A Biased Tour of Geophysical Inversion

Beno Gutenberg Lecture,

American Geophysical Union, Fall 2020 Meeting.

Malcolm Sambridge

Research School of Earth Sciences,

Australian National University

Canberra ACT, Australia.

The lecture is named to honor the life and work of renowned seismologist Beno Gutenberg and is presented during the AGU Fall Meeting

My tour guides:

Thomas Bodin, Jean Braun, Ross Brodie, Bill Compston, Huw Davies, Eric Debayle, Kerry Gallagher, Oli Gudmundsson, Juerg Hauser, Rhys Hawkins, Heiner Igel, Andy Jackson, Brian Kennett, Kurt Lambeck, Herb McQueen, Hugh O'Neill, Nick Rawlinson, Anya Reading, David Robinson, Michelle Salmon, Erdinc Saygin, Roel Snieder, Hrvoje Tkalčić, Andrew Valentine and Paul Williamson.

…who have helped show me the way.

Contributors to this presentation: Matthias Scheiter, Buse Turunçtur and Andrew Valentine.

…a thank you to our sponsors

Australian National University
NATURAM PRIMUM COGNOSCERE RERUM

CSIRO

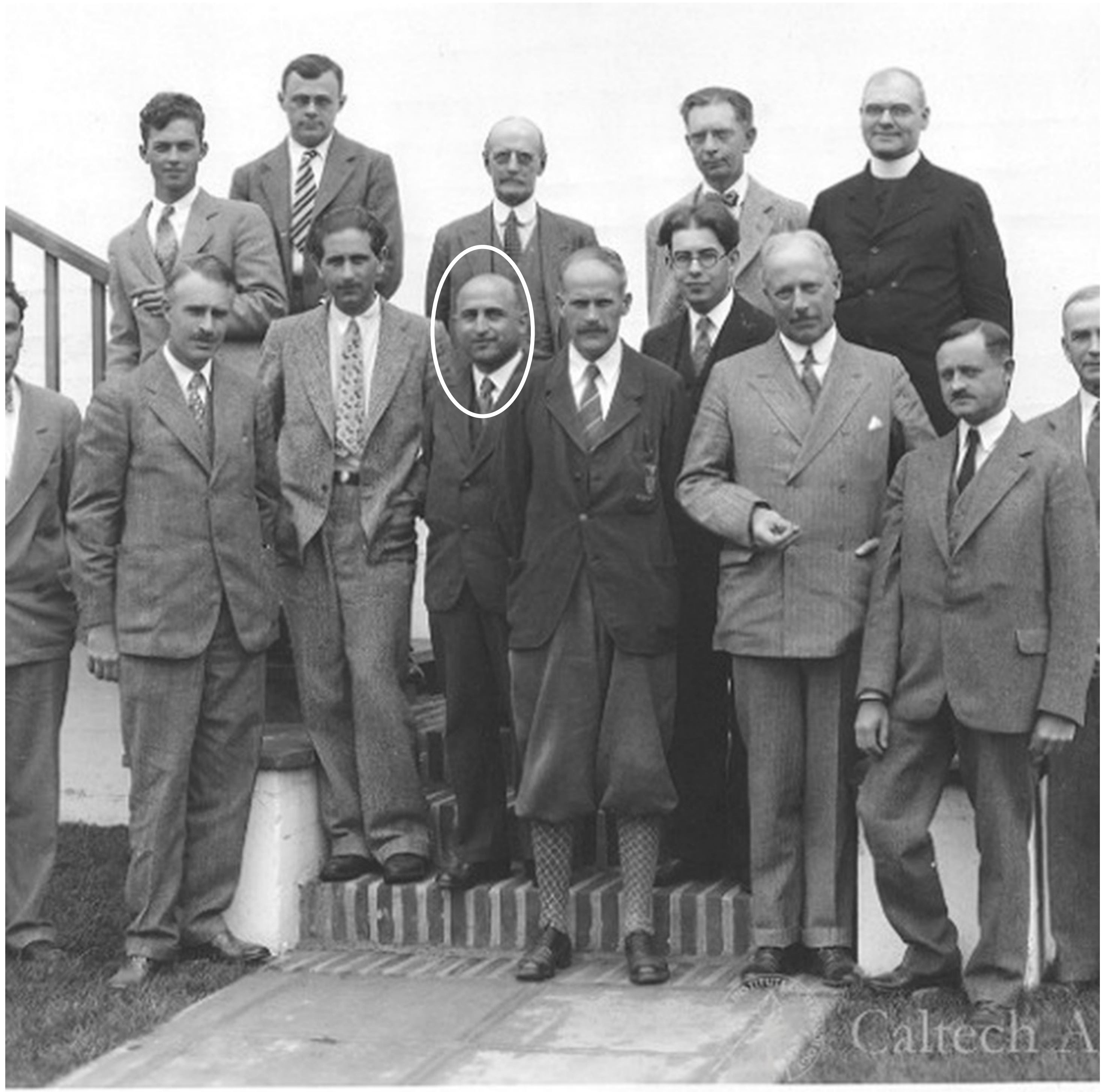
AuScope

Australian Government
Geoscience Australia

inlab.edu.au

Australian Government
Australian Research Council

World authorities at the Seismo Lab, 1929.



L-R: front: Archie King, L. H. Adams, Hugo Benioff, Beno Gutenberg, Harold Jeffreys, Charles Richter, Arthur Day, Harry Wood, Ralph Arnold, John P. Buwalda; Top: Alden Waite, Perry Byerly, Harry Reid, John Anderson, Fr. J. P. Macelwane.

Courtesy Andrew Valentine & Brian Kennett.

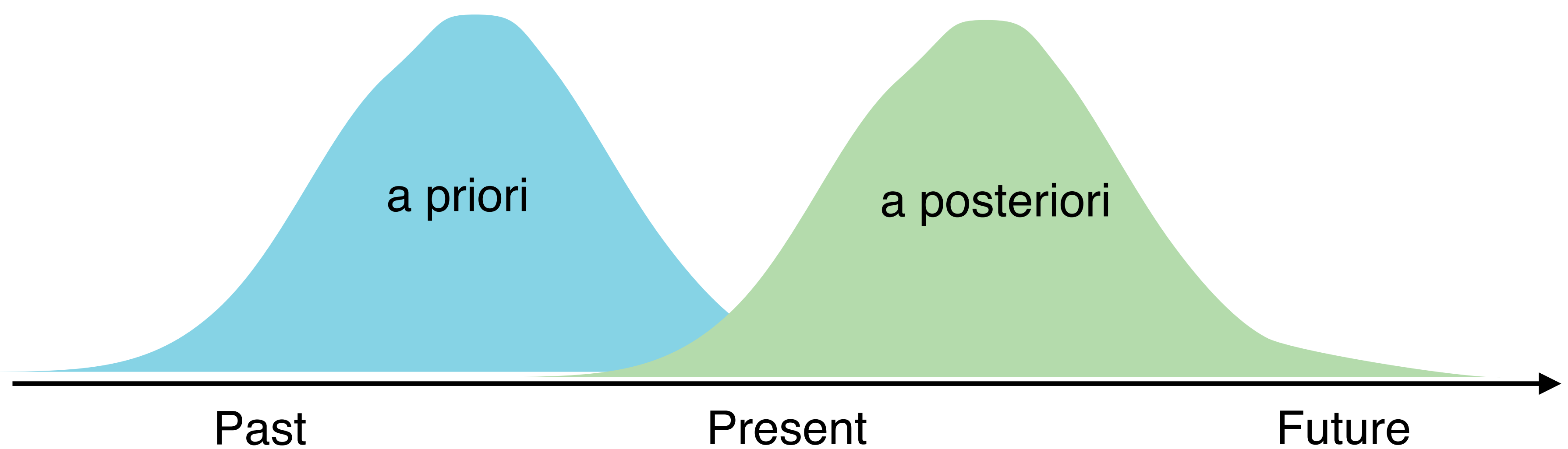
"A **Biased Tour** of Geophysical Inversion"

Definitions:

Biased - having a preference for one thing over another.

Tour - a journey for pleasure in which several different places are visited.

What this talk was about when I wrote the abstract… …and what it is actually about.

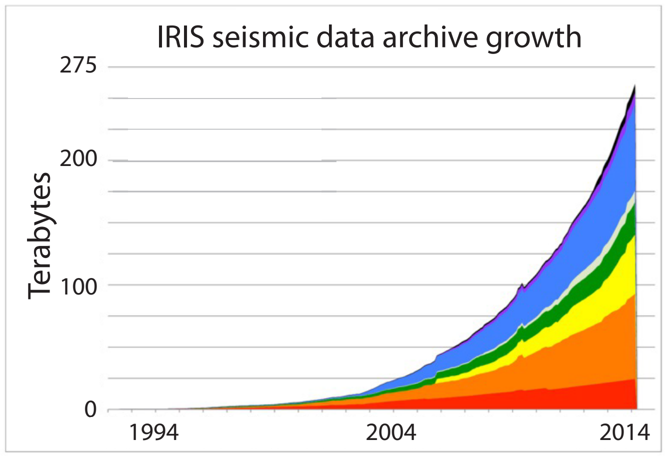


This will be a dynamic talk: http://www.inlab.edu.au/gutenberg

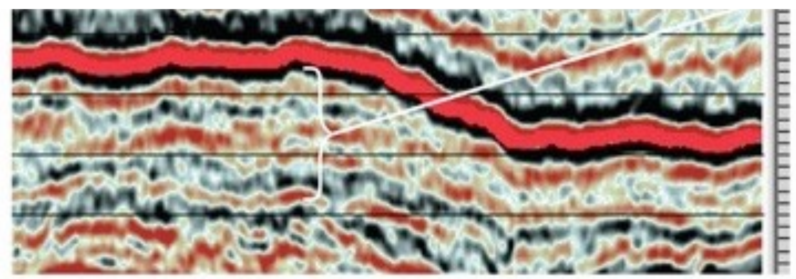
Download slides, animations, give feedback, see updates and more….

Thank you Andrew Valentine

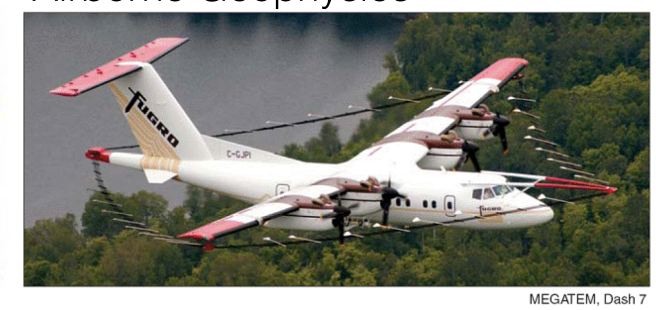
Inverse problems: all shapes and sizes

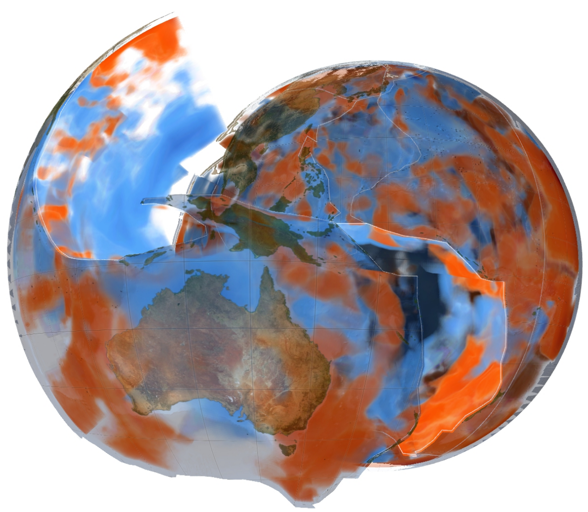


Reflection seismics



Airborne Geophysics

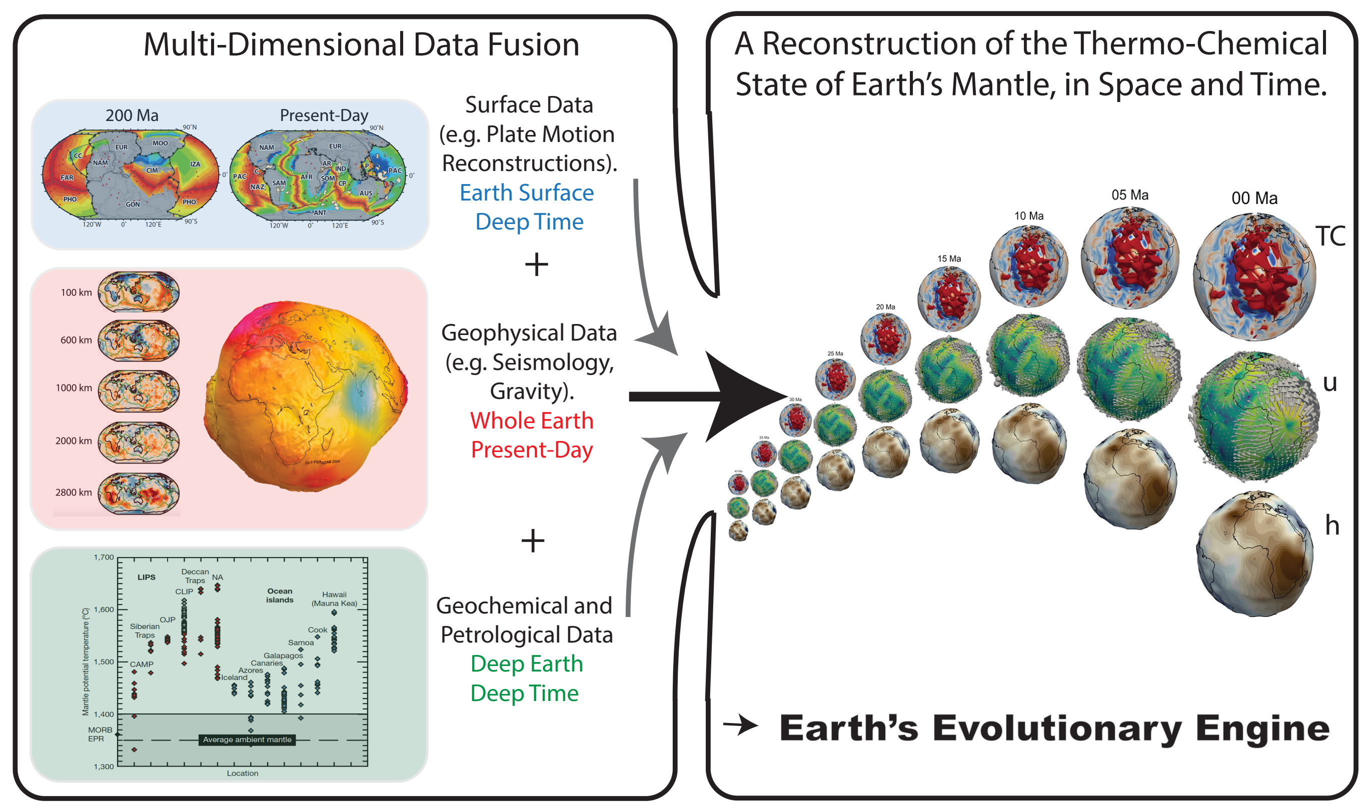




Mantle seismic imaging

Figure Courtesy Christophe Zaroli, Rhys Hawkins

Inverse geodynamics



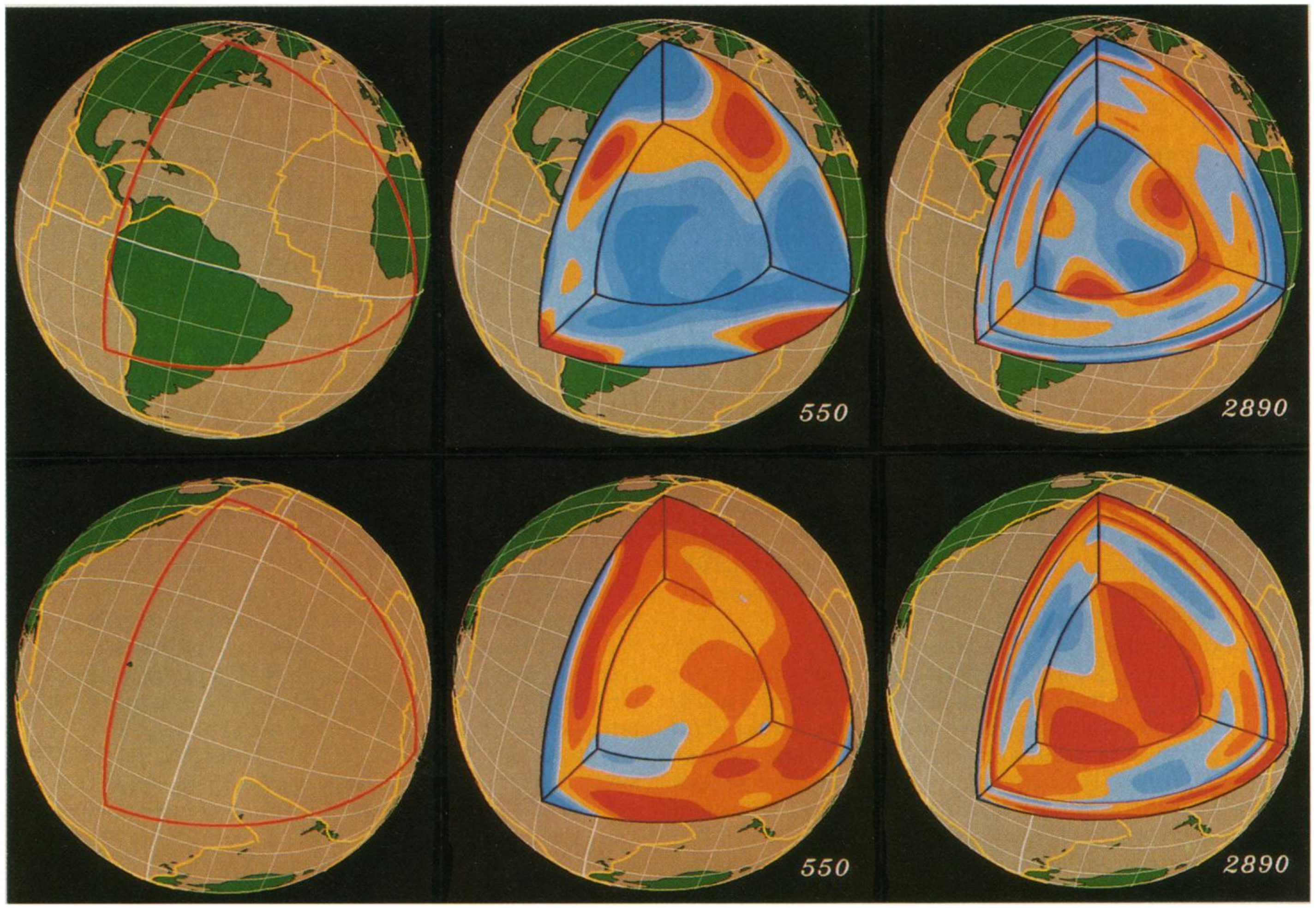
Bunge et al. (2003), Rhodri Davies (2020, pers comm)

Where to from here?

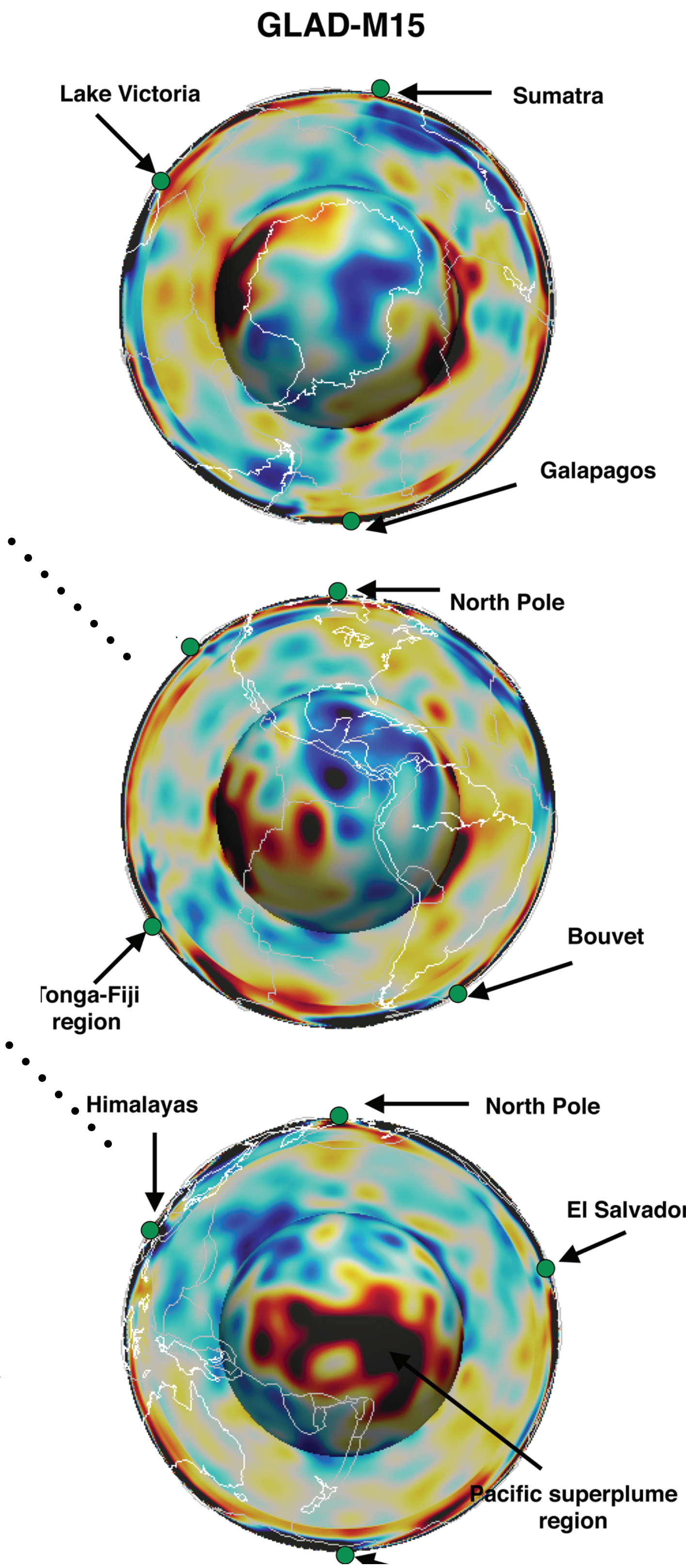
`The purpose of models is not to fit the data but to sharpen the questions.' -S. Karlin, 1983.

A visit to seismic imaging

Windows into the Earth over 30 years



Dziewonski & Woodhouse (1987)



Bozdag et al. (2016)

Left: First-generation global tomographic model constructed based fitting of seismograms using a least squares inversion.

2000 surface wave seismograms, 1500 body wave seismograms, 870 source-receiver pairs.

Changes of model scale, data volumes, computational resources.

Right: First-generation global tomographic model constructed based on adjoint tomography, using an iterative full-waveform inversion.

1.2-3.8 million measurements

1.6million CPU-hours+250k GPU-hours

A visit to Compressive Sensing

Compressive sensing is a novel idea for reconstructing time series from an irregularly sampled subset of data. It replaces regular sampling in time with ‘incoherent’ random sampling.

c.f. Nyquist - Shannon’s sampling theorem

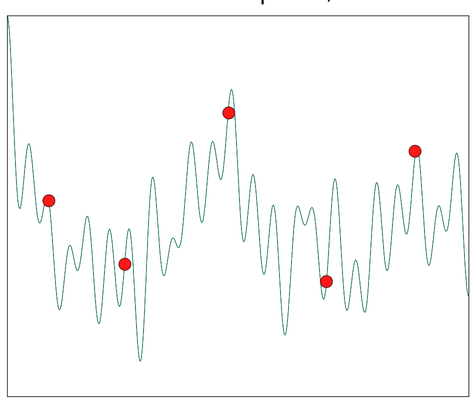
Reduce data misfit

$$\min_{\mathbf{m}} \ ||\mathbf{d} - G\mathbf{m}||_2 + \lambda ||\mathbf{m}||_1$$

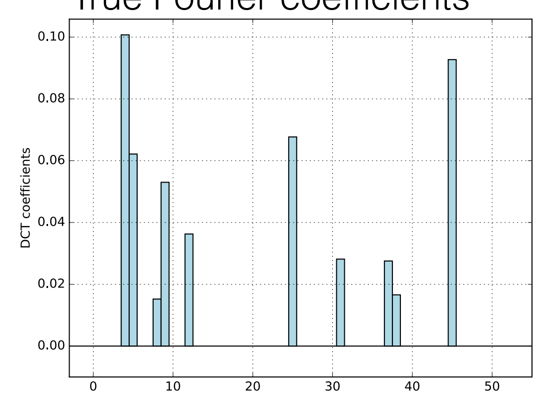
Encourage sparsity

Time series

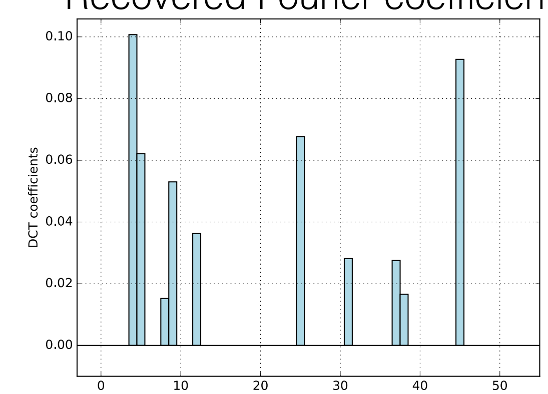
Random time samples, **10%** of total.



True Fourier coefficients

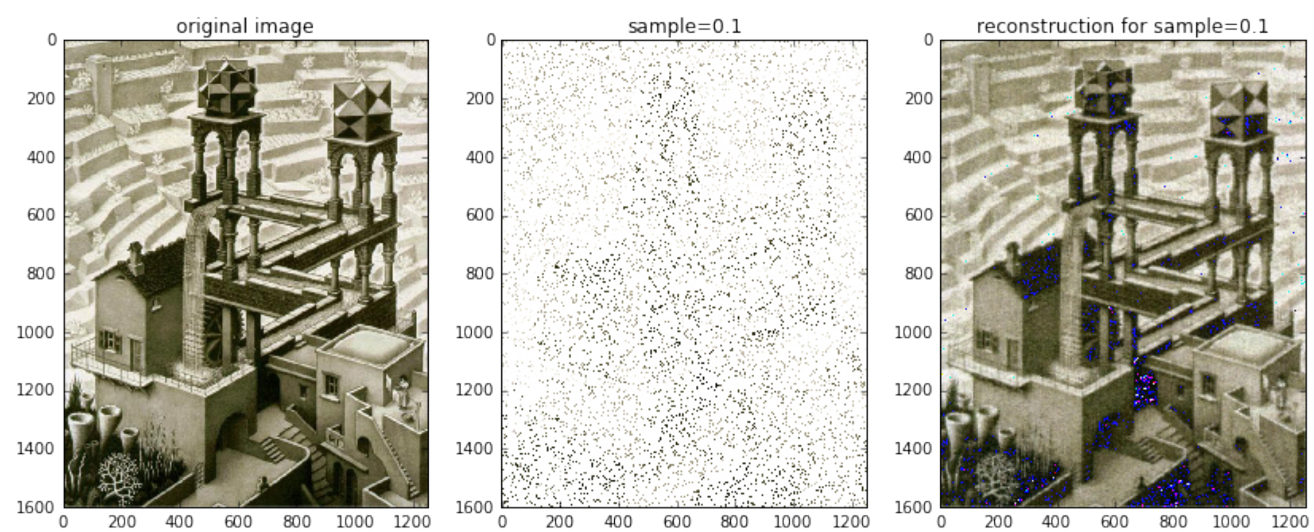


Recovered Fourier coefficients



Should we build a compressive seismometer?

Imaging



Waterfall by M. C. Escher (1200 x1600)

From Robert Taylor’s pyrunner blog

…but Earth structure is not known to be sparse?

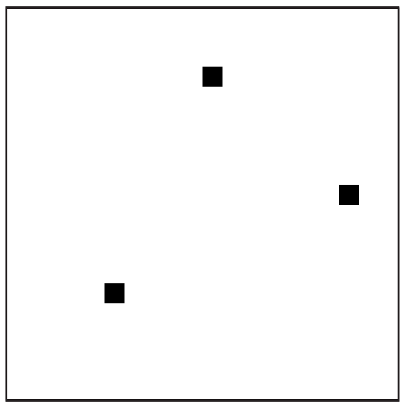
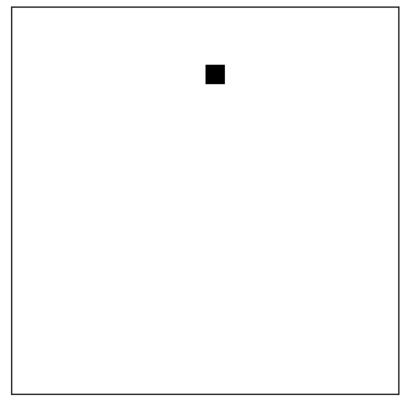
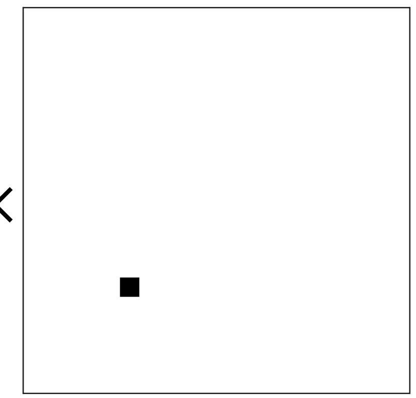
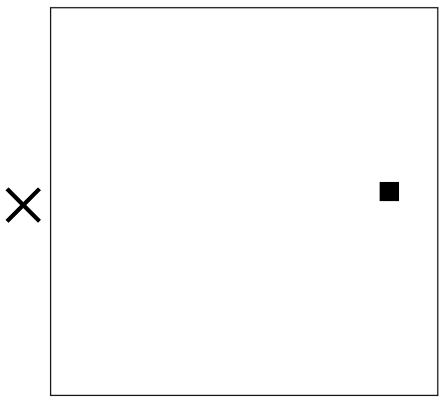
CS refs: Candes et al. (2006), Candes & Tao (2006), Donoho (2006), Candes and Watin (2008), Hermann et al. (2008, 2009)

L1 inversion: Scales et al. (1998); Simons et al. (2011), Loris et al. (2012), Charléty et al. (2013).

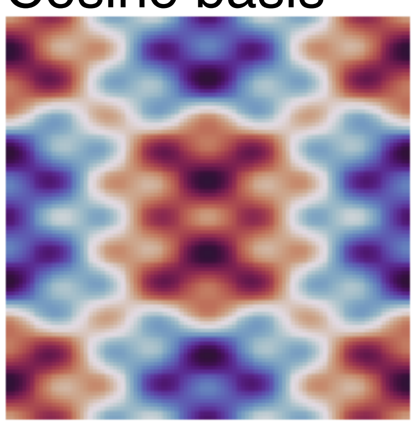
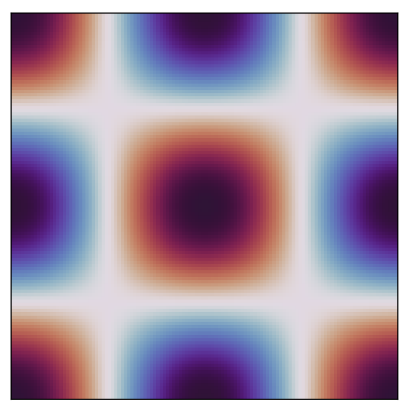
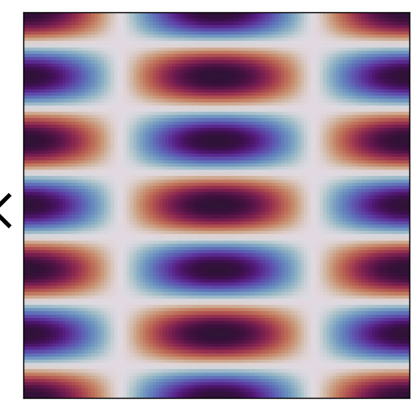
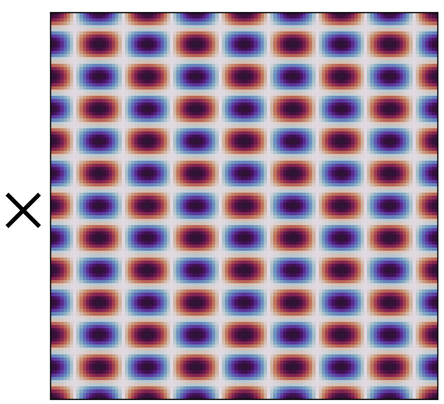
A visit to: Overcomplete tomography

Sparse models $m(\mathbf{x}) = \sum_i^M \alpha_i \phi_i(\mathbf{x})$

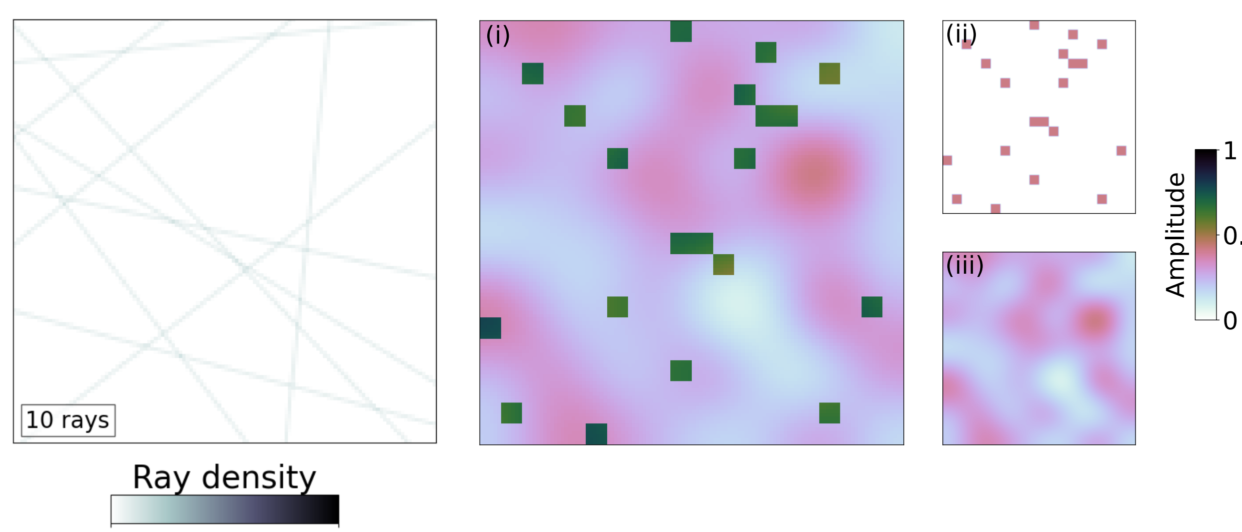
Pixel basis

 $=$ $1.0\times$  $+1.0\times$  $-1.0\times$ 

Cosine basis

 $=$ $0.9\times$  $-0.3\times$  $-0.3\times$ 

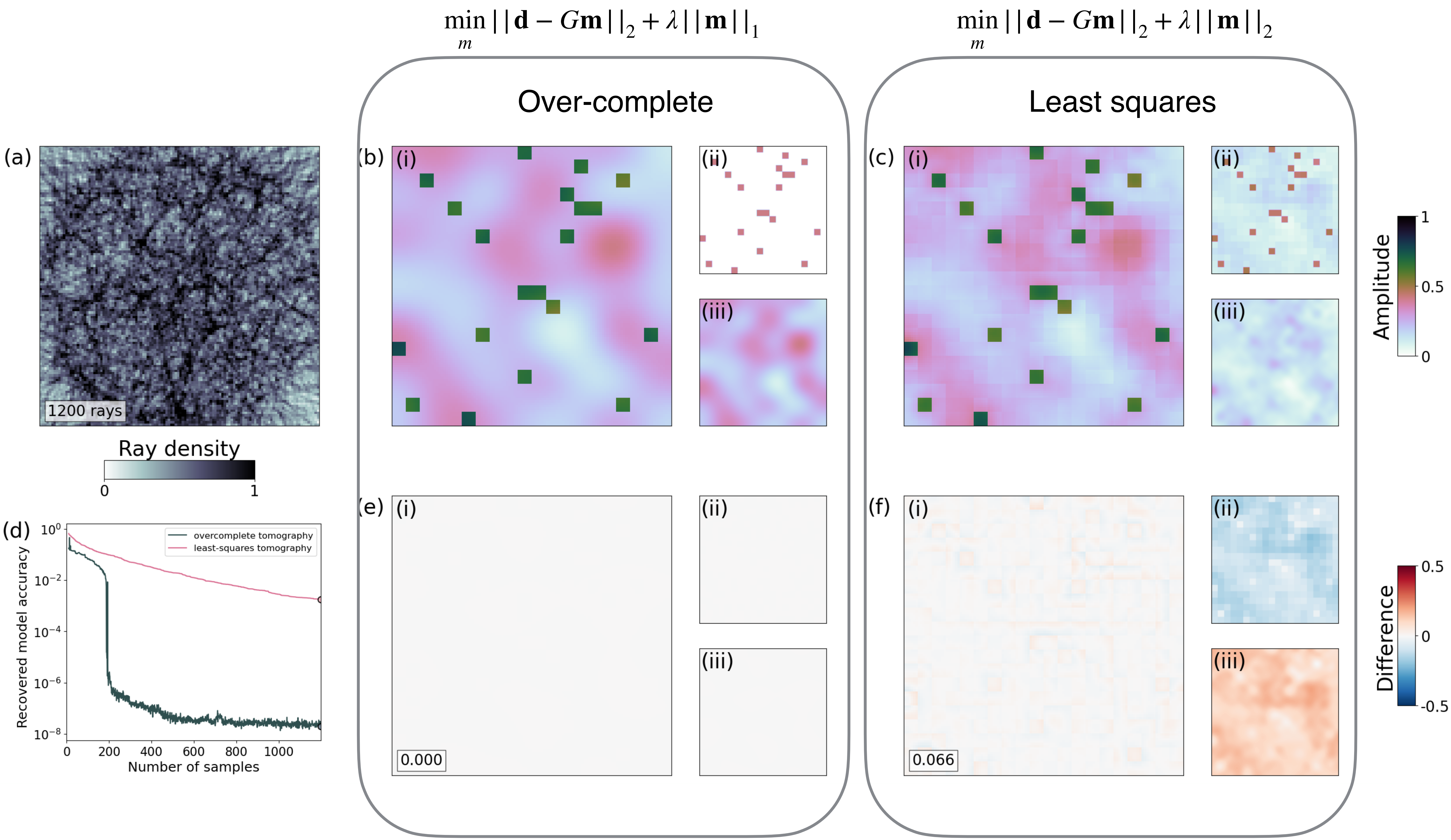
$$m(\mathbf{x}) = \sum_i^M \alpha_i \phi_i(\mathbf{x}) + \sum_j^M \beta_j \psi_j(\mathbf{x})$$



From the Ph.D. studies of Buse Turunçtur

Co-supervised with Andrew Valentine.

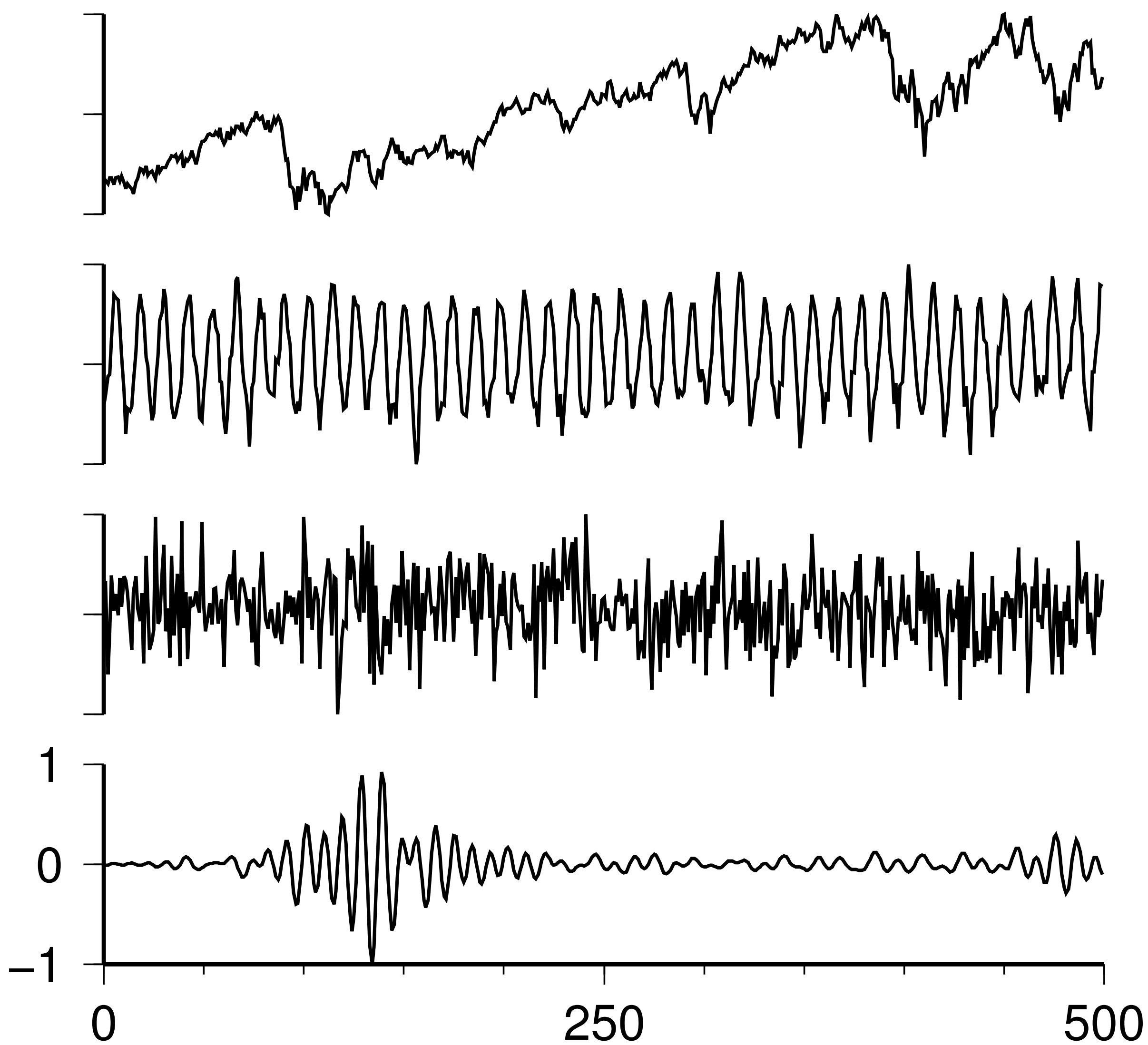
An example of Overcomplete X-ray tomography



From the Ph.D. studies of Buse Turunçtur

A visit to Machine Learning

Which is the seismogram?



Stock prices

Temperature in Birmingham
over 50 years

White noise

Seismogram

From Valentine & Trampert (2012)

Humans can recognise seismic signals. The way computers do it implicitly defines some misfit criteria.

An adversarial inversion framework

Neural Network Generator

Competition

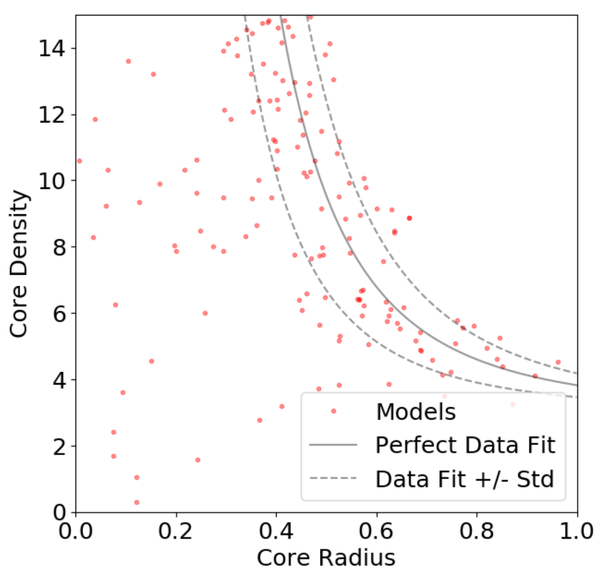
(GAN)

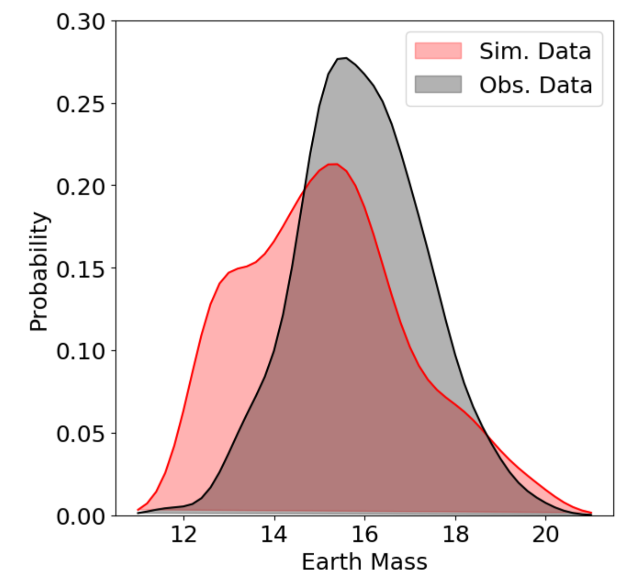
Neural Network Discriminator

Model Ensemble

Forward Operator

Data Distribution





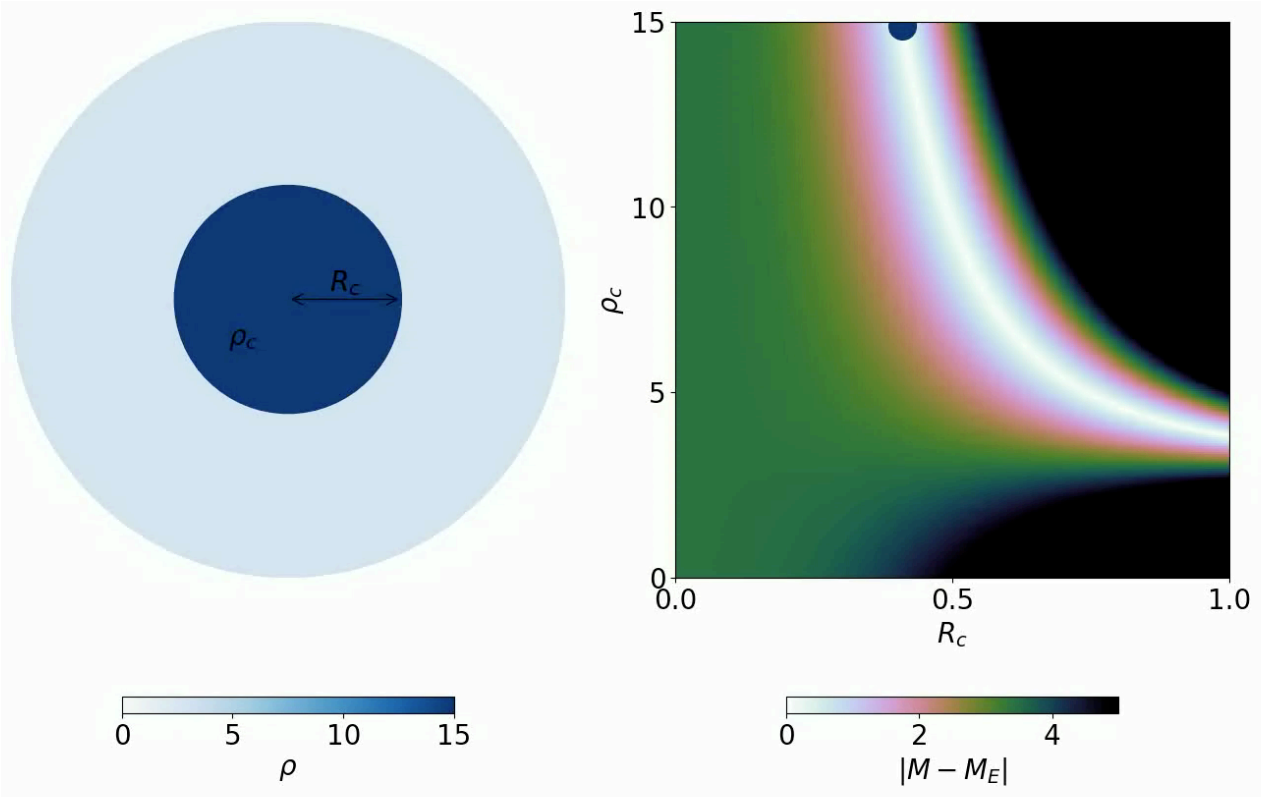
Computer scientists call this a
Generative Adversarial Network (GAN)

From the Ph.D. studies of Matthias Scheiter
Co-supervised with Andrew Valentine.

A toy problem

Determine the density and radius of the core from the Earth's mass

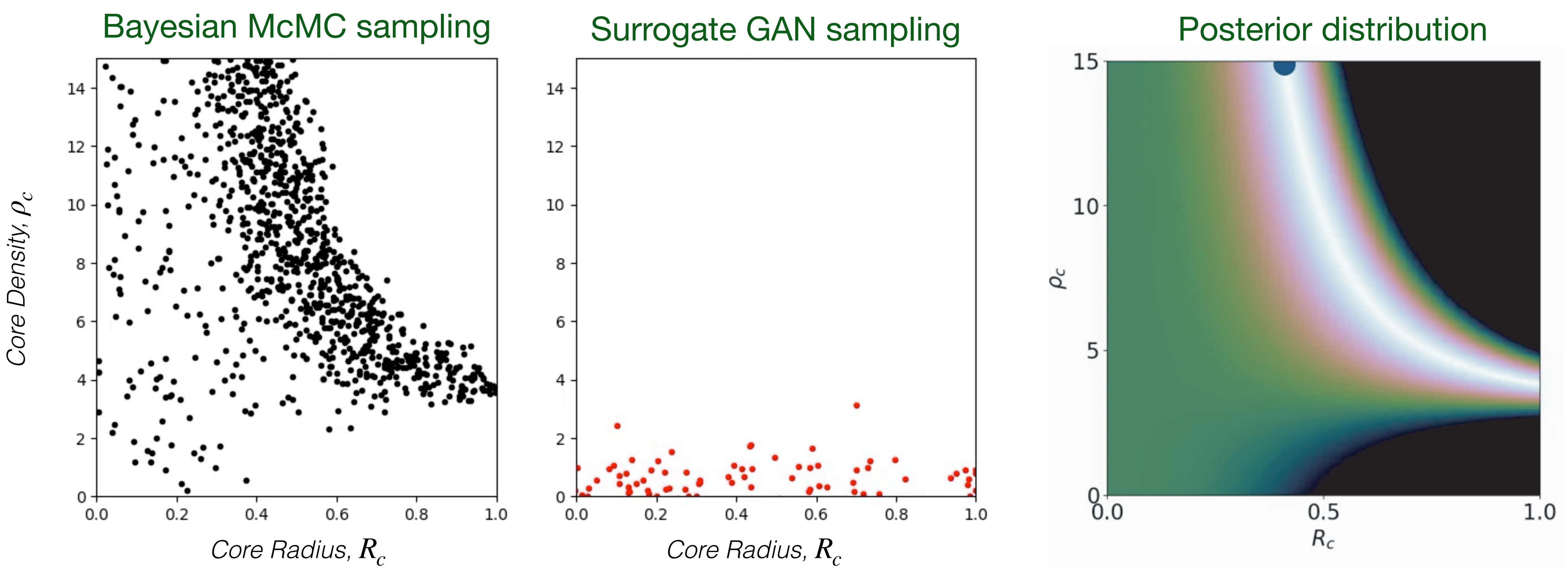
$$M_E = \frac{4}{3}\pi(R_c^3\rho_c + (R_E^3 - R_c^3)\rho_s)$$



A nonlinear, under-determined inference problem.

From the Ph.D. studies of Matthias Scheiter

Surrogate Bayesian sampling



Data: $M_E = 16$ Unknowns: (R_c, ρ_c)

Adversarial balance between Generator and Discriminator drives the GAN to replicate McMC sampling

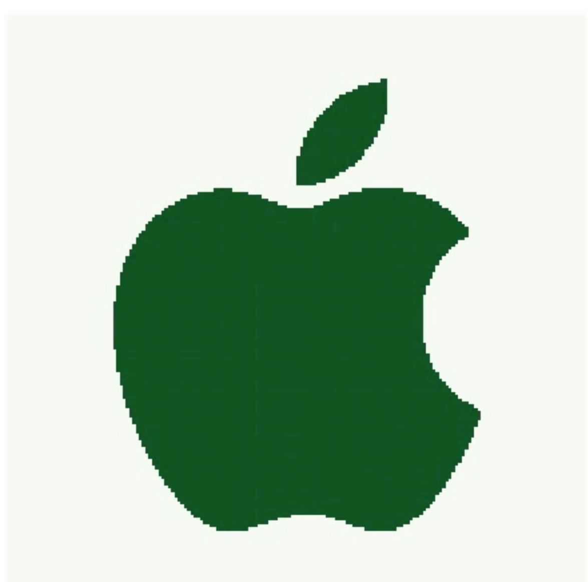
From the Ph.D. studies of Matthias Scheiter

A visit to Optimal Transport

A mathematical topic that originated in the 19th century that has yielded two Fields medals and a Nobel prize. A vast literature exists, from mathematics to computer science.

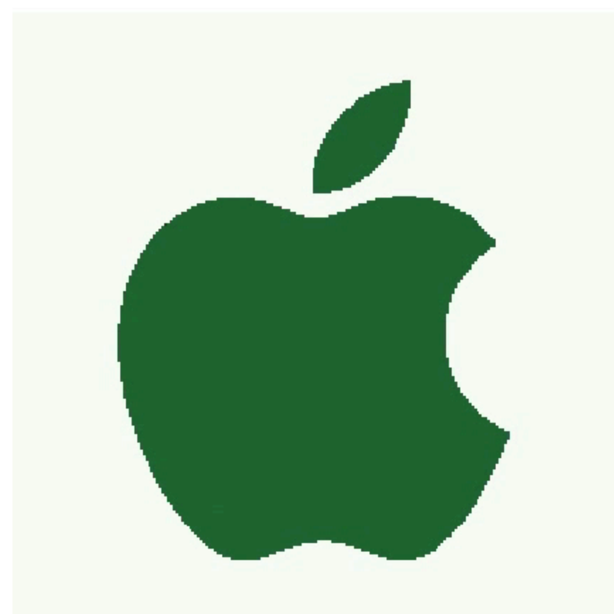
Introduced into Geophysics by Engquist and Froese (2014).

Source



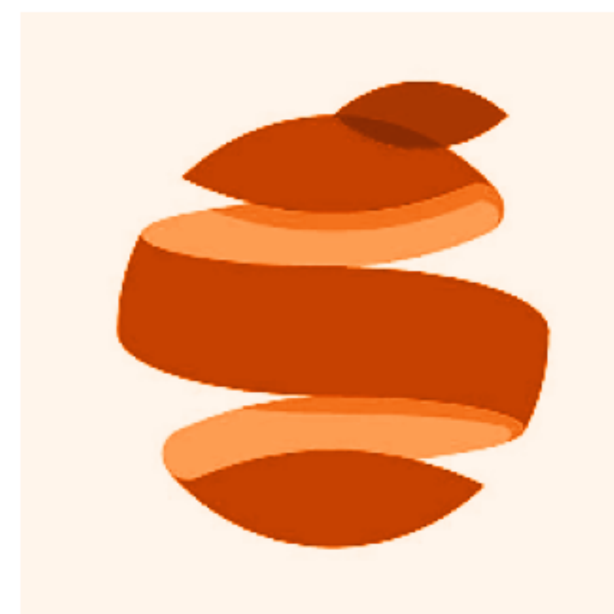
Linear

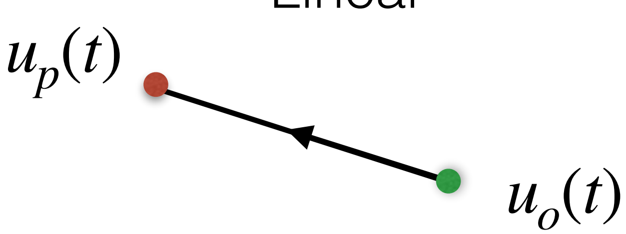
Source



Optimal transport

Target



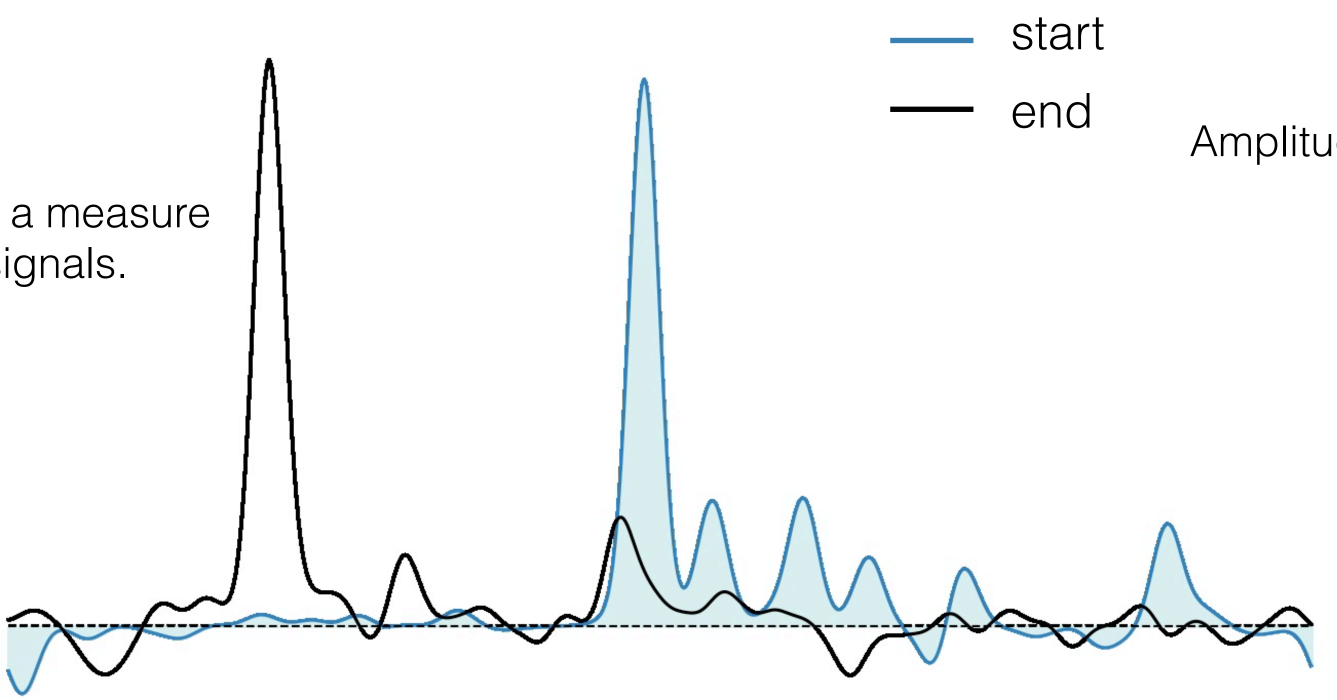


Makes use of Sinkhorn Convolutional Wasserstein algorithm of Solomon et al. (2015)
Rémi Flamary and Nicolas Courty, POT Python Optimal Transport library, 2017. https://pythonot.github.io/

Waveform misfits: Least Squares and OT

Transformations between two seismic waveforms.

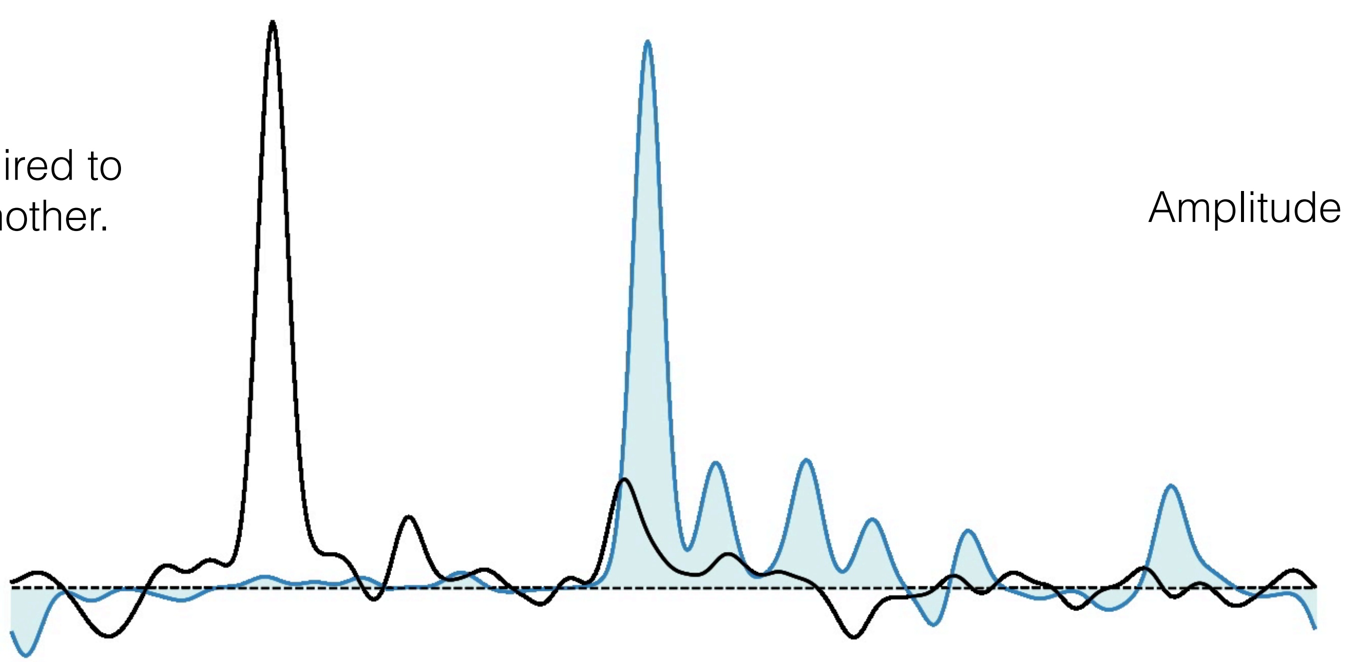
Uses the vertical separation as a measure of distance between two signals.



Amplitude, u changes at fixed time, t

$$\int (u_{obs}(t) - u_{pred}(t))^2 dt$$

Uses the minimum work required to transform one signal onto another.



Amplitude, u and time, t change together.

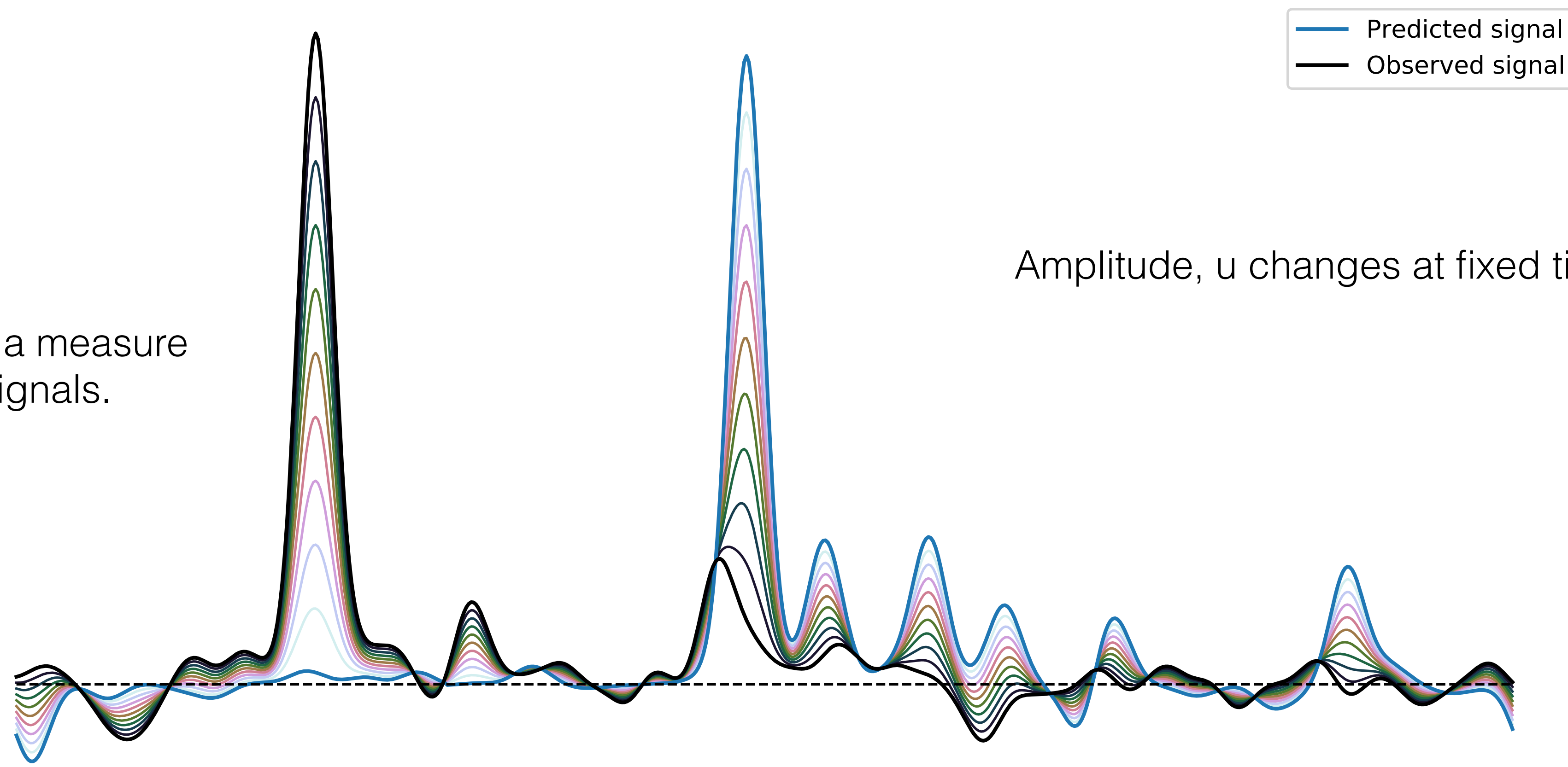
Joint work with Andy Jackson and Andrew Valentine.

Measuring the distance between complex objects

Least squares

Uses the vertical separation as a measure of distance between two signals.

Amplitude, u changes at fixed time, t



Optimal transport

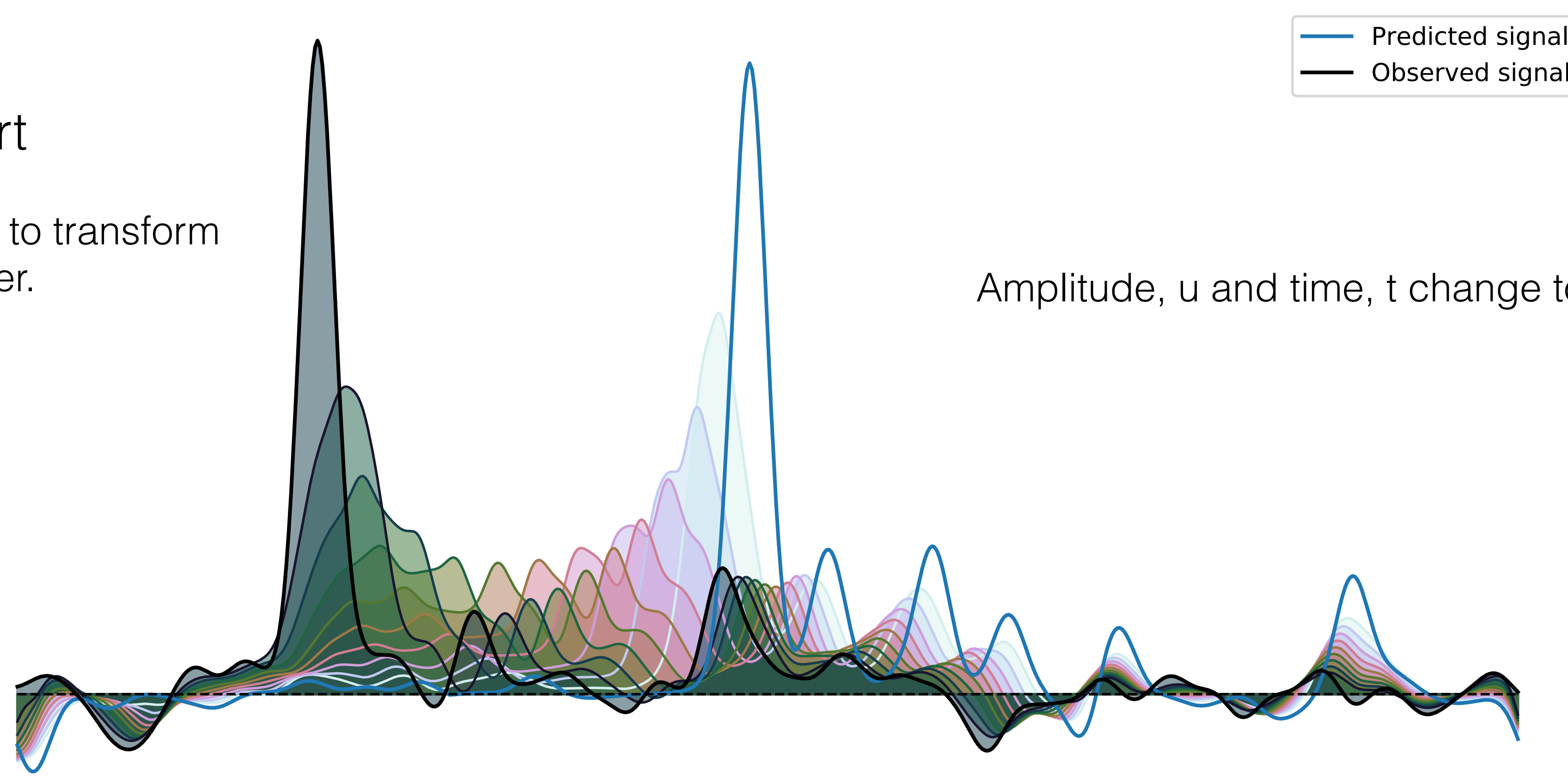
Uses the minimum work required to transform one signal onto another.

W_1 = distance x mass

W_2^2 = $(\text{distance})^2$ x mass

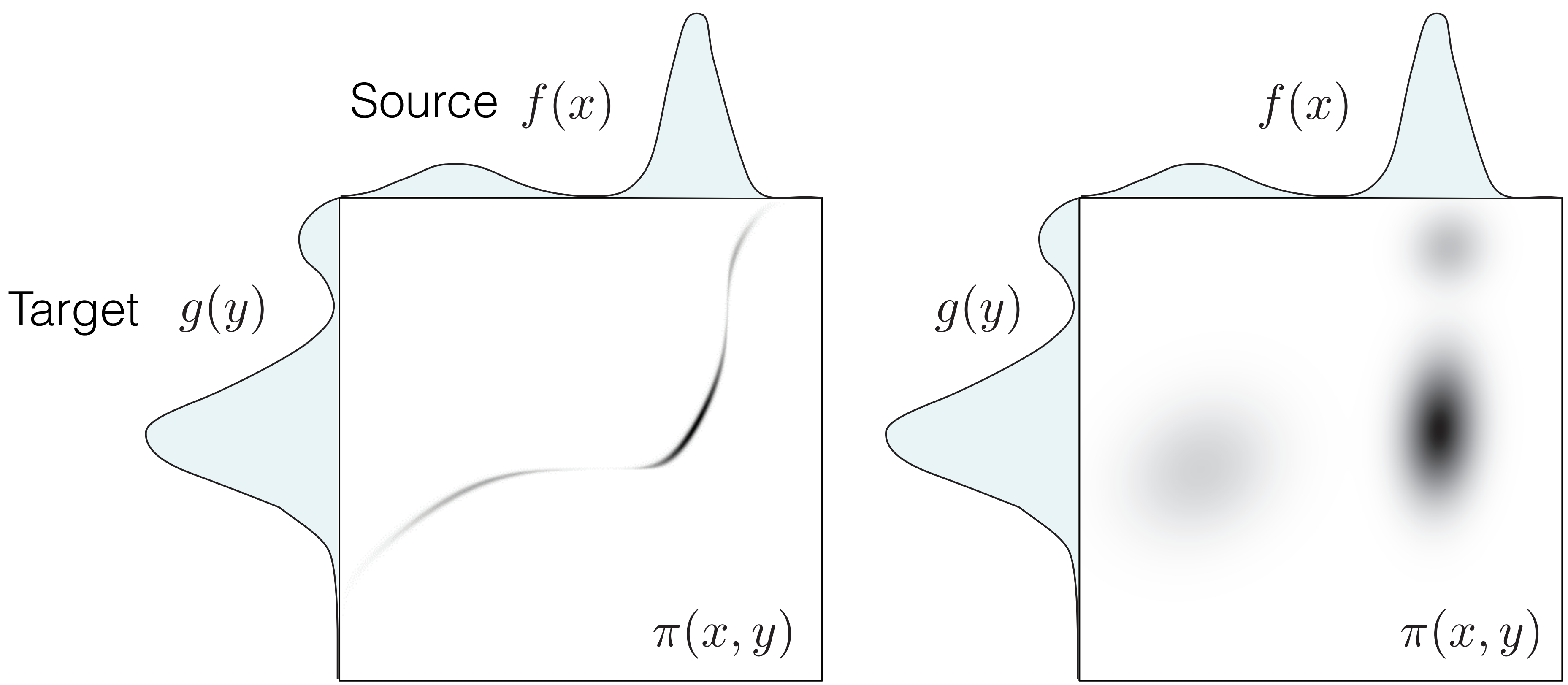
W_p^p = $(\text{distance})^p$ x mass

Amplitude, u and time, t change together.



Optimal transport maps one PDF onto another

Linear programming formulation of Kantorovich (1942). Solve for transport plan $\pi_{i,j}$



$$\min_{\pi(x_i, y_j)} W_p^p = \sum_{i,j} c_{i,j}\pi_{i,j}, \quad \sum_i \pi_{i,j} = f(x_i), \quad \sum_j \pi_{i,j} = g(y_i)$$

W_1 = distance x mass

$\pi(x, y)$ = Transport plan

$c(x, y)$ = distance between x and y

$$W_p^p = (\text{Distance})^{\text{p}} \times (\text{mass})$$

W_2^2 = (distance)^2 x mass

Figure adapted from Solomon et al. (2015)

Optimal transport in seismic waveform inversion

A number of groups have applied variants of OT in geophysics, primarily to full waveform inversion (FWI) in exploration seismology.

Engquist and Froese (2014); Engquist et al. (2016); Yang and Engquist (2018); Yang et al. (2018); - Monge-Ampere PDE solver (p=2)

Me´tivier et al. (2016 a,b,c,d); Me´tivier et al. (2018 a,b); Me´tivier et al. (2019); Yong et al. (2018) - Dual formulation optimisation (p=1)

Hedjazian et al. (2019) - Seismic receiver functions; *Huang et al. (2019)* - Gravity inversion.

Books and lecture notes:

Villani (2003, 2008); Ambrosio (2003); Santambrogio (2015).

Approaches differ between studies:

- Solution method for Wasserstein distance, W_p, and also p value.
- Transform of seismic trace to a Probability Density Function (PDF).
- 1D OT Trace by trace or 2D reflection image.

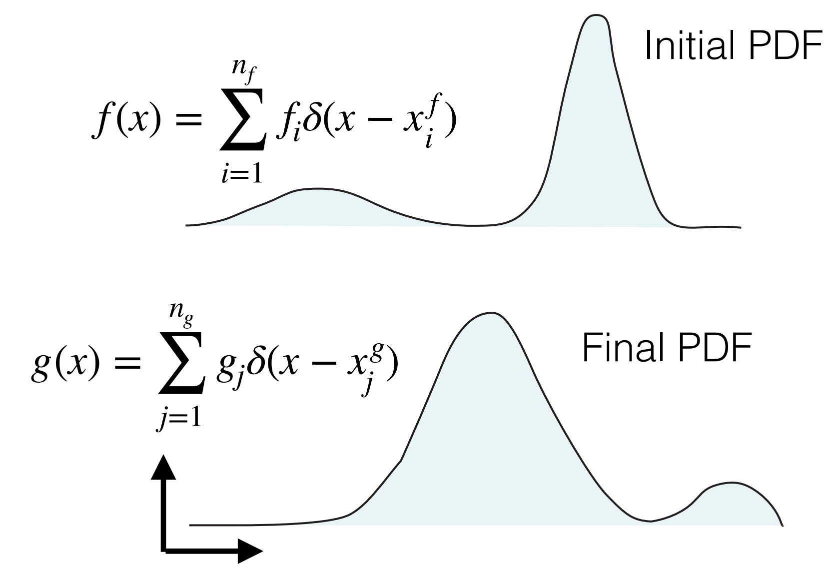
These are all **open** issues.

OT solutions in 1D

Analytical solution for 1D continuous case in terms of inverse CDFs (Villani, 2003)

$$W_p(f,g) = \int_0^1 |F^{-1} - G^{-1}|^p \, dy$$

Point mass representation



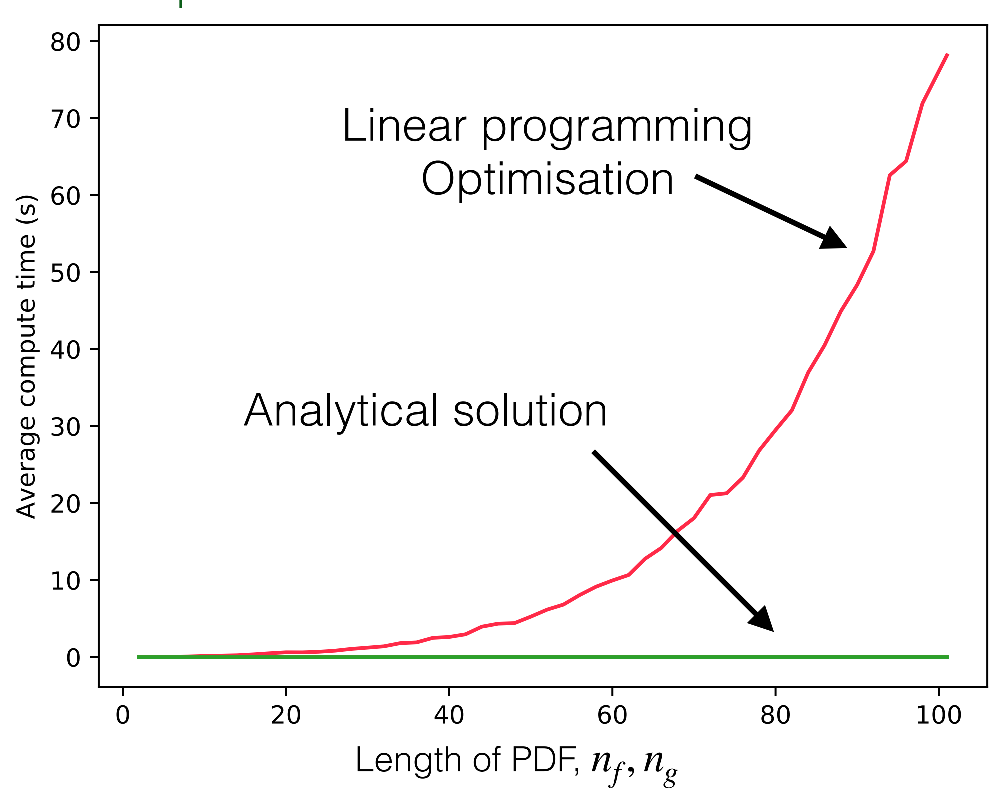
Our 1D discrete solution

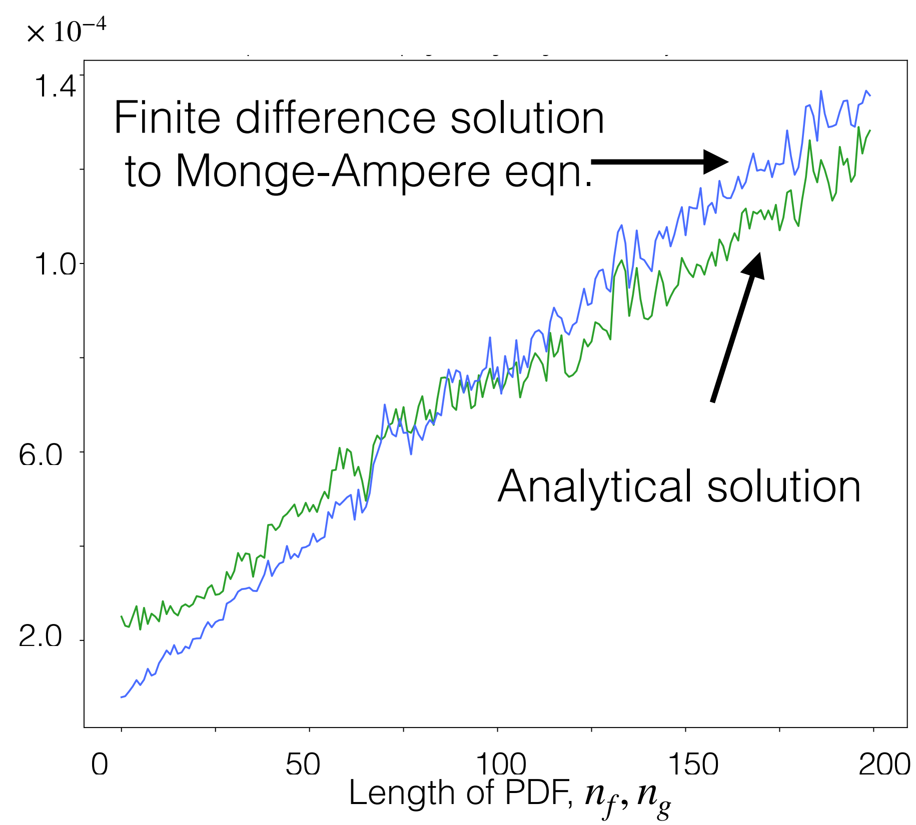
$$\boxed{W_p(f,g) = [\Delta \mathbf{z}^T \Delta \mathbf{y}]^{1/p}}$$

where $\mathbf{z}(x^f, x^g)$ depends only on point mass locations and $\mathbf{y}(f_i, g_j)$ depends only on point mass weights.

Exact for any p. Requires just one sort and a vector dot product. Exact derivatives available, $\partial W_p / \partial f_i$

Relative computational costs





Averaged compute time
over 10 random
1D problems

How to convert a waveform into a PDF?

Many approaches have been applied to turn a time series into a positive PDF.

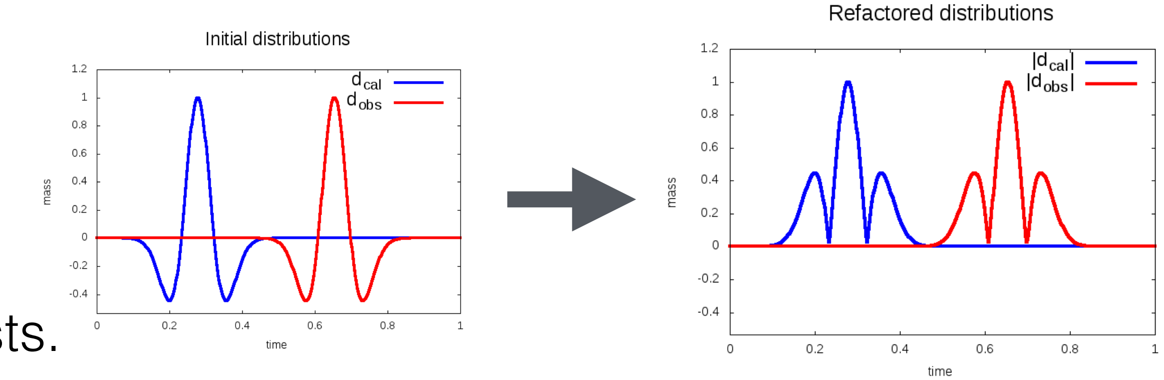
- Addition method: Add positive constant to f(x) and g(x)

$$\tilde{W}_p(f,g) := W(f+\alpha, g+\alpha)$$

Issues: Loss of convexity w.r.t. time shifts; Transformation becomes local. => Reject.

- Take absolute values. Quite common solution.

Issues: Loss of polarity information in signal.
In FWI results in no sensitivity to impedance contrasts.



$$\tilde{W}_p(f,g) := W(|f|,|g|)$$

- Like with like: Separately transform +ve to +ve and -ve to -ve.

Issues: Artificial decorrelation between +ve and -ve parts. Loss of mass conservation.

- Global strategy: Mix +ve and -ve parts between f and g.

$$\tilde{W}_p(f,g) := W_p(f^+ + g^-, f^- + g^+)$$

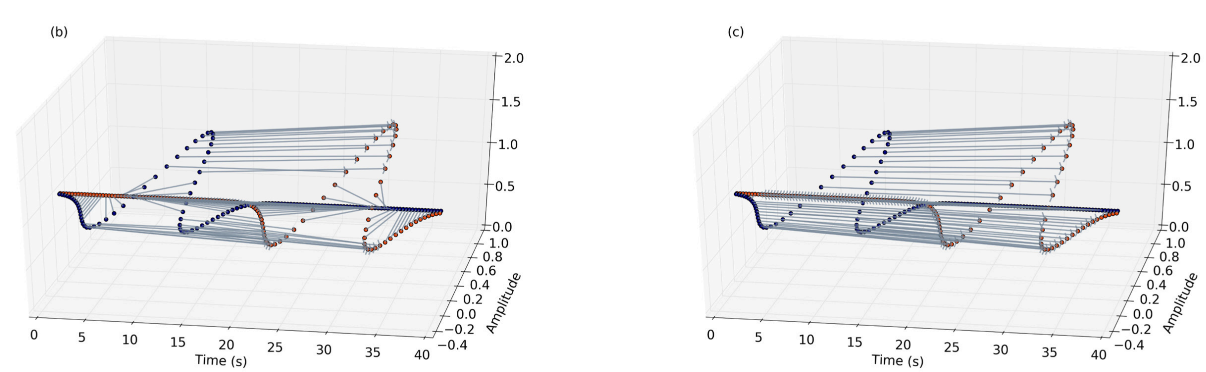
Mainini (2012)

Issues: Ensures preservation of mass conservation;
If time signals are separated in time will map f+ to f- and g+ to g-.

- Represent waveforms 2D point clouds :

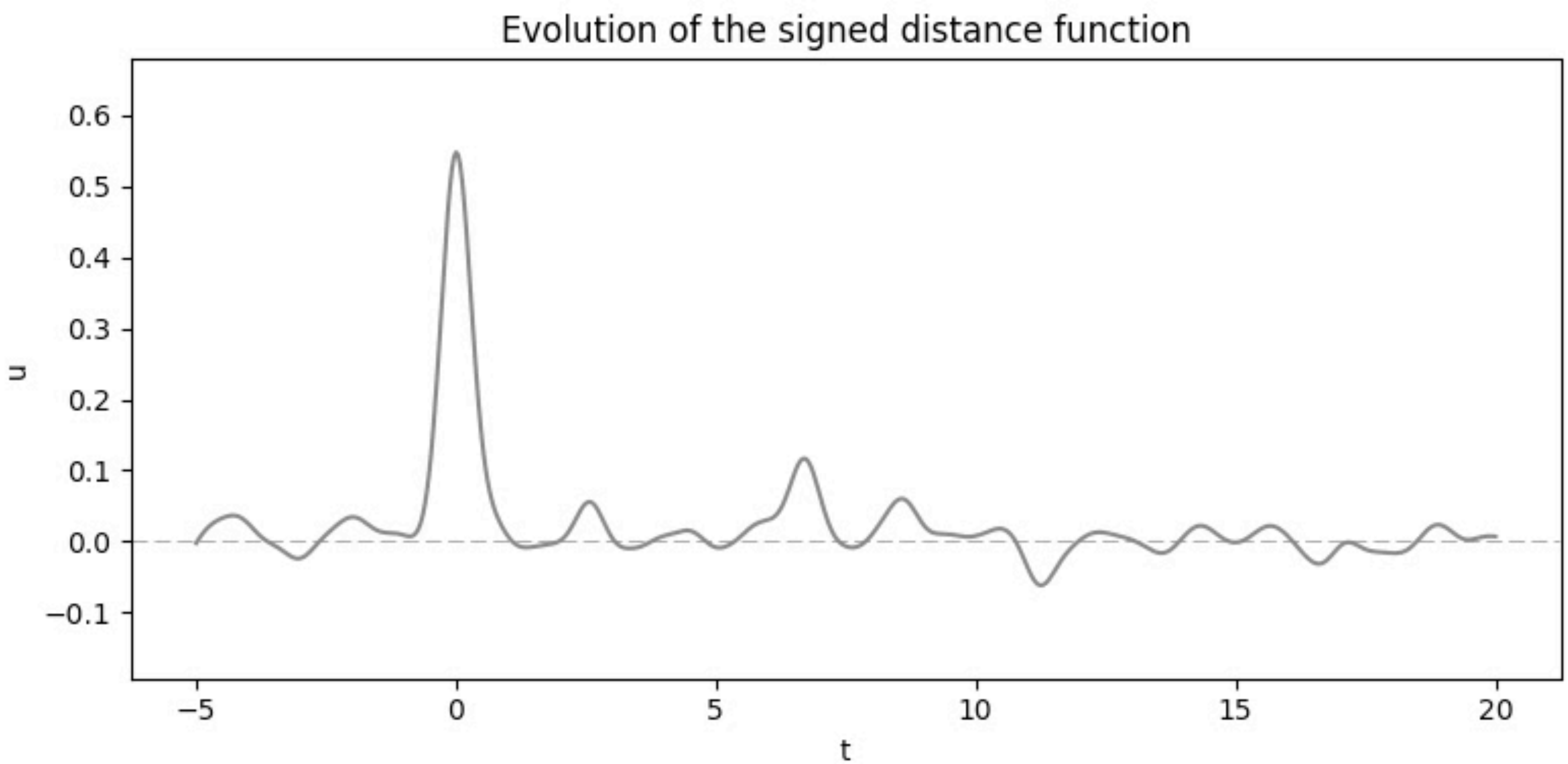
Metivier et al. (2018)

Can map signal zero line.
Outcome depends on distance metric



Our way: Create a 2D `Fingerprint' from 1D waveform

Step 1: Rather than treating the +ve and -ve parts of a time series, $u(t)$, differently, we create a 2-D positive function, $d(u, t)$, which is the minimum distance from (u, t) to the waveform.

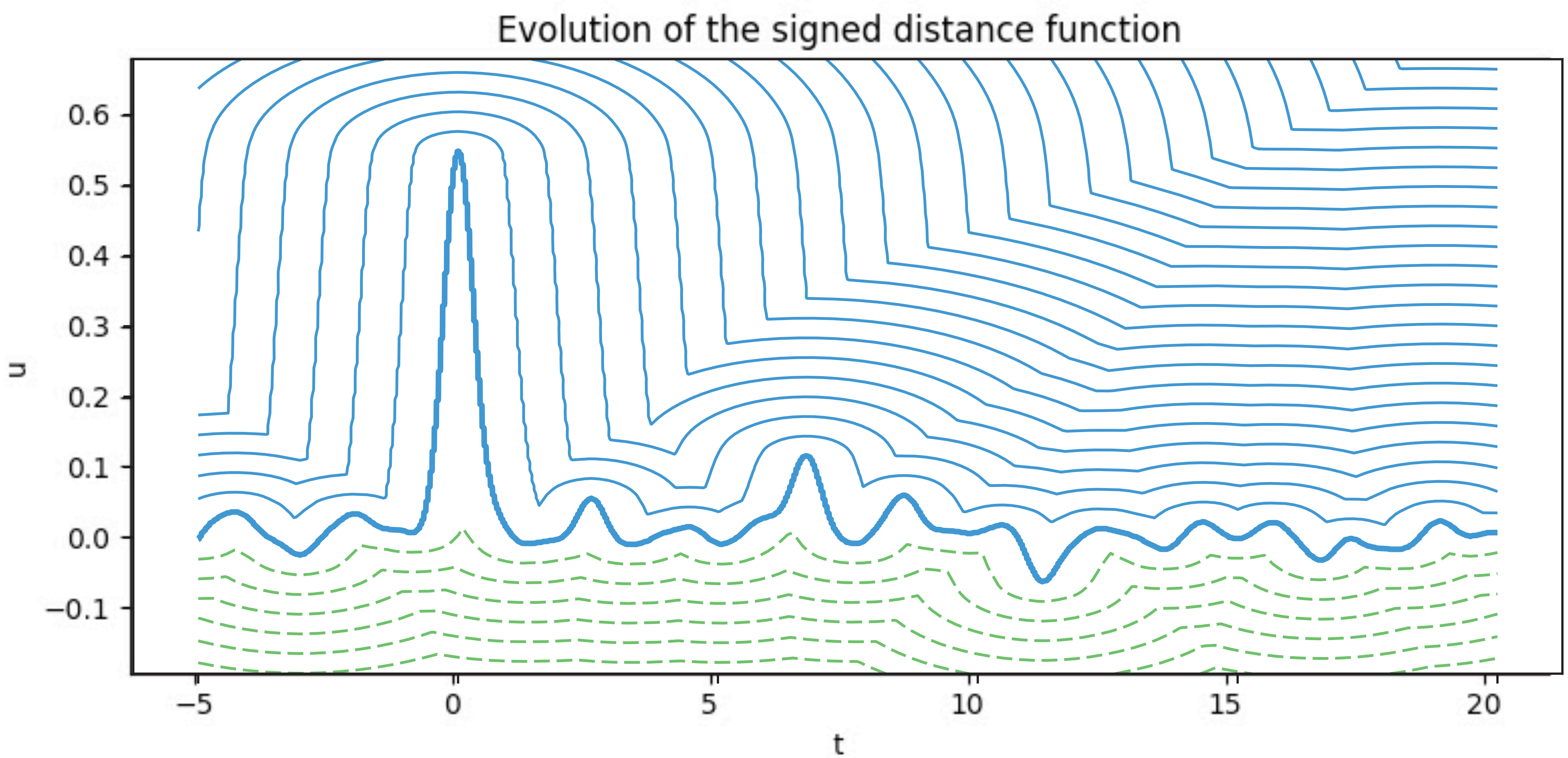


Seismologists will recognise this calculation because it is identical to that of propagating a seismic wavefront, initially in position $u(t)$, both upward and downward according to Huygens' principle (geometric ray theory) in a medium with constant velocity.

Efficiently done with the Fast Marching method of Sethian (1996,1999); Sethian and Popovici, (1999); Rawlinson and Sambridge, (2004).

Our way: Create a 2D `Fingerprint' from 1D waveform

Step 1: Rather than treating the +ve and -ve parts of a time series, $u(t)$, differently, we create a 2-D positive function, $d(u,t)$, which is the minimum distance from (u,t) to the waveform.

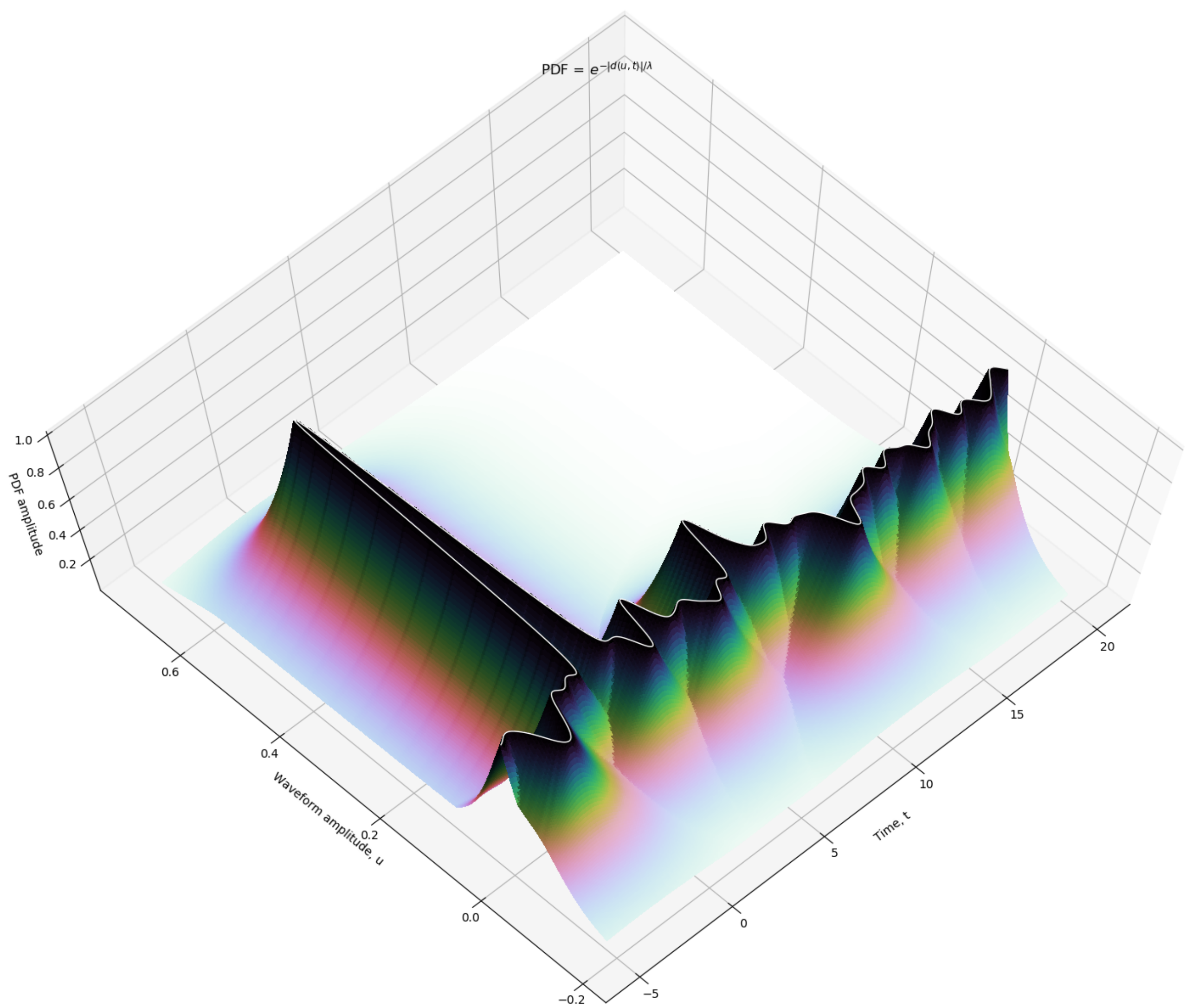


Seismologists will recognise this calculation because it is identical to that of propagating a seismic wavefront, initially in position $u(t)$, both upward and downward according to Huygens' principle (geometric ray theory) in a medium with constant velocity.

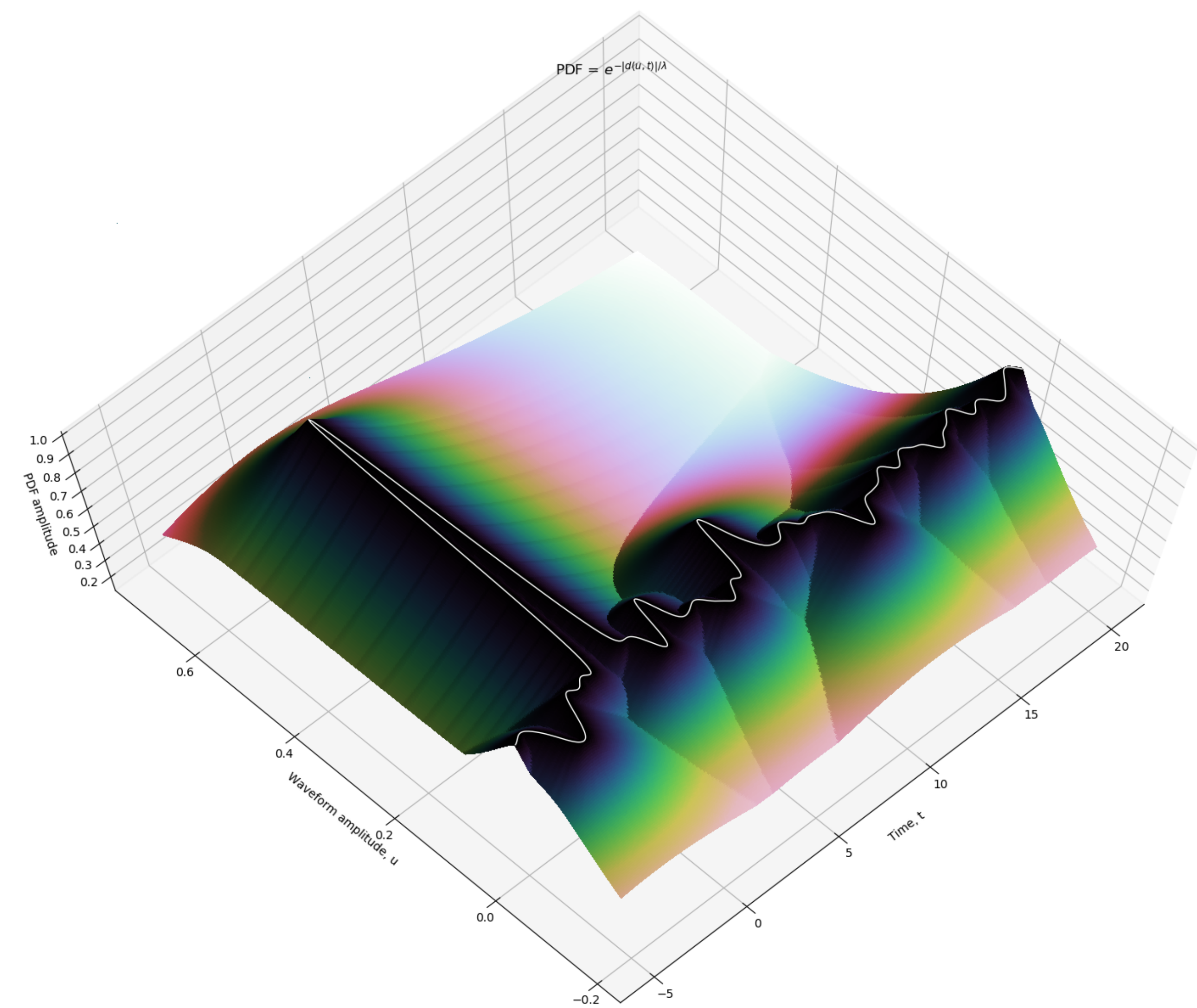
Efficiently done with the Fast Marching method of Sethian (1996,1999); Sethian and Popovici, (1999); Rawlinson and Sambridge, (2004).

Our way: Create a 2D `Fingerprint' from 1D waveform

Step 2: Take the exponential of the distance function, $\phi(u,t) = e^{-d(u,t)/\lambda}$



Small λ

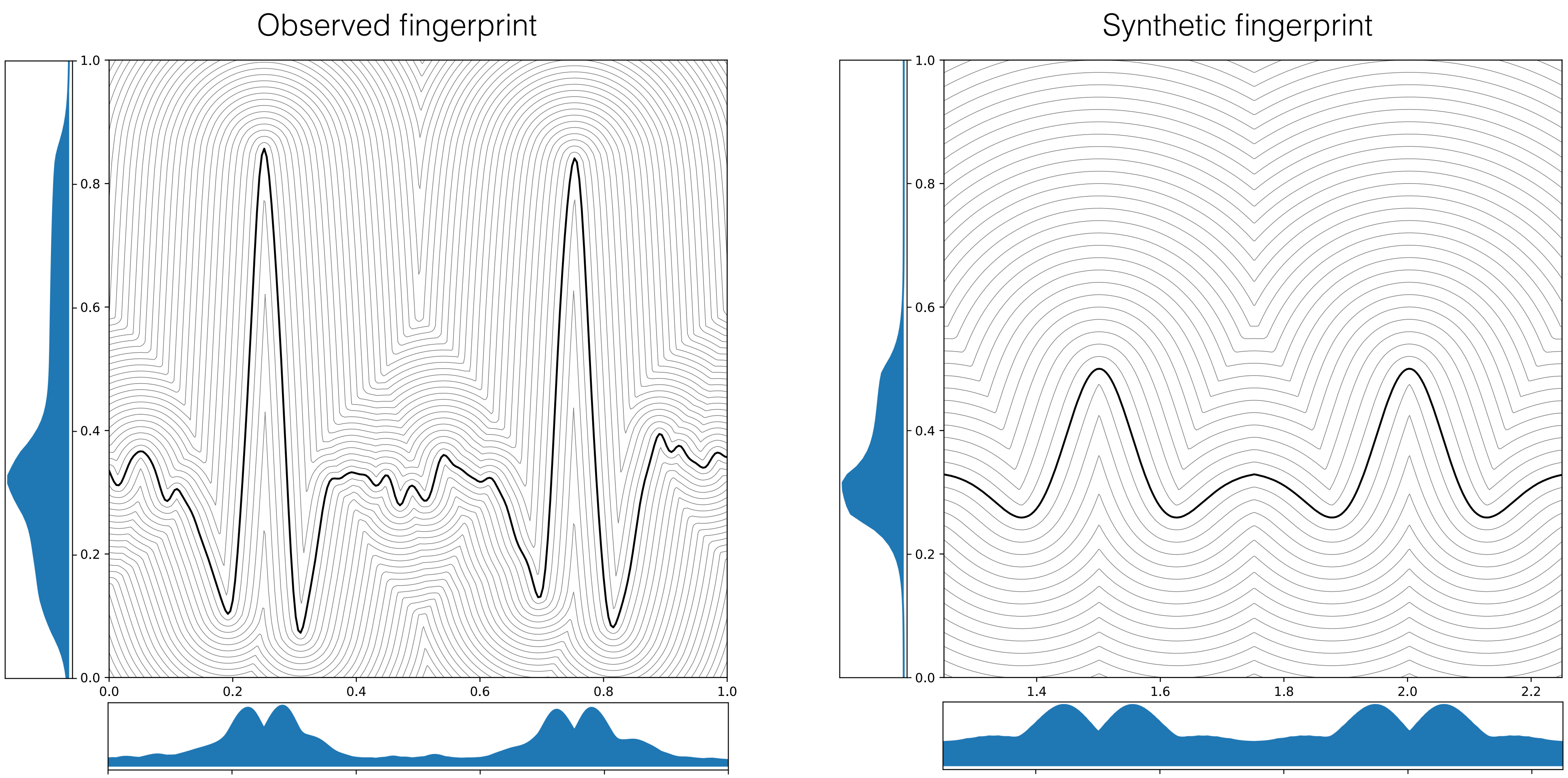


Large λ

Marginal Wasserstein in 2D

Our waveform misfit becomes the Wasserstein distance between observed and synthetic PDFs

We sum over each axis, and average Wasserstein distances between 1D Marginals.

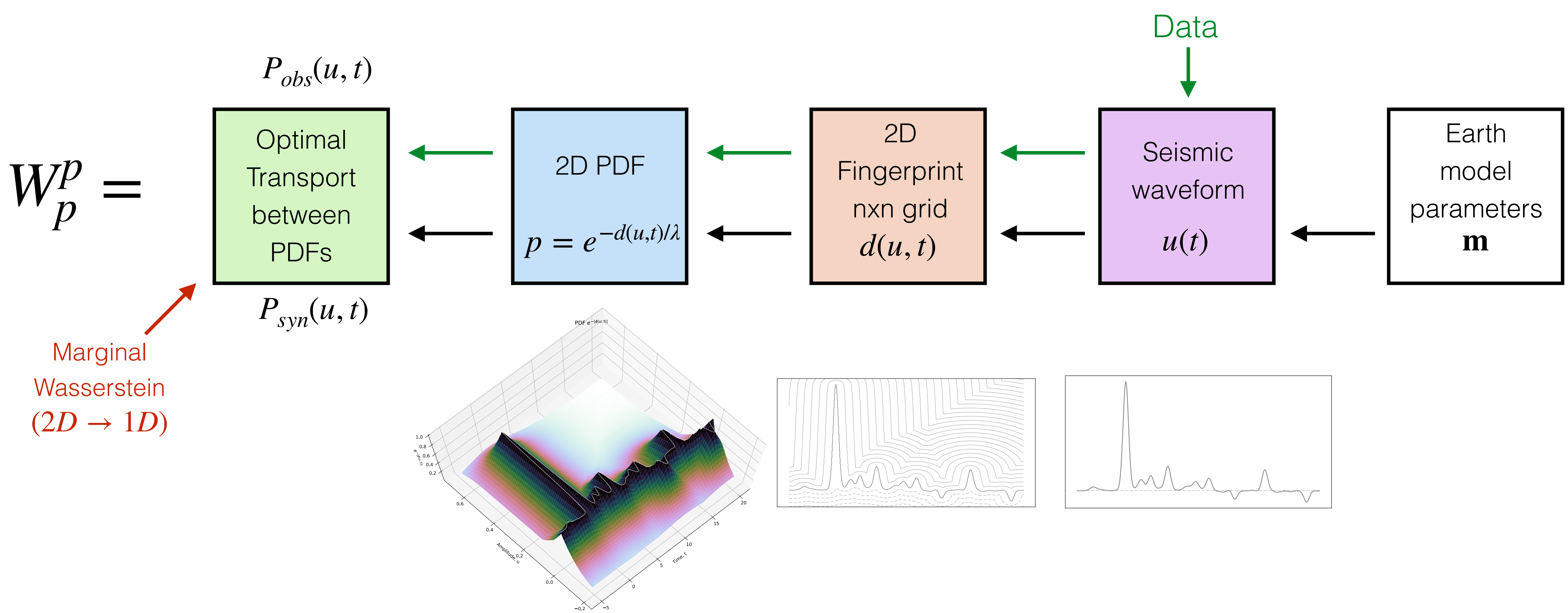


Advantages:

- Takes advantage of 1D analytical OT solutions.
- Handles different time windows about predicted and observed waveforms.
- Computational cost scales with n rather than n^2 for $n \times n$ grid. Faster than Sliced Wasserstein with similar results.
- Derivatives $\partial W_p^p / \partial u$ can be calculated.

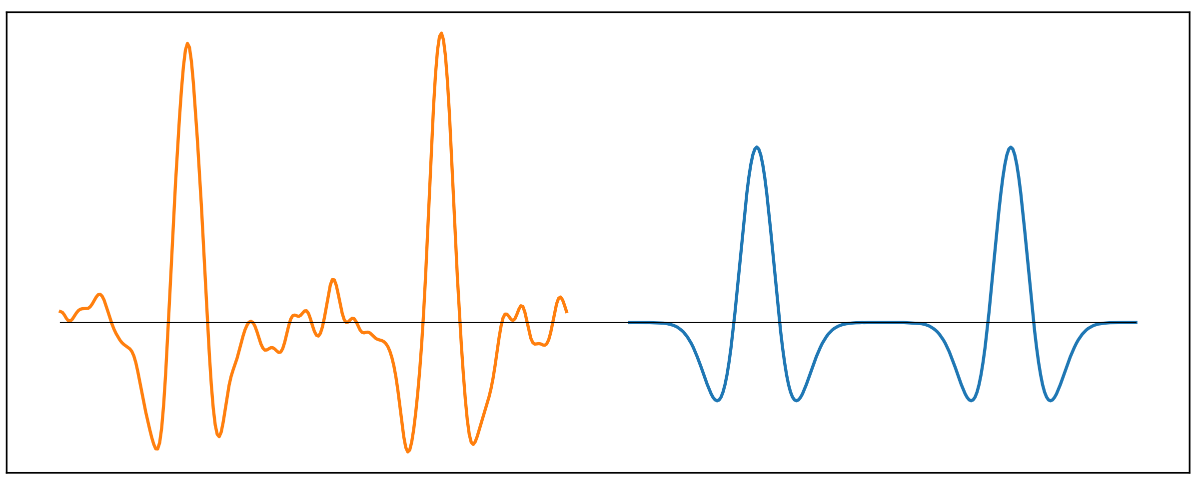
Computation of the Wasserstein distance between seismic fingerprints

Breaking down the calculation into 4 steps:

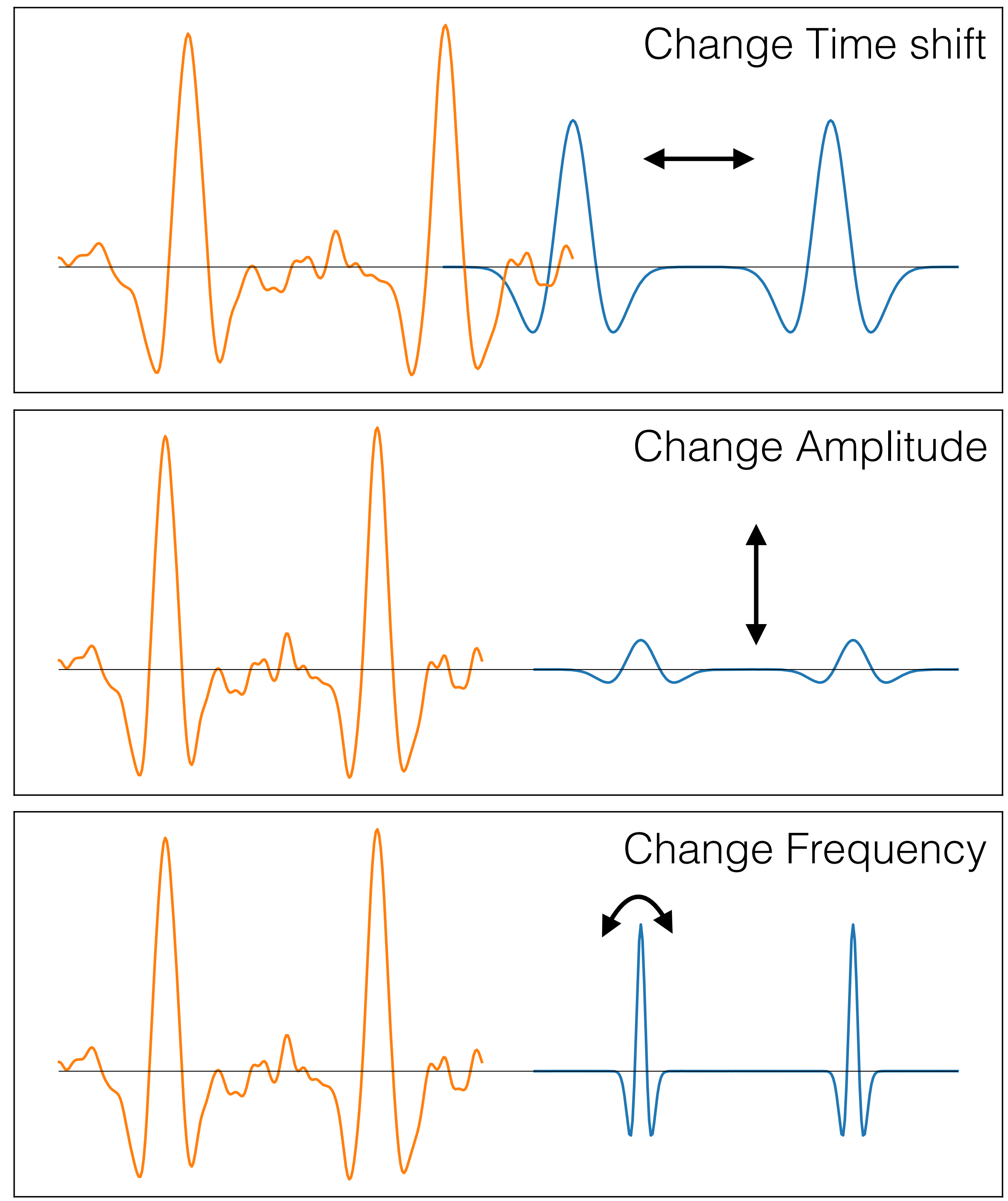


A toy problem: Double Ricker wavelet fitting

Fit the noisy waveform by adjusting three parameters



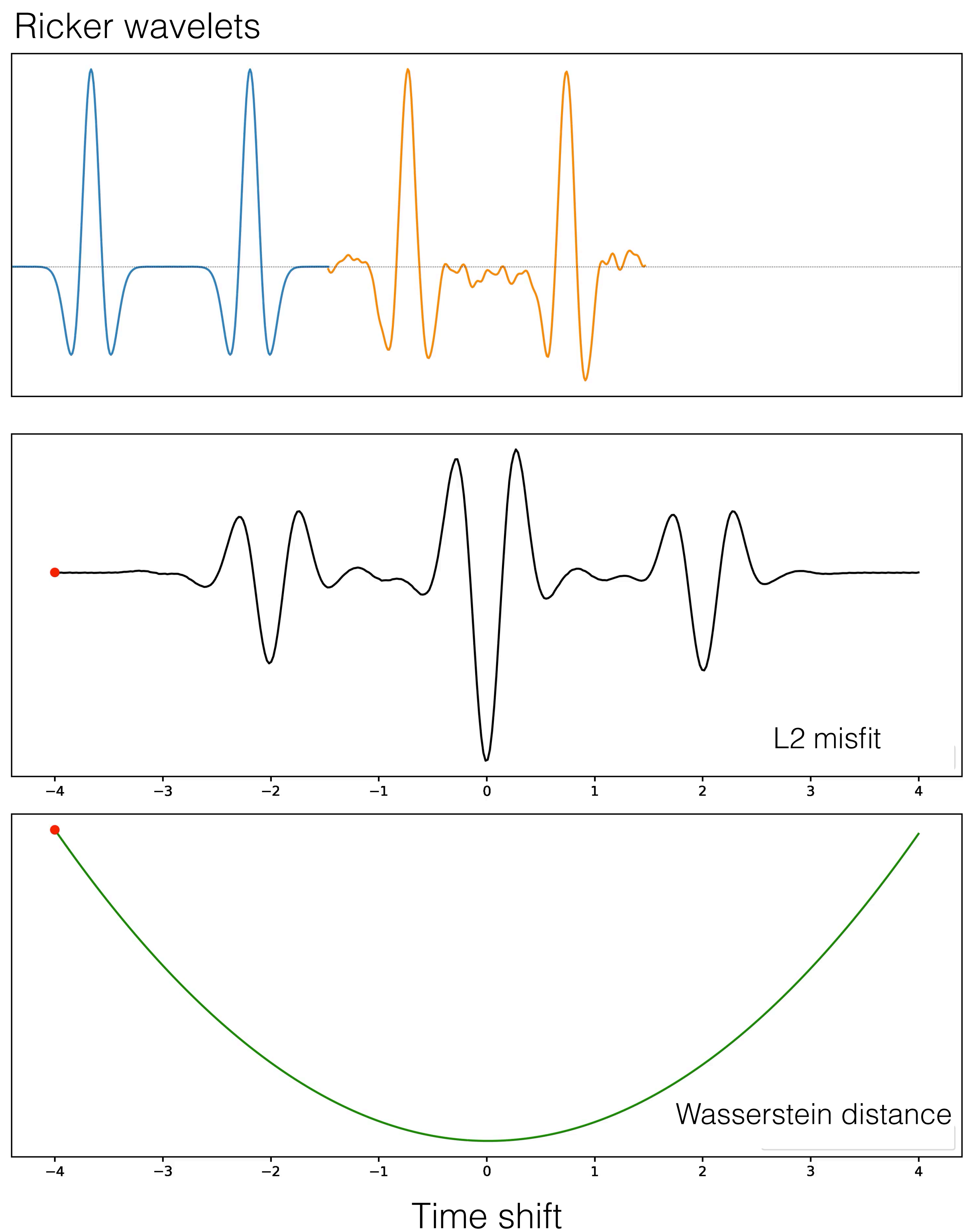
Observed Predicted



Noise is $N(\mu, \sigma^2)$

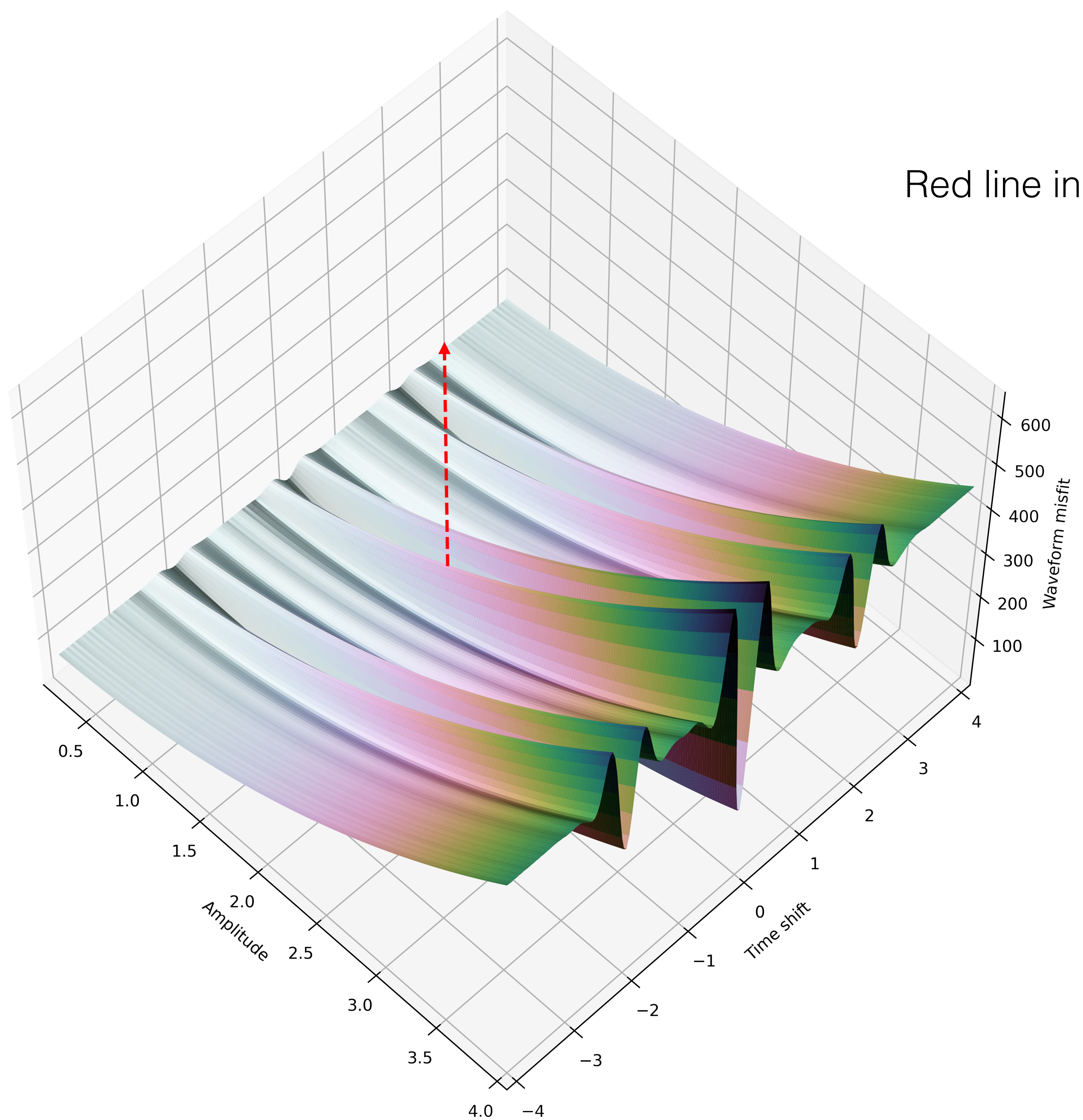
$\mu =$ 5% of maximum Ricker amplitude
$\sigma =$ 50% of maximum Ricker period

Least squares misfit and Wasserstein distance between a pair of double Ricker wavelets



After Engquist and Froese (2014)

L2 waveform misfit surface

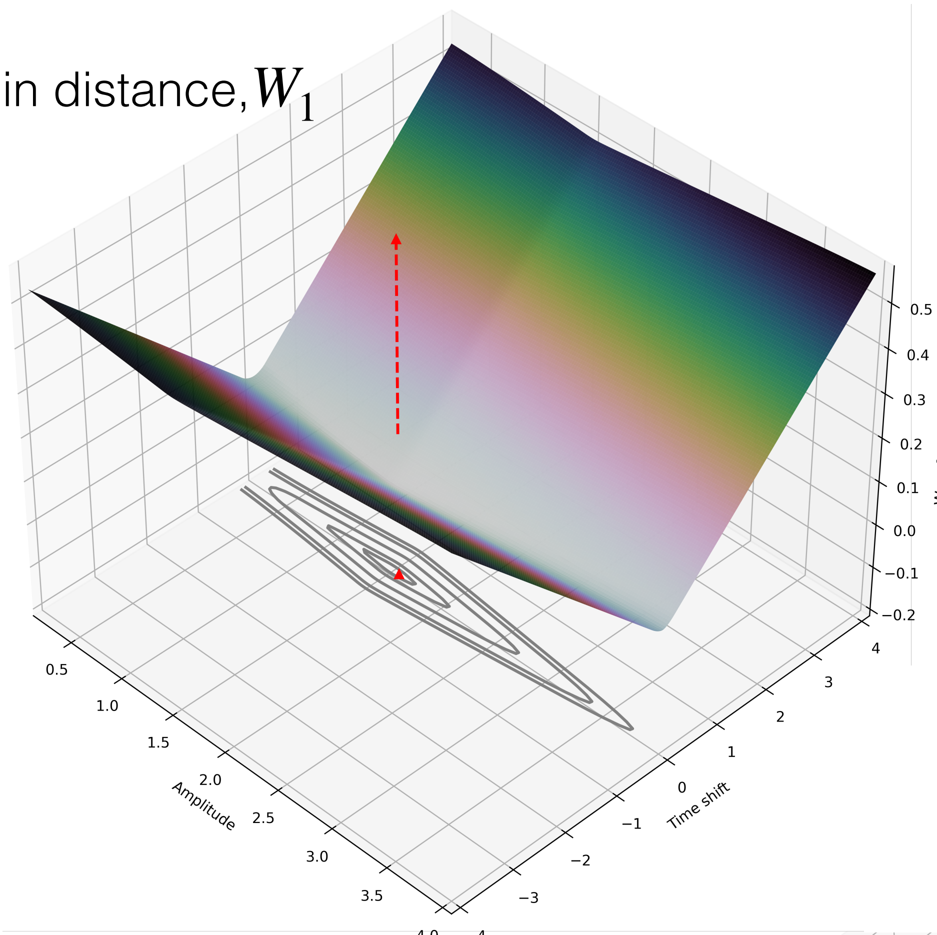


Red line indicates global minimum

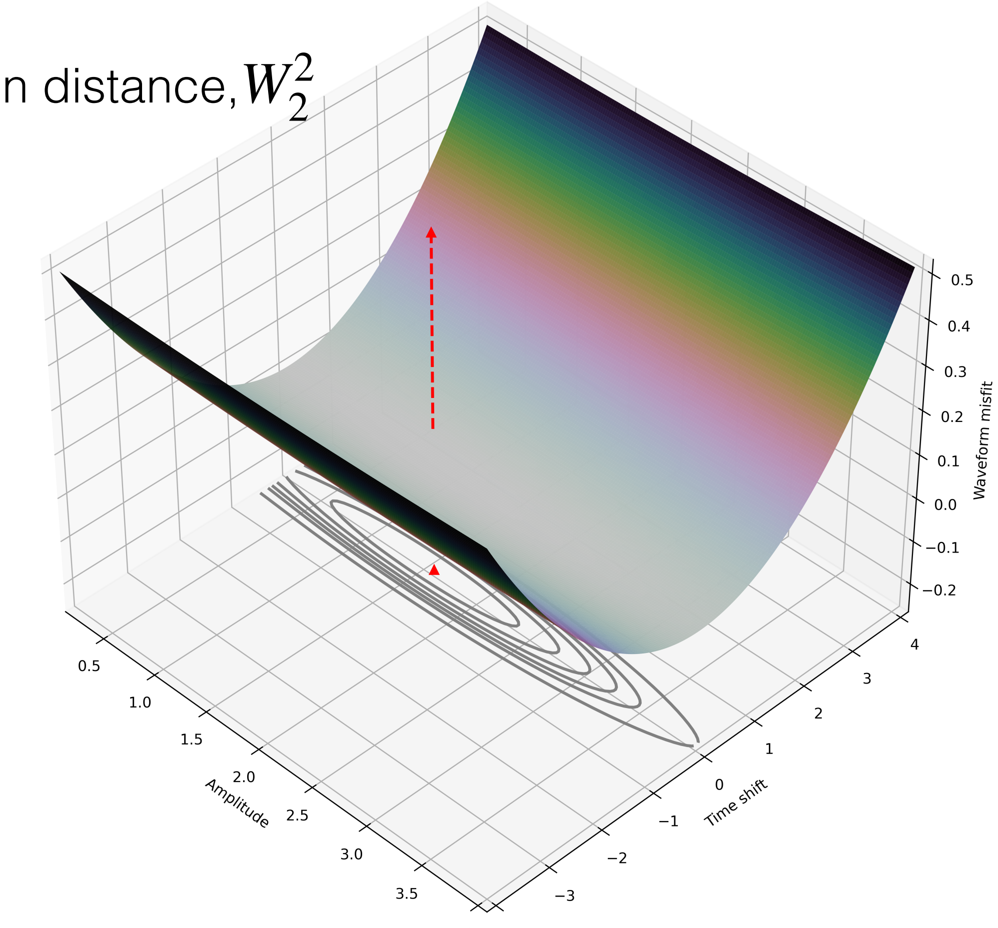
Least squares waveform misfit as a function of
Time shift and Amplitude parameters

Wasserstein and L2 waveform misfit surfaces

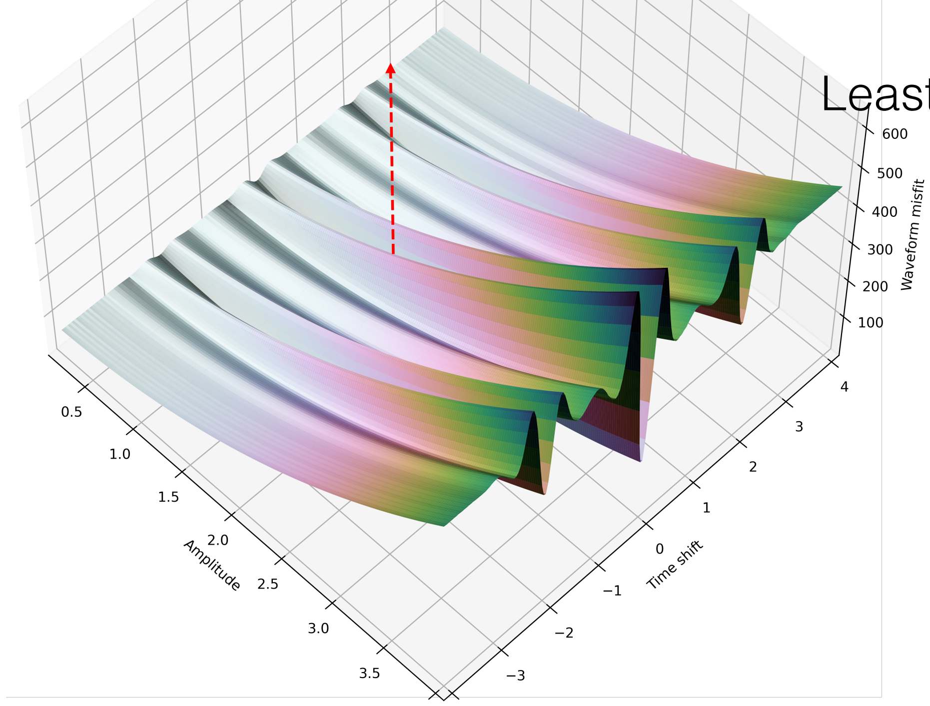
Wasserstein distance, W_1



Wasserstein distance, W_2^2



Least squares misfit, L_2



Wasserstein based on
2D fingerprint PDF

Red line indicates global minimum

Calculating derivatives of Wasserstein distance

For the Wasserstein distance to be useful within an optimisation framework it is advantageous to be able to calculate derivatives of the Wasserstein distance, W_p^p (p=1 or 2), with respect to underlying Earth model parameters.

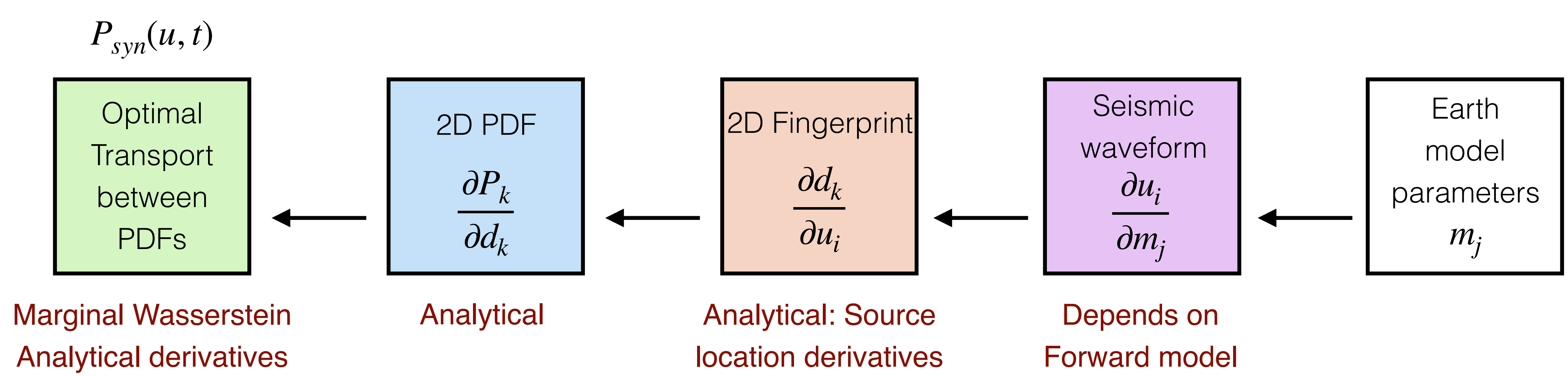
(*This is the discrete analogue of the Adjoint method in full waveform inversion with least squared waveform misfits.)*

$$\frac{\partial W_p^p}{\partial m_j}, (j = 1,\ldots, M).$$

Other authors implementing OT to FWI have all done something similar with details depending on the choices made in applying Optimal Transport, e.g. Finite Difference solution of Monge-Ampere equations (Enquist and co workers, 2014-) or constrained optimisation using the Monge-Kantorovich dual formulation (Metivier and co-workers 2016-).

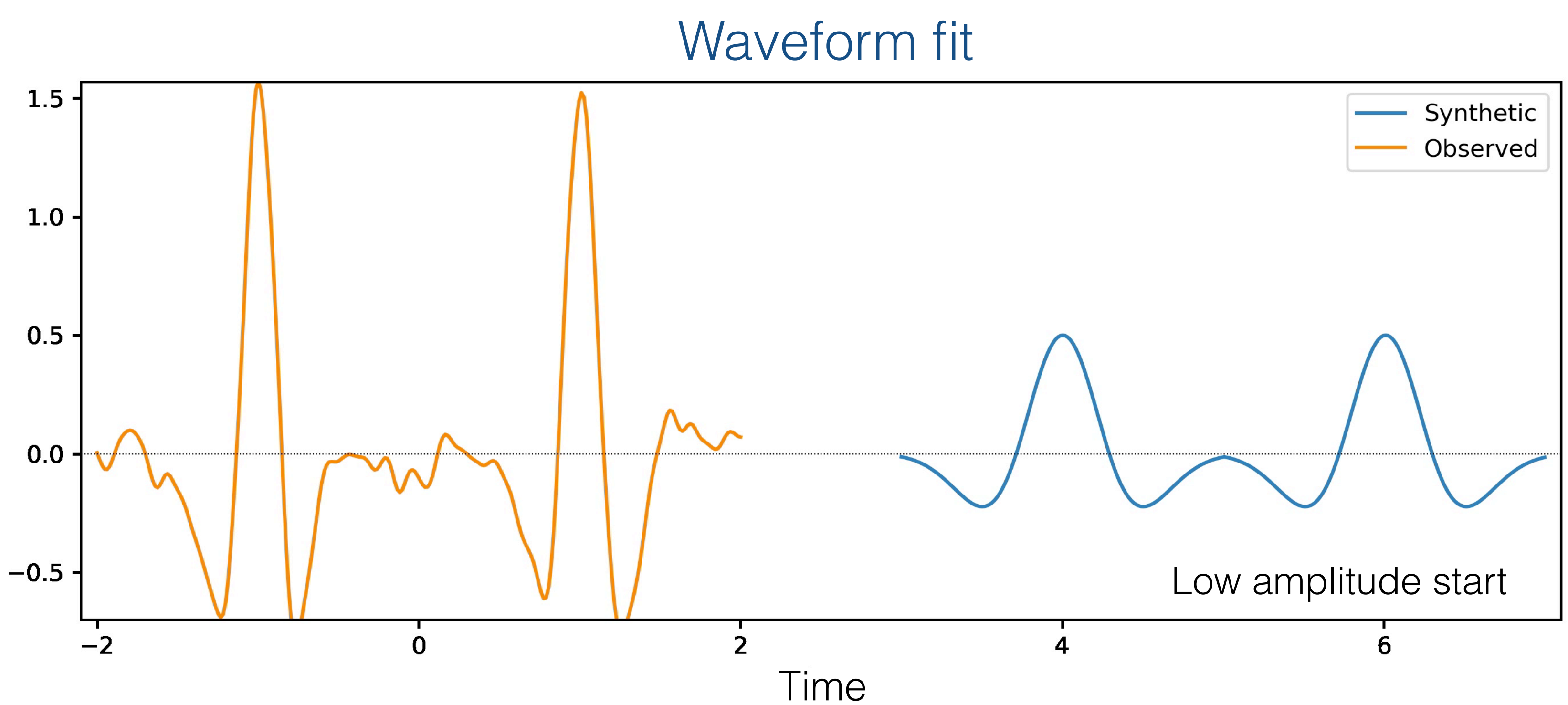
In our case

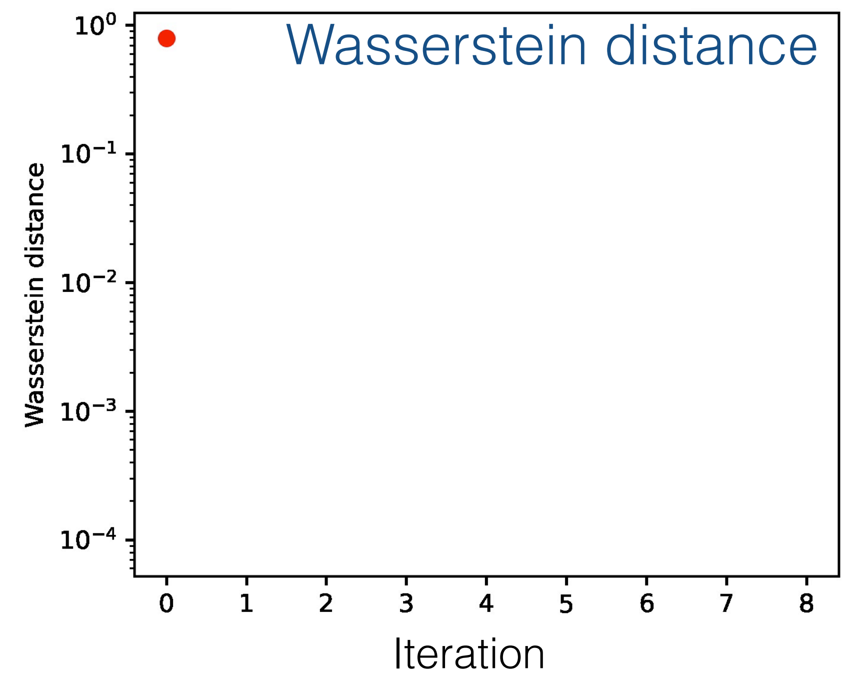
$$\frac{\partial W_p^p}{\partial m_j} =$$



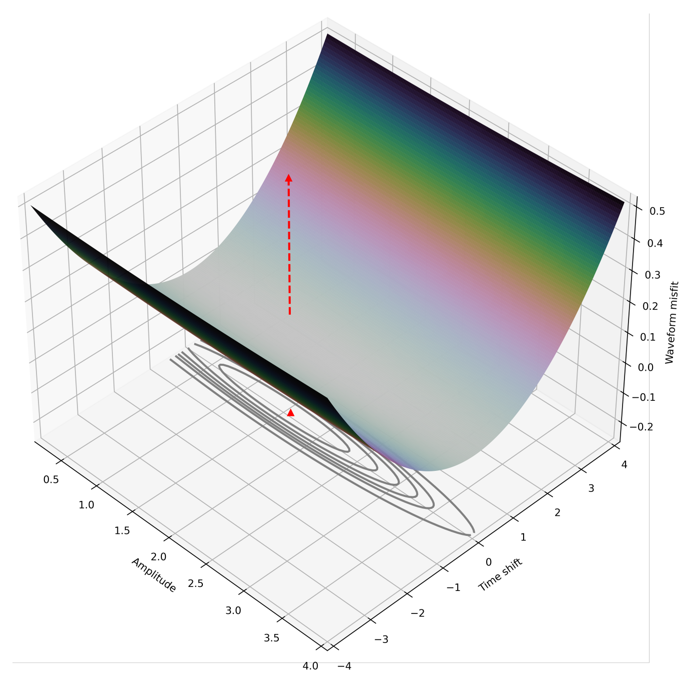
We apply the **chain rule** of partial differentiation across each intermediate set of variables.

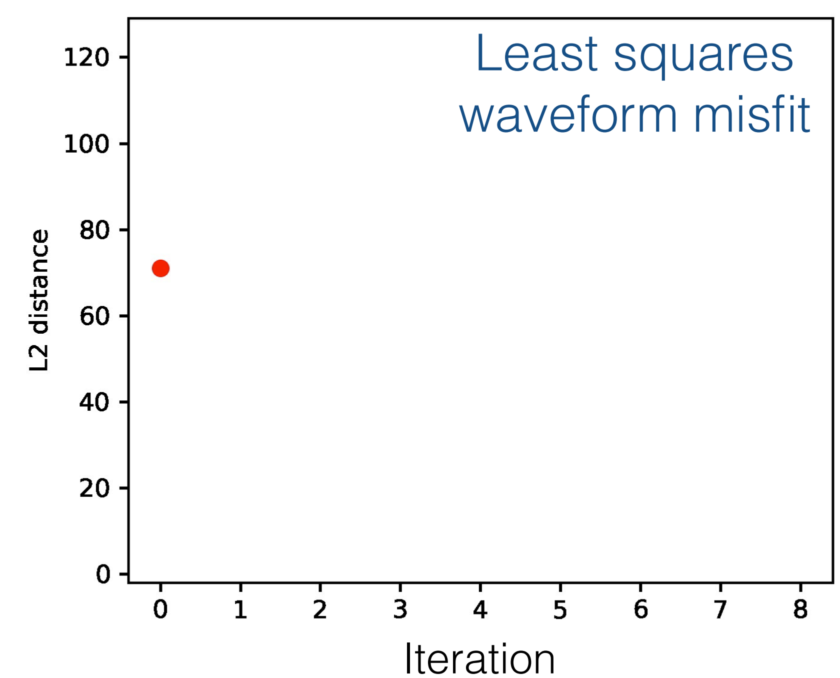
Minimizing the Wasserstein distance W_2^2



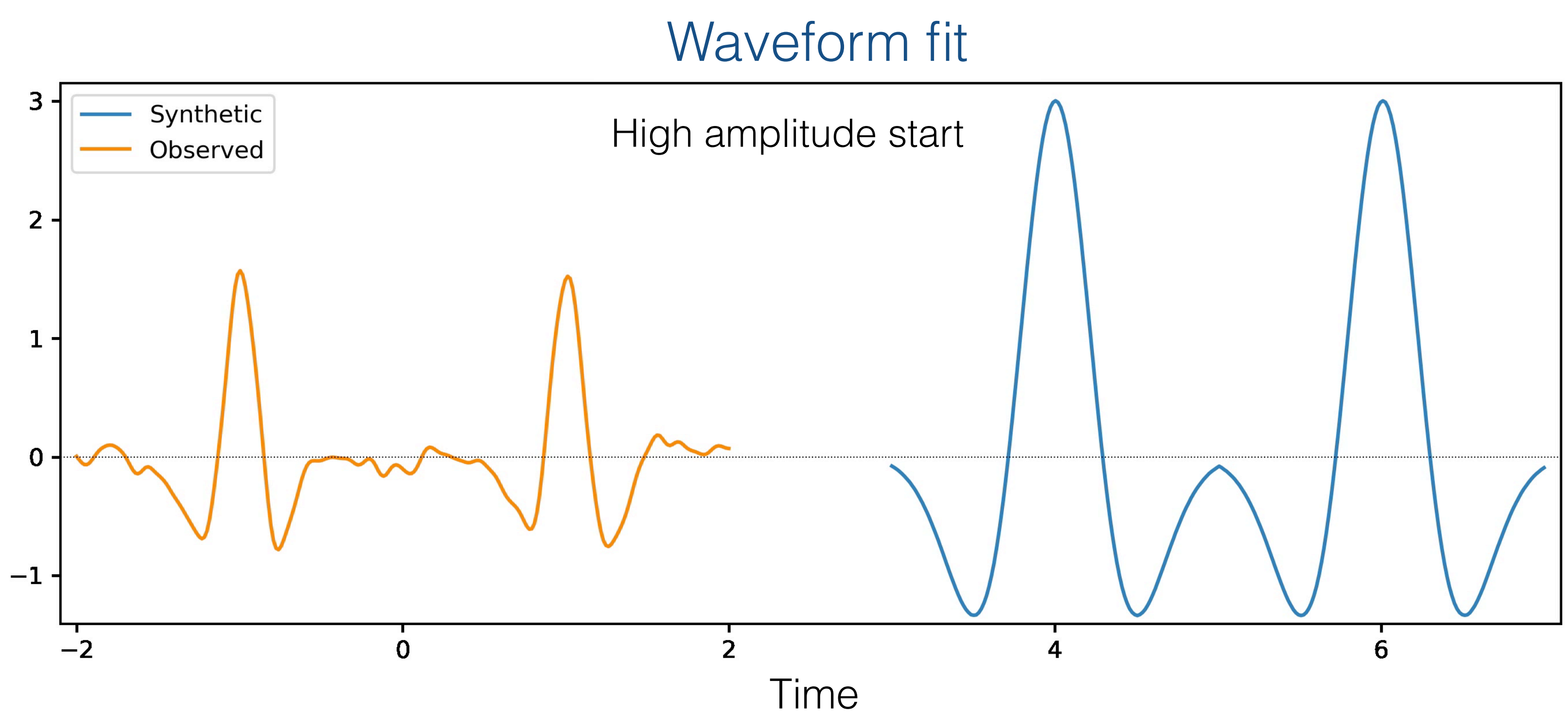


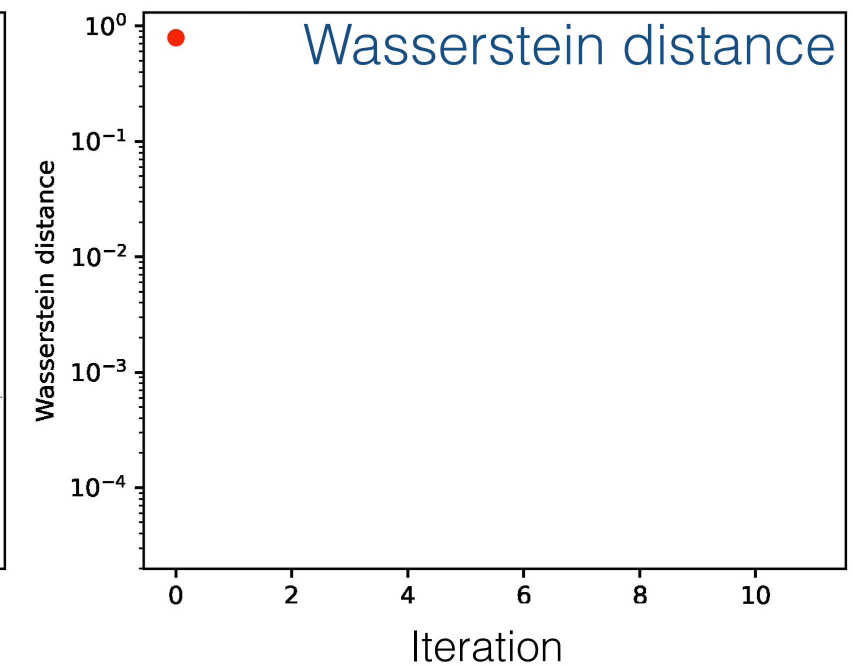
Wasserstein based on
2D PDF of fingerprint



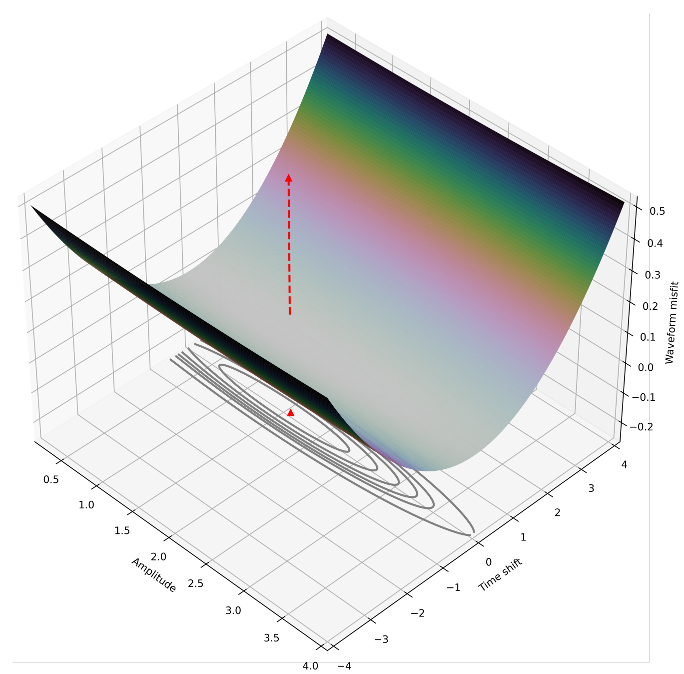


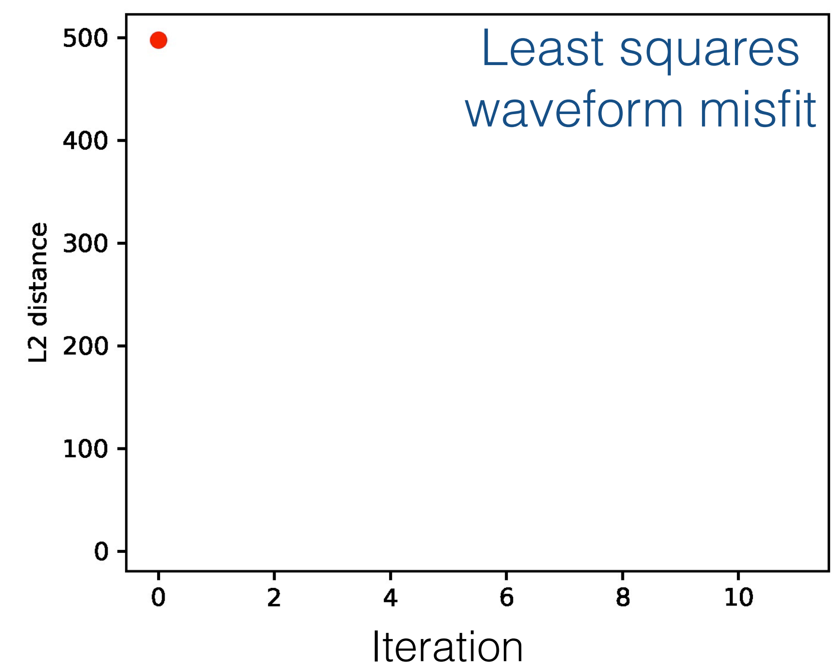
Minimizing the Wasserstein distance W_2^2





Wasserstein based on
2D PDF of fingerprint





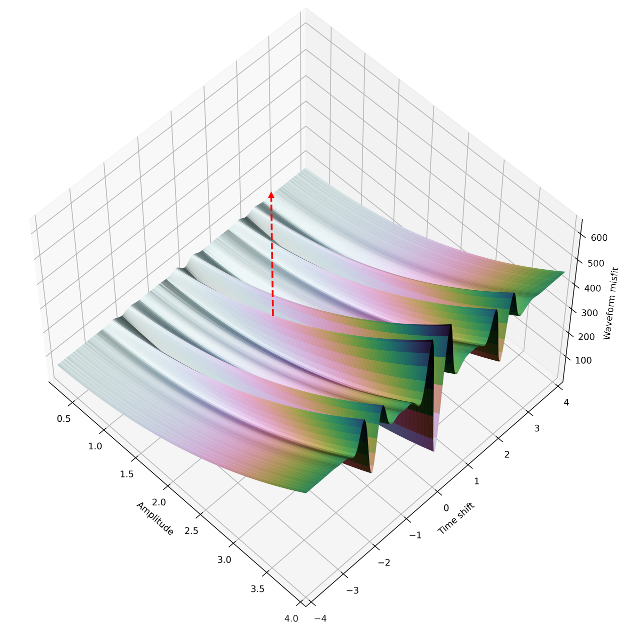
Biased conclusions

- There remain **unexplored applications** of sparsity:
  Over-complete imaging; 4D tomography,…

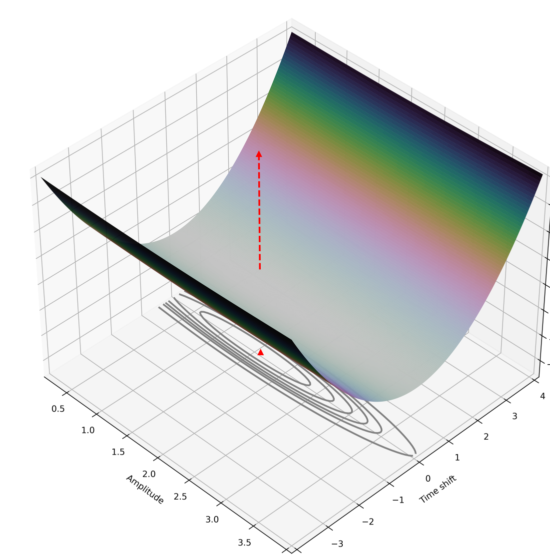
- Machine Learning has great potential in inversion.
  A surrogate for forward and inverse problems, many possibilities

- New approaches to optimal transport in highly nonlinear inversion
  Fitting of waveforms, surfaces and more…

A **key** element is to **pose** inversion questions
in a way that utilises that potential…



Least squares misfit



Wasserstein distance

My life tour guides:

Julie, Callum, Cameron and Isobel.