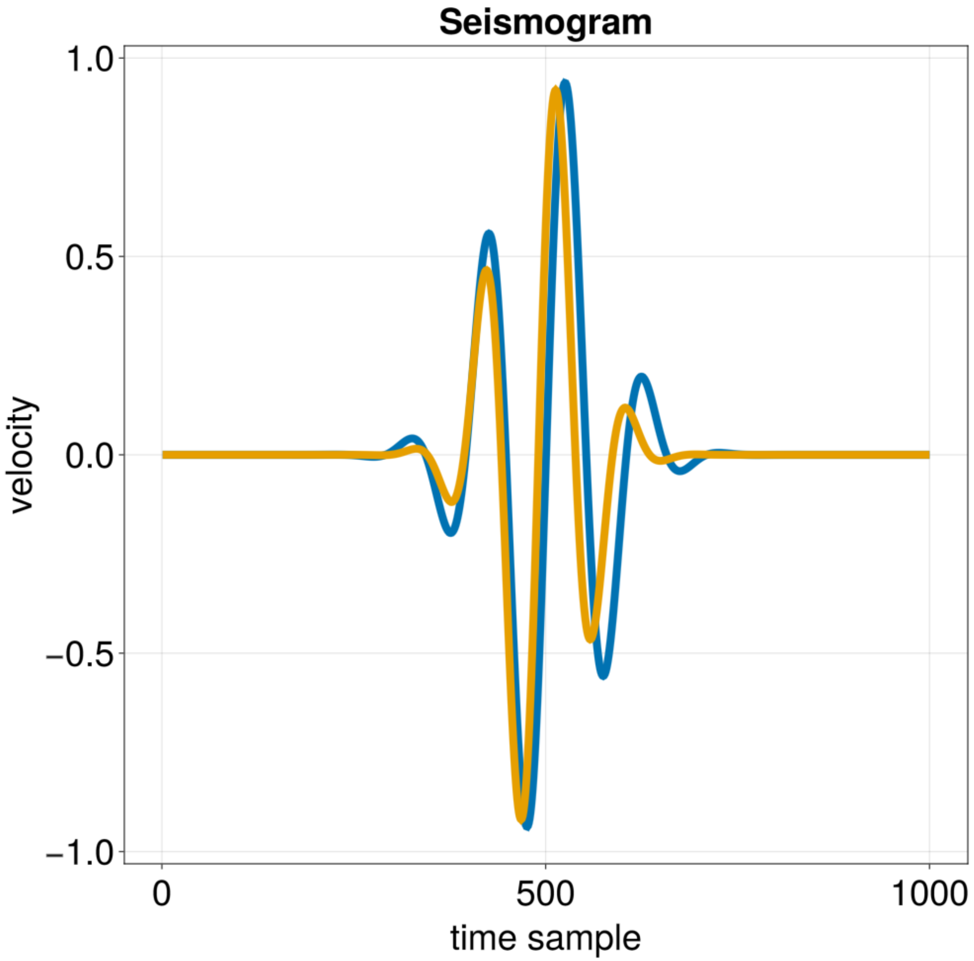


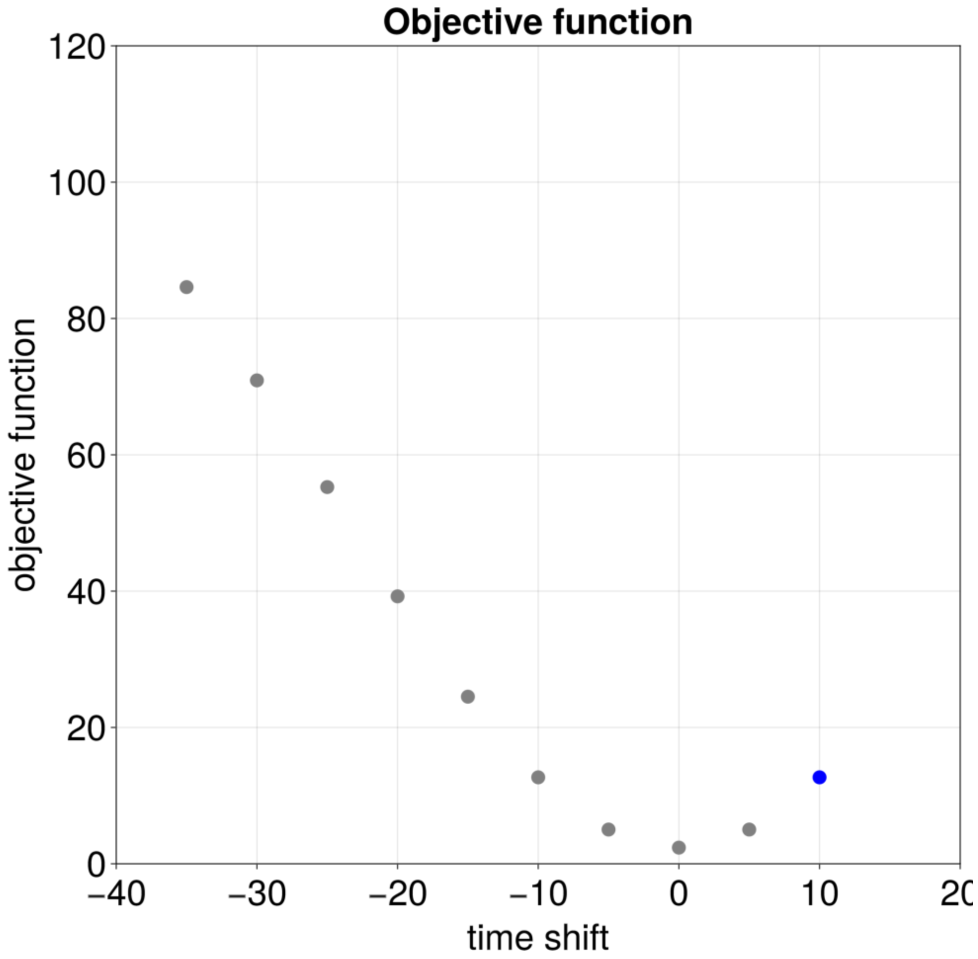


Objective functions

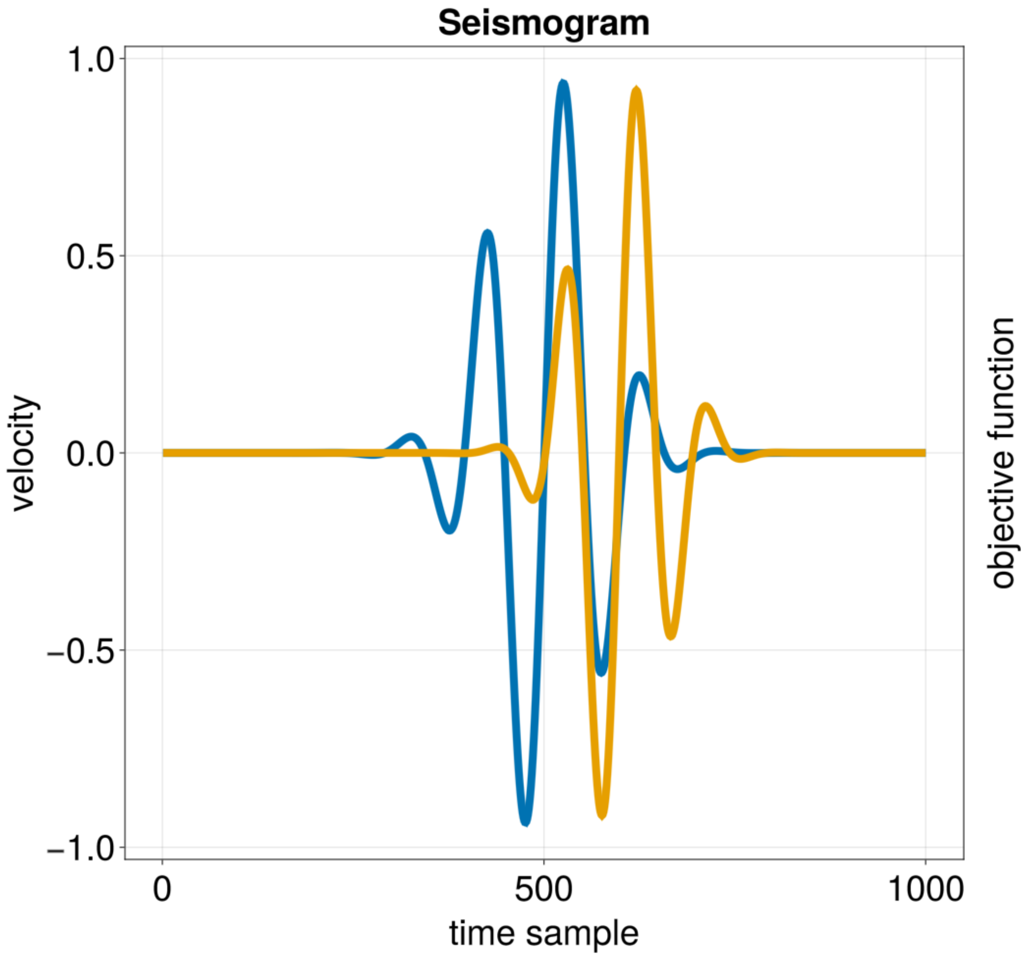


Objective functions

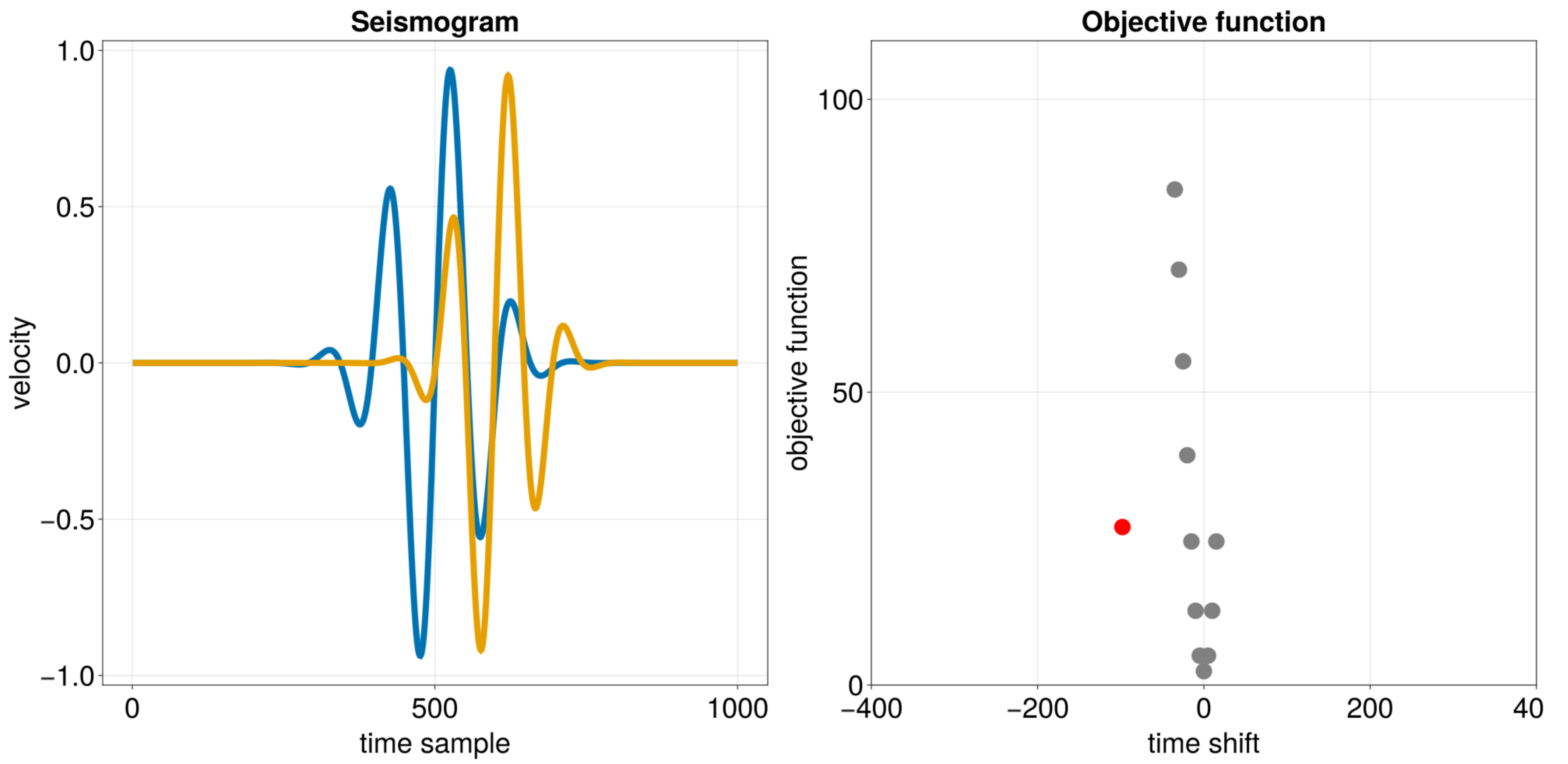




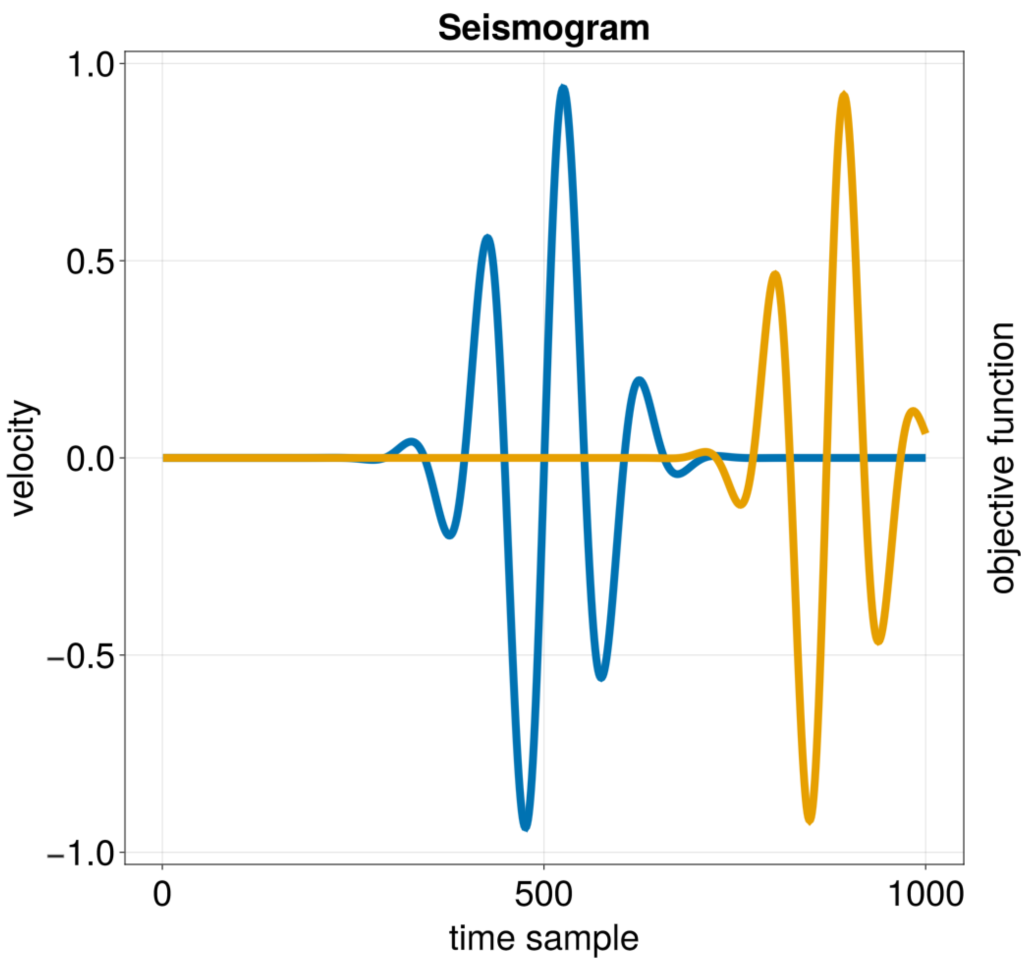
Objective functions



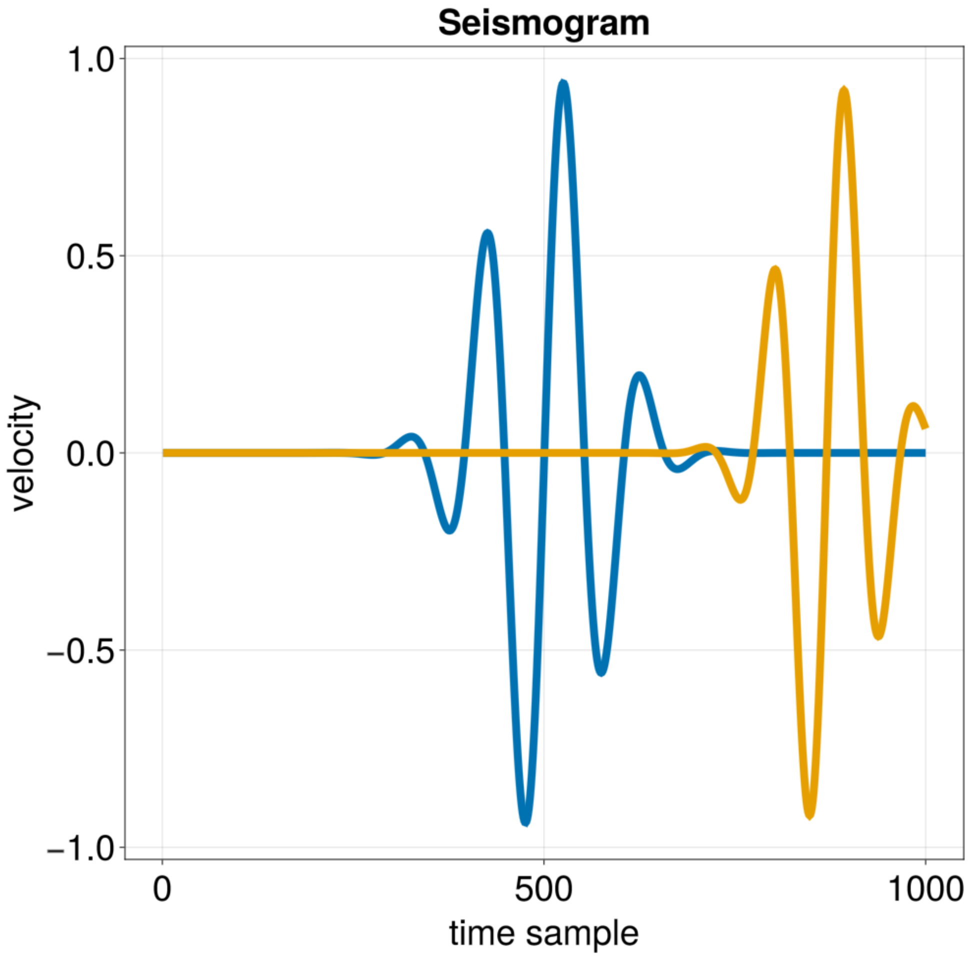
Objective functions

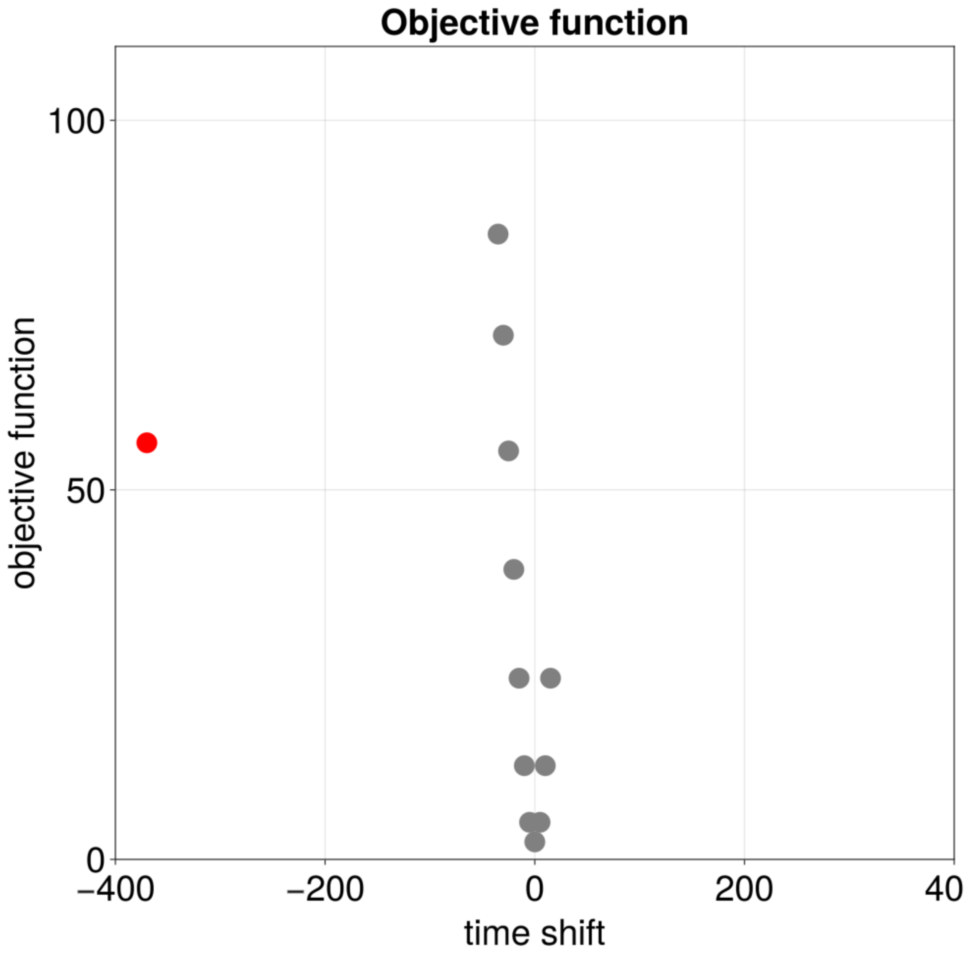


Objective functions

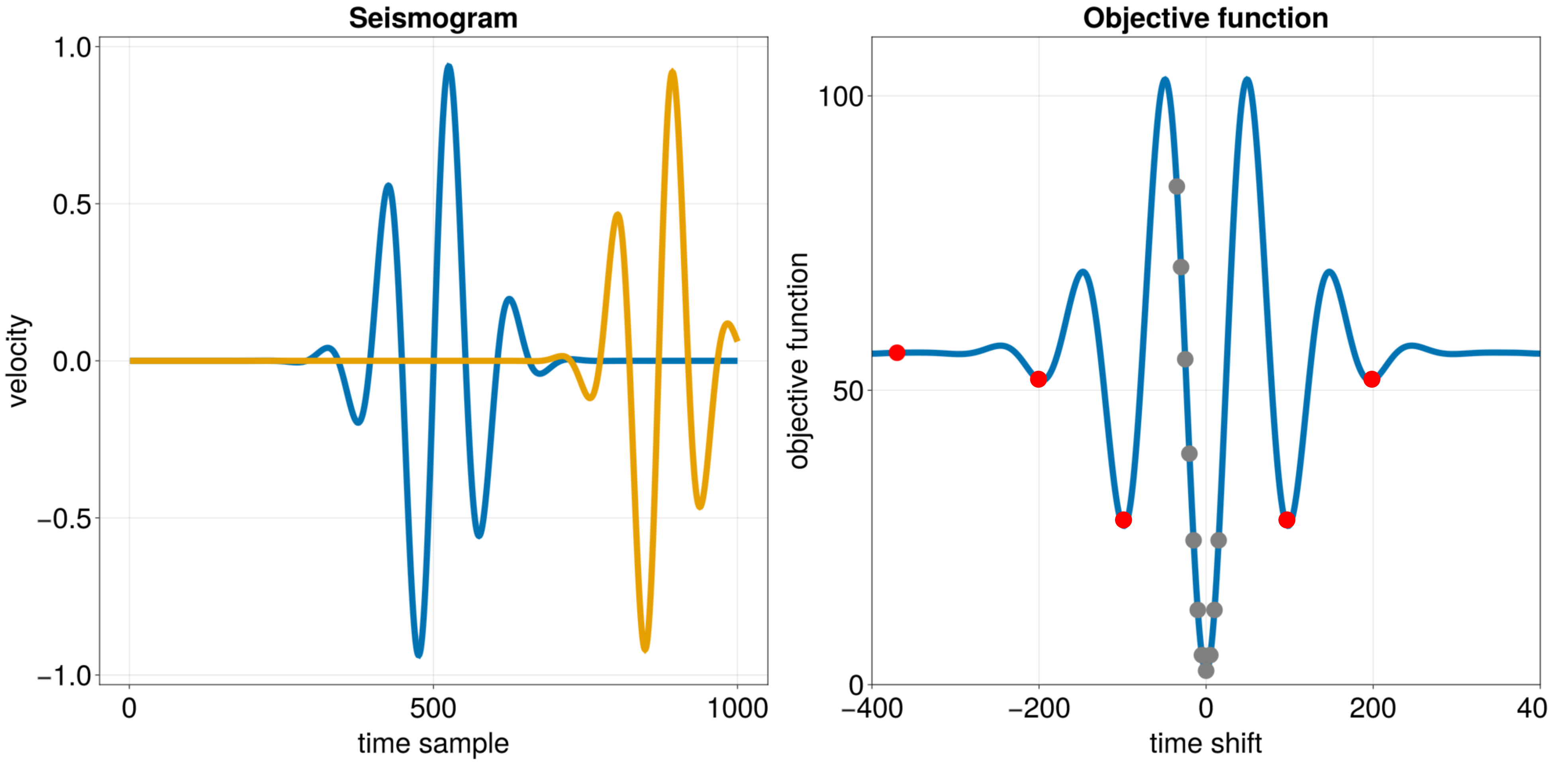


Objective functions





Objective functions



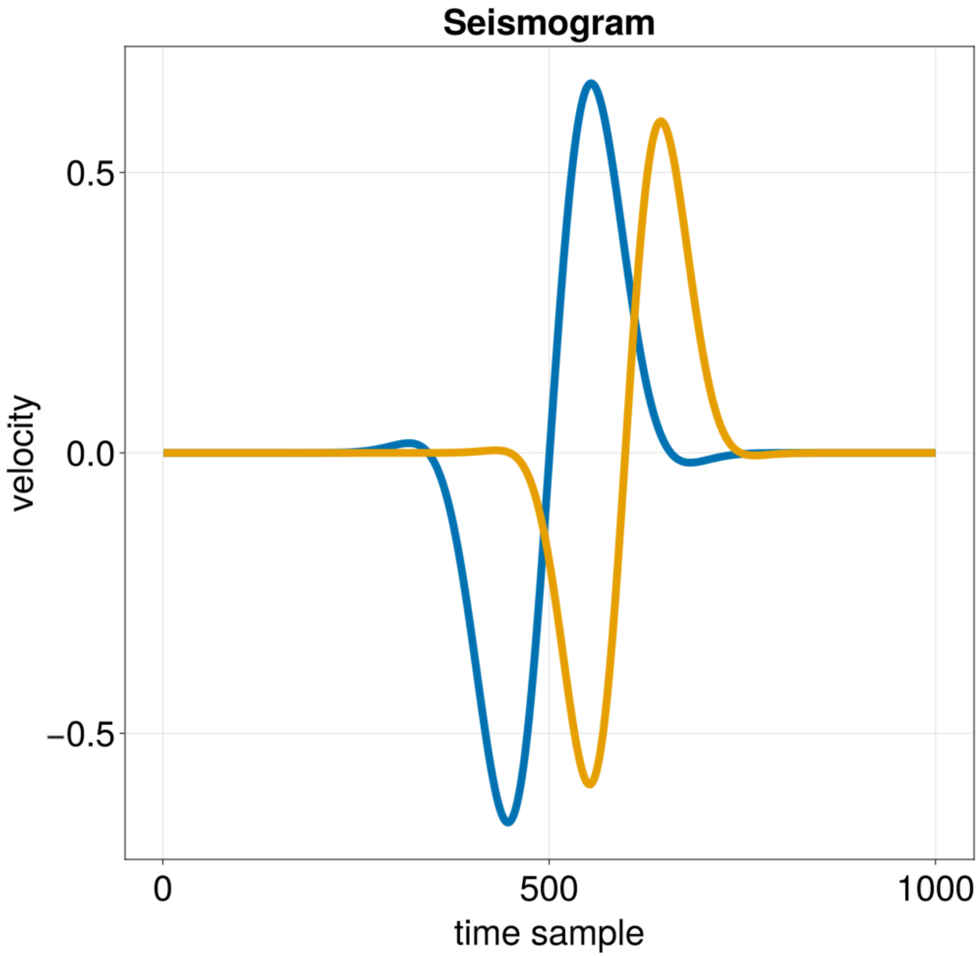
Cycle-skipping and insensitivity

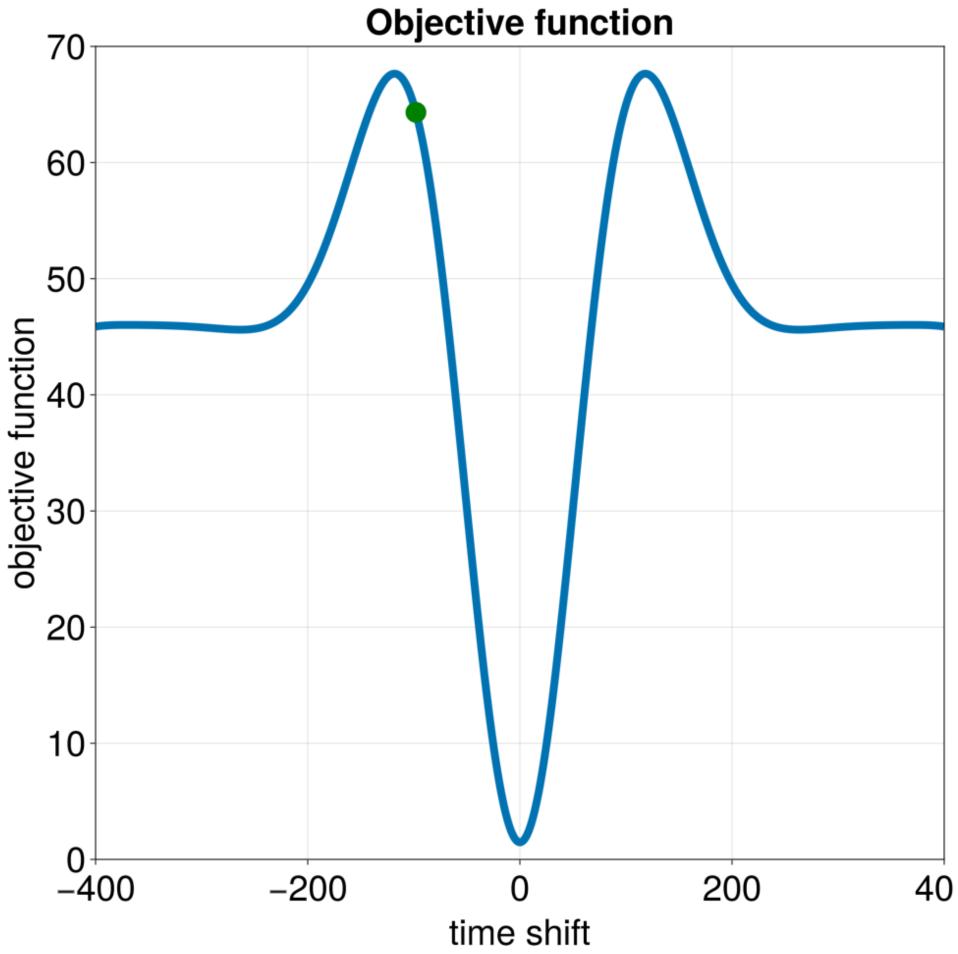
For models that poorly match the measured data, the L^2 misfit does not represent well “better” and “worse” data matches

Cycle-skipping, where mismatched peaks and troughs align, introduces significant nonlinearity

We also see insensitivity when measurements do not overlap with synthetics

Multi-scale inversion



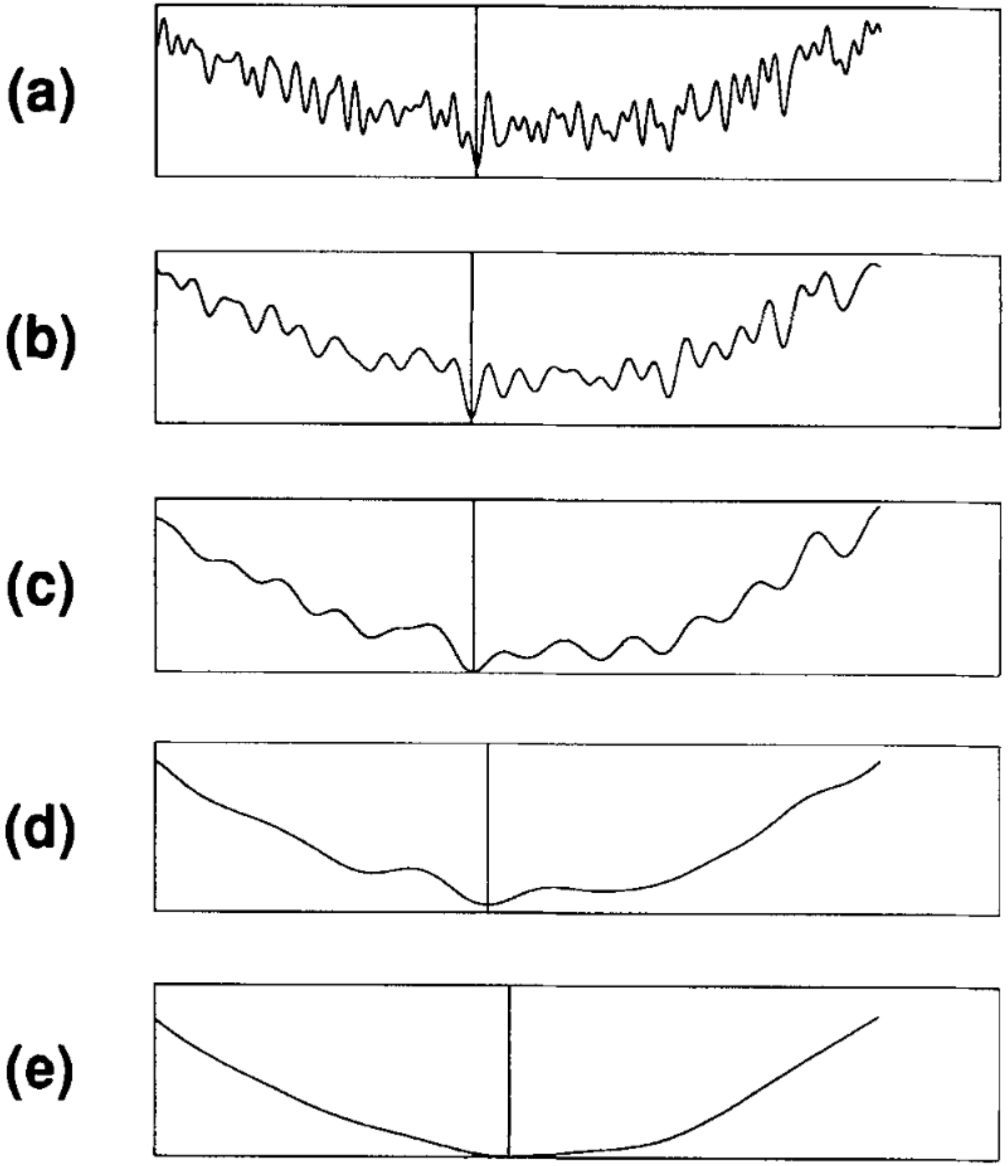


Multi-scale inversion

Lower frequencies are cheap to model

Low frequency data have better convexity properties

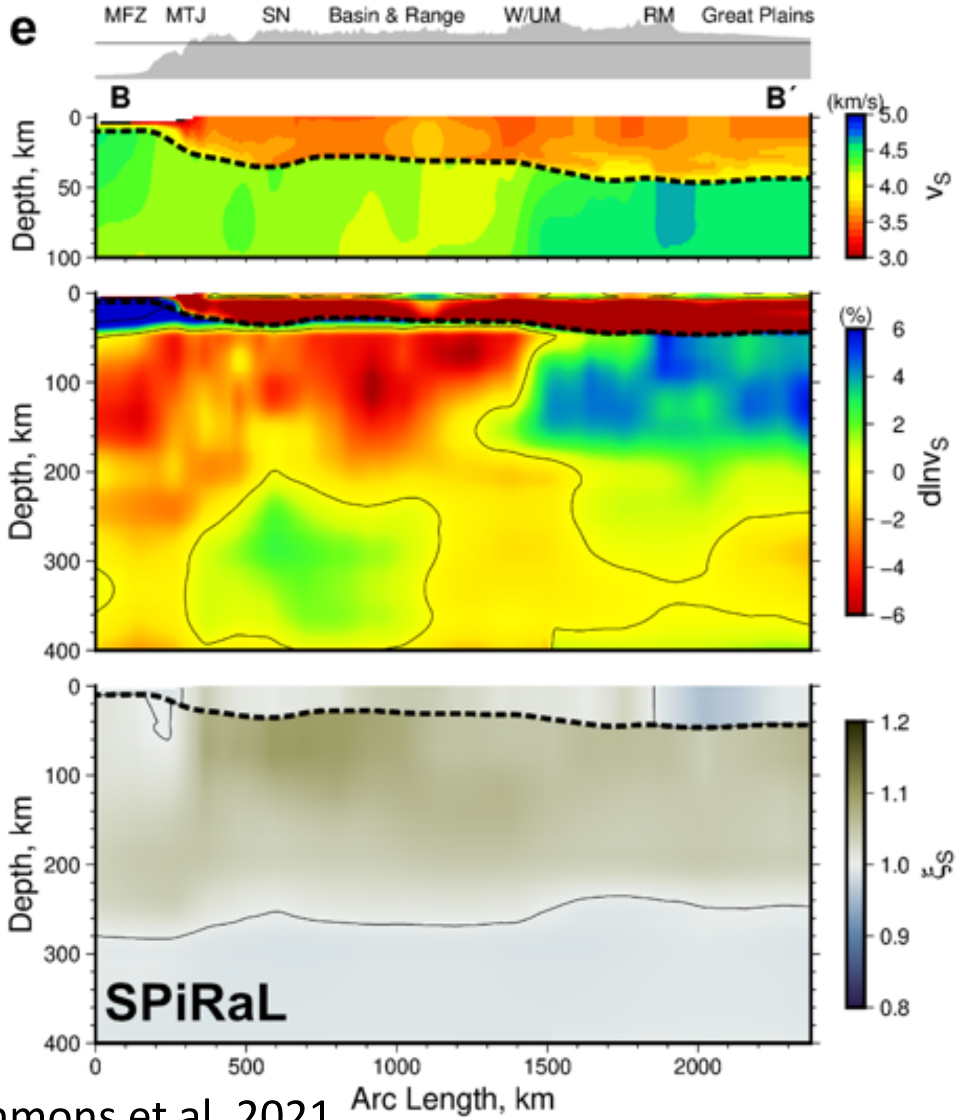
By inverting lower frequencies first, we may be able to start closer to the global minimum at higher bands



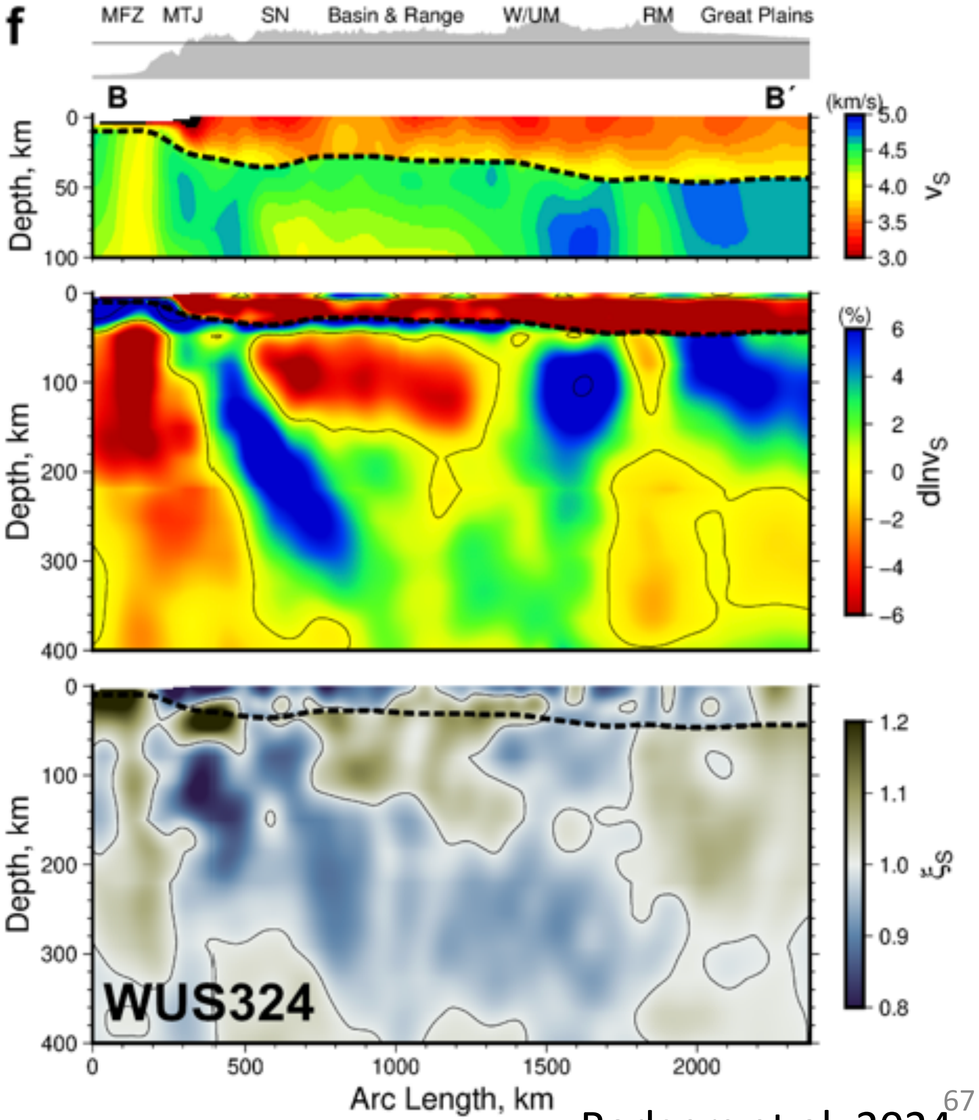
Bunks et al. 1995

Traveltime

Full-waveform

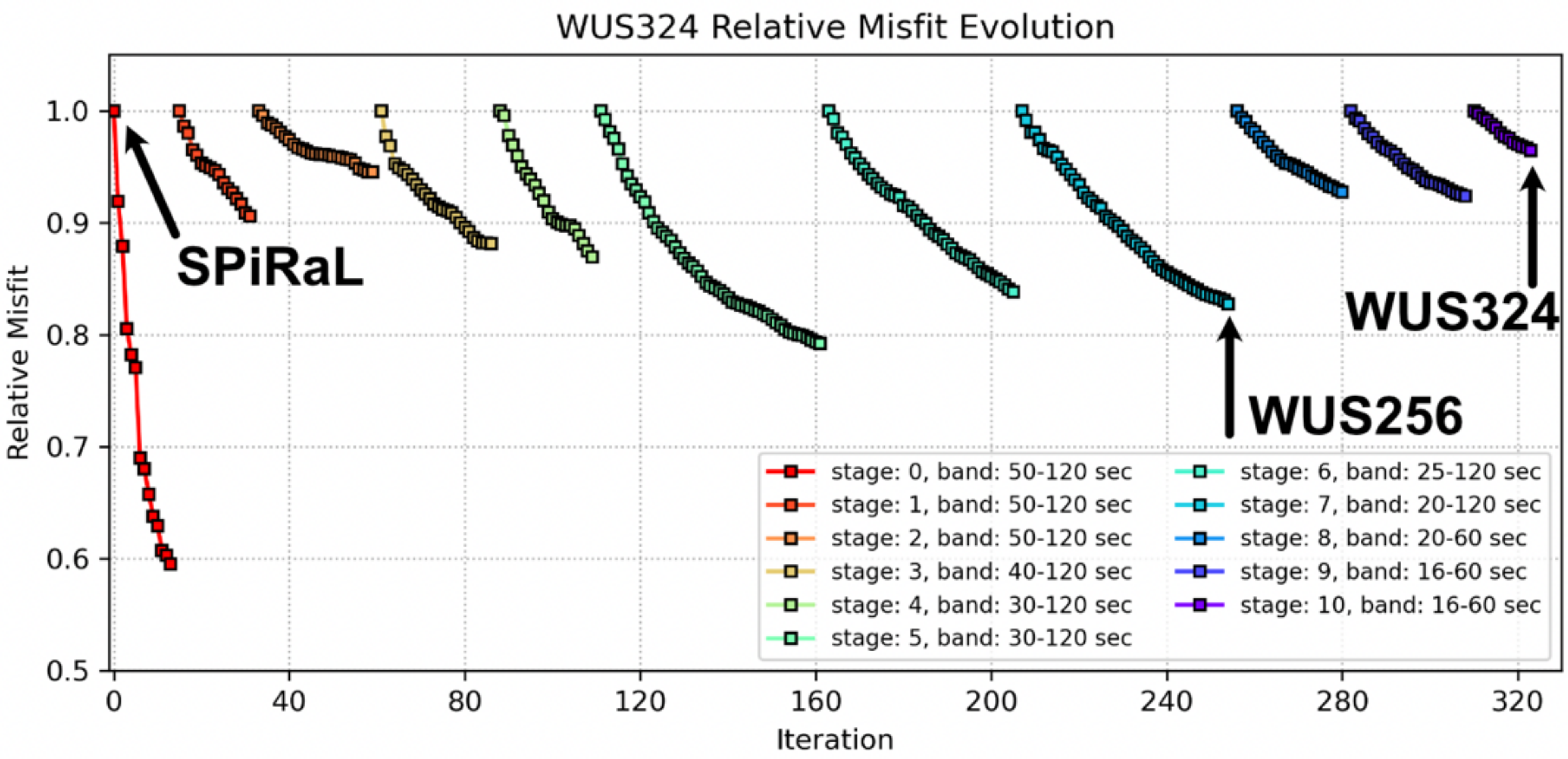


Simmons et al. 2021



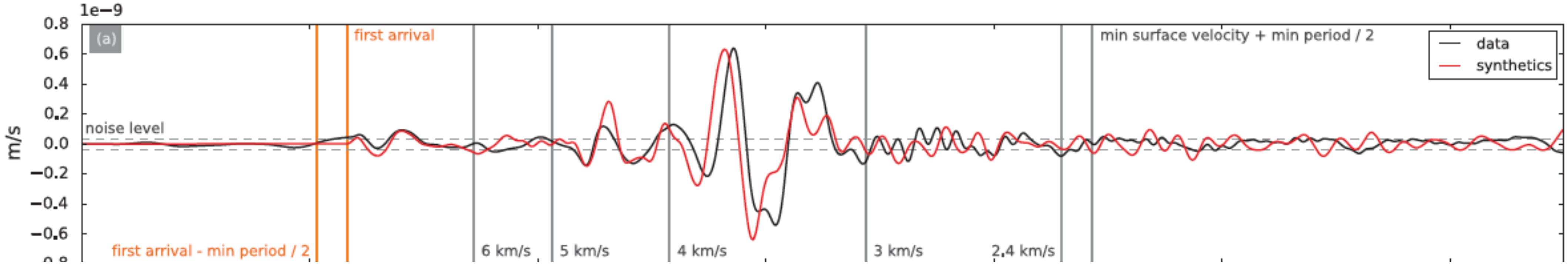
Rodgers et al. 2024

Multi-scale inversion



Rodgers et al. 2024

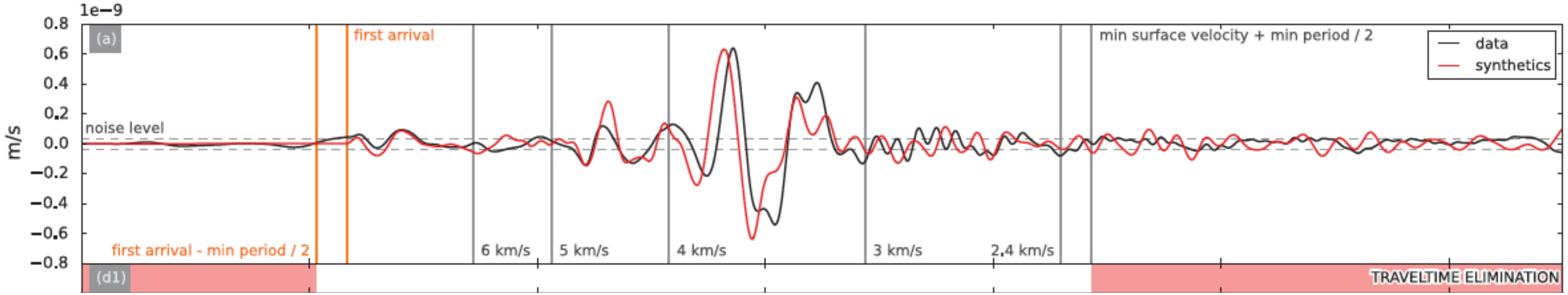
Windowing



Which data to consider?

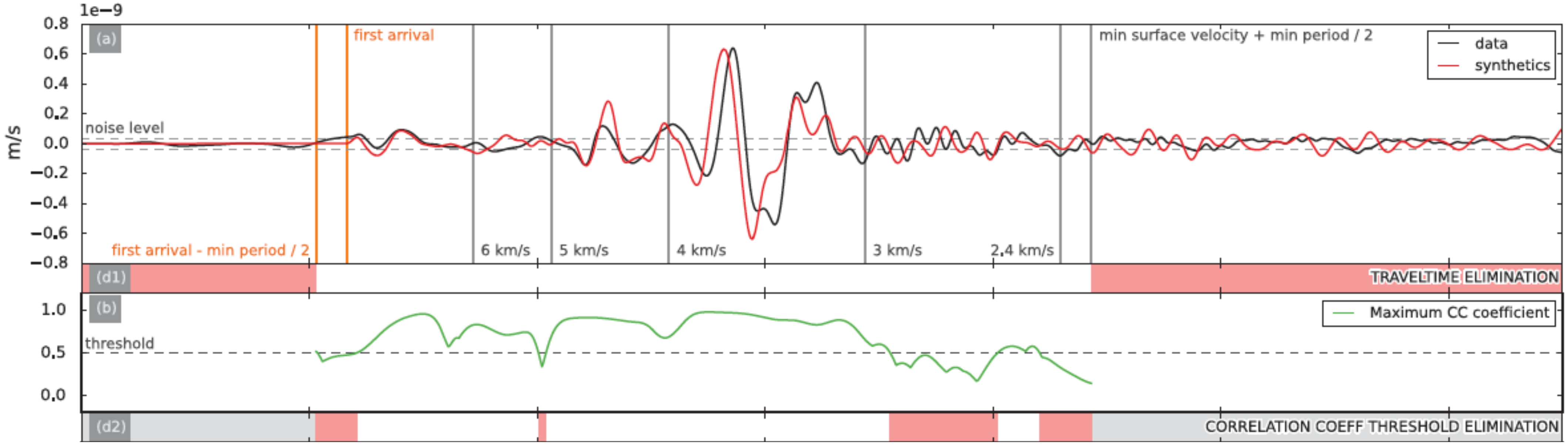
Krischer et al. 2015

Windowing



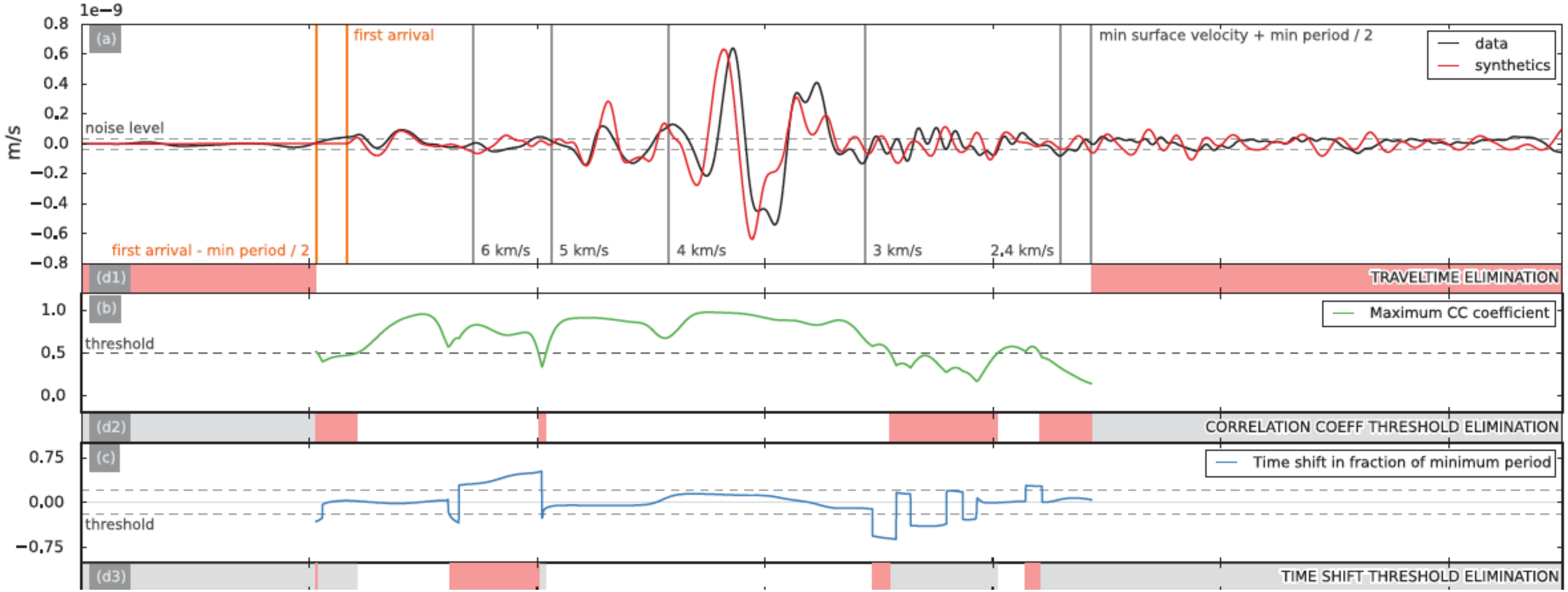
At what times should we be able to see the earthquake?

Windowing



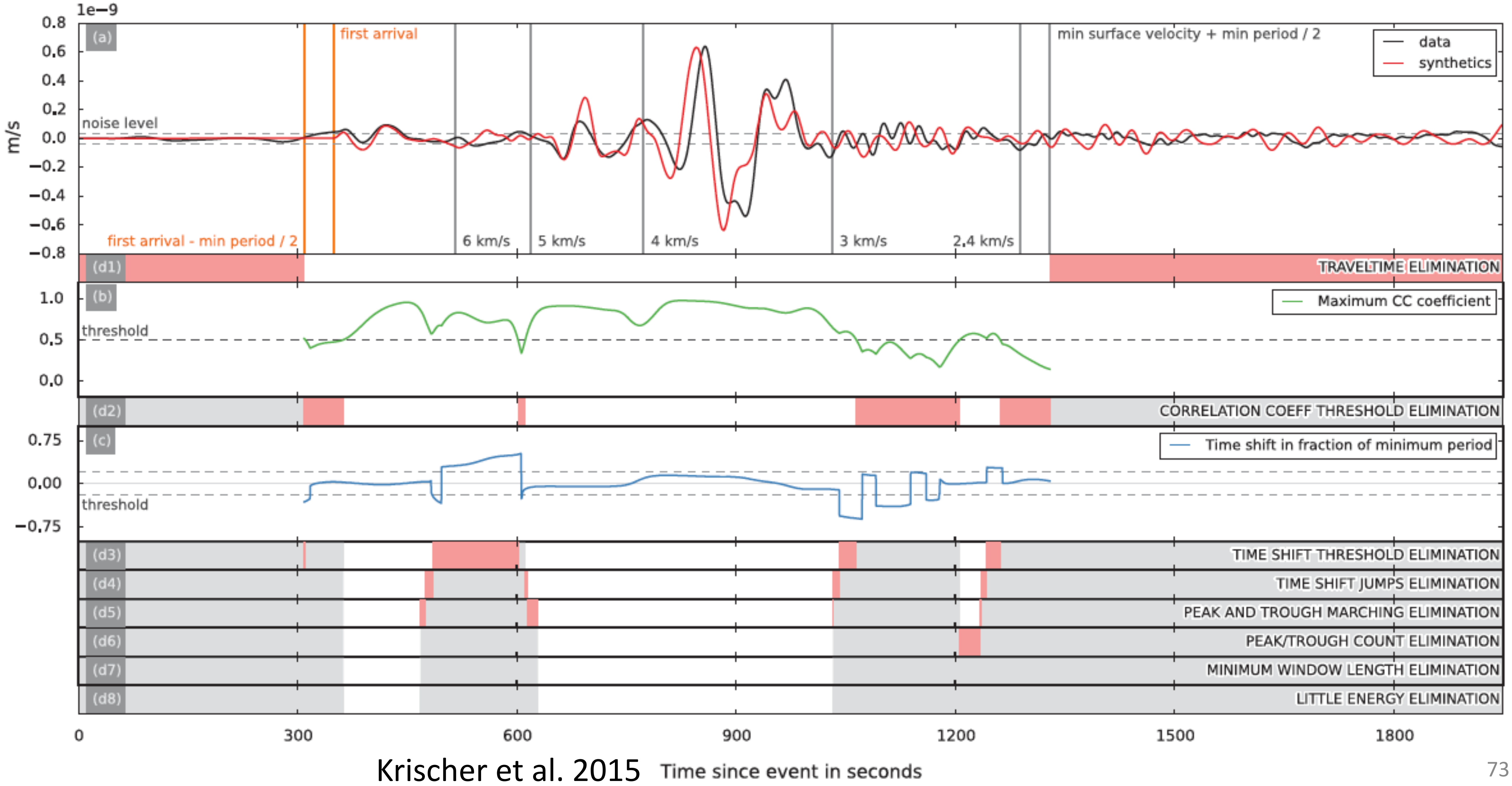
Does the synthetic correlate well with the data at some time shift?

Windowing



Is that time-shift close enough to prevent cycle-skipping?

Windowing



Krischer et al. 2015

Seismic amplitudes

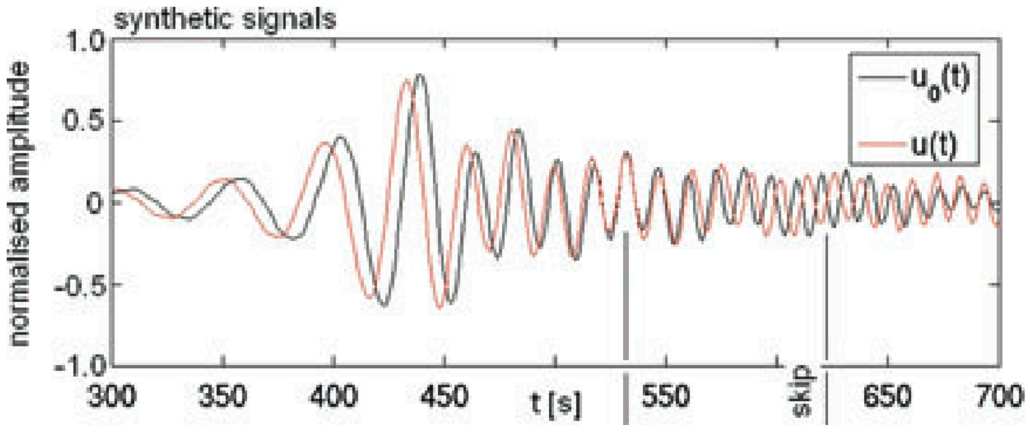
Measurements and simulations of seismic amplitudes can be unreliable

Measurements can be highly dependent on the Earth properties very close to the sensor, which may be unknown

Simulations are highly dependent on the source strength, which may not be known accurately

Phase-based objective functions

Fichtner, et al. 2008



If we neglect amplitudes, only phase information remains in our inversion

The phase of the full data set has a very nonlinear relation to our time series

Localized phase information can make a better objective

Gabor transform

A Fourier transform can be defined as

$$F(\omega) = \int f(t) e^{i\omega t} dt$$

The Gabor transform is effectively a windowed Fourier transform:

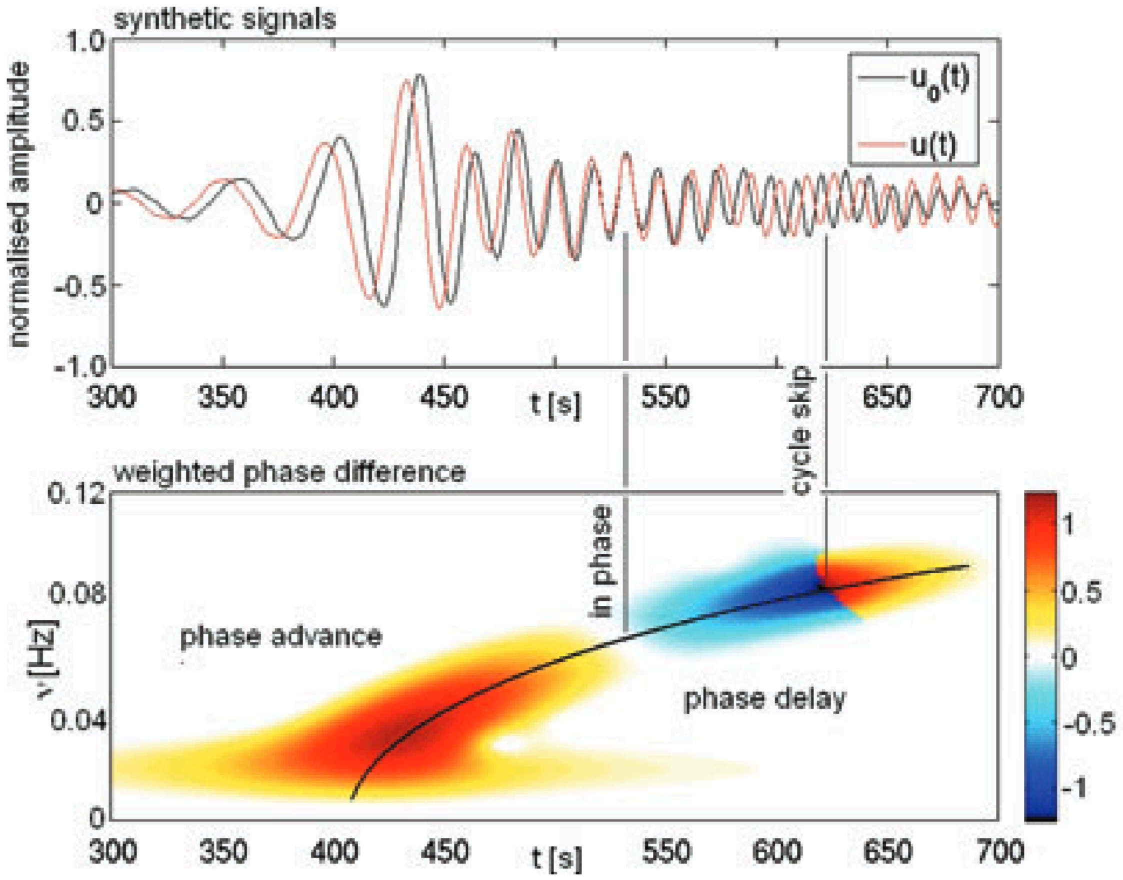
$$G(\omega, \tau) = \int f(t) e^{i\omega t} e^{-\alpha (t-\tau)^2} dt$$

This allows us to compare spectral information in a window

Time-frequency phase misfit

- The time-frequency phase misfit considers only phase differences between signals in time-frequency space

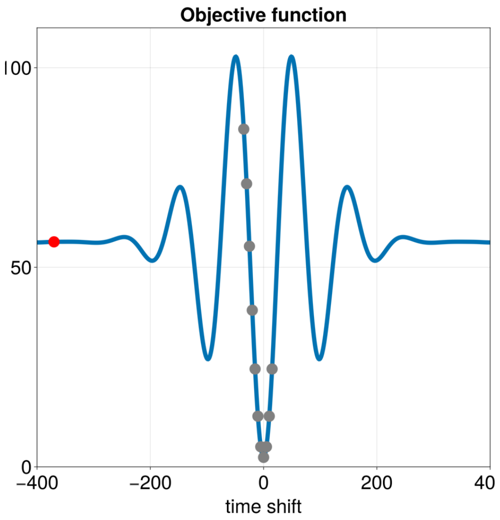
Fichtner, et al. 2008



Graph-space optimal transport

Time-frequency phase misfits can allow us to neglect amplitudes, but they don't help with cycle-skipping and insensitivity

Another suite of objective functions including graph-space optimal transport try to solve this problem

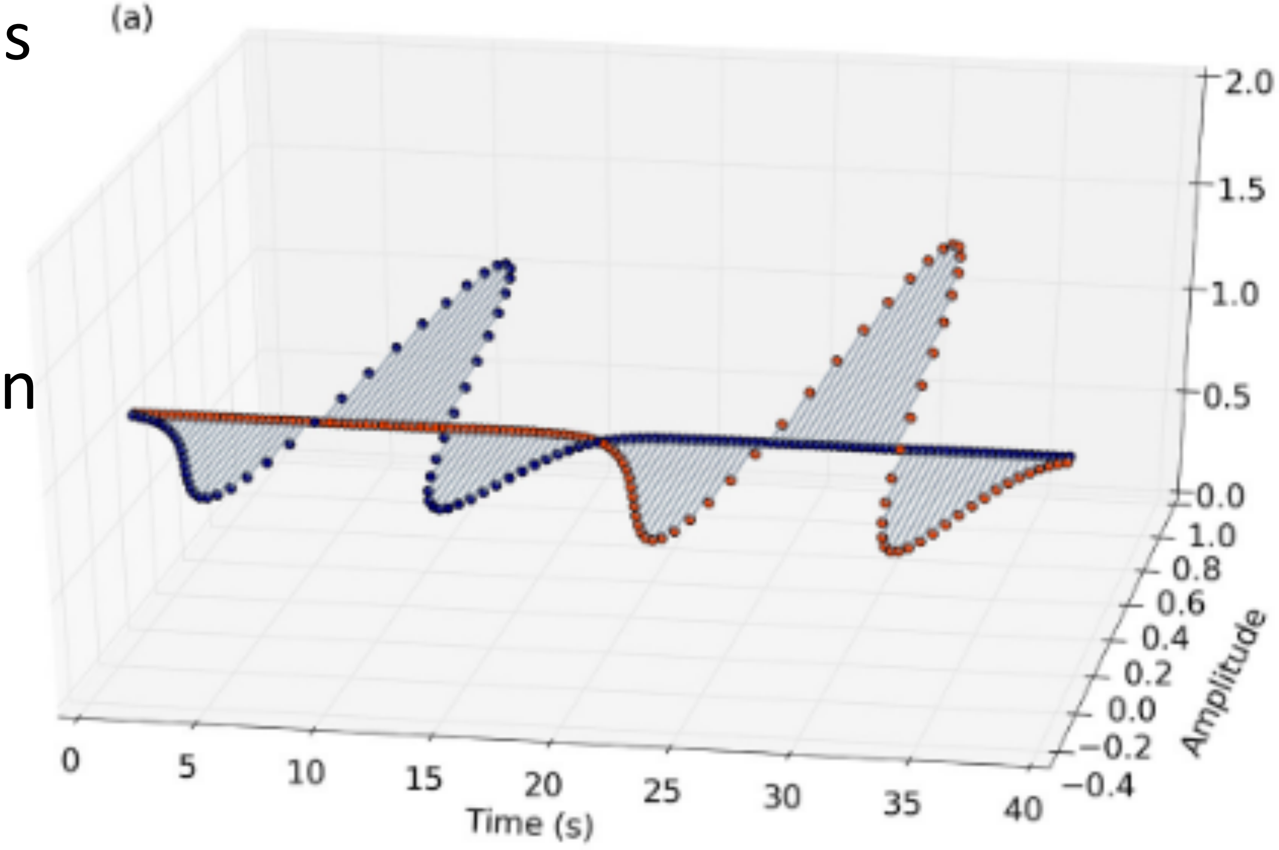


Graph-space optimal transport

An L^2 misfit can be thought of as a cost of moving each synthetic data point to match each measured data point

In L^2, these moves can only be in the measurement value

Metivier, et al. 2019



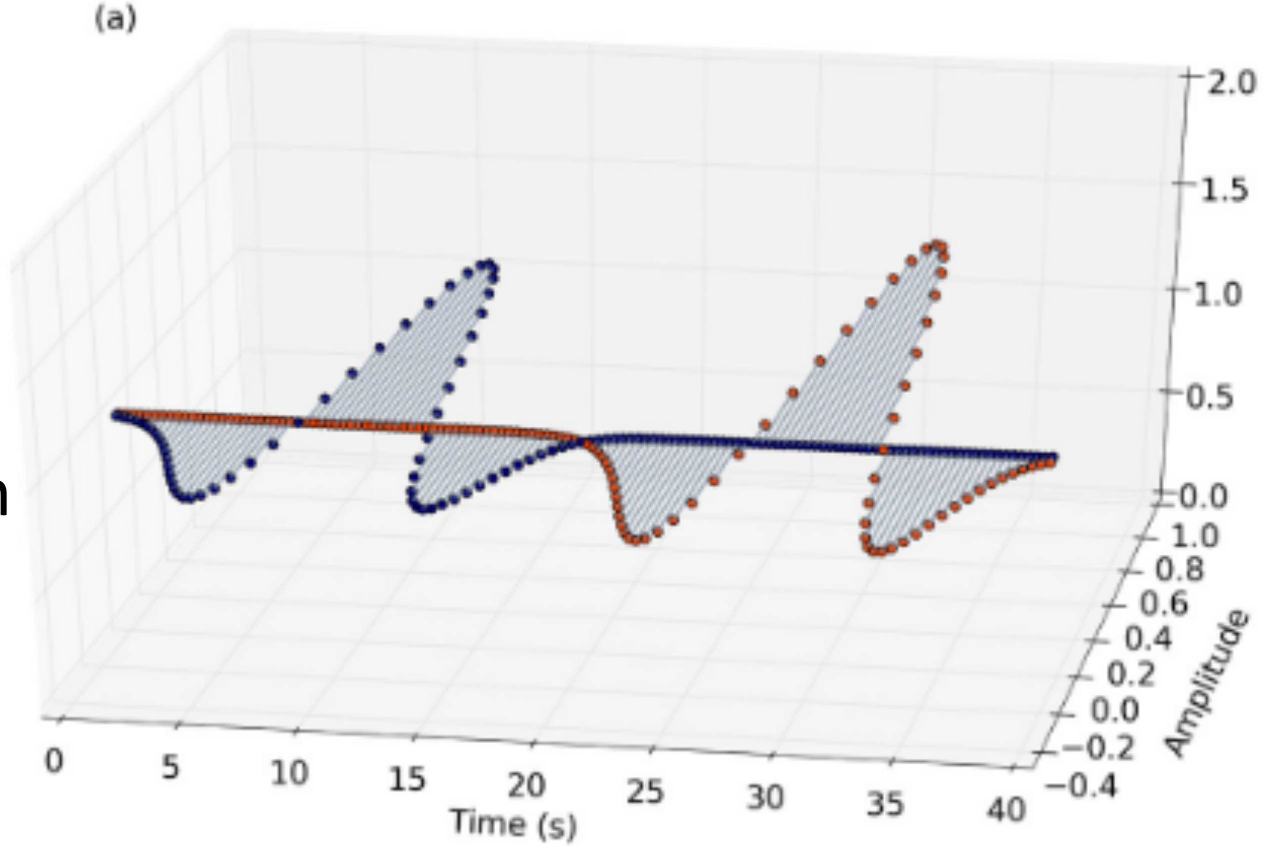
Graph-space optimal transport

Metivier, et al. 2019

GSOT allow for moves in **both dimensions**

Many moves become possible, so GSOT counts only the cost of the most efficient redistribution

By changing the cost of moves in time, better convexity can be achieved



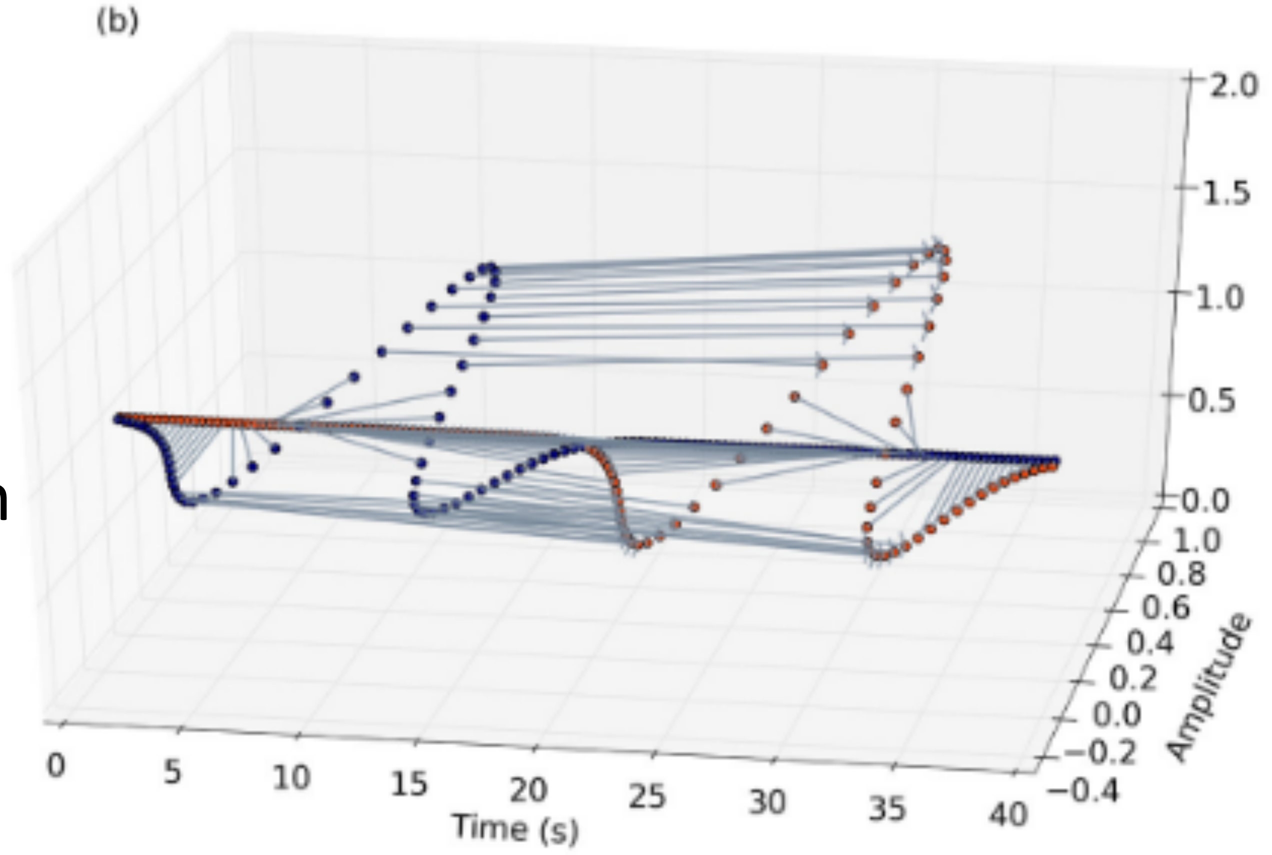
Graph-space optimal transport

GSOT allow for moves in **both dimensions**

Many moves become possible, so GSOT counts only the cost of the most efficient redistribution

By changing the cost of moves in time, better convexity can be achieved

Metivier, et al. 2019



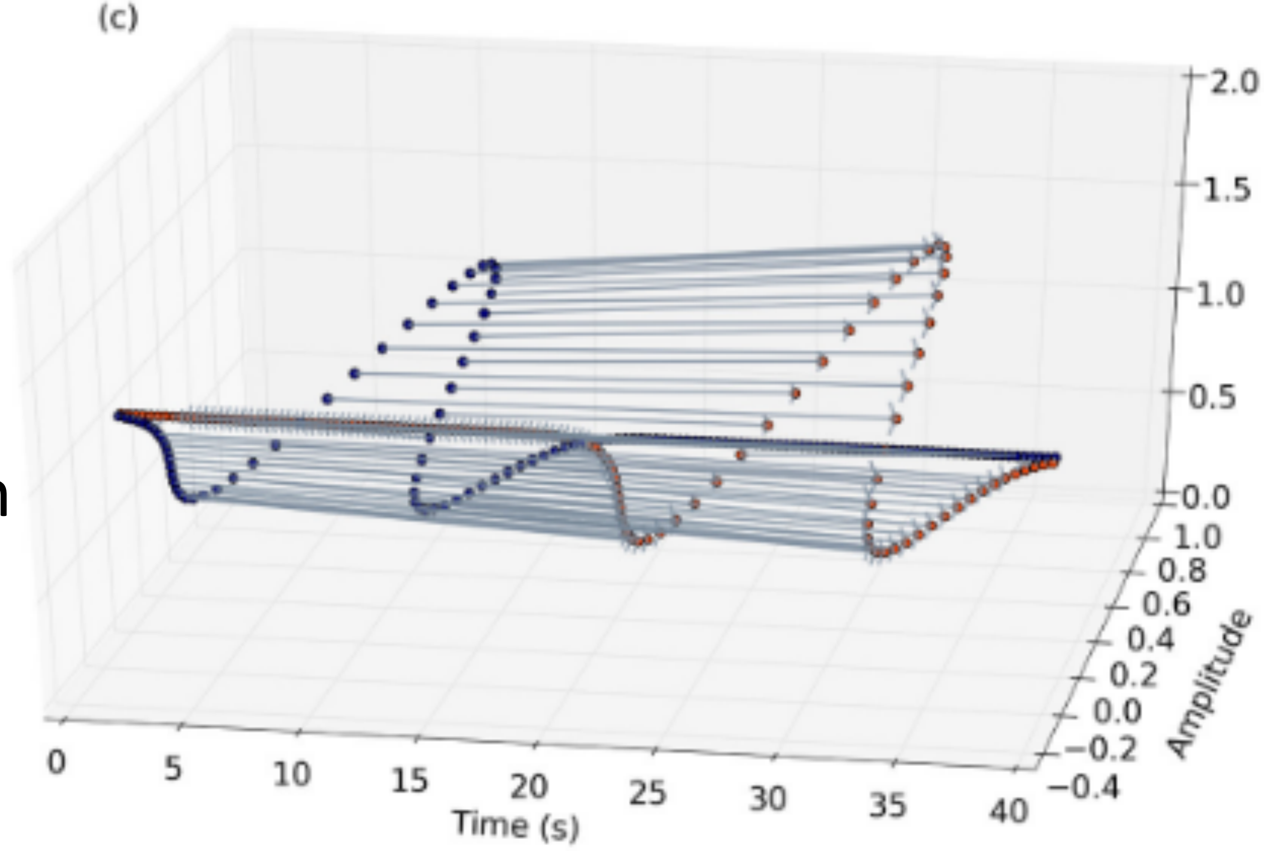
Graph-space optimal transport

GSOT allow for moves in **both dimensions**

Many moves become possible, so GSOT counts only the cost of the most efficient redistribution

By changing the cost of moves in time, better convexity can be achieved

Metivier, et al. 2019



GSOT

When time-shifts are cheap, we have a very convex objective function

This gets sharper (and better-resolved) as time-shifts become more expensive

Ideally, we optimize over a series with increasing time-shift cost

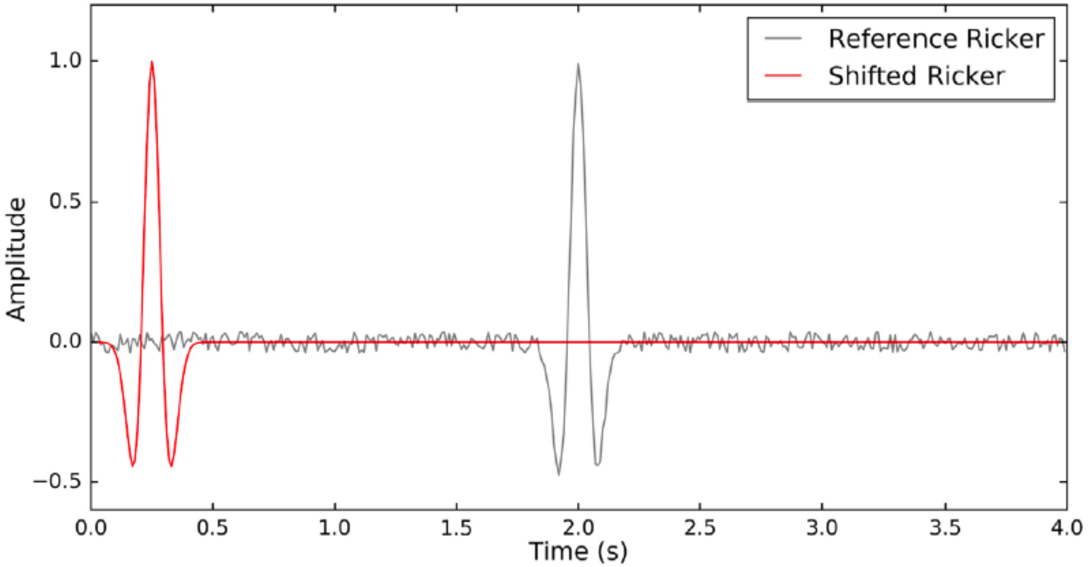
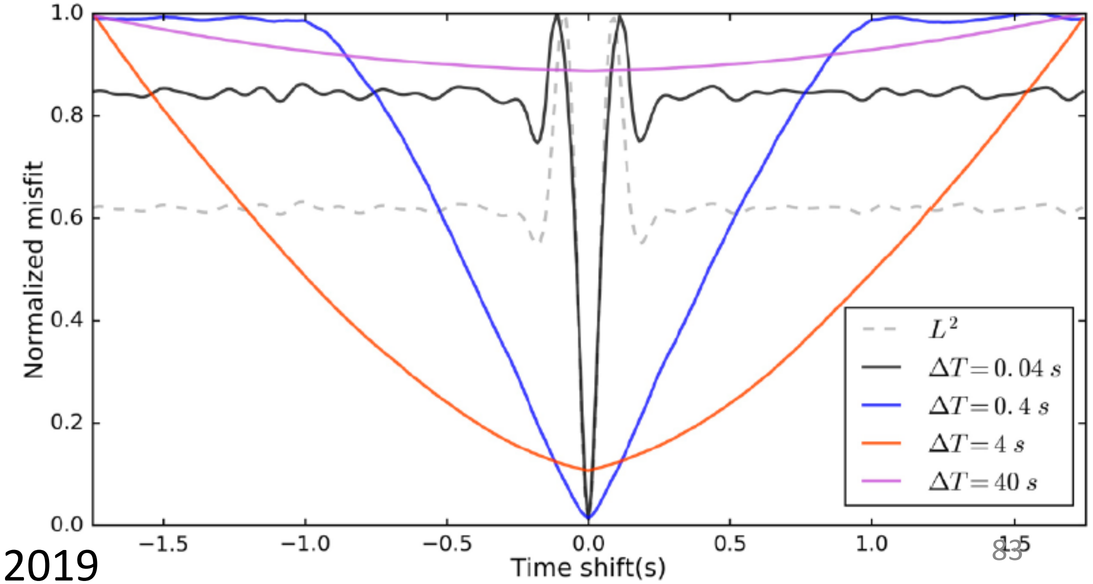


Figure 2. Reference Ricker function in solid gray line. Shifted in time Ricker functi in solid black line.



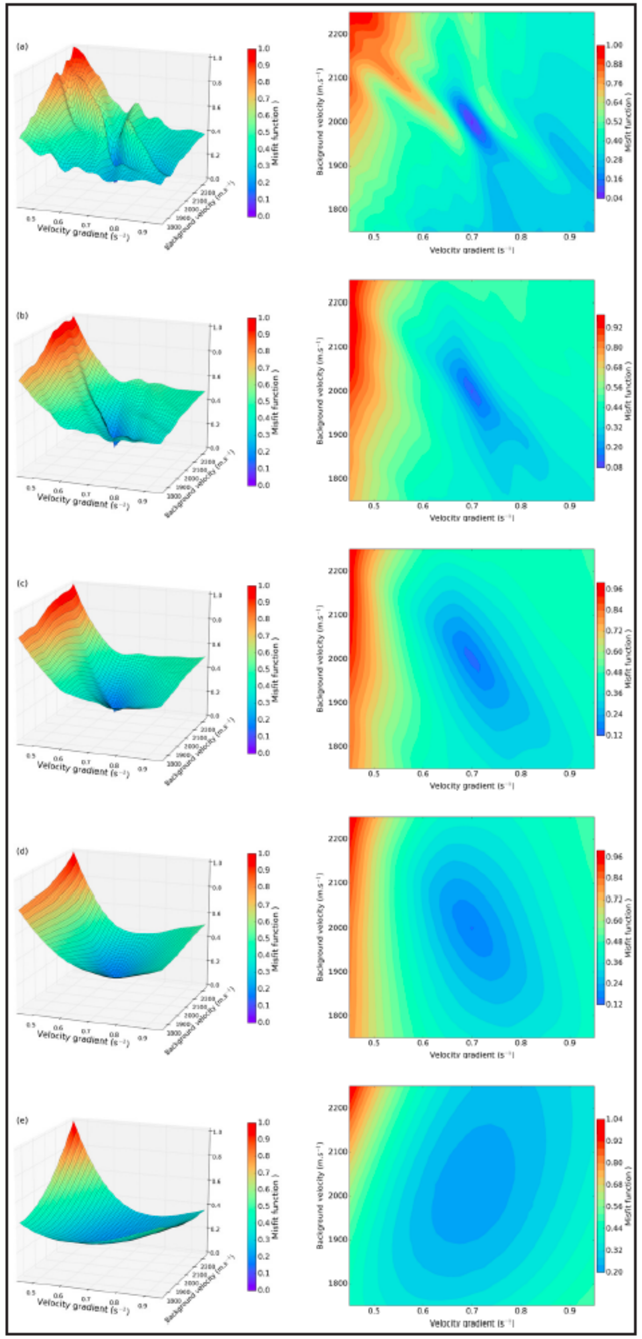
Metivier, et al. 2019

GSOT

When time-shifts are cheap, we have a very convex objective function

This gets sharper (and better-resolved) as time-shifts become more expensive

Ideally, we optimize over a series with increasing time-shift cost



Metivier, et al. 2019

Outline

1. Overview of full-waveform inversion
2. Numerical modelling of seismic waves
3. Objective function
4. Optimization
5. Reducing computational cost
6. Uncertainty quantification

Optimization

1. Gradient calculation
2. Prior information
3. l-BFGS

Full-waveform inversion: why is it difficult?

- Objective function, ϕ, defines better / worse models
- Iterative solutions driven by derivatives of ϕ
- $\phi(\boldsymbol{m}^{*} + \Delta\boldsymbol{m}) = \phi(\boldsymbol{m}^{*}) + \frac{d\phi}{d\mathbf{m}}\Delta\boldsymbol{m} + \frac{1}{2}\Delta\boldsymbol{m}^{T}\,\frac{d^{2}\phi}{d\boldsymbol{m}^{2}}\Delta\boldsymbol{m} + \ O(\Delta\boldsymbol{m}^{3})$

Full-waveform inversion: why is it difficult?

$$\phi = \phi(R\boldsymbol{u}(\boldsymbol{m}), \boldsymbol{d})$$

Full-waveform inversion: why is it difficult?

$$\phi = \phi(R\boldsymbol{u}(\boldsymbol{m}), \boldsymbol{d})$$

$$\frac{d\phi}{d\boldsymbol{m}} = \left(R\,\frac{d\boldsymbol{u}}{d\boldsymbol{m}}\right)^T \frac{\partial\phi}{\partial\boldsymbol{u}}$$

Full-waveform inversion: why is it difficult?

$$\phi = \phi(R\boldsymbol{u}(\boldsymbol{m}), \boldsymbol{d})$$

$$\frac{d\phi}{d\boldsymbol{m}} = \left(R\,\frac{d\boldsymbol{u}}{d\boldsymbol{m}}\right)^T \frac{\partial\phi}{\partial\boldsymbol{u}}$$

Uh oh

Full-waveform inversion: why is it difficult?

$$\phi = \phi(R\boldsymbol{u}(\boldsymbol{m}), \boldsymbol{d})$$

$$\frac{d\phi}{d\boldsymbol{m}} = \left(R\frac{d\boldsymbol{u}}{d\boldsymbol{m}}\right)^T \frac{\partial\phi}{\partial\boldsymbol{u}}$$

Uh oh

For 10^7 model parameters the wave propagation cost of FD estimation is untenable

Full-waveform inversion: why is it possible?

The inversion problem is made feasible by use of the adjoint-state method

This allows the derivatives with respect to any number of model parameters to be calculated at the cost of a single additional wave-propagation

Full-waveform inversion: why is it possible?

$$L = \phi(R\boldsymbol{u}, \boldsymbol{d}) + (F(\boldsymbol{m})\boldsymbol{u} - S)^{\dagger}\boldsymbol{\lambda}$$

Full-waveform inversion: why is it possible?

$$L = \phi(R\boldsymbol{u}, \boldsymbol{d}) + (F(\boldsymbol{m})\boldsymbol{u} - S)^{\dagger}\boldsymbol{\lambda}$$

$$L(\boldsymbol{u}^{*}) = \phi(R\boldsymbol{u}(\boldsymbol{m}), \boldsymbol{d})$$

Full-waveform inversion: why is it possible?

$$L = \phi(R\boldsymbol{u}, \boldsymbol{d}) + (F(\boldsymbol{m})\boldsymbol{u} - S)^{\dagger}\boldsymbol{\lambda}$$

$$\frac{d\phi}{d\boldsymbol{m}} = \frac{dL}{d\boldsymbol{m}}(\boldsymbol{u}^{*}) = \frac{\partial L}{\partial m}(\boldsymbol{u}^{*}) + \frac{\partial L}{\partial \boldsymbol{u}^{*}} \frac{d\boldsymbol{u}^{*}}{d\boldsymbol{m}}$$

Full-waveform inversion: why is it possible?

$$L = \phi(R\boldsymbol{u}, \boldsymbol{d}) + (F(\boldsymbol{m})\boldsymbol{u} - S)^{\dagger}\boldsymbol{\lambda}$$

$$\frac{d\phi}{d\boldsymbol{m}} = \frac{dL}{d\boldsymbol{m}}(\boldsymbol{u}^{*}) = \frac{\partial L}{\partial m}(\boldsymbol{u}^{*}) + \frac{\partial L}{\partial \boldsymbol{u}^{*}}\frac{d\boldsymbol{u}^{*}}{d\boldsymbol{m}}$$

$$\frac{\partial L}{\partial \boldsymbol{u}^{*}}(\boldsymbol{\lambda}^{*}) = \boldsymbol{0} = \frac{\partial \phi}{\partial \boldsymbol{u}^{*}} + F(\boldsymbol{m})^{\dagger}\boldsymbol{\lambda}^{*}$$

Full-waveform inversion: why is it possible?

$$L = \phi(R\boldsymbol{u}, \boldsymbol{d}) + (F(\boldsymbol{m})\boldsymbol{u} - S)^{\dagger}\boldsymbol{\lambda}$$

$$\frac{d\phi}{d\boldsymbol{m}} = \frac{dL}{d\boldsymbol{m}}(\boldsymbol{u}^*) = \frac{\partial L}{\partial m}(\boldsymbol{u}^*) + \frac{\partial L}{\partial \boldsymbol{u}^*}\frac{d\boldsymbol{u}^*}{d\boldsymbol{m}}$$

$$\frac{\partial L}{\partial \boldsymbol{u}^*}(\boldsymbol{\lambda}^*) = \boldsymbol{0} = \frac{\partial \phi}{\partial \boldsymbol{u}^*} + F(\boldsymbol{m})^{\dagger}\boldsymbol{\lambda}^*$$

$$F(\boldsymbol{m})^{\dagger}\boldsymbol{\lambda}^* = -\frac{\partial \phi}{\partial \boldsymbol{u}^*}$$

Full-waveform inversion: why is it possible?

$$L = \phi(R\boldsymbol{u}, \boldsymbol{d}) + (F(\boldsymbol{m})\boldsymbol{u} - S)^{\dagger}\boldsymbol{\lambda}$$

$$\frac{d\phi}{d\boldsymbol{m}} = \frac{dL}{d\boldsymbol{m}}(\boldsymbol{u}^*) = \frac{\partial L}{\partial m}(\boldsymbol{u}^*) + \mathbf{0}$$

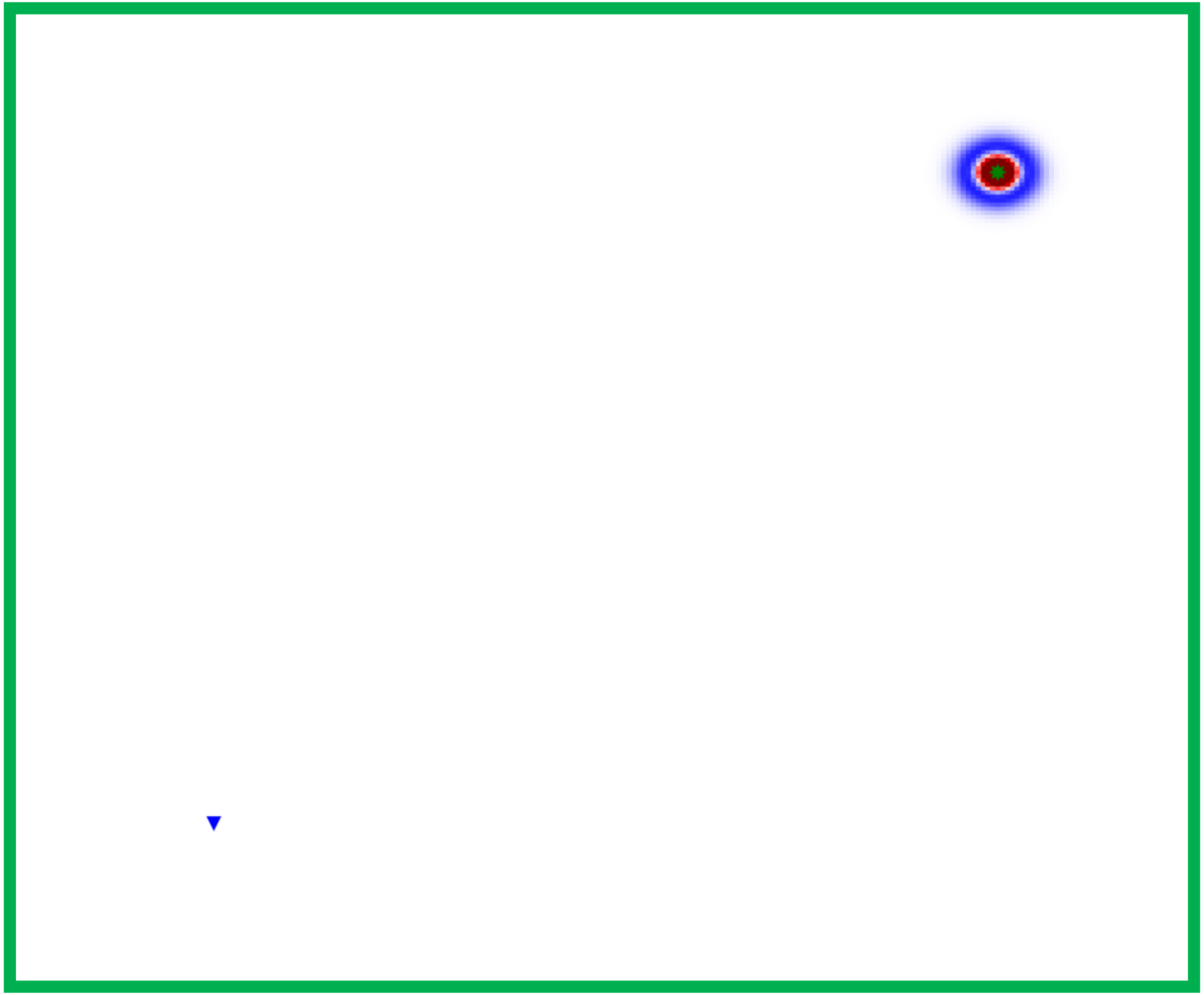
Full-waveform inversion: why is it possible?

$$L = \phi(R\boldsymbol{u}, \boldsymbol{d}) + (F(\boldsymbol{m})\boldsymbol{u} - S)^{\dagger}\boldsymbol{\lambda}$$

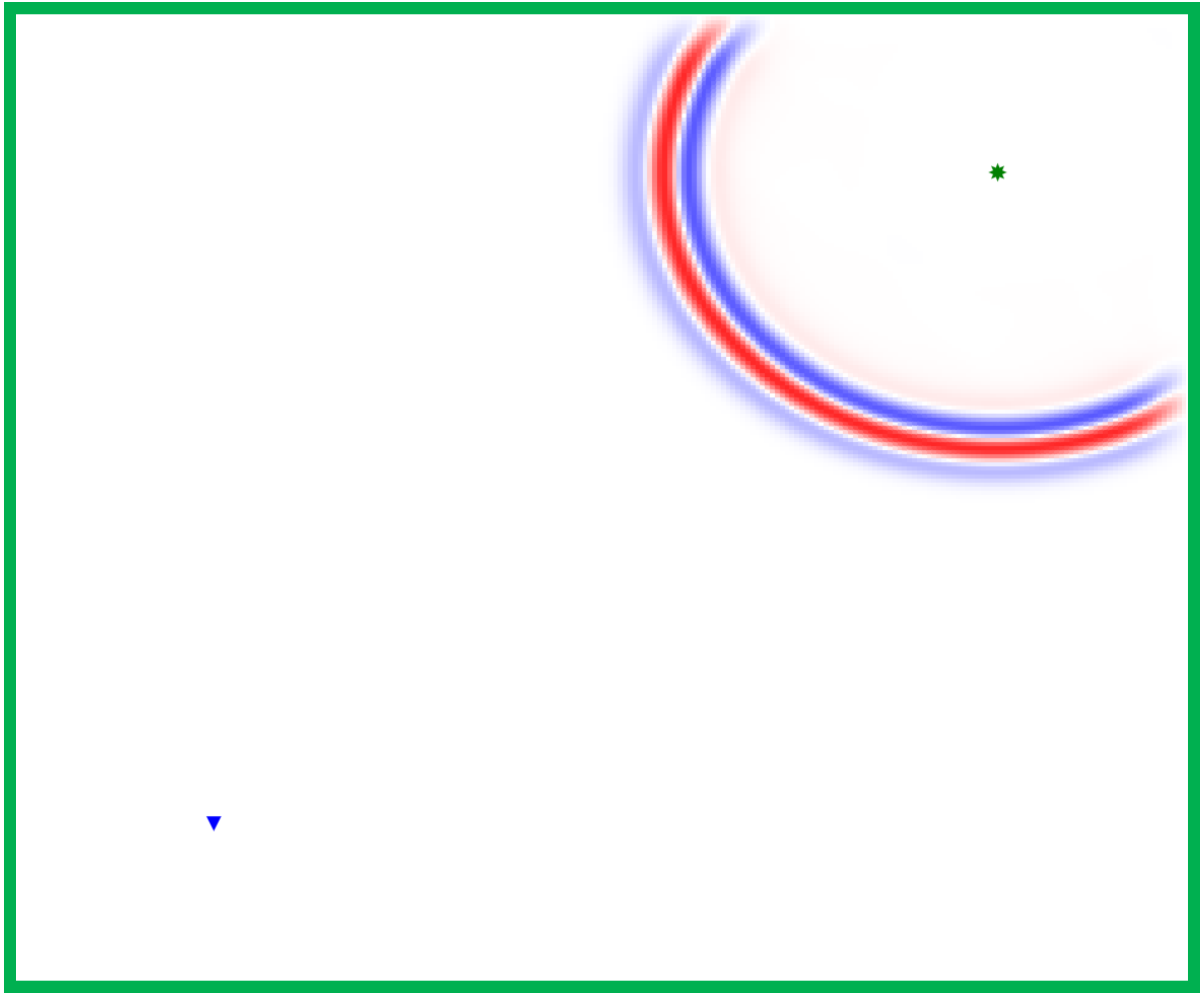
$$\frac{d\phi}{d\boldsymbol{m}} = \frac{dL}{d\boldsymbol{m}}(\boldsymbol{u}^{*}) = \frac{\partial L}{\partial m}(\boldsymbol{u}^{*}) + \mathbf{0}$$

$$\frac{d\phi}{d\boldsymbol{m}}(u^{*}) = \left(\frac{\partial F}{\partial \boldsymbol{m}}\boldsymbol{u}^{*}\right)^{\dagger} \boldsymbol{\lambda}^{*}$$

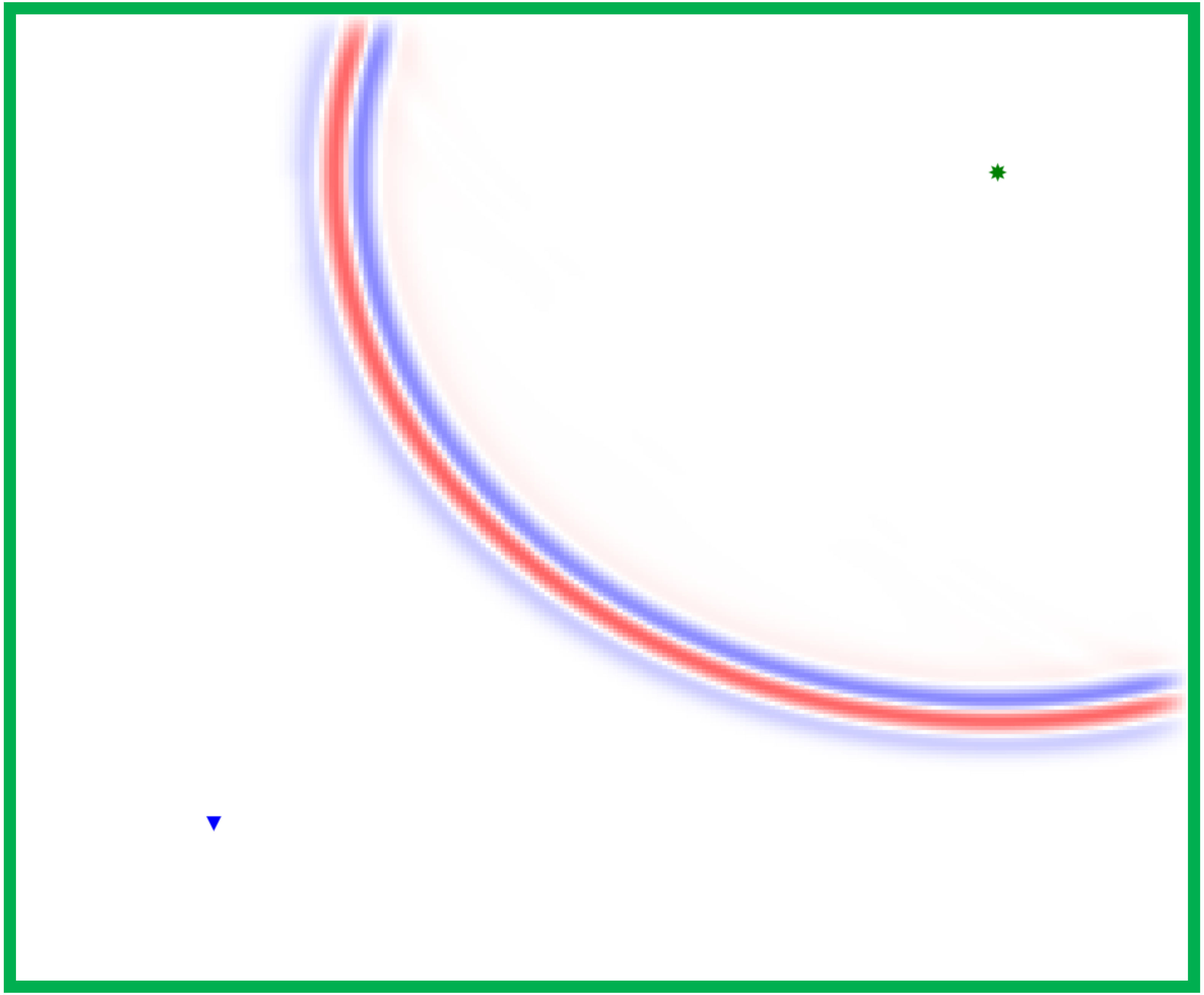
Forward



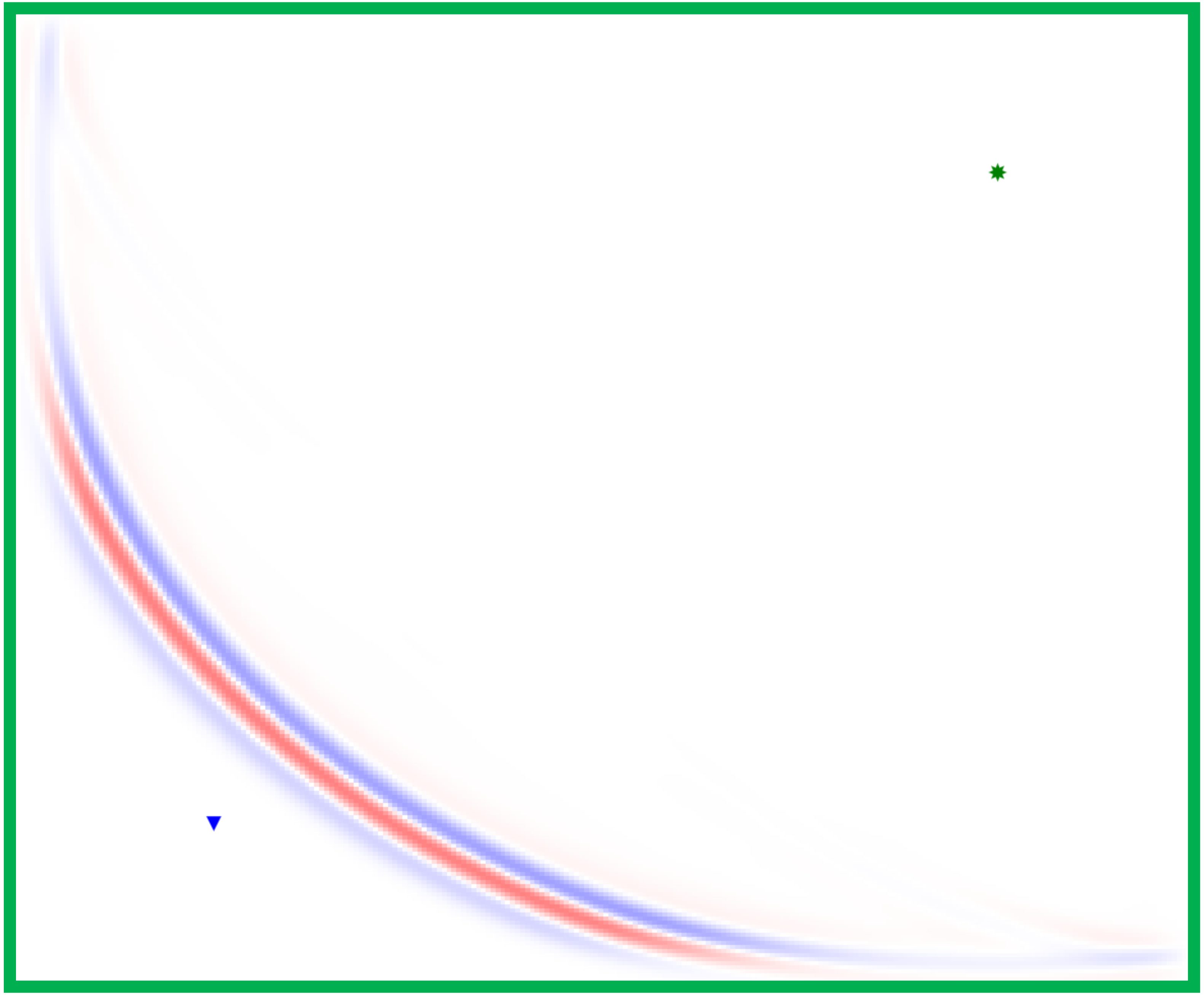
Forward



Forward

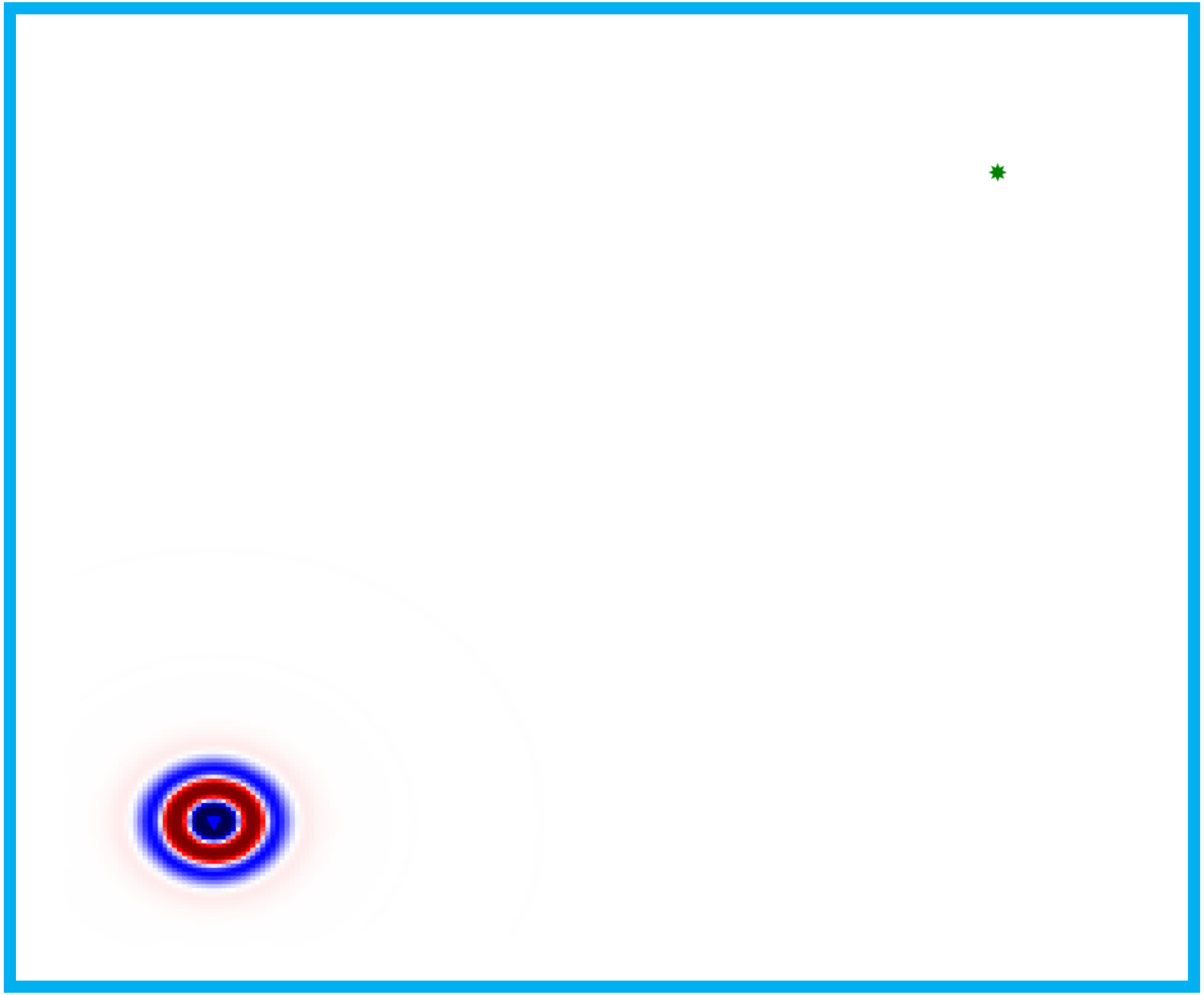


Forward

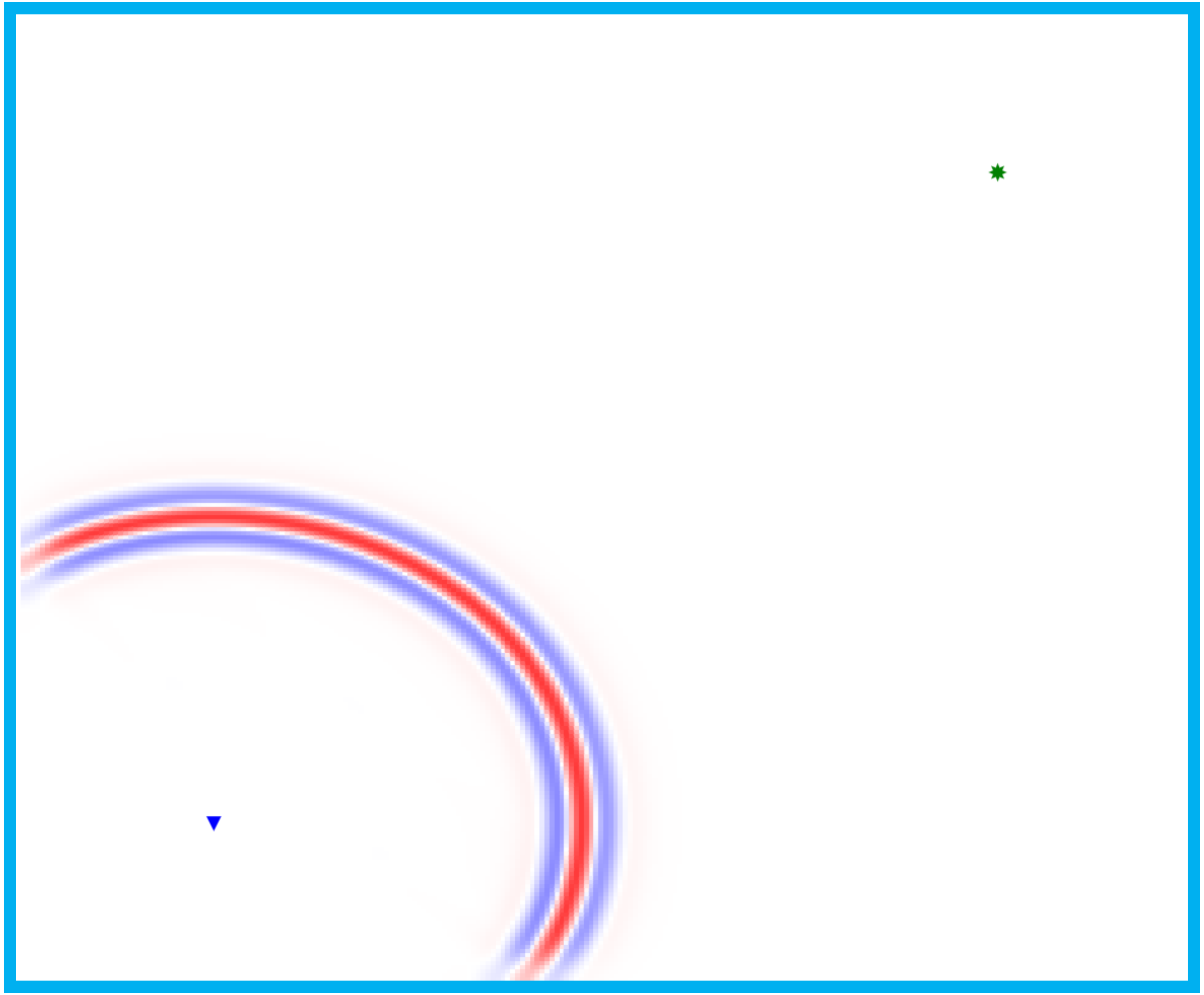


Forward

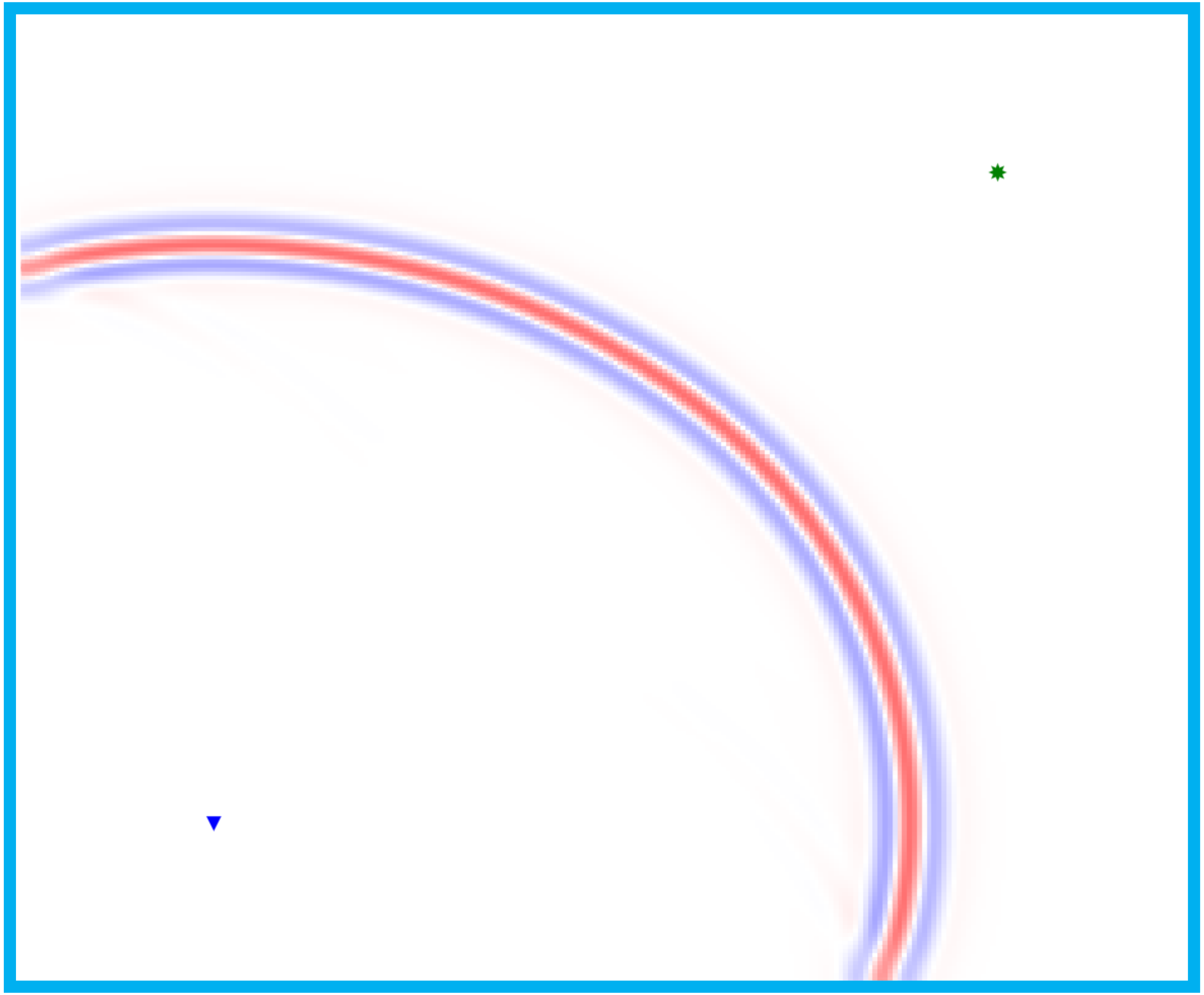
Adjoint



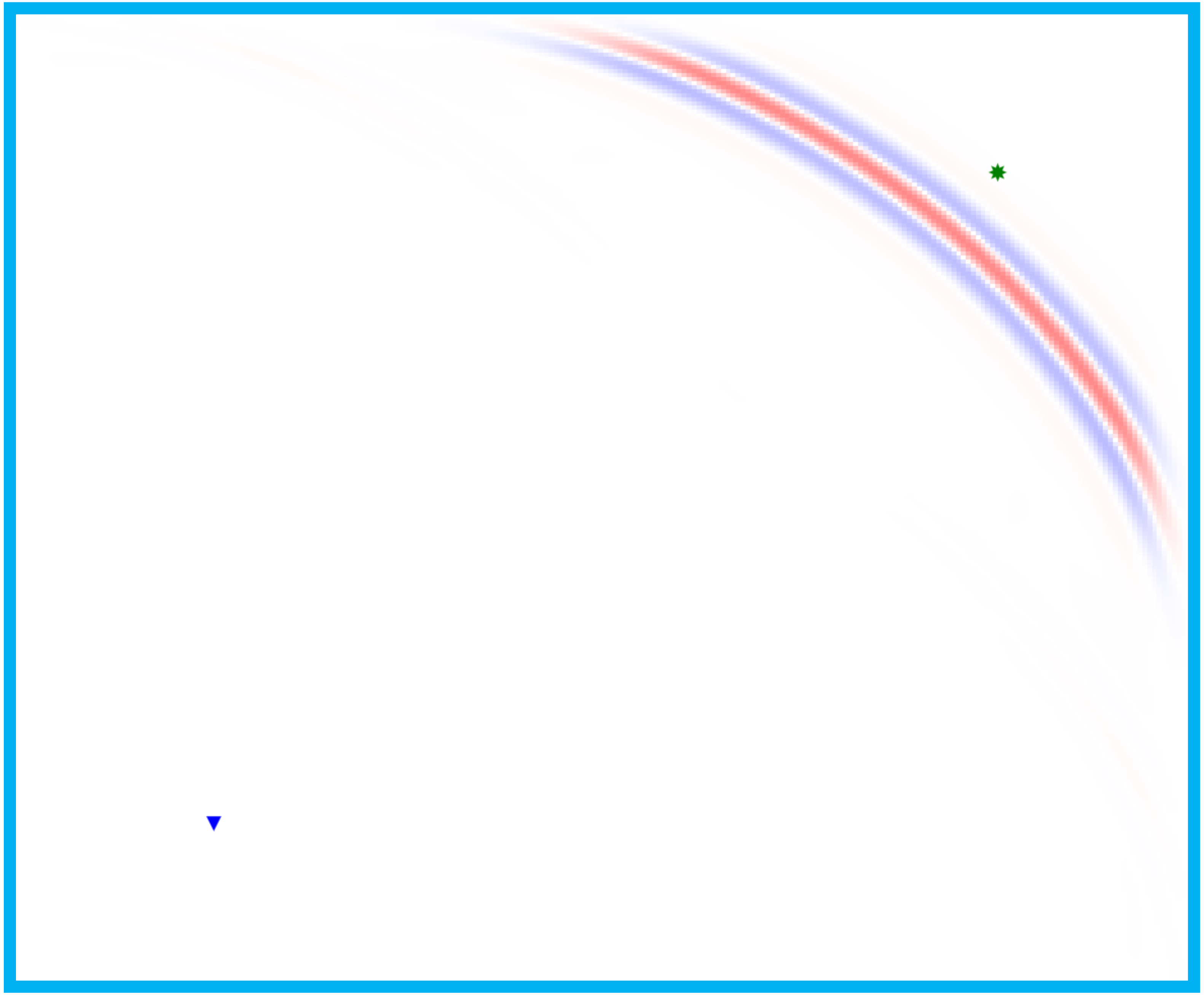
Adjoint



Adjoint



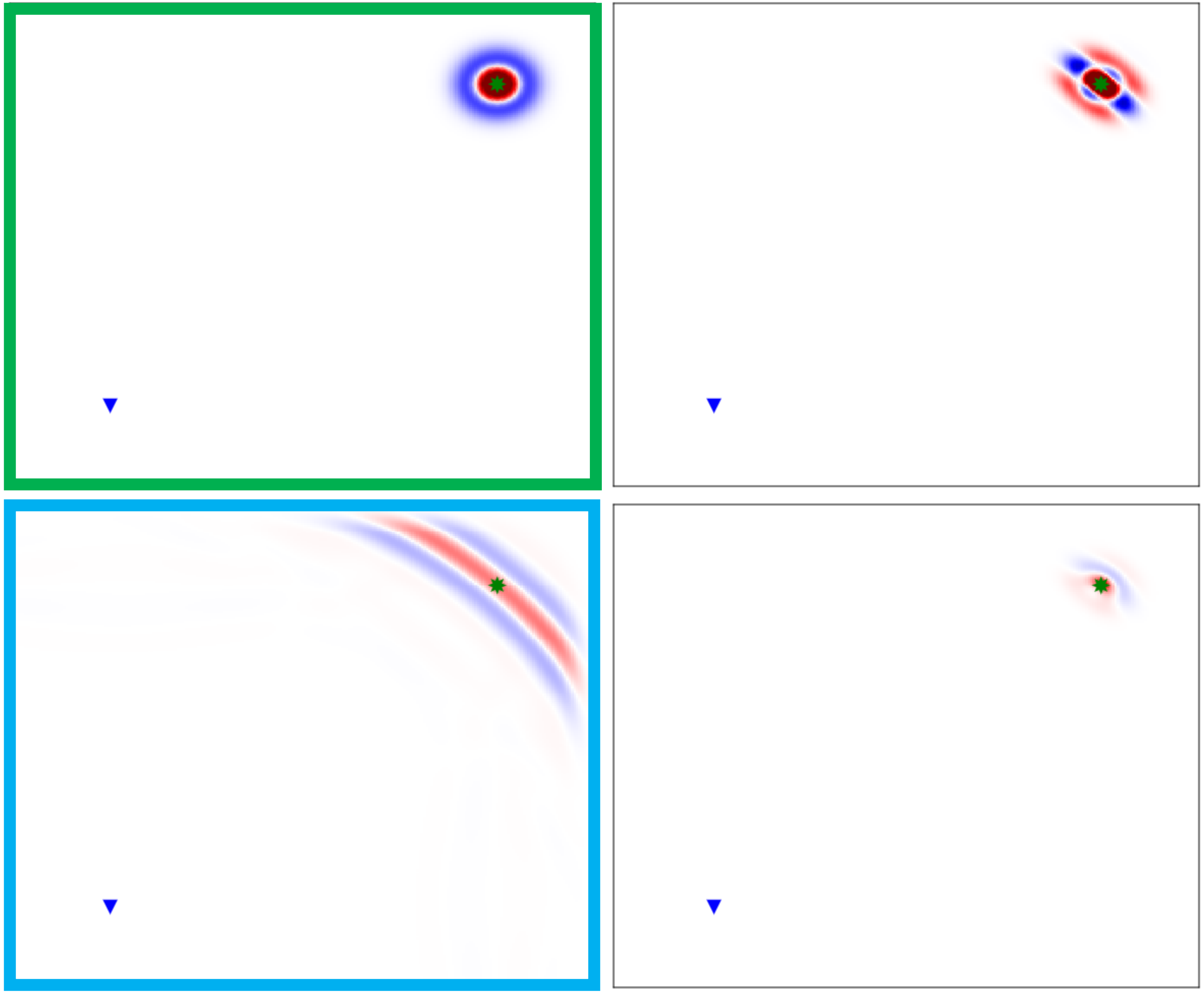
Adjoint



Adjoint

$u(t_F)$

$\lambda(T - t_F)$

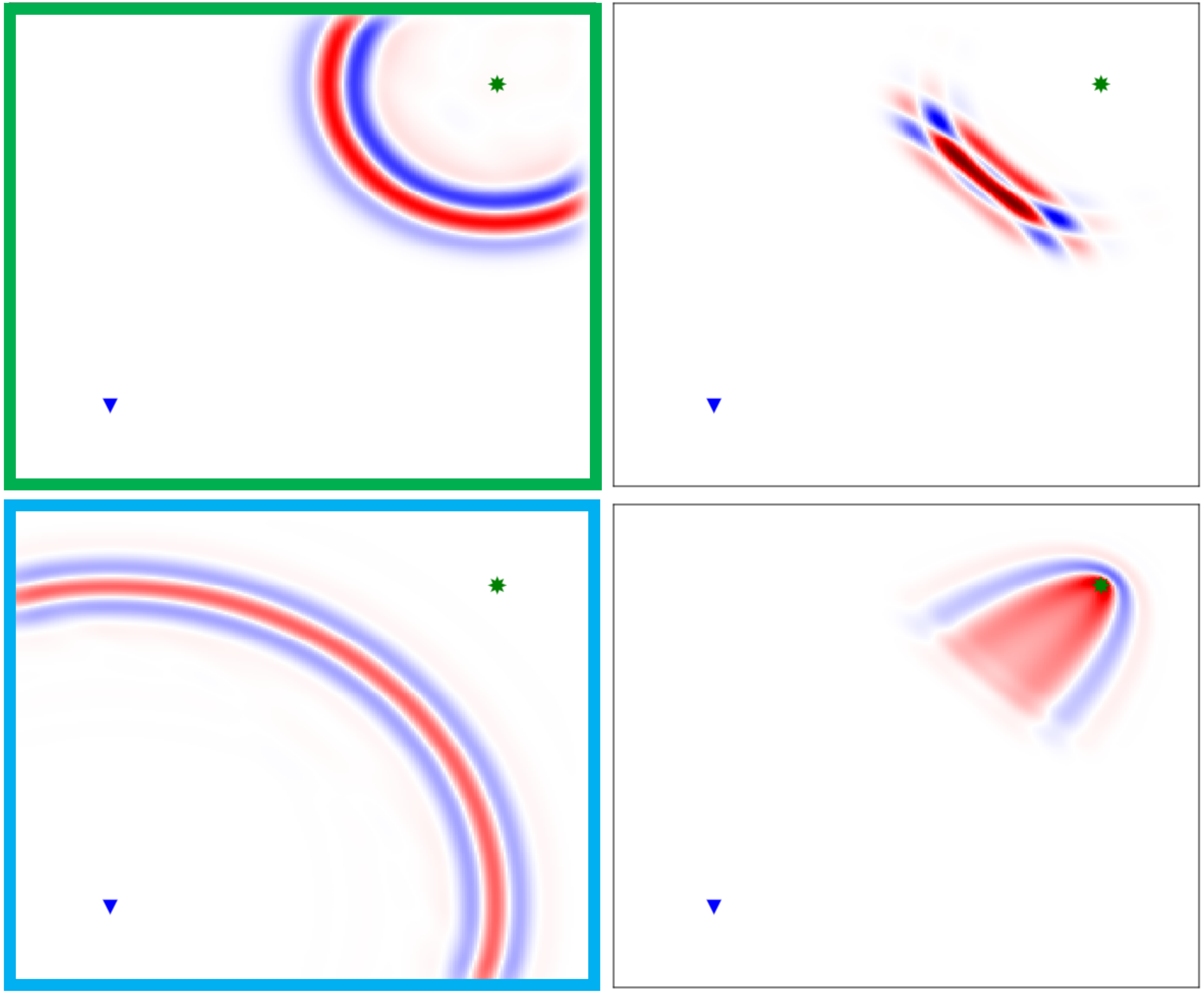


$u(t_F) * \lambda(T - t_F)$

$\sum_{n=1}^{t_F} u(t_n) * \lambda(T - t_n)$

$u(t_F)$

$\lambda(T - t_F)$

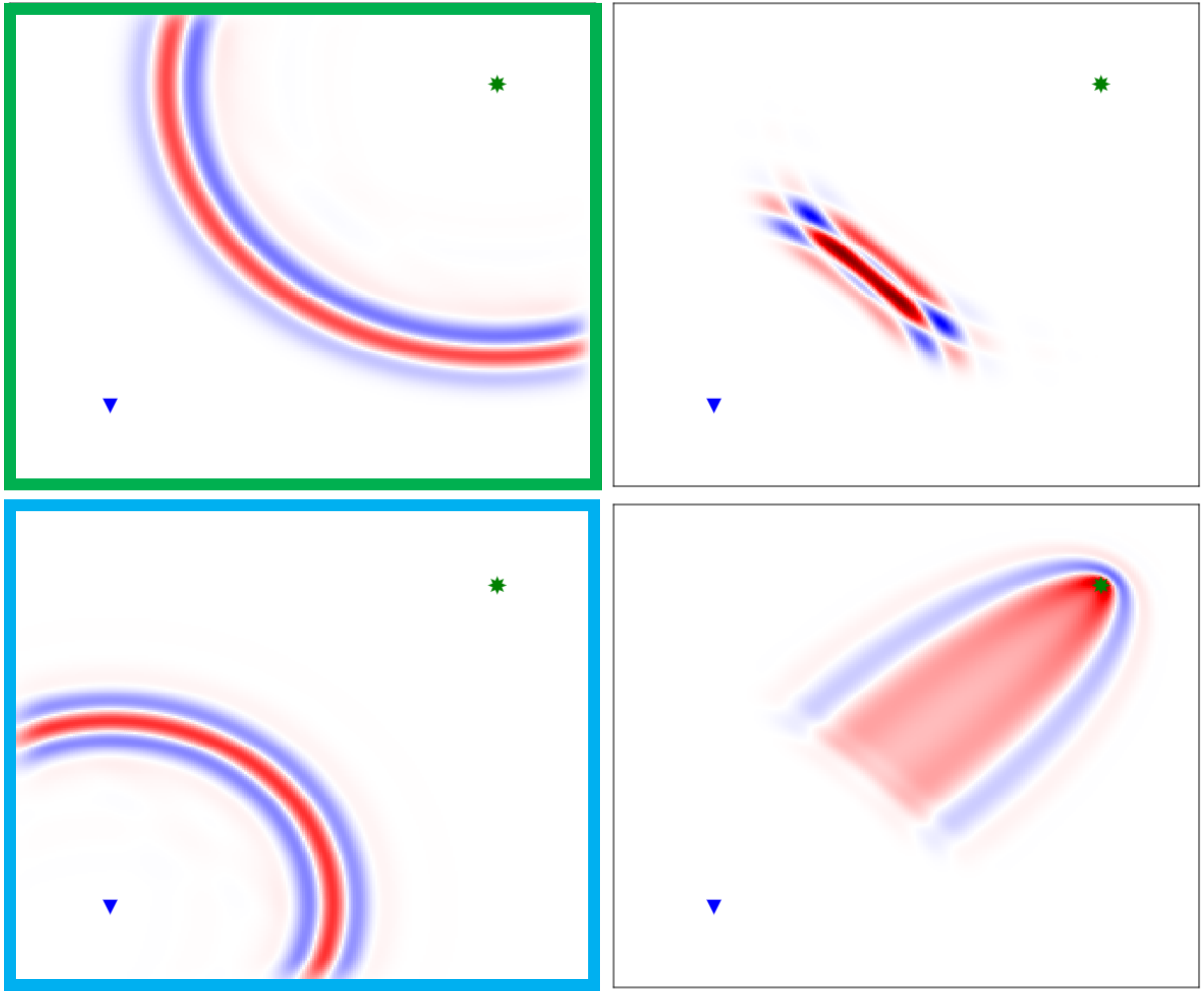


$u(t_F) * \lambda(T - t_F)$

$\sum_{n=1}^{t_F} u(t_n) * \lambda(T - t_n)$

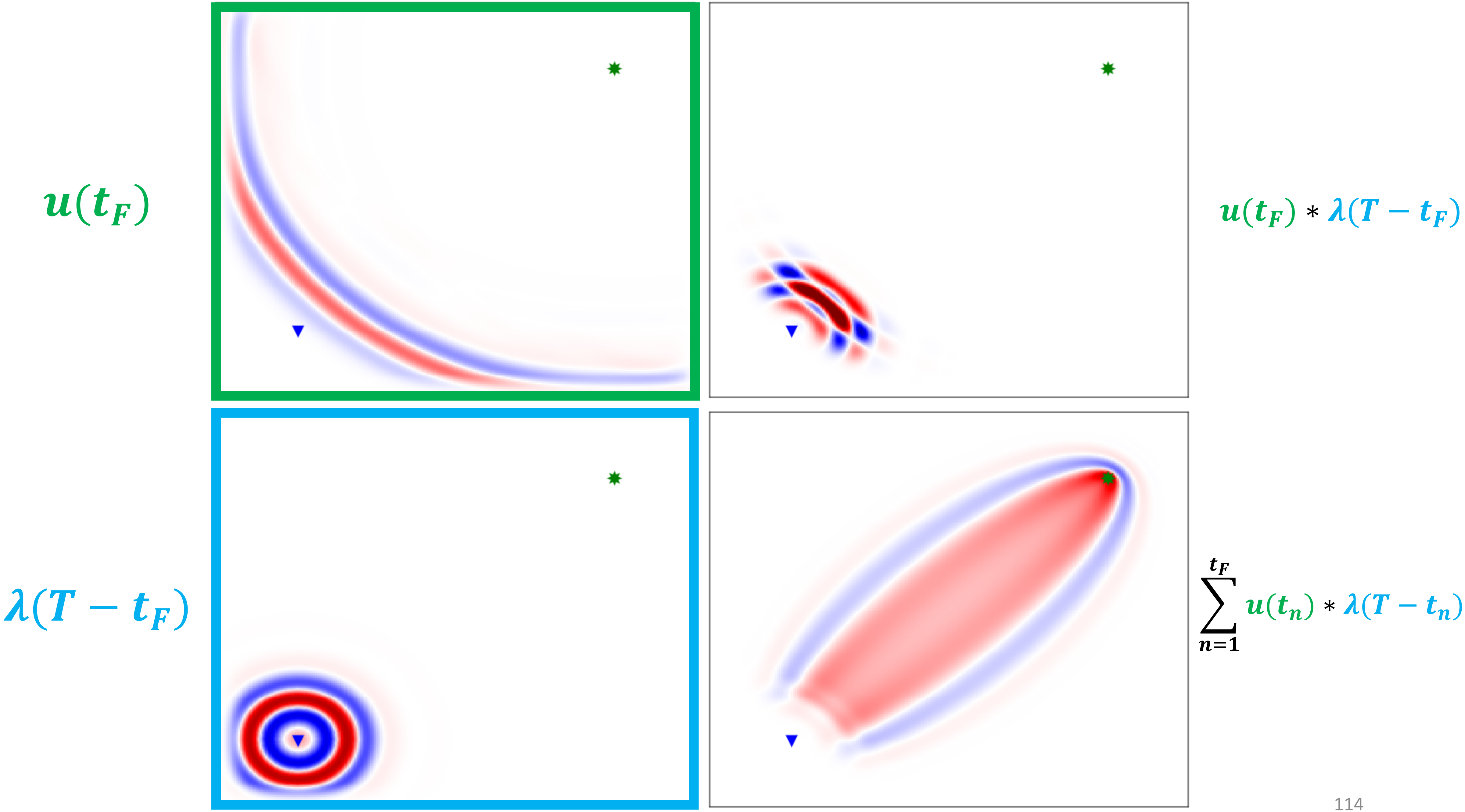
$u(t_F)$

$u(t_F) * \lambda(T - t_F)$



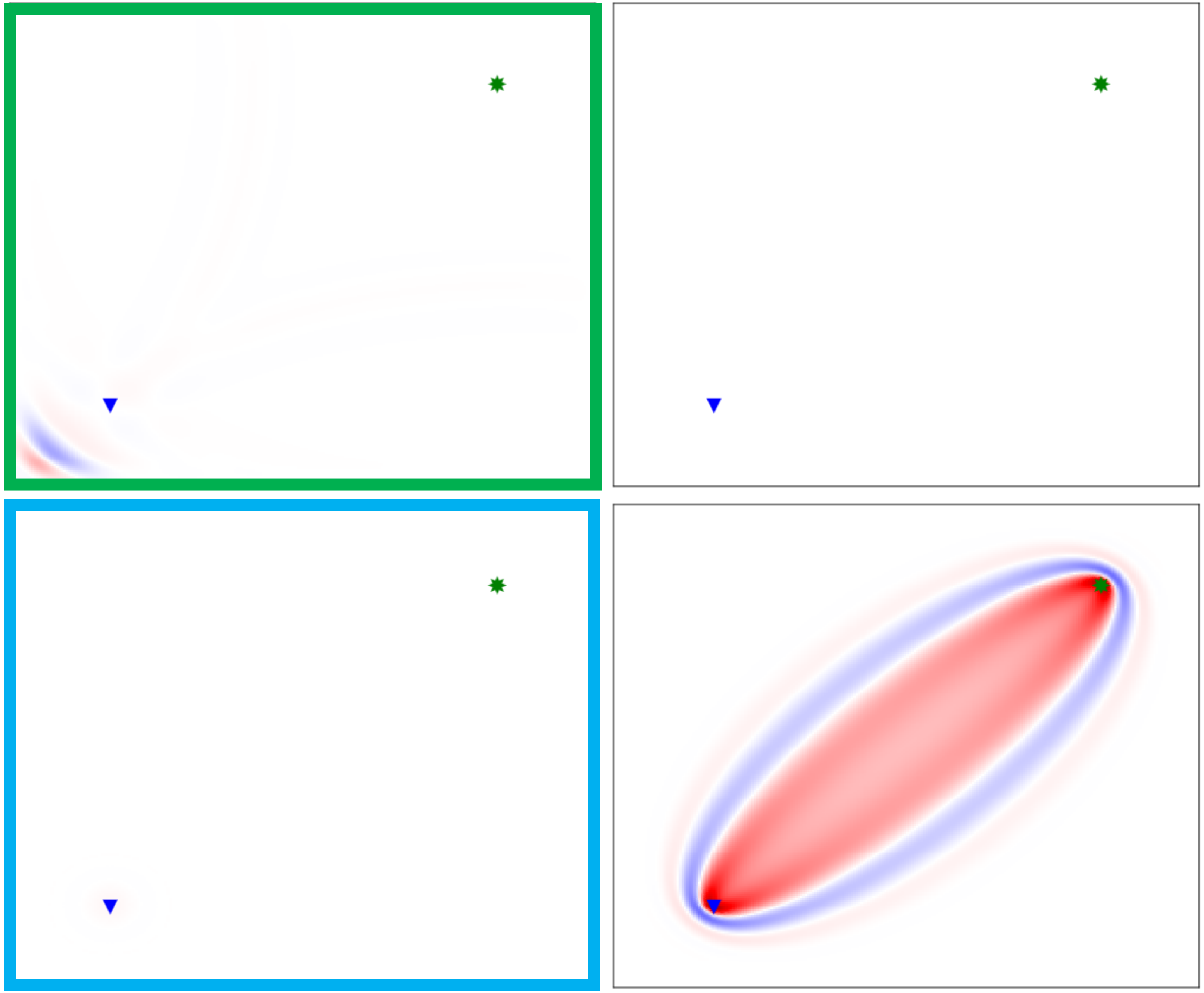
$\lambda(T - t_F)$

$\sum_{n=1}^{t_F} u(t_n) * \lambda(T - t_n)$


$u(t_F)$
$u(t_F) * \lambda(T - t_F)$
$\lambda(T - t_F)$
$\sum_{n=1}^{t_F} u(t_n) * \lambda(T - t_n)$

$$u(t_F)$$

$$u(t_F) * \lambda(T - t_F)$$

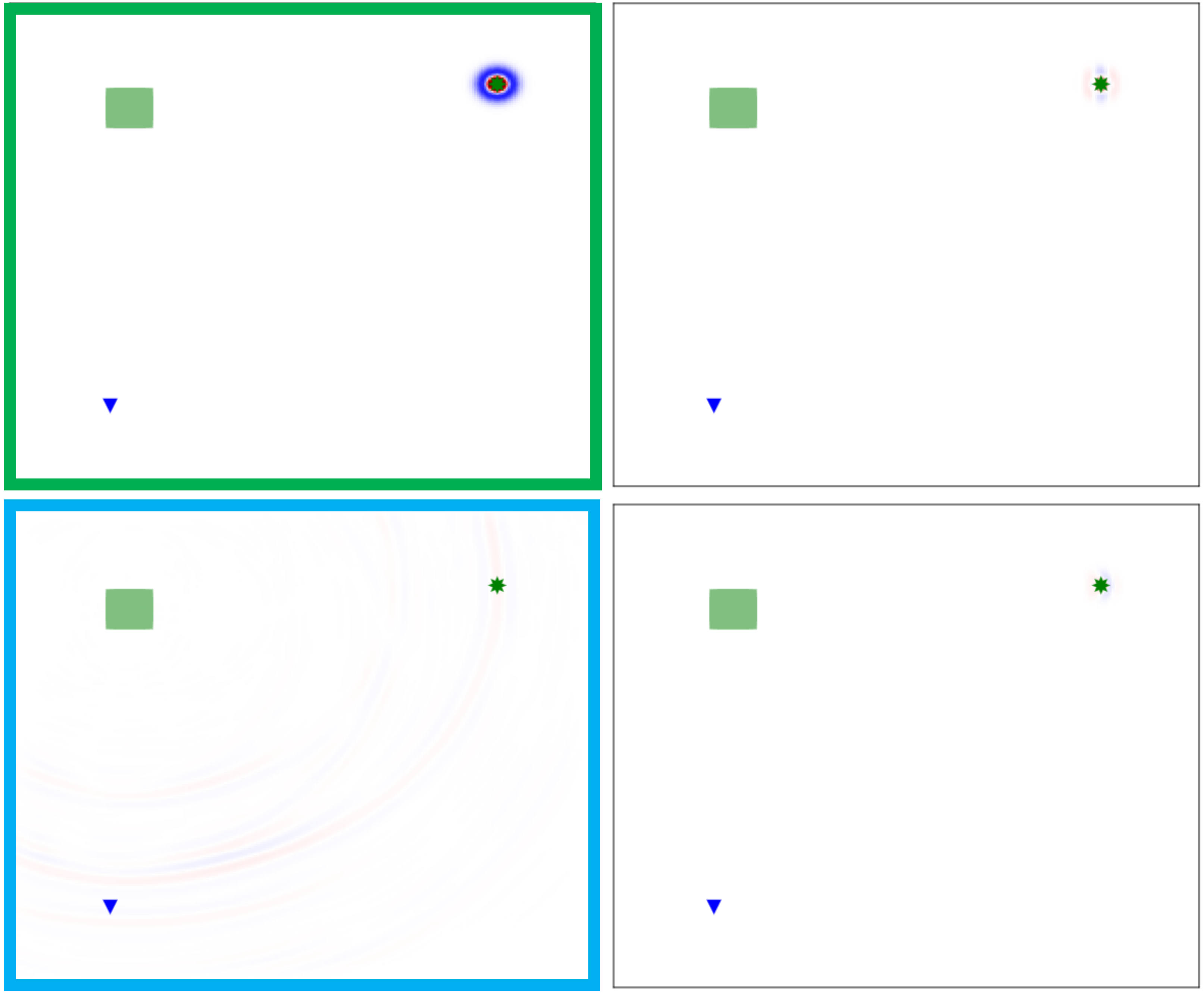


$$\lambda(T - t_F)$$

$$\sum_{n=1}^{t_F} u(t_n) * \lambda(T - t_n)$$

$u(t_F)$

$\lambda(T - t_F)$

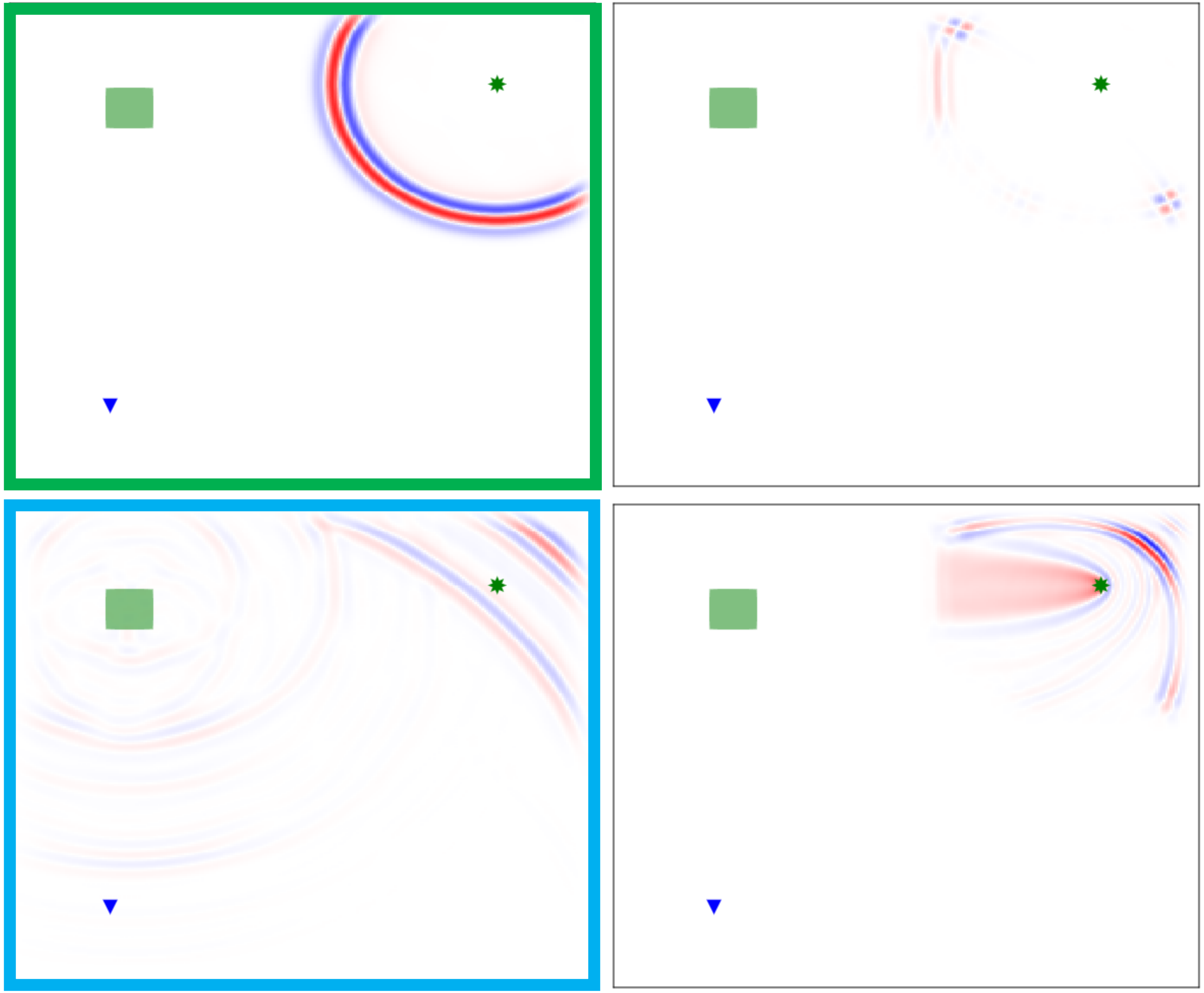


$u(t_F) * \lambda(T - t_F)$

$\sum_{n=1}^{t_F} u(t_n) * \lambda(T - t_n)$

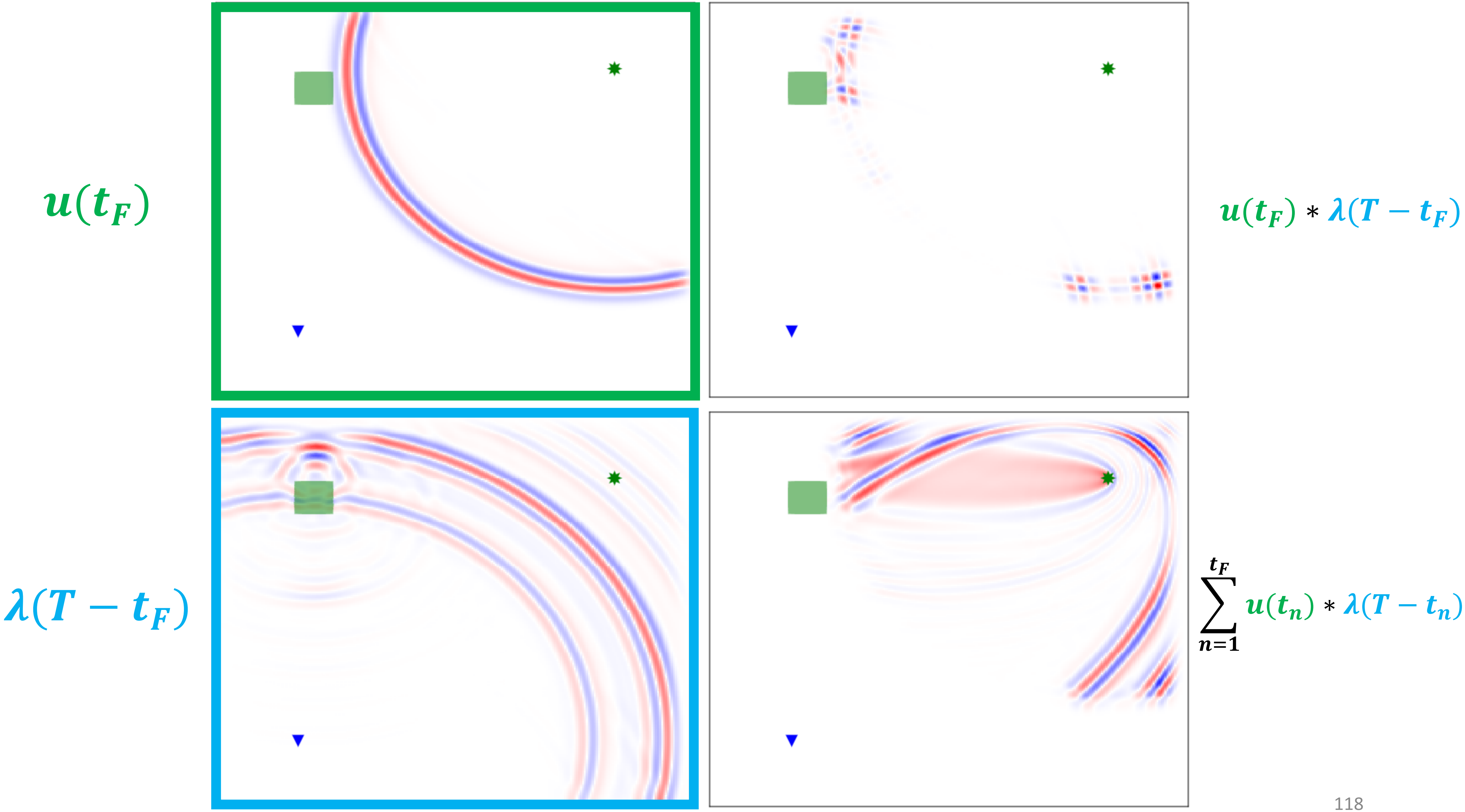
$u(t_F)$

$u(t_F) * \lambda(T - t_F)$

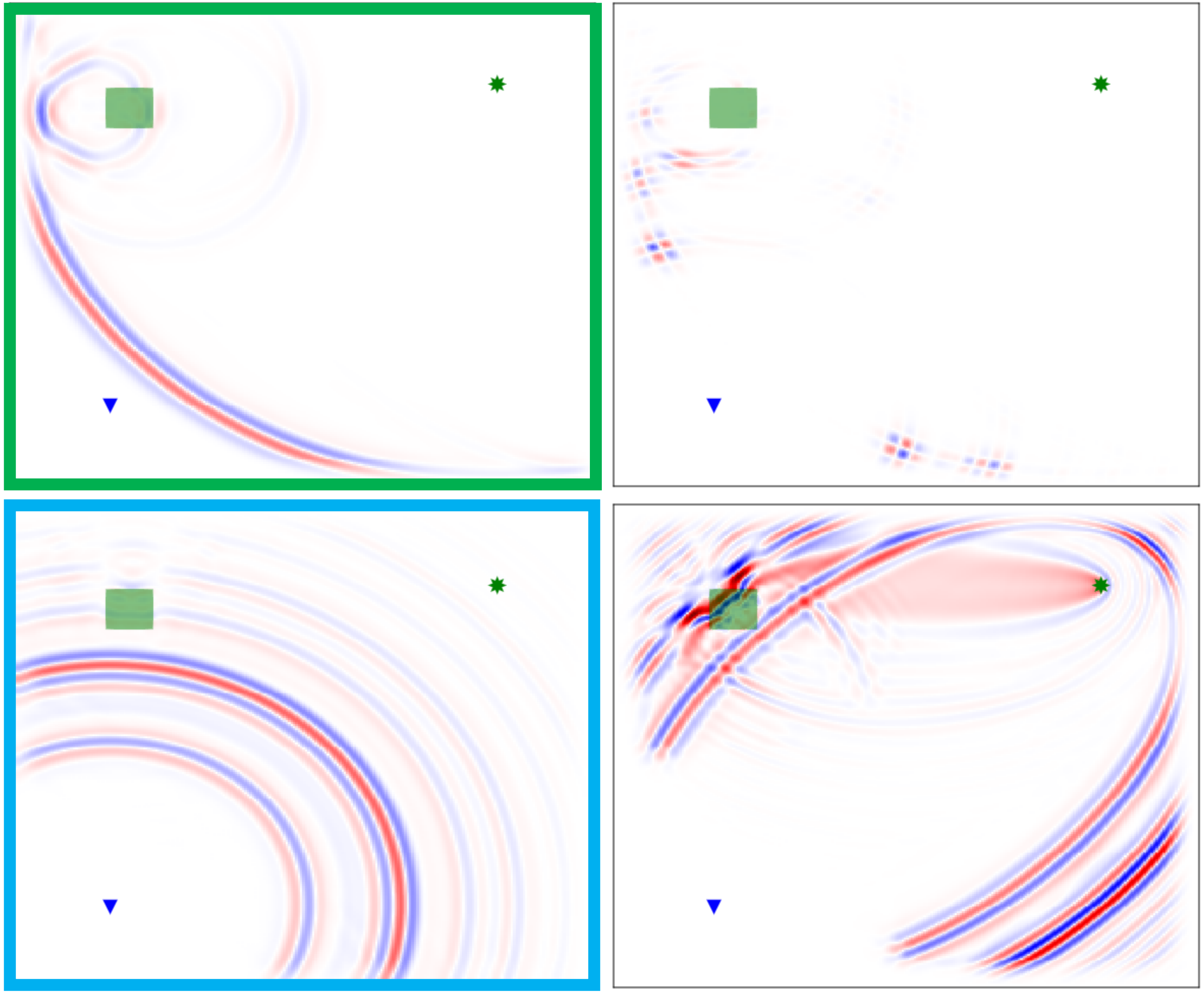


$\lambda(T - t_F)$

$\sum_{n=1}^{t_F} u(t_n) * \lambda(T - t_n)$


$u(t_F)$
$u(t_F) * \lambda(T - t_F)$
$\lambda(T - t_F)$
$\sum_{n=1}^{t_F} u(t_n) * \lambda(T - t_n)$

$u(t_F)$

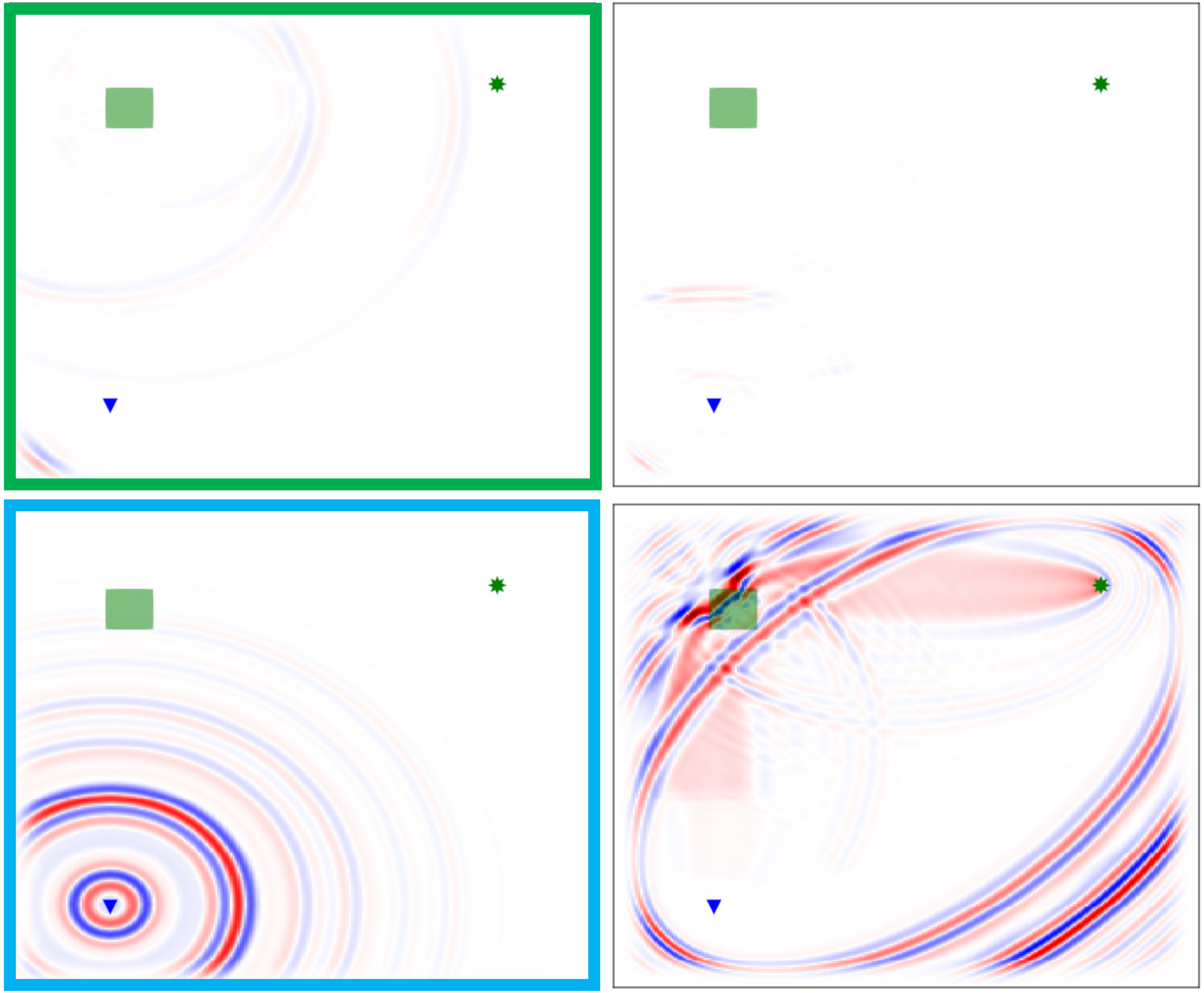


$u(t_F) * \lambda(T - t_F)$

$\lambda(T - t_F)$

$\sum_{n=1}^{t_F} u(t_n) * \lambda(T - t_n)$

$$u(t_F)$$



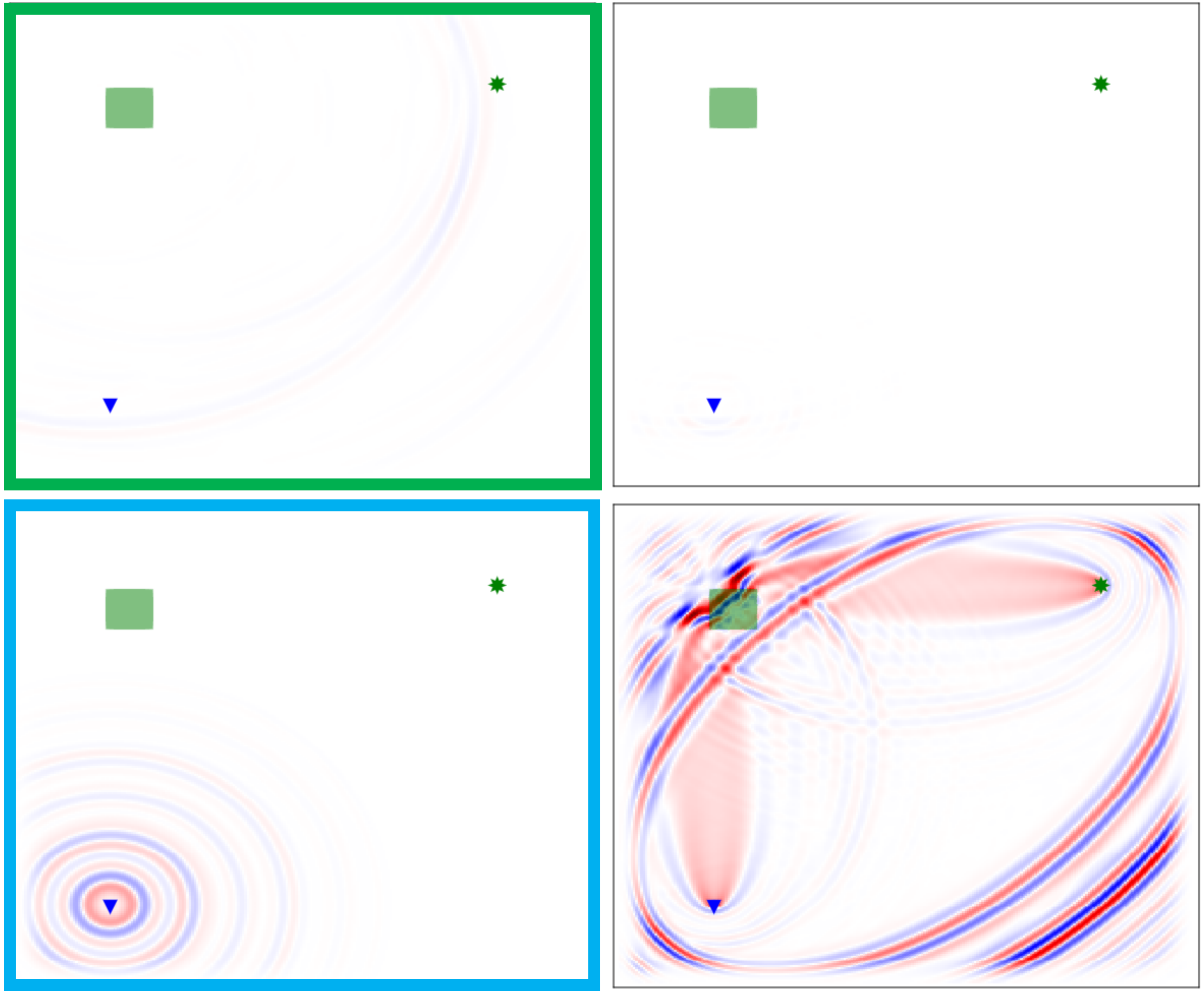
$$u(t_F) * \lambda(T - t_F)$$

$$\lambda(T - t_F)$$

$$\sum_{n=1}^{t_F} u(t_n) * \lambda(T - t_n)$$

$u(t_F)$

$u(t_F) * \lambda(T - t_F)$

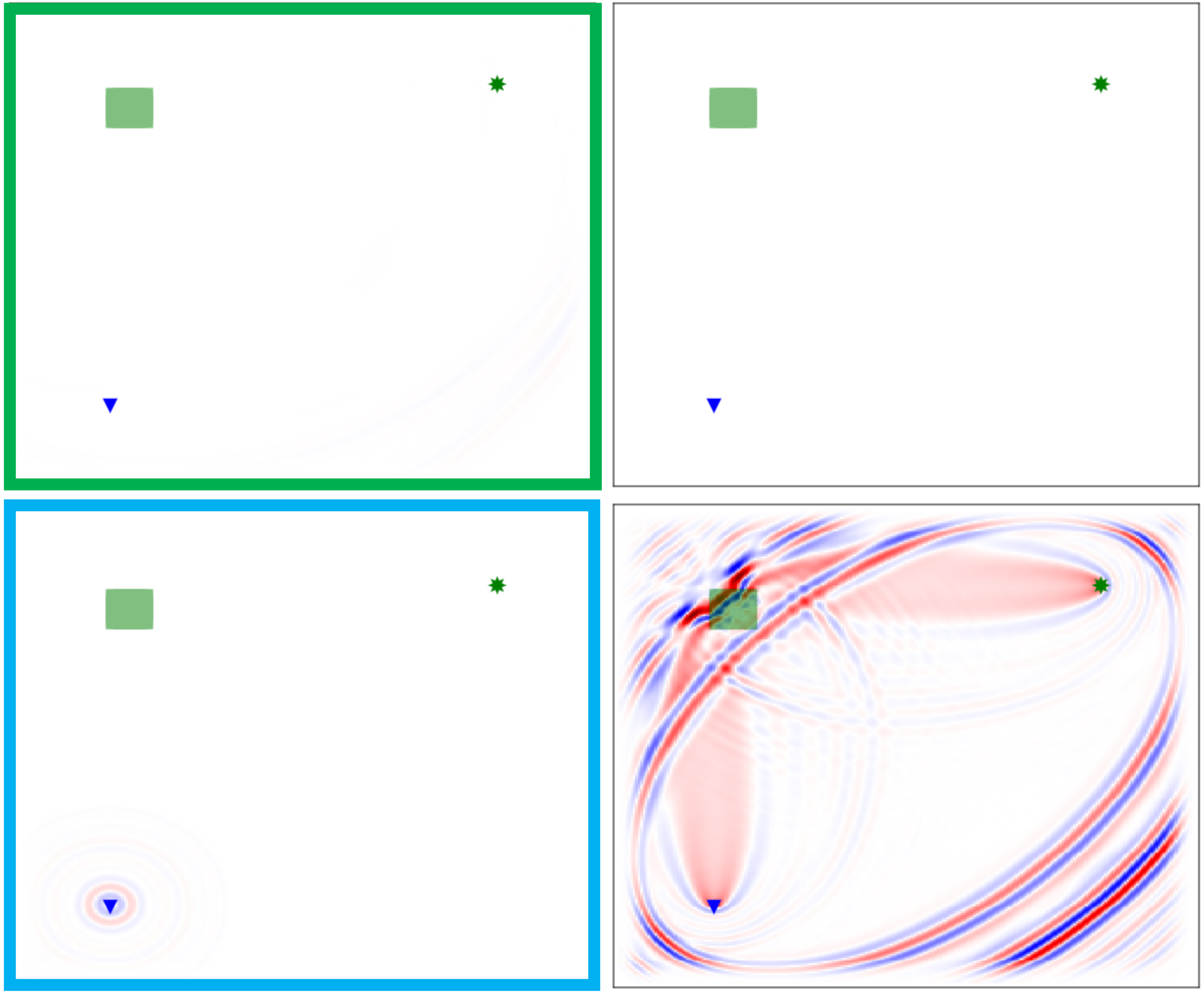


$\lambda(T - t_F)$

$\sum_{n=1}^{t_F} u(t_n) * \lambda(T - t_n)$

$u(t_F)$

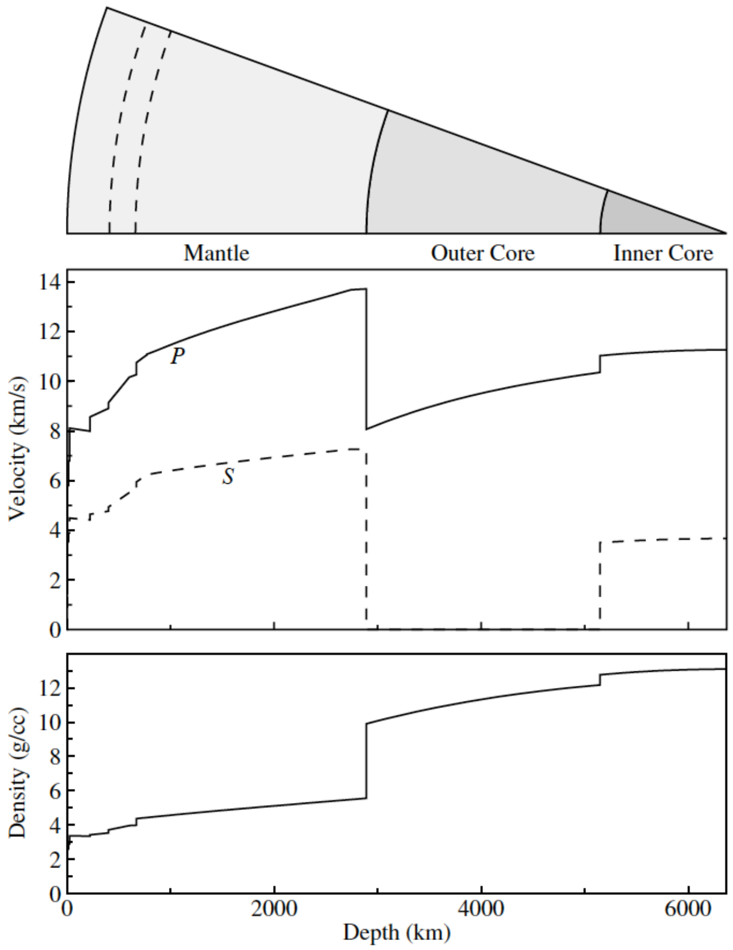
$u(t_F) * \lambda(T - t_F)$



$\lambda(T - t_F)$

$\sum_{n=1}^{t_F} u(t_n) * \lambda(T - t_n)$

Prior information



Earth's 1D elastic structure is known to a very good degree of approximation

Seismic inversion is used to recover the changes from this trend

These are relatively small in amplitude, but significant for structural insights

l-BFGS

$$g(m^* + \Delta m) = g(m^*) + H\Delta m + O(\Delta m^2)$$

To find a minimum, then

$$\Delta m \approx H^{-1} g$$

l-BFGS

Unfortunately, H has a dimension the square of the model, and isn’t accessible in real problems

l-BFGS

Unfortunately, H has a dimension the square of the model, and isn't accessible in real problems

We instead estimate H based on what we learn about ϕ and g during the inversion

l-BFGS

Unfortunately, H has a dimension the square of the model, and isn't accessible in real problems

We instead estimate H based on what we learn about ϕ and g during the inversion

Specifically, we try to find an estimate of the inverse of H, B such that

$B = B^T$ and $B\boldsymbol{\Delta g} = \boldsymbol{\Delta m}$

l-BFGS

Unfortunately, H has a dimension the square of the model, and isn't accessible in real problems

We instead estimate H based on what we learn about ϕ and g during the inversion

Specifically, we try to find an estimate of the inverse of H, B such that

$B = B^T$ and $B\mathbf{\Delta g} = \mathbf{\Delta m}$

Many B's satisfy this condition, so we add the condition

$$B = \underset{B}{\operatorname{argmin}} ||B - B_0||$$

l-BFGS

- Even the memory requirements of this reduced problem (one model and gradient per iteration) can be large enough to cause problems
- l-BFGS manages this by considering only a finite number of prior models and gradients

l-BFGS

0^{th} approximation of $\mathbf{H} \rightarrow \mathbf{H}_0=\mathbf{I}=\mathbf{M}$

l-BFGS

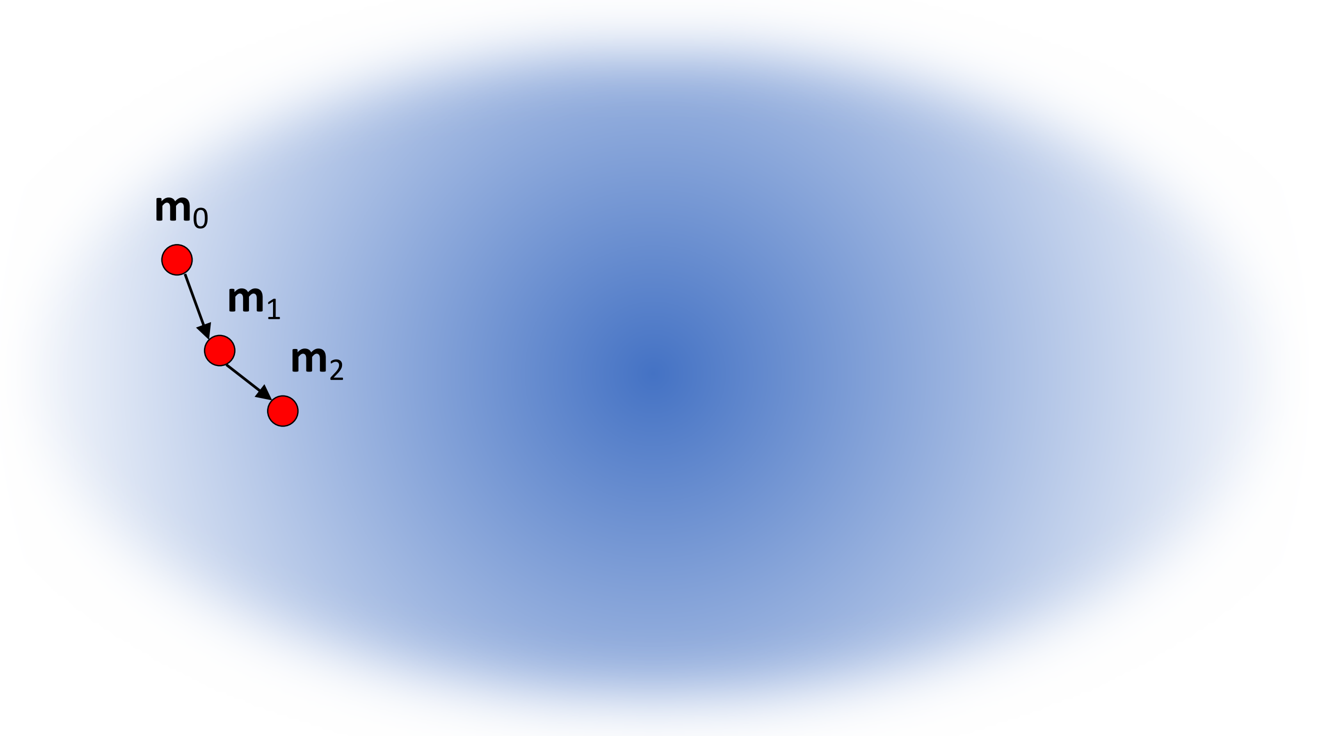
1^{st} approximation of $\mathbf{H} \rightarrow \mathbf{H}_1 = \mathbf{M}$

based on $\mathbf{m}_0$, $\mathbf{m}_1$

l-BFGS

2^{nd} approximation of **H** → $\mathbf{H}_2$=**M**

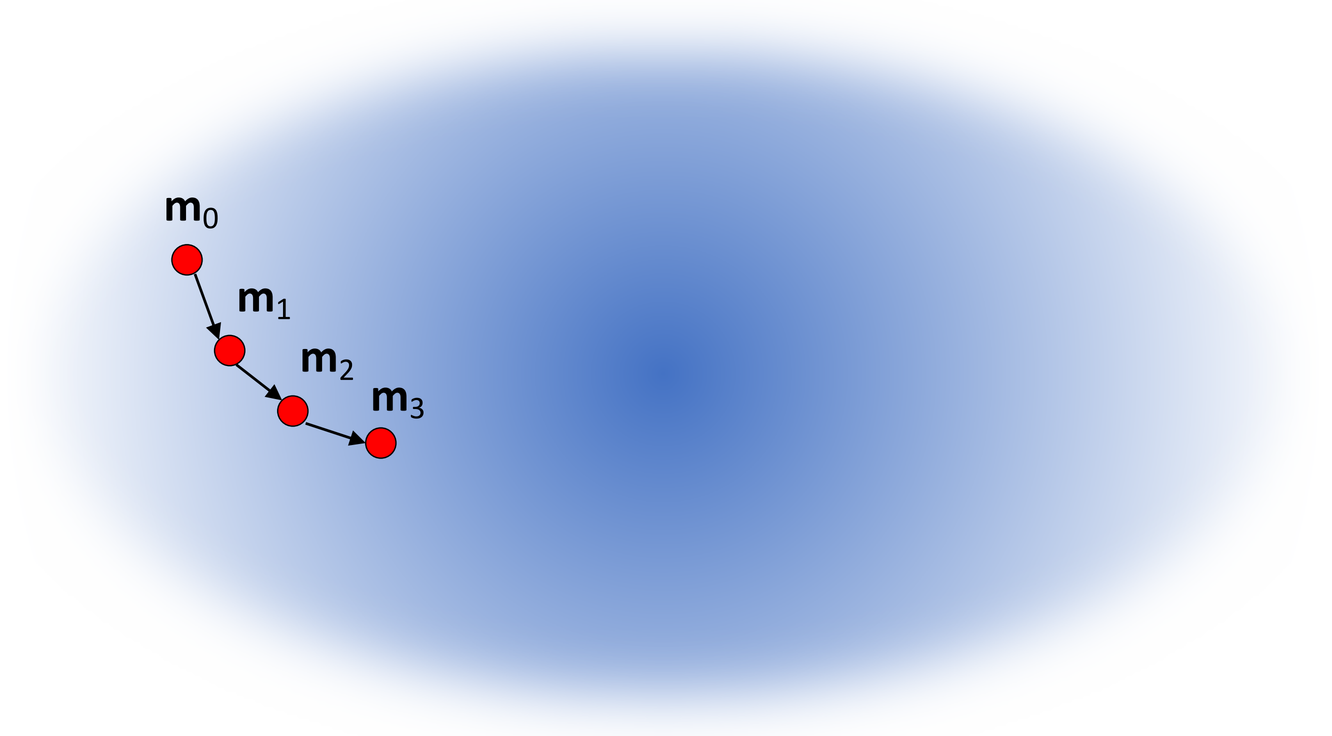
based on $\mathbf{m}_0$, $\mathbf{m}_1$, $\mathbf{m}_2$



l-BFGS

3^{rd} approximation of $\mathbf{H} \rightarrow \mathbf{H}_3 = \mathbf{M}$

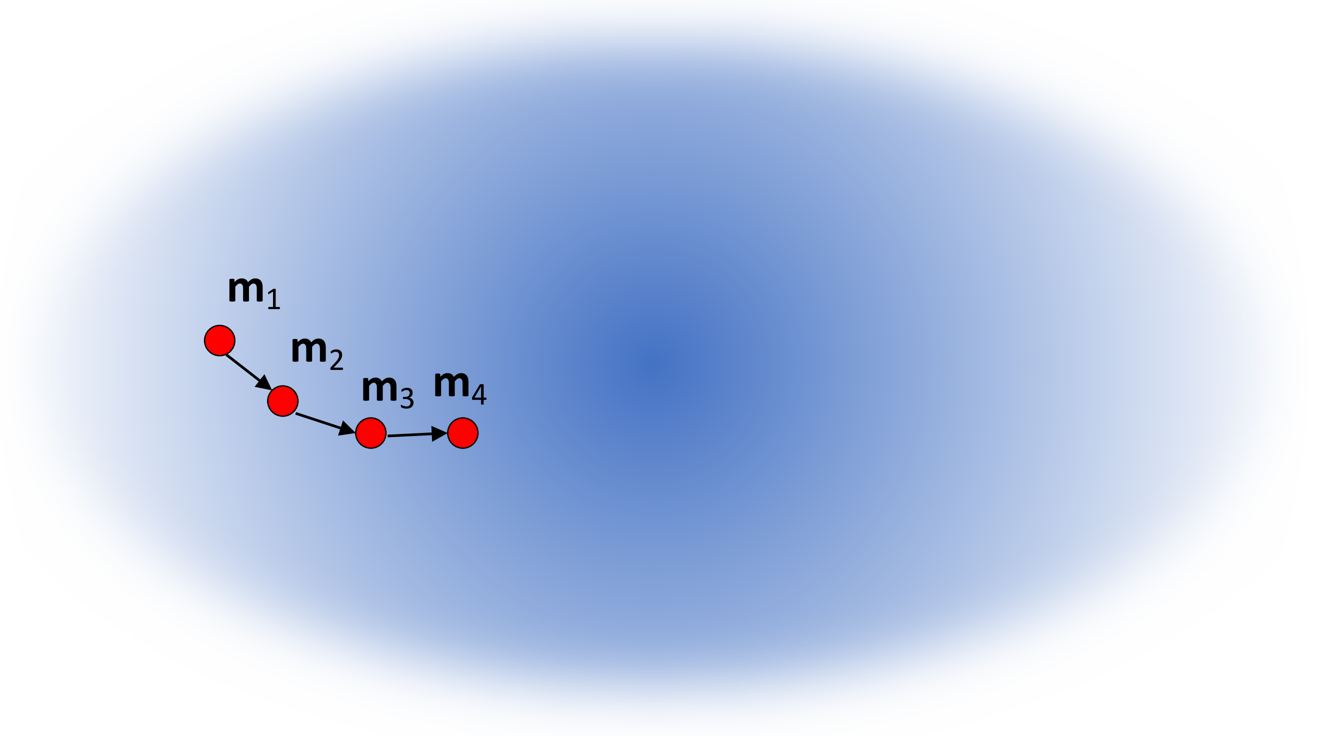
based on $\mathbf{m}_0$, $\mathbf{m}_1$, $\mathbf{m}_2$, $\mathbf{m}_3$



l-BFGS

4^{th} approximation of $\mathbf{H} \rightarrow \mathbf{H}_4 = \mathbf{M}$

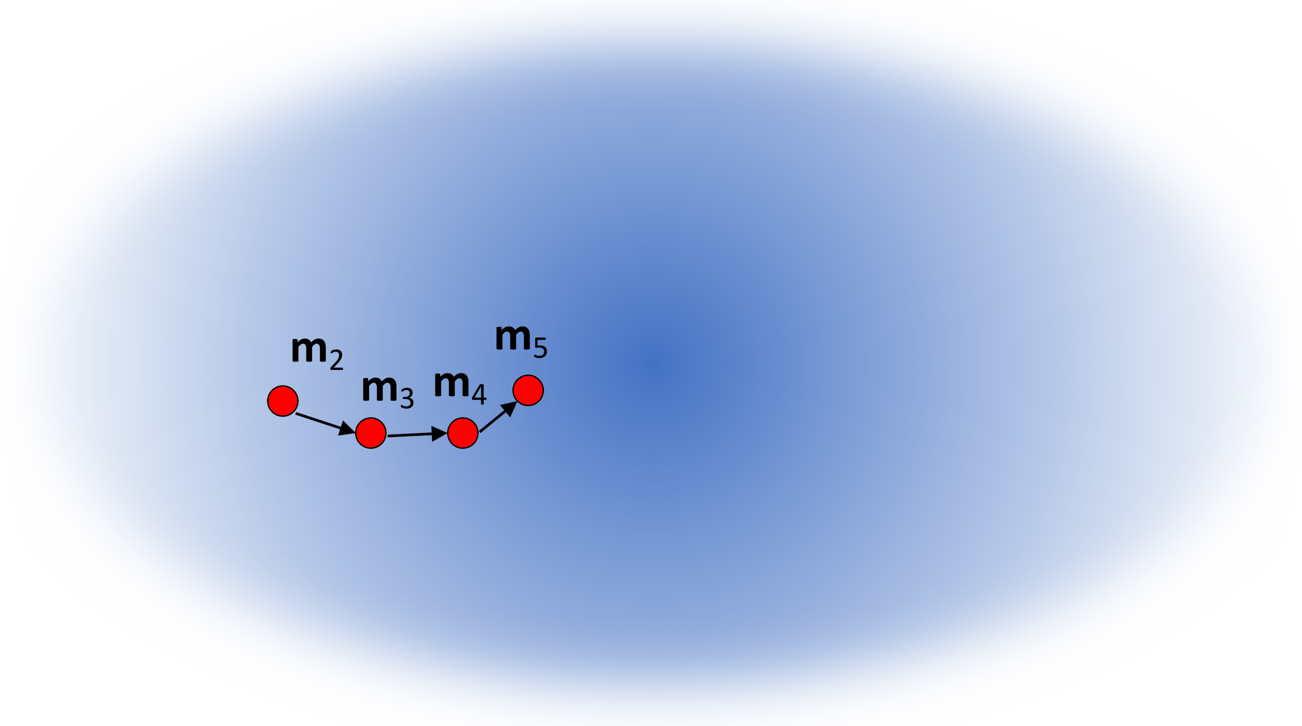
based on $\mathbf{m}_1$, $\mathbf{m}_2$, $\mathbf{m}_3$, $\mathbf{m}_4$



l-BFGS

5^{th} approximation of $\mathbf{H} \rightarrow \mathbf{H}_5 = \mathbf{M}$

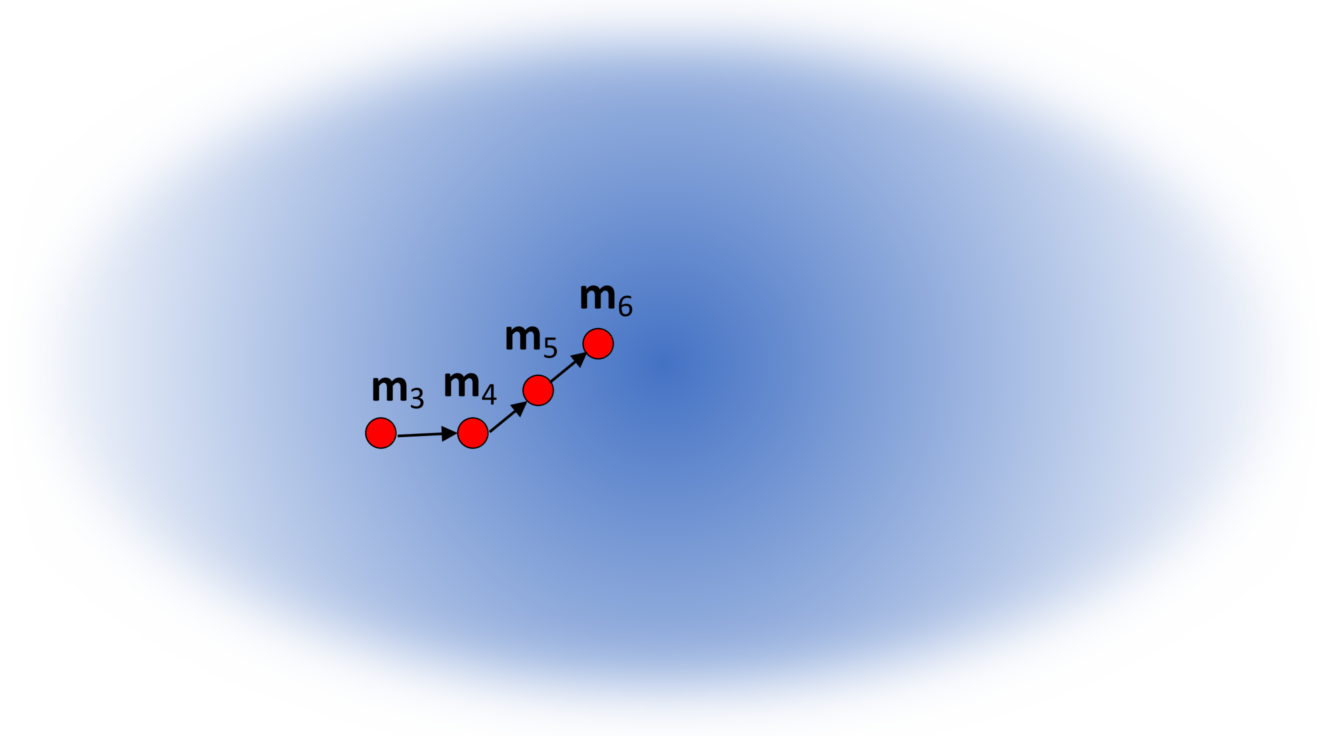
based on $\mathbf{m}_2$, $\mathbf{m}_3$, $\mathbf{m}_4$, $\mathbf{m}_5$



l-BFGS

6^{th} approximation of $\mathbf{H} \rightarrow \mathbf{H}_6 = \mathbf{M}$

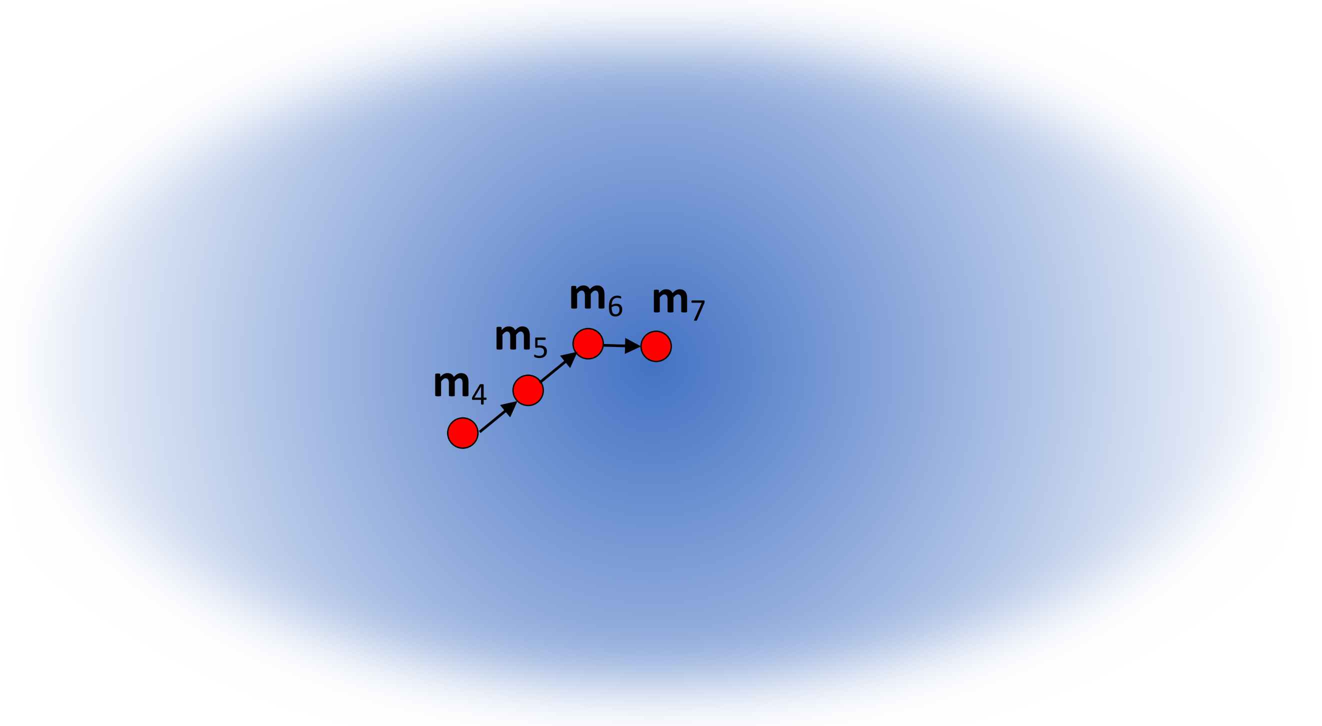
based on $\mathbf{m}_3$, $\mathbf{m}_4$, $\mathbf{m}_5$, $\mathbf{m}_6$



l-BFGS

7^{th} approximation of $\mathbf{H} \rightarrow \mathbf{H}_7 = \mathbf{M}$

based on $\mathbf{m}_4$, $\mathbf{m}_5$, $\mathbf{m}_6$, $\mathbf{m}_7$



Outline

1. Overview of full-waveform inversion
2. Numerical modelling of seismic waves
3. Objective function
4. Optimization
5. Reducing computational cost
6. Uncertainty quantification

Improving efficiency

1. Wavefield-adaptive meshes
2. Mini-batches
3. Source-stacking

Improving efficiency

General modelling of lots of seismic wavefields is expensive

What can we do to reduce cost?

Improving efficiency

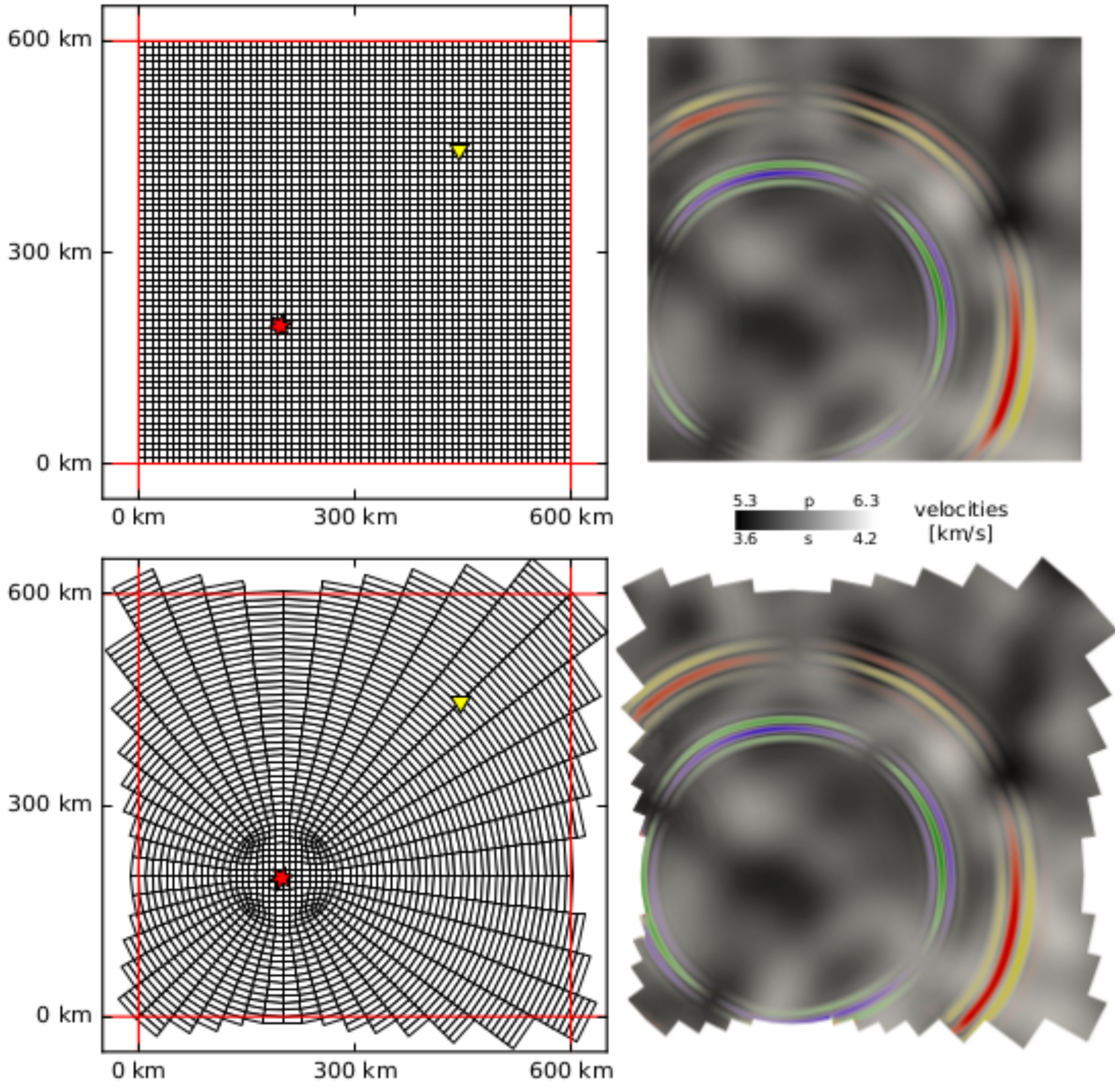
General modelling of lots of seismic wavefields is expensive

What can we do to reduce cost?

Broadly, we can

1. Model less generally
2. Model fewer wavefields

Wavefield-adaptive meshes



Regular finite-element meshes

- certain number of elements per minimum wavelength
- ensure reasonable numerical accuracy

Wavelength is anisotropic

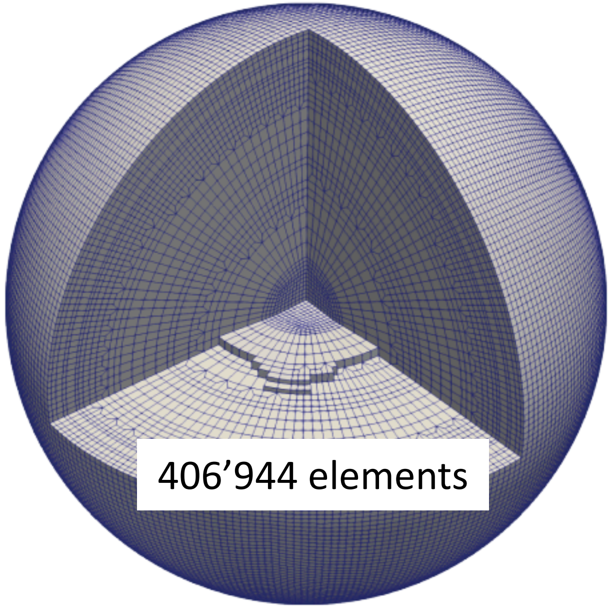
- Wavefield varies **rapidly parallel** to propagation direction.
- Wavefield varies **slowly perpendicular** to propagation direction.

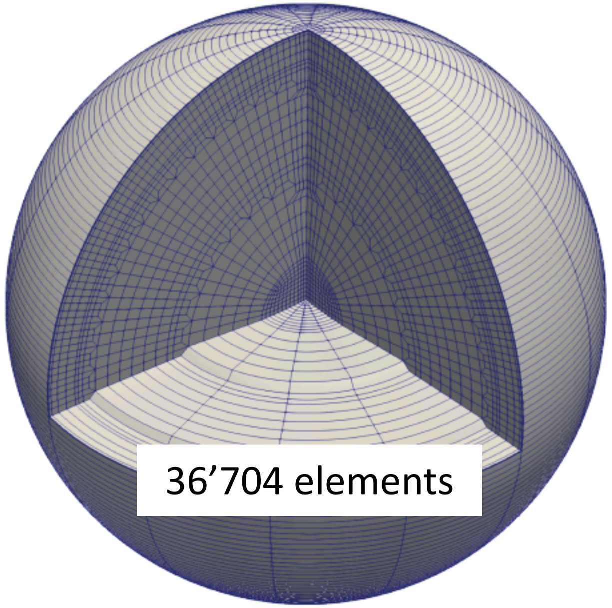
Reduce number of elements with anisotropic meshes

- **Complexity-adapted mesh**.
- ≈8 times less elements [1’250 vs. 10’000].
- Number of azimuthal elements can be adapted to medium complexity

Fichtner, pers. comm.

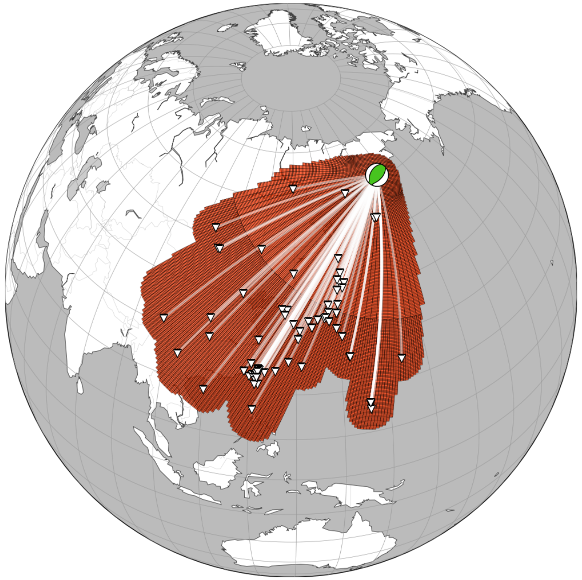
Wavefield-adaptive meshes

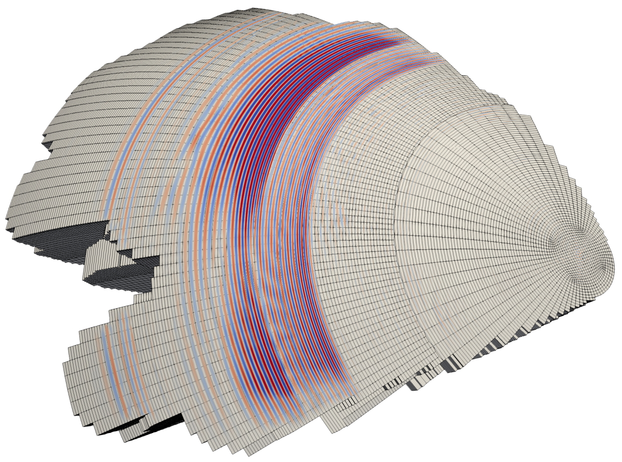




For complex media with high-amplitude perturbations, this approach must sacrifice either cost or efficiency

For relatively small amplitude wave-scattering, it is very efficient





Fichtner, pers. comm.

Mini-batch optimization

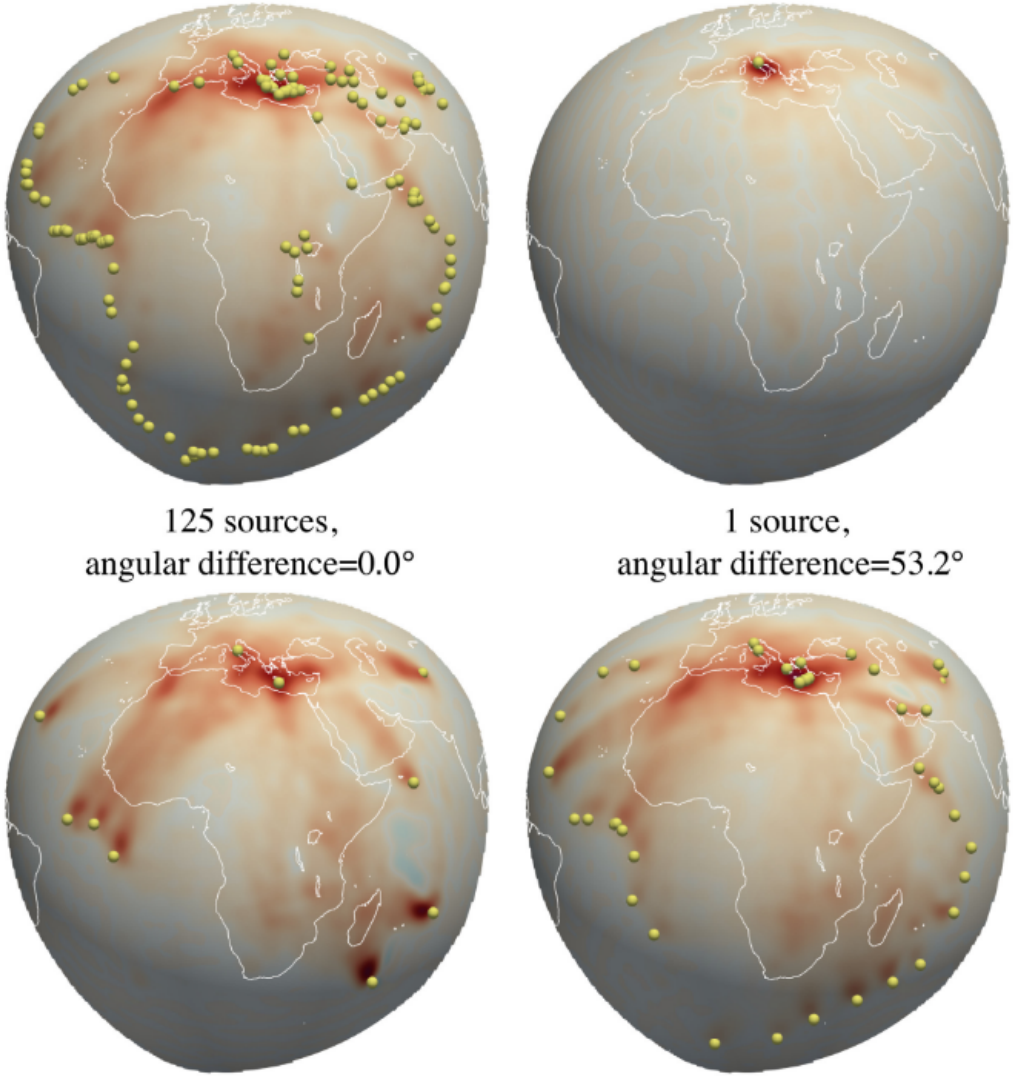
Waves traversing the same media give us similar information

Much of the information contained in our is redundant

We can do optimization with a subset of sources and achieve similar results

Mini-batch optimization

Similar gradients can be estimated using only a fraction of the earthquakes



125 sources,
angular difference=0.0°

1 source,
angular difference=53.2°

10 sources,
angular difference=27.8°

40 sources,
angular difference=8.8°

Van Herwaarden et al. 2020

Stochastic optimization

Optimization proceeds very quickly far from the solution with stochastic optimization

Closer to the solution, slowness and non-convergence become issues

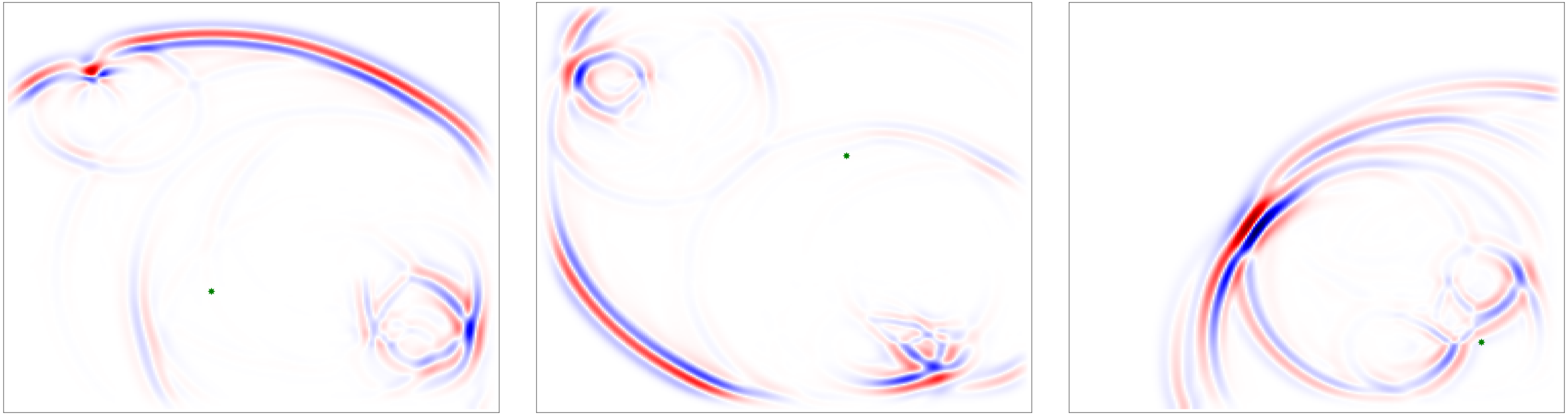
Source stacking

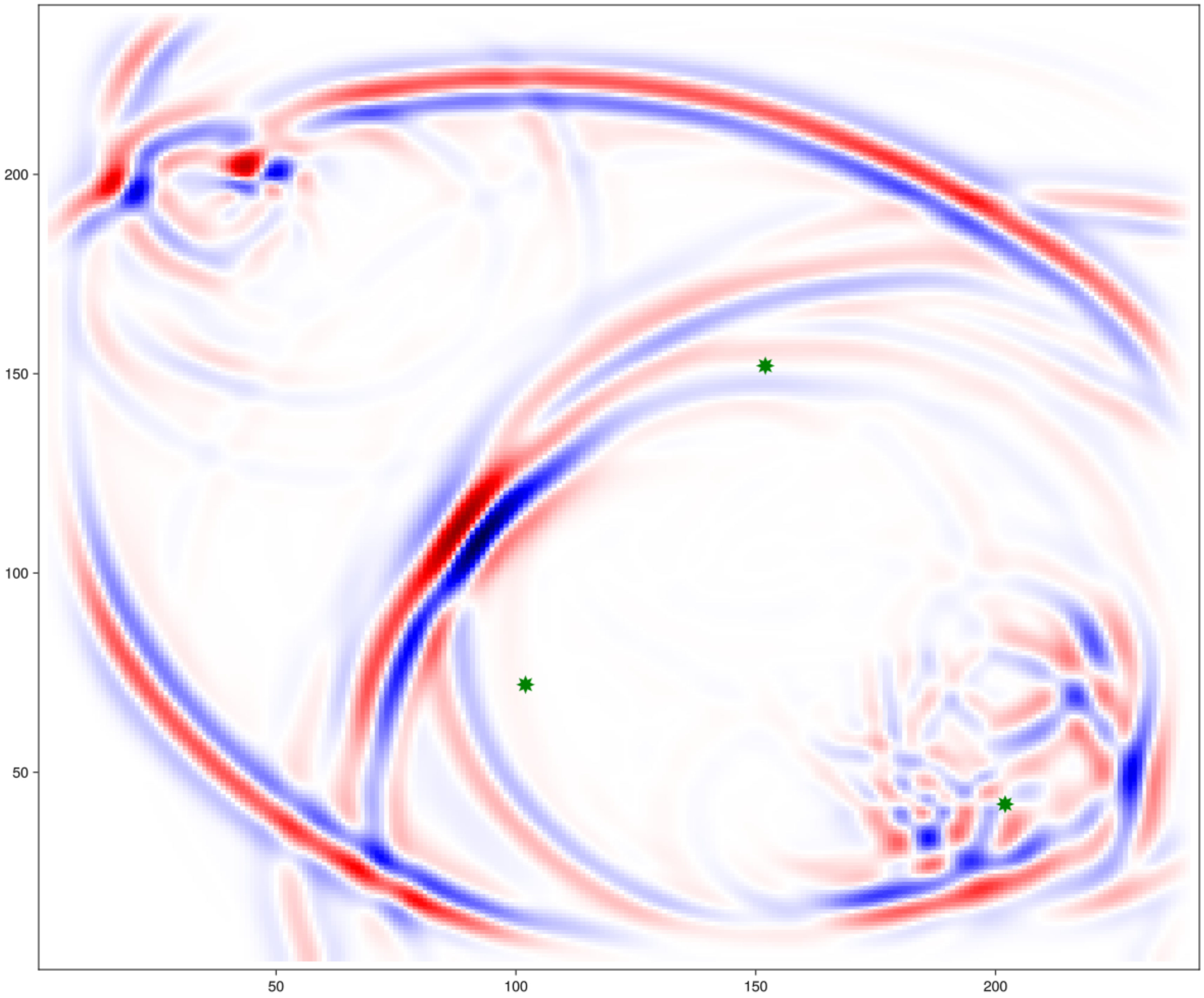
Our wavefield is linear with respect to source amplitudes

We can treat sums of measurements as data and simulate multiple sources in a single simulation

This increases efficiency, but introduces the possibility of cross-talk

Source stacking




200
150
100
50
50
100
150
200

Source stacking - encoding

Normally, we measure and simulate $\boldsymbol{u_n}(t)$

Source stacking - encoding

Normally, we measure and simulate $\boldsymbol{u_n}(t)$

Suppose we replace this with $\boldsymbol{u_n}(t) * \boldsymbol{p_n}(t)$, with $\boldsymbol{p_n}(t) * \boldsymbol{p_m}(t) \approx \delta(t)\delta_{nm}$

Source stacking - encoding

Normally, we measure and simulate $\boldsymbol{u_n}(t)$

Suppose we replace this with $\boldsymbol{u_n}(t) * \boldsymbol{p_n}(t)$, with $\boldsymbol{p_n}(t) * \boldsymbol{p_m}(t) \approx \boldsymbol{q}(t)\delta_{nm}$

Then, we can limit the cross-talk between different sources in a stack

Source stacking - encoding

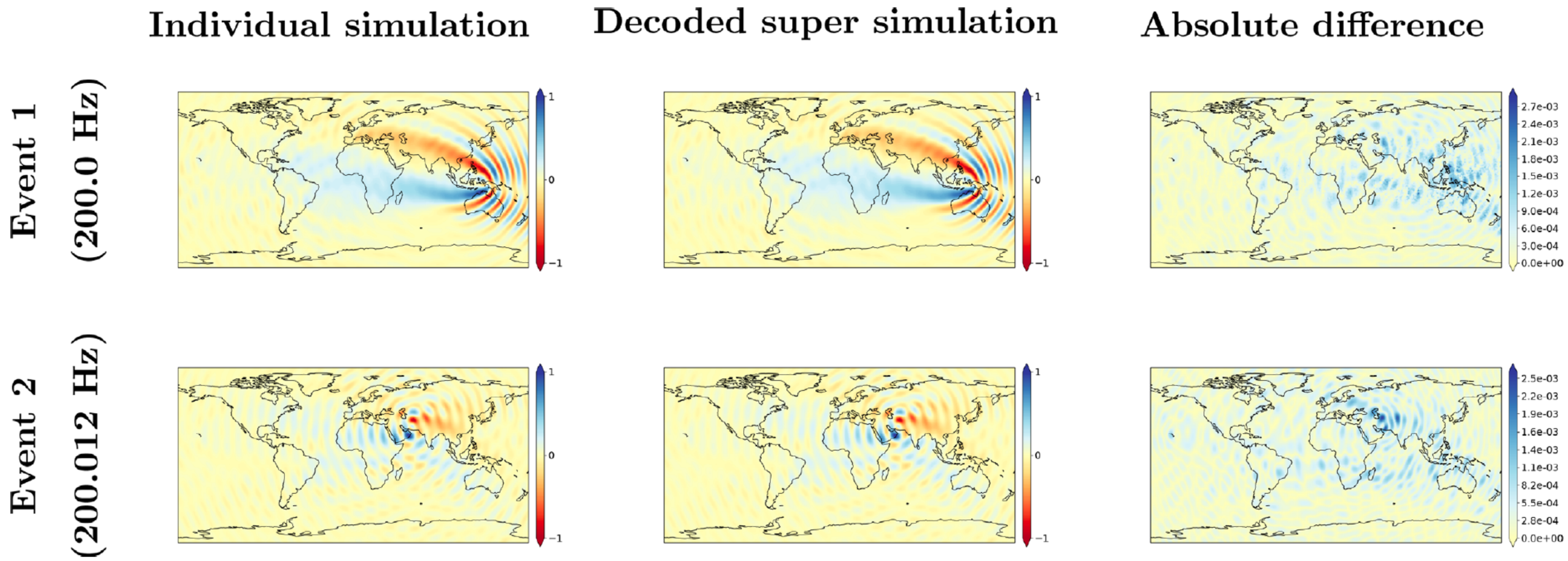
Normally, we measure and simulate $\boldsymbol{u_n}(t)$

Suppose we replace this with $\boldsymbol{u_n}(t) * \boldsymbol{p_n}(t)$, with $\boldsymbol{p_n}(t) * \boldsymbol{p_m}(t) \approx \boldsymbol{q}(t)\delta_{nm}$

Then, we can limit the cross-talk between different sources in a stack

This process is called **encoding**

Source stacking - encoding



Tromp and Bachmann, 2019

Limitations

Each of these approaches is best suited for improving convergence to a reasonable model

The details of a model can be difficult to get with these efficient approaches

Often a two-stage procedure is used, in which a fast approach is followed by a slower, more accurate one

Outline

1. Overview of full-waveform inversion
2. Numerical modelling of seismic waves
3. Objective function
4. Optimization
5. Reducing computational cost
6. Uncertainty quantification

Uncertainty quantification

1. Bayesian approaches
2. Shuttling approaches
3. Curvature approaches
4. Brute-force hypothesis testing

Bayesian uncertainty quantification

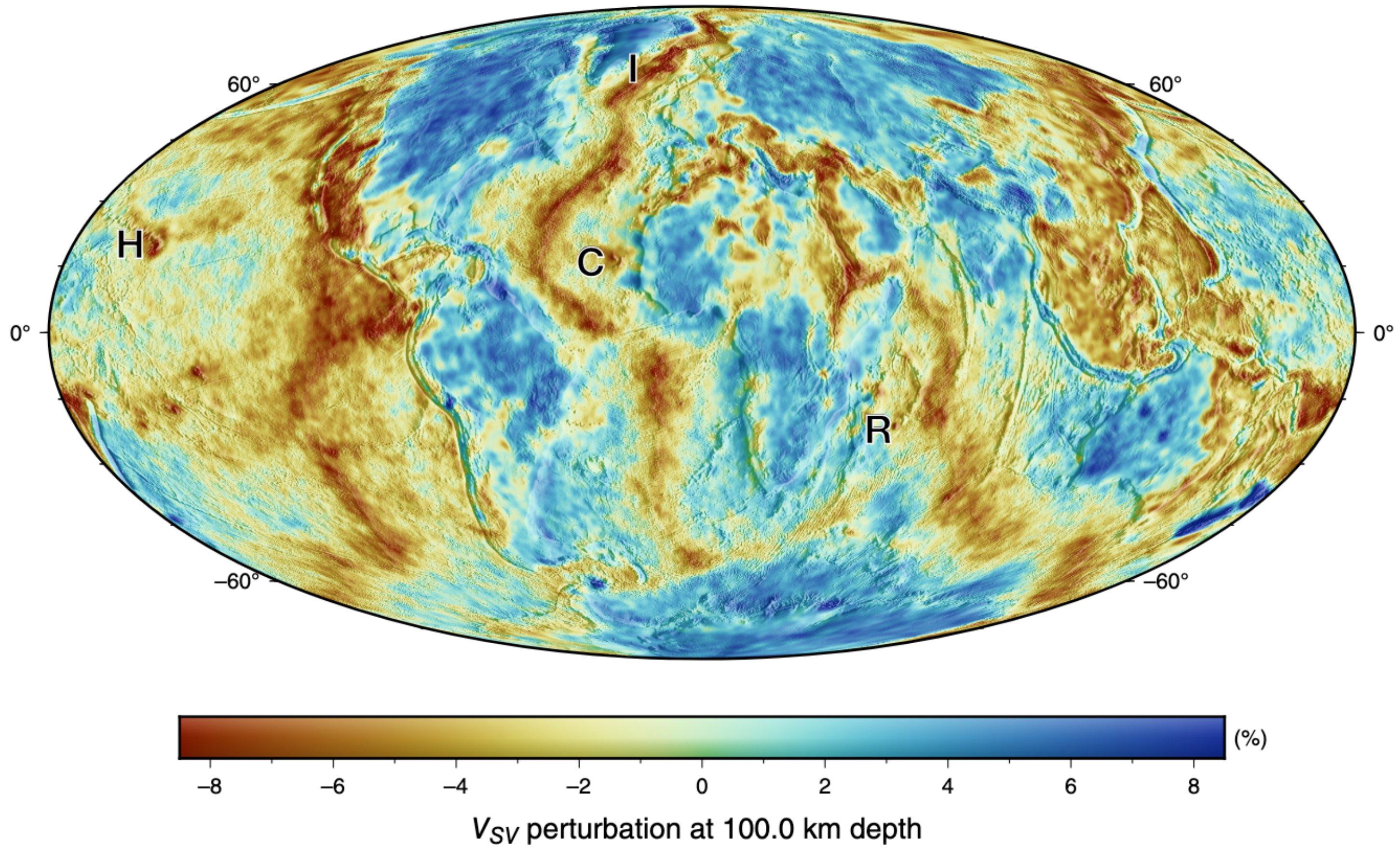
The probability of a given model is closely tied to its objective function value

Bayesian approaches attempt to comprehensively map out a probability density in model space

This solves the uncertainty problem almost completely, but typically comes at **very** large computational cost

Deterministic inversion

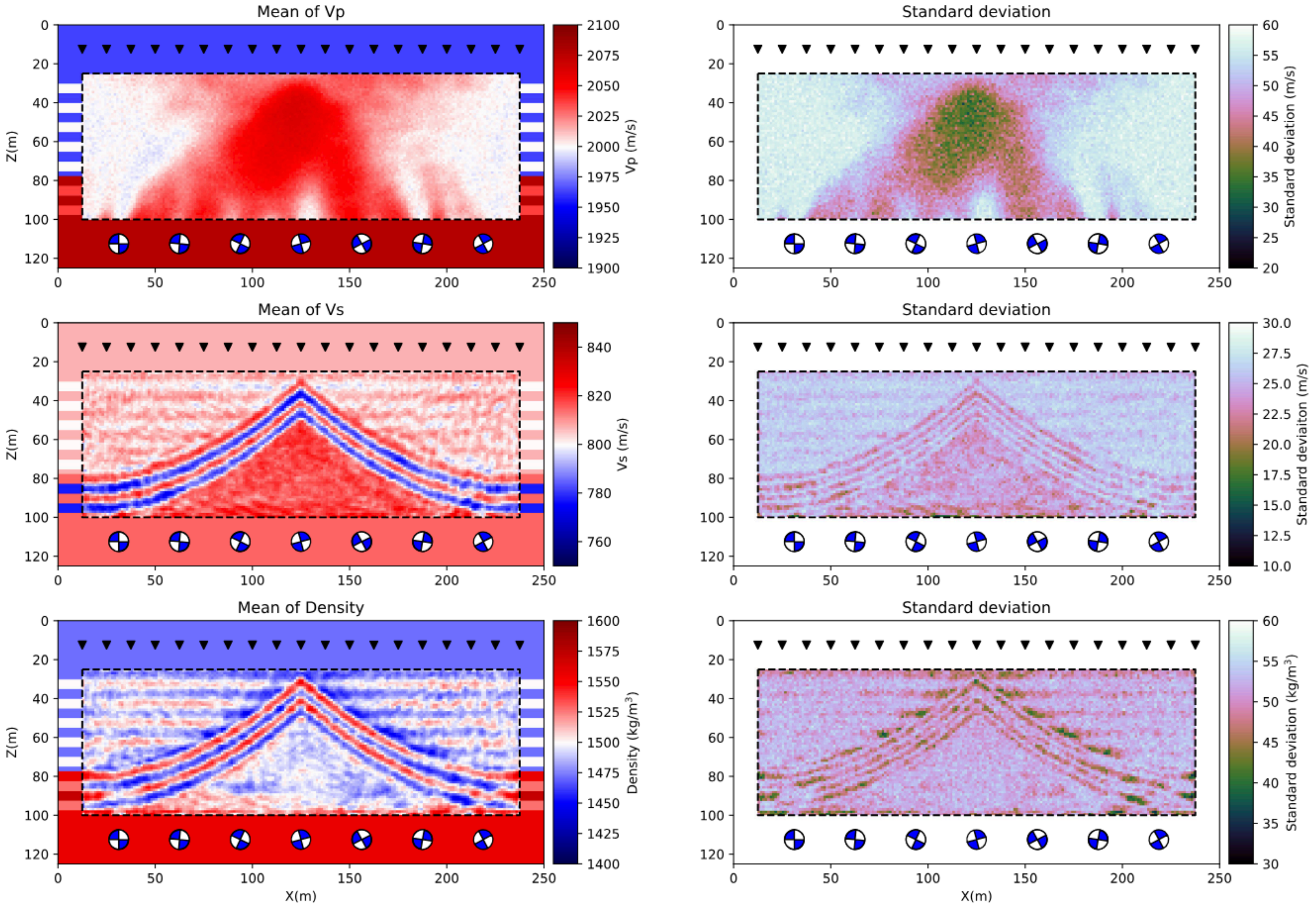
305
forward
and
adjoint



Thrastarson et al. 2024

SVGD

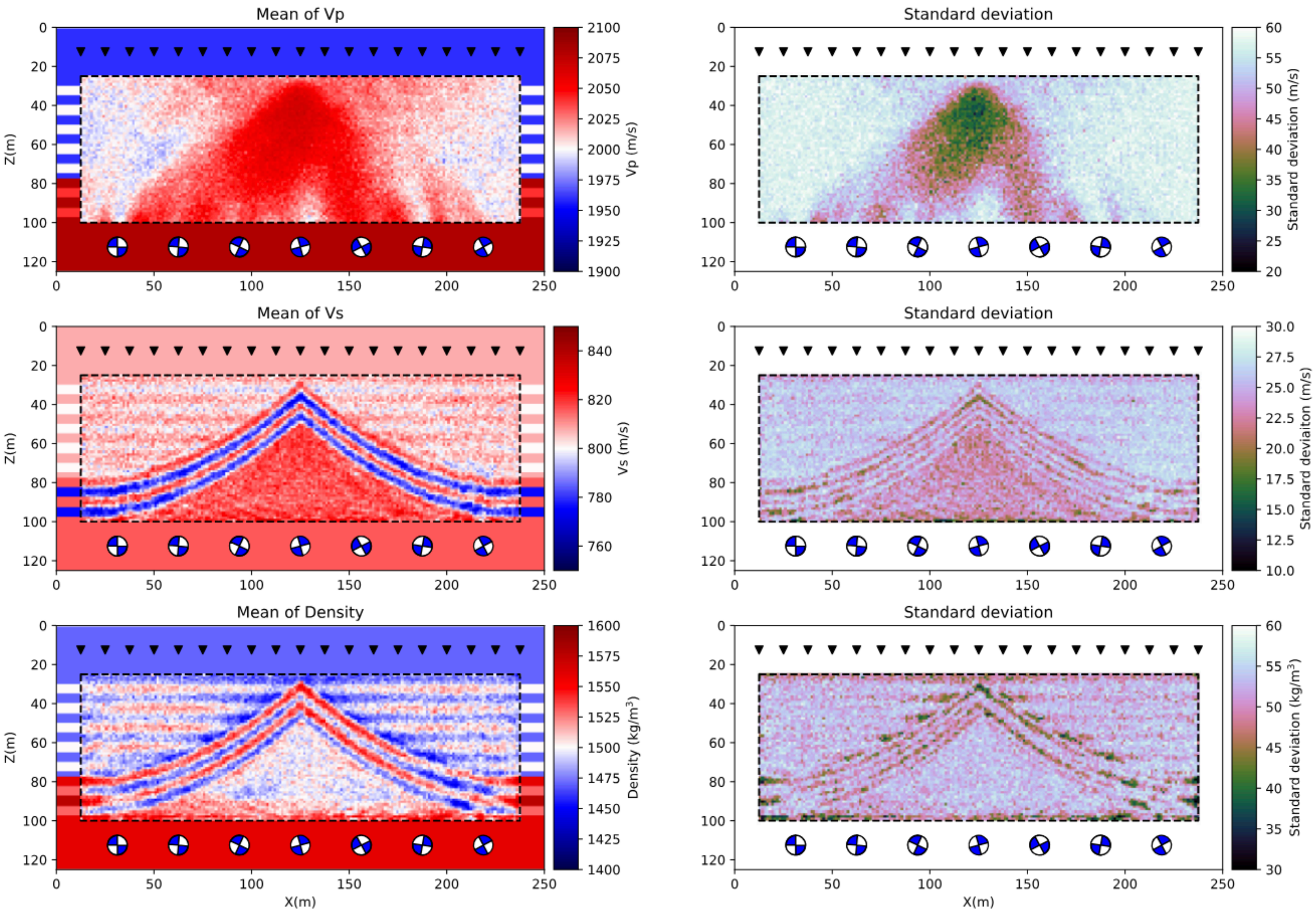
360,000 forward and adjoint



Zhang et al. 2020

HMC

130,000 forward and adjoint



Gebraad et al. 2020

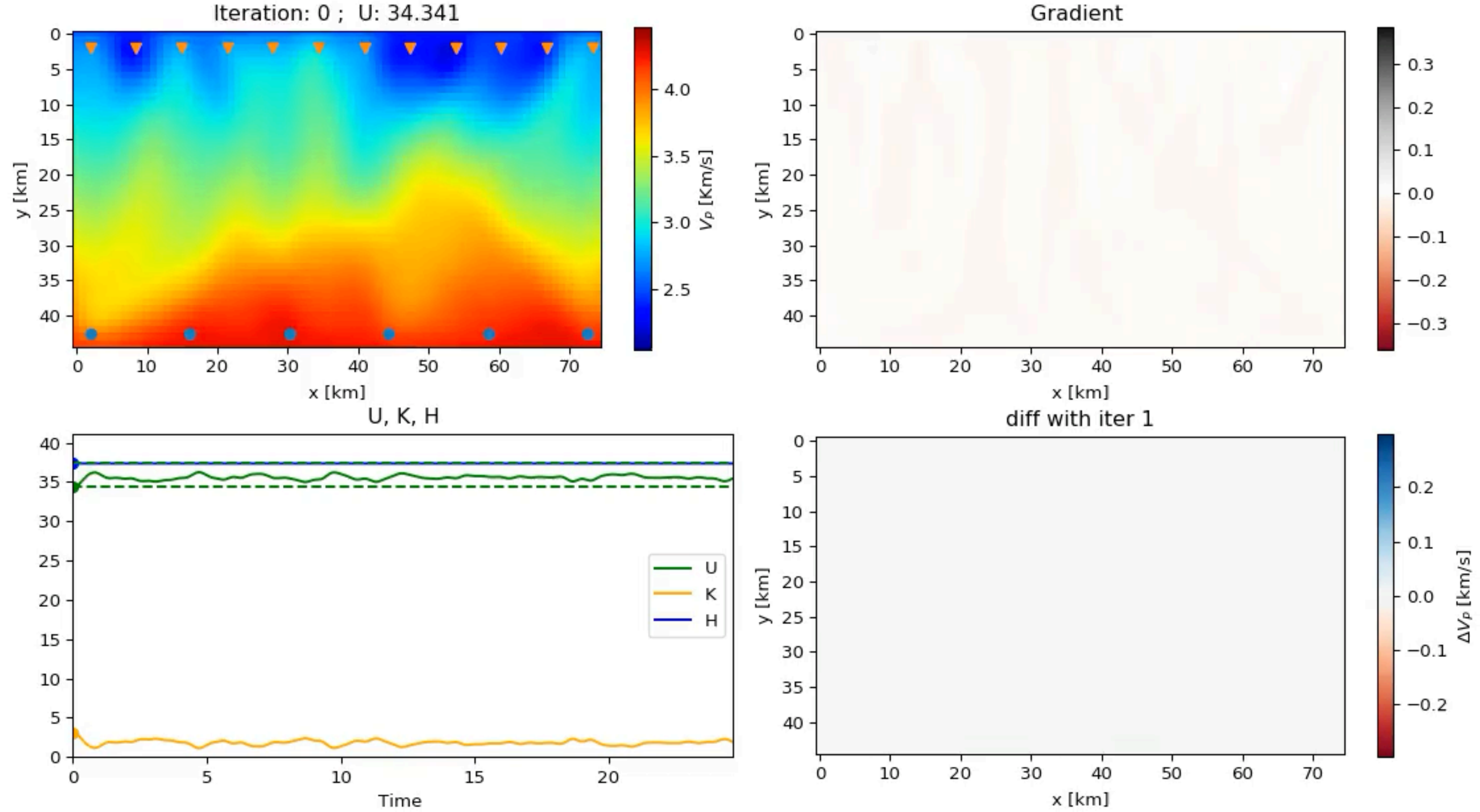
Nullspace shuttling

The inversion nullspace is the set of models which satisfy our model and priors “acceptably”

Characterizing the nullspace gives us a less expensive, but less complete form of UQ

Nullspace shuttling

Zunino, pers. comm.



Curvature estimation

At the minimum of the objective function,

$$\phi(\boldsymbol{m} + \Delta\boldsymbol{m}) = \phi(\boldsymbol{m}) + \boldsymbol{g}\Delta\boldsymbol{m} + \frac{1}{2}\Delta\boldsymbol{m}^T H \Delta\boldsymbol{m} + O(\Delta\boldsymbol{m}^3)$$

Curvature estimation

At the minimum of the objective function,

$$\phi(\boldsymbol{m} + \Delta\boldsymbol{m}) = \phi(\boldsymbol{m}) + \boldsymbol{g}\Delta\boldsymbol{m} + \frac{1}{2}\Delta\boldsymbol{m}^T H \Delta\boldsymbol{m} + O(\Delta\boldsymbol{m}^3)$$

Curvature estimation

At the minimum of the objective function,

$$\phi(\boldsymbol{m} + \Delta\boldsymbol{m}) \approx \phi(\boldsymbol{m}) + \frac{1}{2}\Delta\boldsymbol{m}^T H \Delta\boldsymbol{m}$$

Curvature estimation

At the minimum of the objective function,

$$\phi(\boldsymbol{m} + \Delta\boldsymbol{m}) \approx \phi(\boldsymbol{m}) + \frac{1}{2}\Delta\boldsymbol{m}^T H \Delta\boldsymbol{m}$$

So, if we know both the minimum of the objective function and the Hessian, we can characterize the uncertainty!

Curvature estimation

At the minimum of the objective function,

$$\phi(\boldsymbol{m} + \Delta\boldsymbol{m}) \approx \phi(\boldsymbol{m}) + \frac{1}{2}\Delta\boldsymbol{m}^T H \Delta\boldsymbol{m}$$

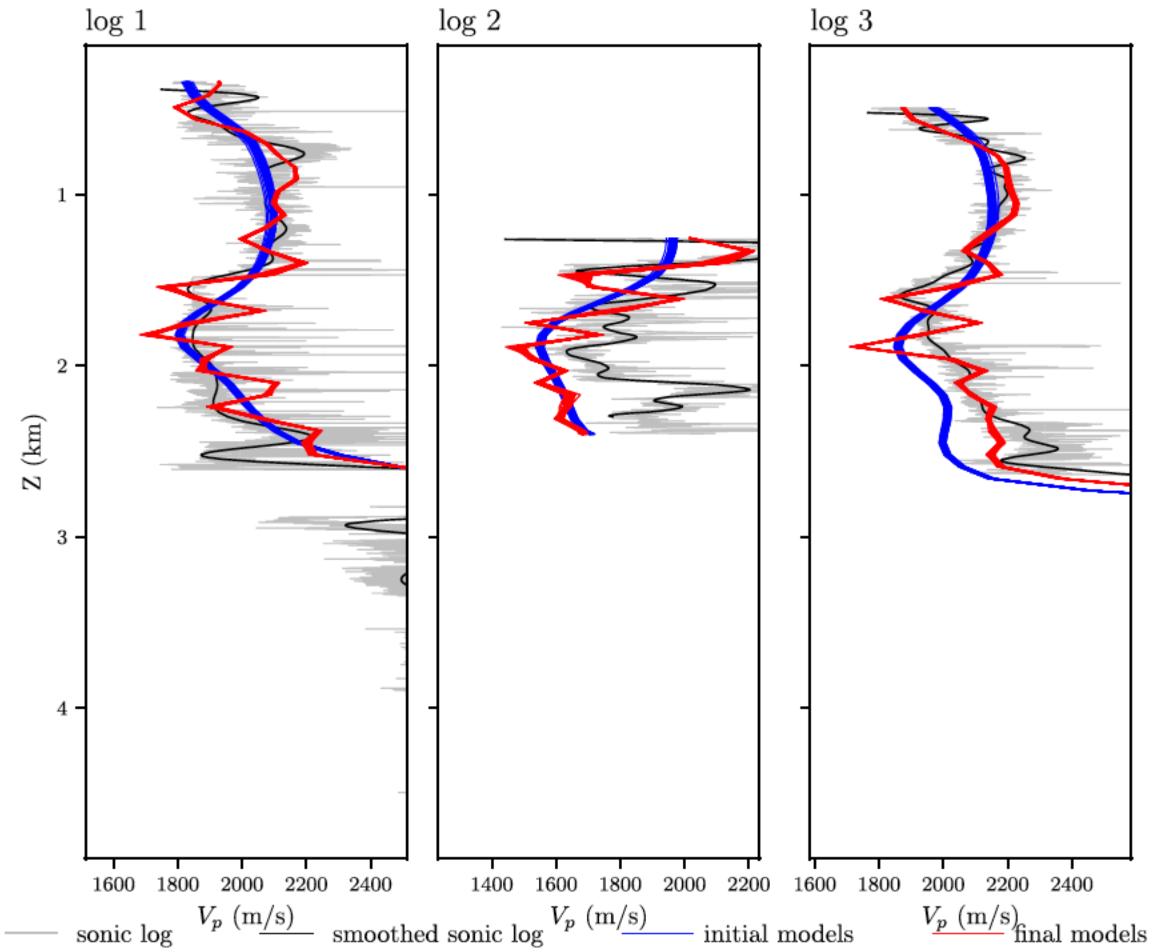
So, if we know both the minimum of the objective function and the Hessian, we can characterize the uncertainty!

The Hessian is too big to calculate, but we already estimate it!

Uncertainty quantification

Uncertainty analysis based on curvature estimates or ensemble approaches can be achieved at much lower cost

These approaches get an incomplete picture, and tend to chronically underestimate the uncertainties



Hoffman et al. 2024

Brute-force

Unfortunately, cost constraints mean that we most often revert to brute-force hypothesis testing

This means checking the objective function for a model both with and without a given feature to determine which is preferred

The strong tendency of these tests is to prefer the inversion result as our “alternatives” are typically ad hoc

Takeaways

1. Full-waveform inversion tries to build Earth models using the full information content of measurements
2. This requires computationally intensive modelling
3. Inversions are driven by objective functions – these are tricky to define well
4. The adjoint-state method allows us to use gradient-based optimization
5. Computational speedups exist, but always have tradeoffs
6. Uncertainty quantification remains elusive

References

- Simmons, N. A., Myers, S. C., Morency, C., Chiang, A., & Knapp, D. R. (2021). SPiRaL: a multiresolution global tomography model of seismic wave speeds and radial anisotropy variations in the crust and mantle. *Geophysical Journal International*, *227*(2), 1366-1391.
- Hoffmann, A., Brossier, R., Métivier, L., & Tarayoun, A. (2024). Local uncertainty quantification for 3-D time-domain full-waveform inversion with ensemble Kalman filters: application to a North Sea OBC data set. Geophysical Journal International, 237(3), 1353-1383.
- Métivier, L., Brossier, R., Merigot, Q., & Oudet, E. (2019). A graph space optimal transport distance as a generalization of Lp distances: application to a seismic imaging inverse problem. Inverse Problems, 35(8), 085001.
- Rodgers, A. J., Doody, C., & Fichtner, A. (2024). WUS324: Full Waveform Inversion Approaching Convergence Improves Waveform Fits While Imaging Seismic Structure of the Western United States. Authorea Preprints..
- Zhang, X., & Curtis, A. (2020). Variational full-waveform inversion. Geophysical Journal International, 222(1), 406-411.
- Thrastarson, S., van Herwaarden, D. P., Noe, S., Josef Schiller, C., & Fichtner, A. (2024). REVEAL: A global full-waveform inversion model. Bulletin of the Seismological Society of America, 114(3), 1392-1406.
- Zunino, A., personal communication
- Thrastarson, S., Van Herwaarden, D. P., Krischer, L., Boehm, C., van Driel, M., Afanasiev, M., & Fichtner, A. (2022). Data-adaptive global full-waveform inversion. Geophysical Journal International, 230(2), 1374-1393.
- Gebraad, L., Boehm, C., & Fichtner, A. (2020). Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo. Journal of Geophysical Research: Solid Earth, 125(3), e2019JB018428.
- Krischer, L., Fichtner, A., Zukauskaite, S., & Igel, H. (2015). Large-scale seismic inversion framework. Seismological Research Letters, 86(4), 1198-1207.
- Tromp, J., & Bachmann, E. (2019). Source encoding for adjoint tomography. Geophysical Journal International, 218(3), 2019-2044.
- Fichtner, A., Kennett, B. L., Igel, H., & Bunge, H. P. (2008). Theoretical background for continental-and global-scale full-waveform inversion in the time–frequency domain. Geophysical Journal International, 175(2), 665-685.
- van Herwaarden, D. P., Boehm, C., Afanasiev, M., Thrastarson, S., Krischer, L., Trampert, J., & Fichtner, A. (2020). Accelerated full-waveform inversion using dynamic mini-batches. Geophysical Journal International, 221(2), 1427-1438.
- Bunks, C., Saleck, F. M., Zaleski, S., & Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, *60*(5), 1457-1473.