Deep learning for deep Earth:

physics-driven machine learning

Kuangdai Leng

Department of Earth Sciences, University of Oxford

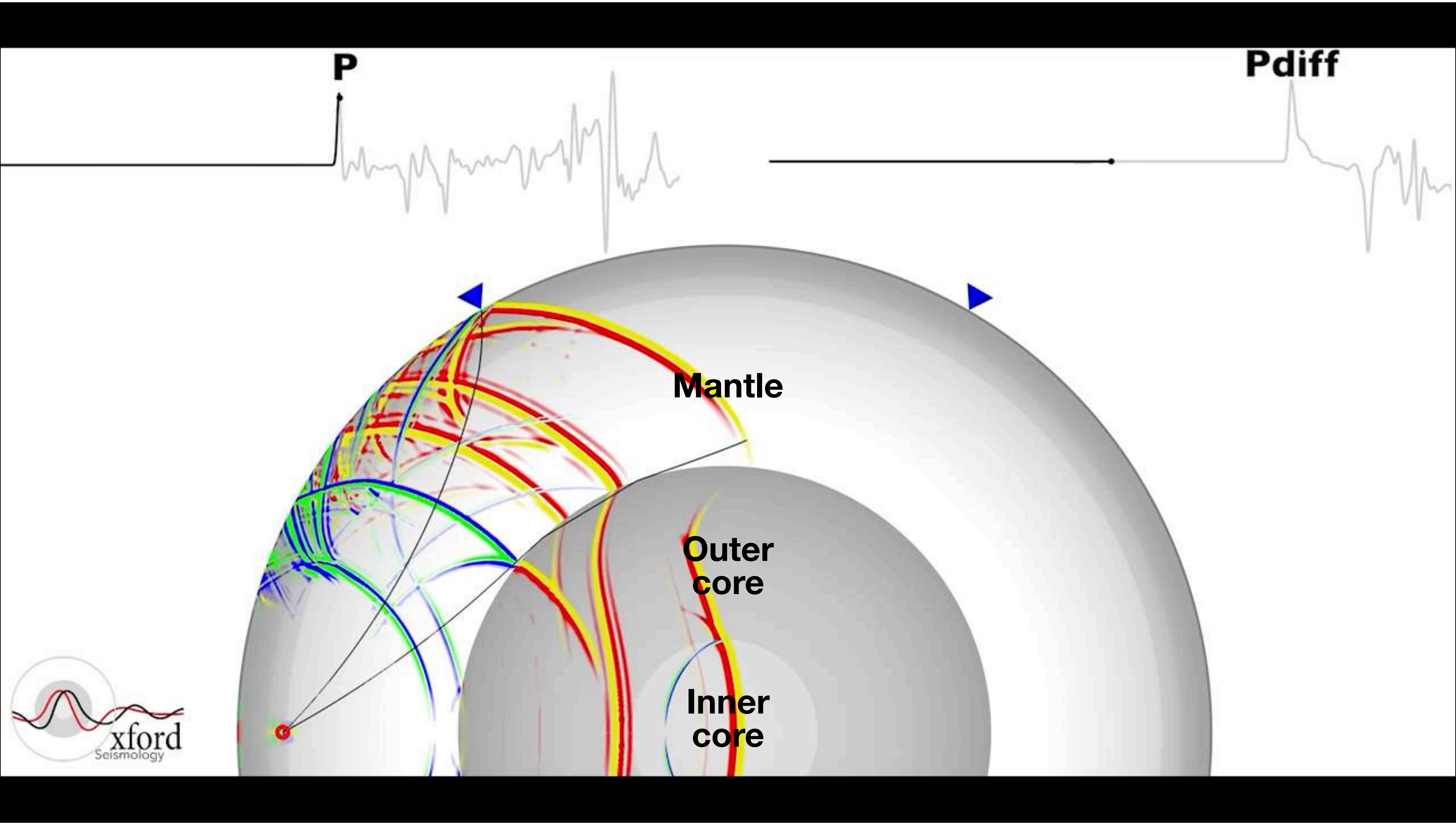
STFC, RAL, Harwell

April 2020

Topics

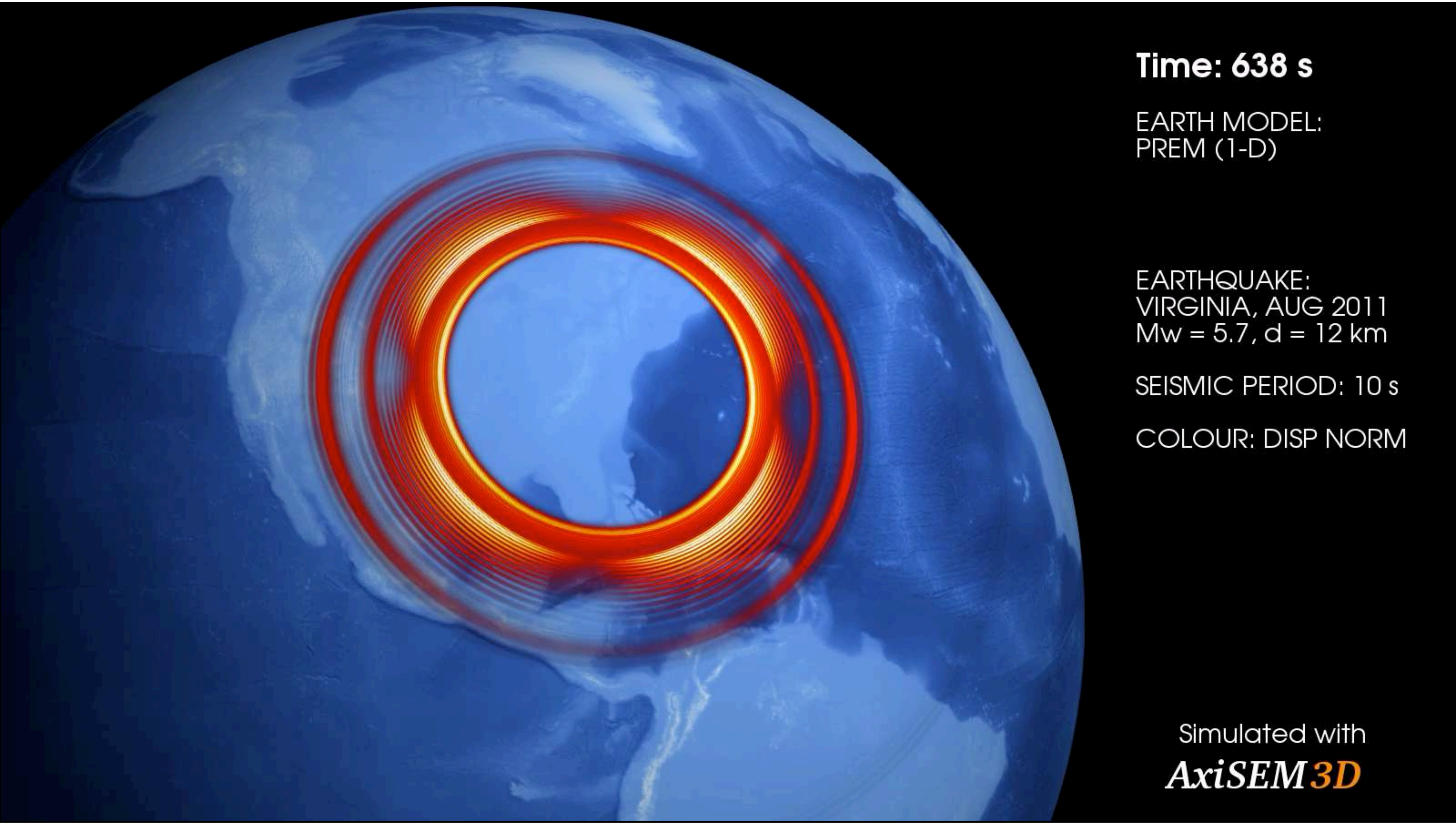
- Global seismology and Earth structure
- “Solving” a PDE-informed system by ML
- Qualifying seismic data by ML

Seismic wave propagation in **1D** Earth



YouTube channel: Seismology Oxford

Seismic wave propagation in **1D** Earth

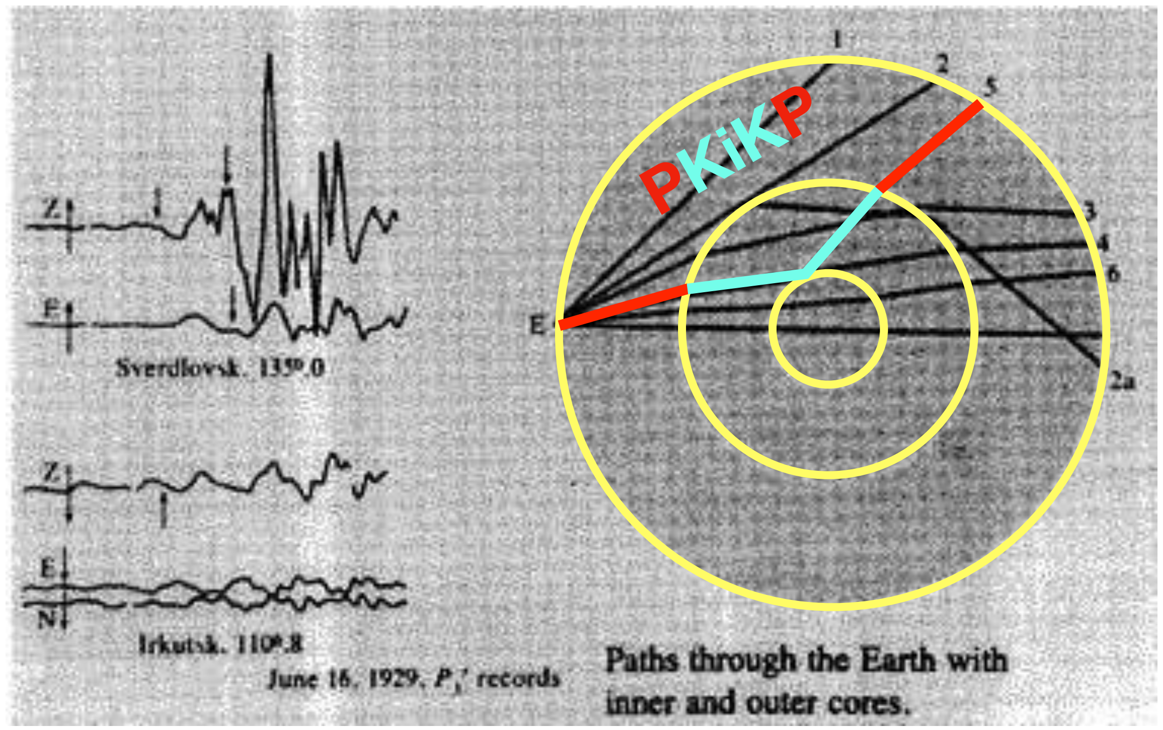


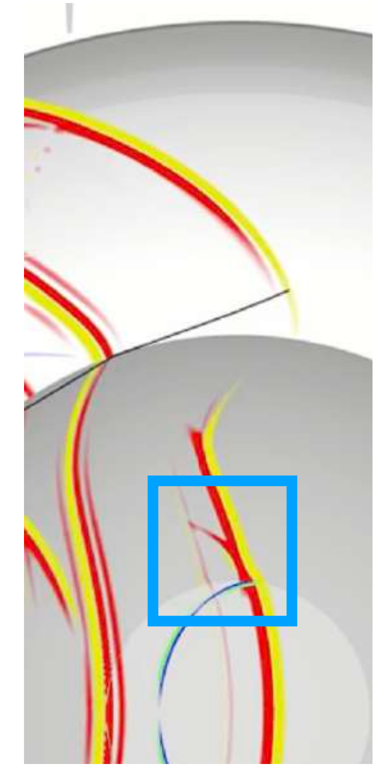
YouTube channel: Seismology Oxford

1D Earth structure

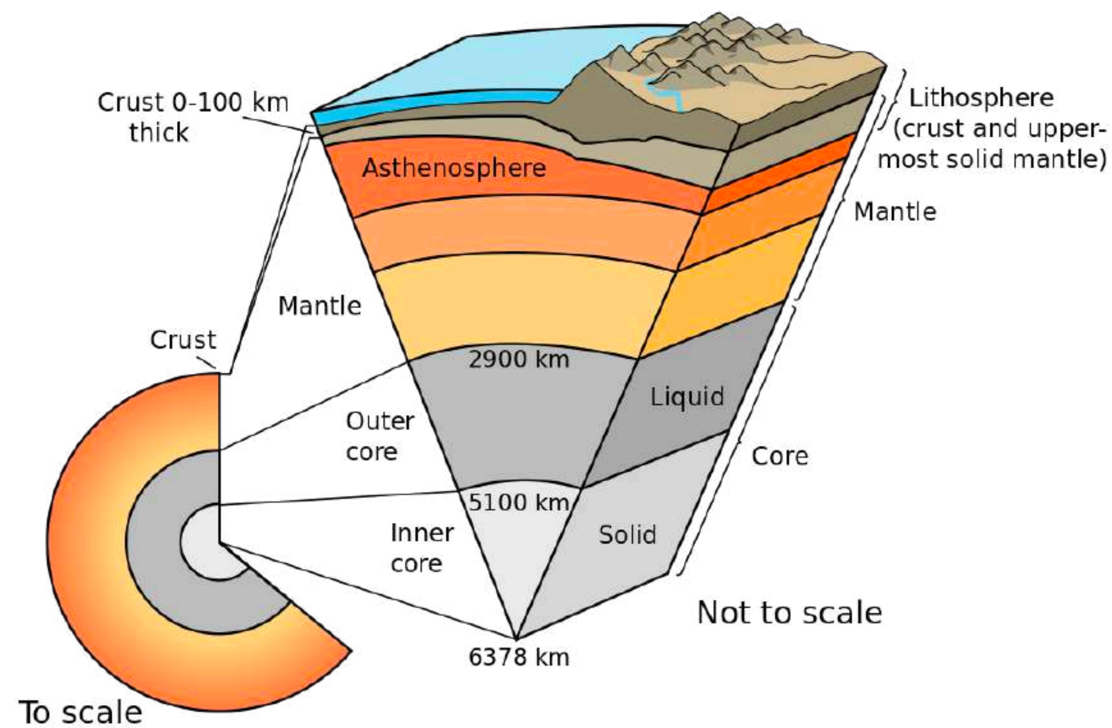
- Discovery of Earth's inner core

Inge Lehmann
(1888–1993)





- Earth's solid interior is **mostly** a layered cake
  - **Crust**: cold and hard, ~25 km on average
  - **Mantle**: hot and soft, ~2,900 km
  - **Outer core**: iron-rich fluid, ~2,400 km
  - **Inner core**: iron–nickel alloy, ~1,250 km



3D Earth structure

- A 1D Earth is dead, but our Earth is quite alive
- Earth’s structure must be 3D, though rather weak
- Plate tectonics and volcanism are the surface expressions of mantle convection

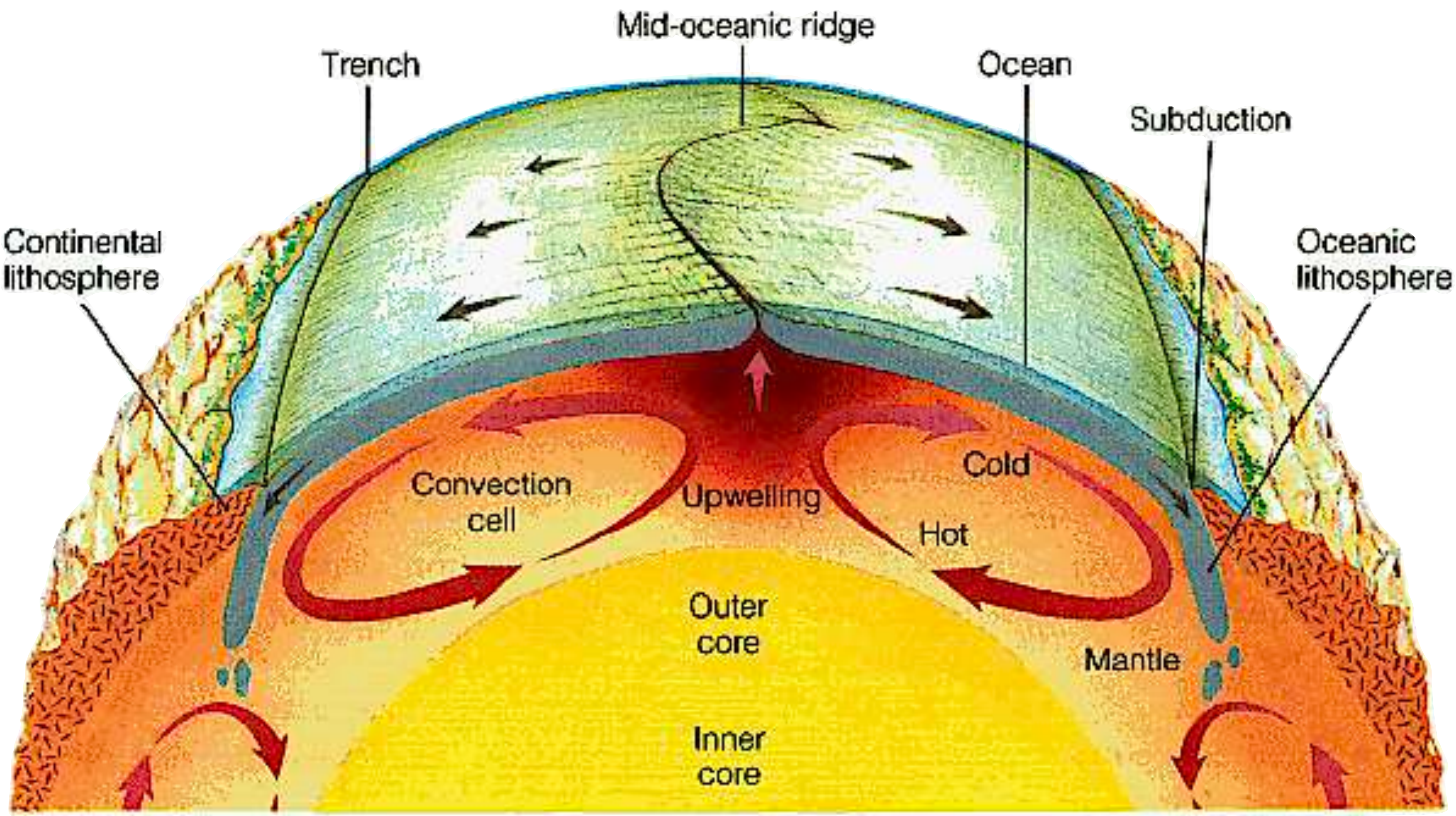
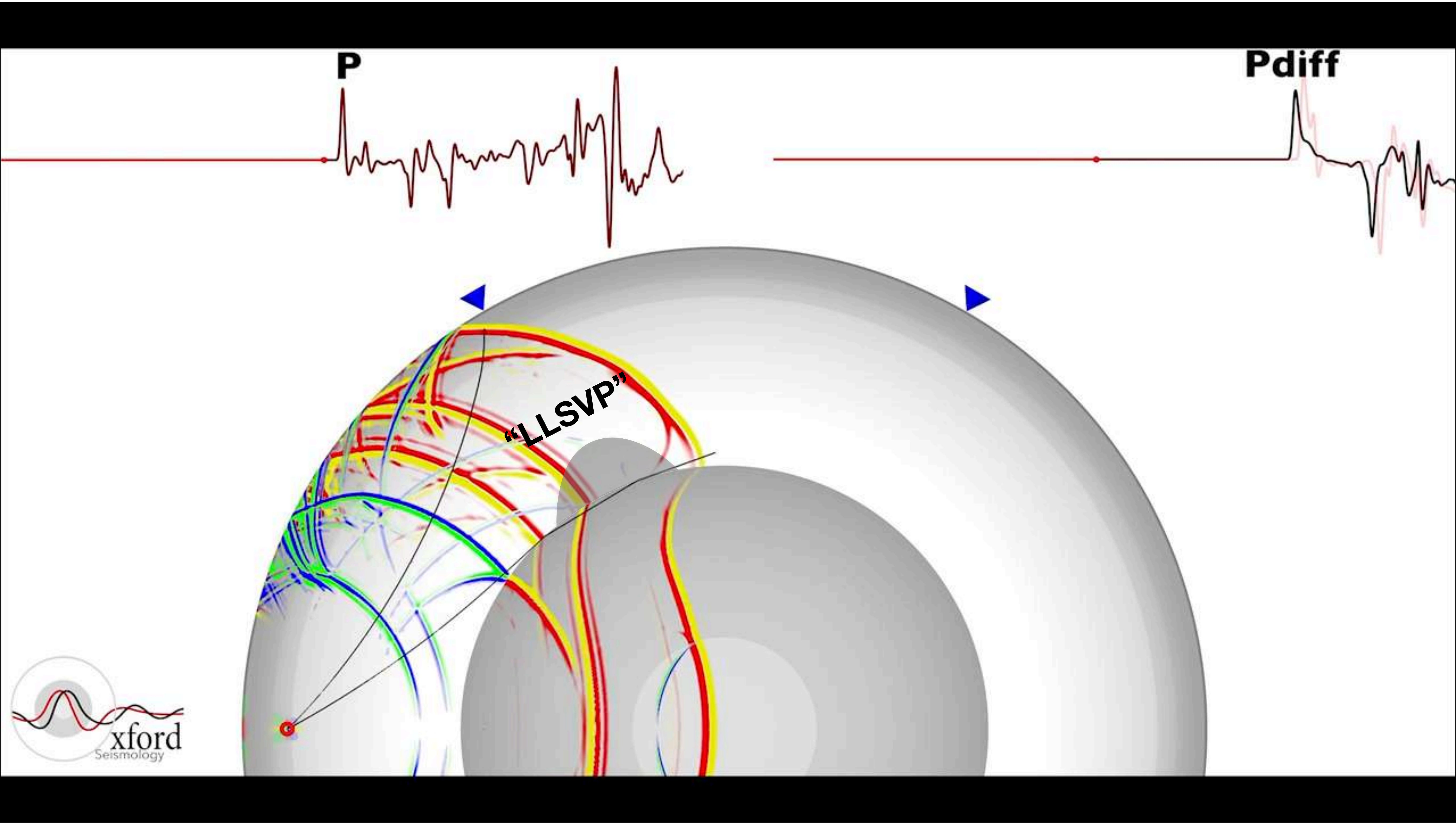


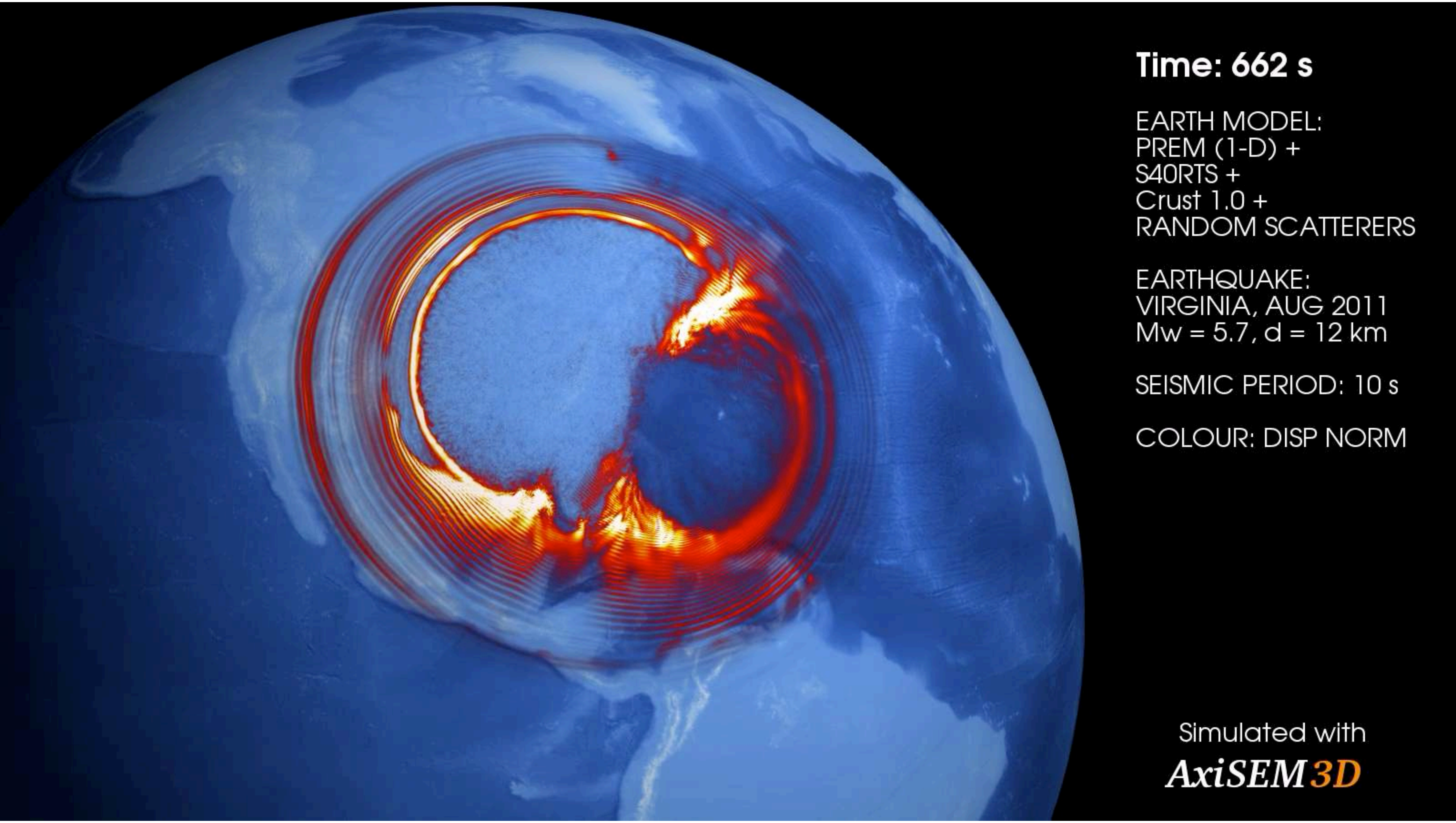
Illustration by Byrd Polar Research Center

Seismic wave propagation in **3D** Earth



YouTube channel: Seismology Oxford

Seismic wave propagation in **3D** Earth



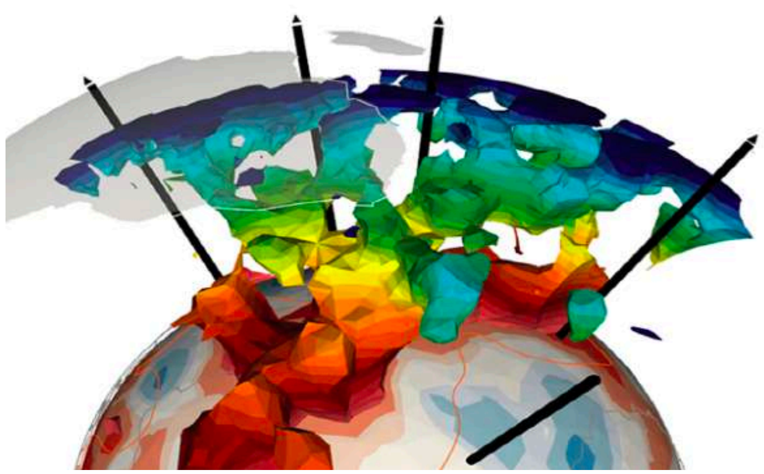
Grand Prize of ARCHER Video Competition (*Leng & Fernando, 2020*)

Forward modelling and inversion

Forward modelling

$$\mathbf{d} = \mathscr{F}(\mathbf{m})$$

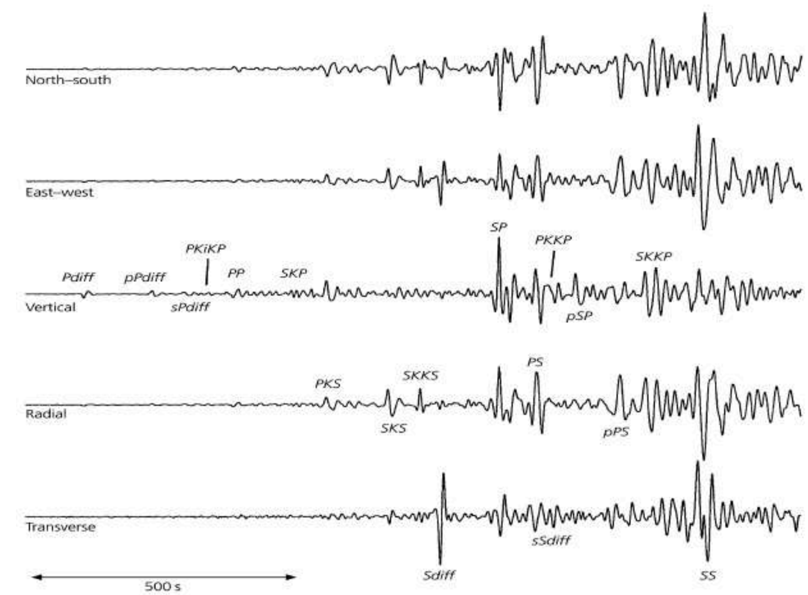
3D Earth model



La Reunion Plume (*Tsekhmistrenko, 2019*)

Basis

Waveform data



Inversion (MAP)

$$\mathbf{m} = \underset{\mathbf{m}}{\operatorname{argmax}} P(\mathbf{m}\,|\,\mathbf{d})$$

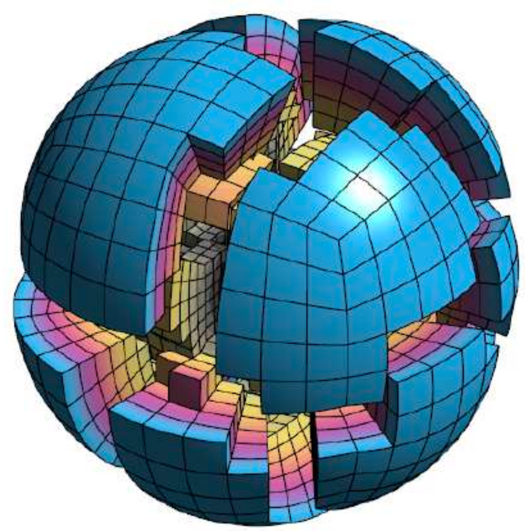
Topics

- Global seismology and Earth structure
- “Solving” a PDE-informed system by ML
- Qualifying seismic data by ML

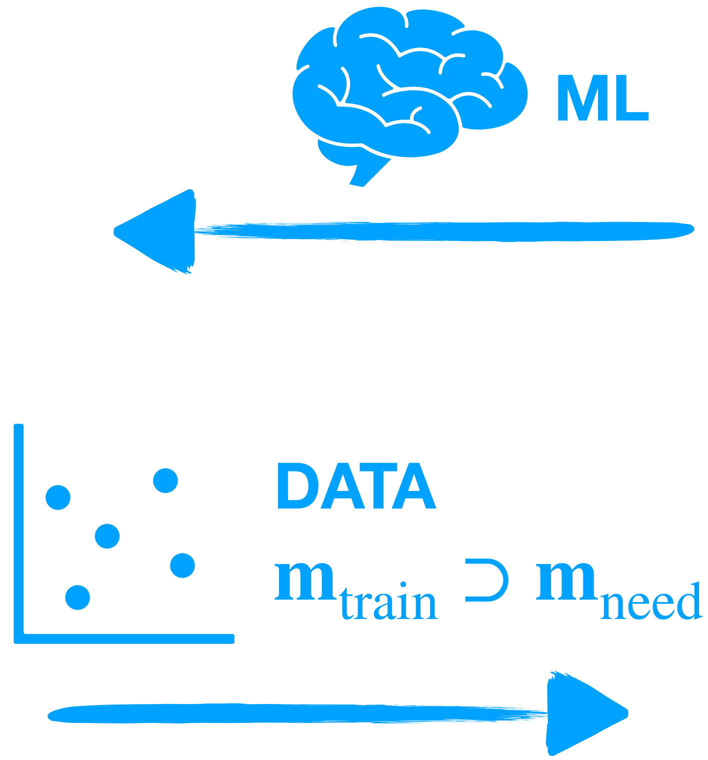
Simulator & Emulator

SIMULATOR

$$\mathbf{d} = \mathscr{F}(\mathbf{m})$$

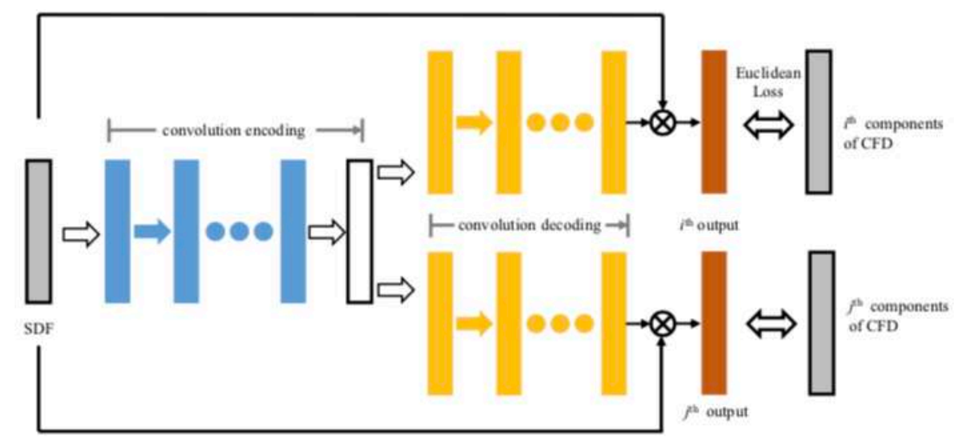


Finite element mesh (*GLVis*)



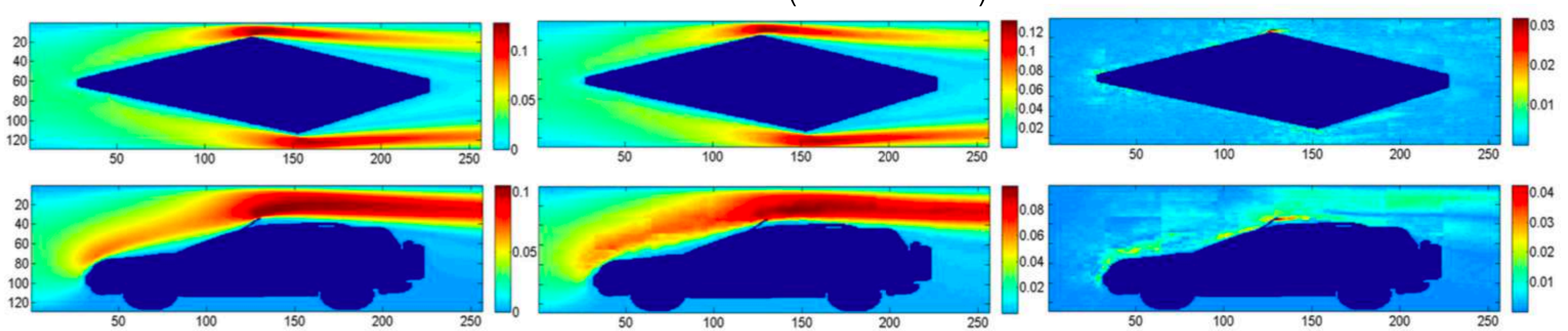
EMULATOR

$$\mathbf{d} \approx F(\mathbf{w}; \mathbf{m})$$



CNN for CDF (*Guo et al., 2016*)

Simulator | CNN emulator (x100 faster) | Error

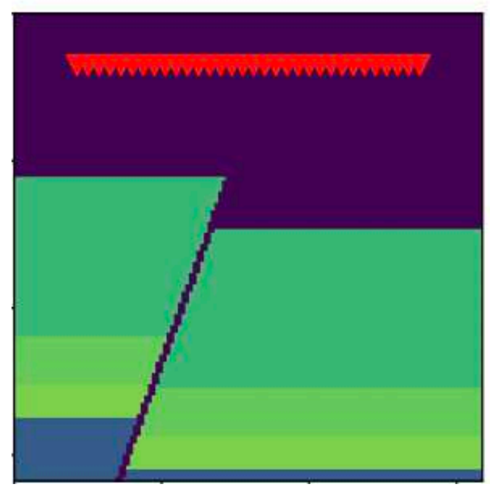


CNN-predicted velocity fields with different obstacles (*Guo et al., 2016*)

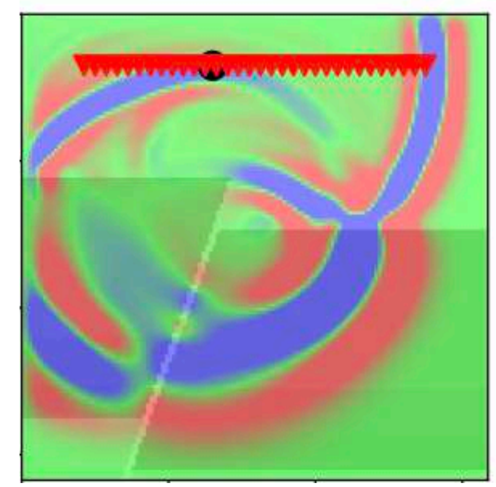
End-to-end emulator

- Acoustic wave modelling (*Moseley et al., 2019*)

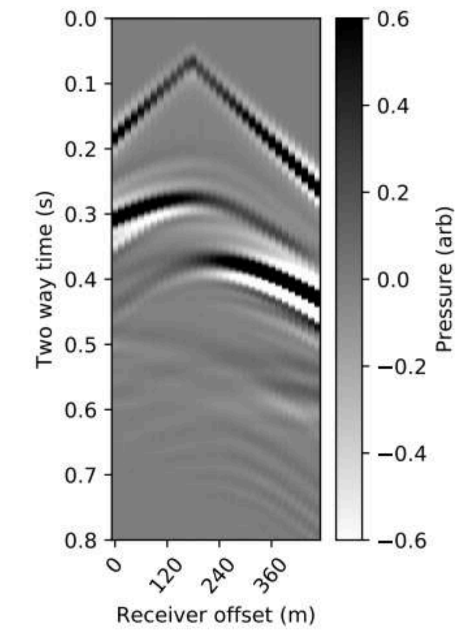
SIMULATOR



Earth model

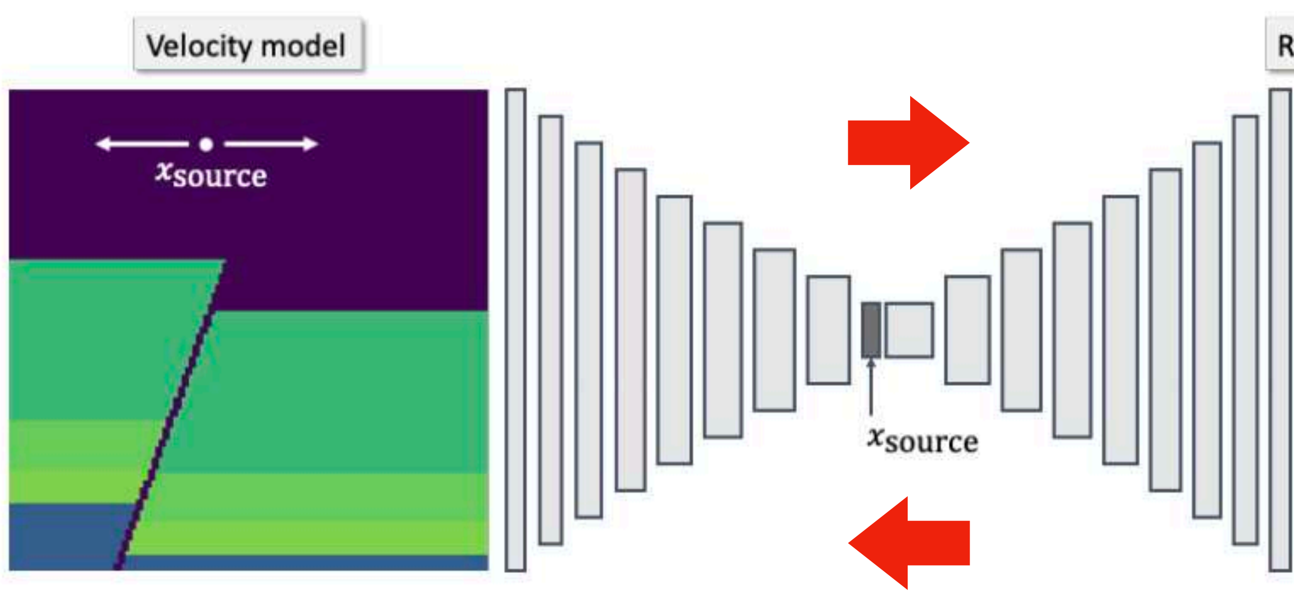


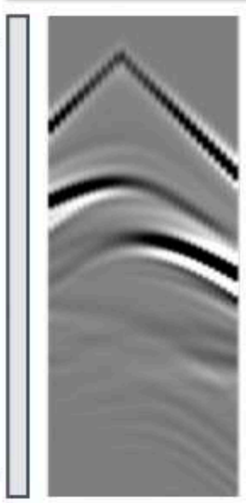
Wavefield by FD
(wave physics)



Seismograms at receivers

EMULATOR

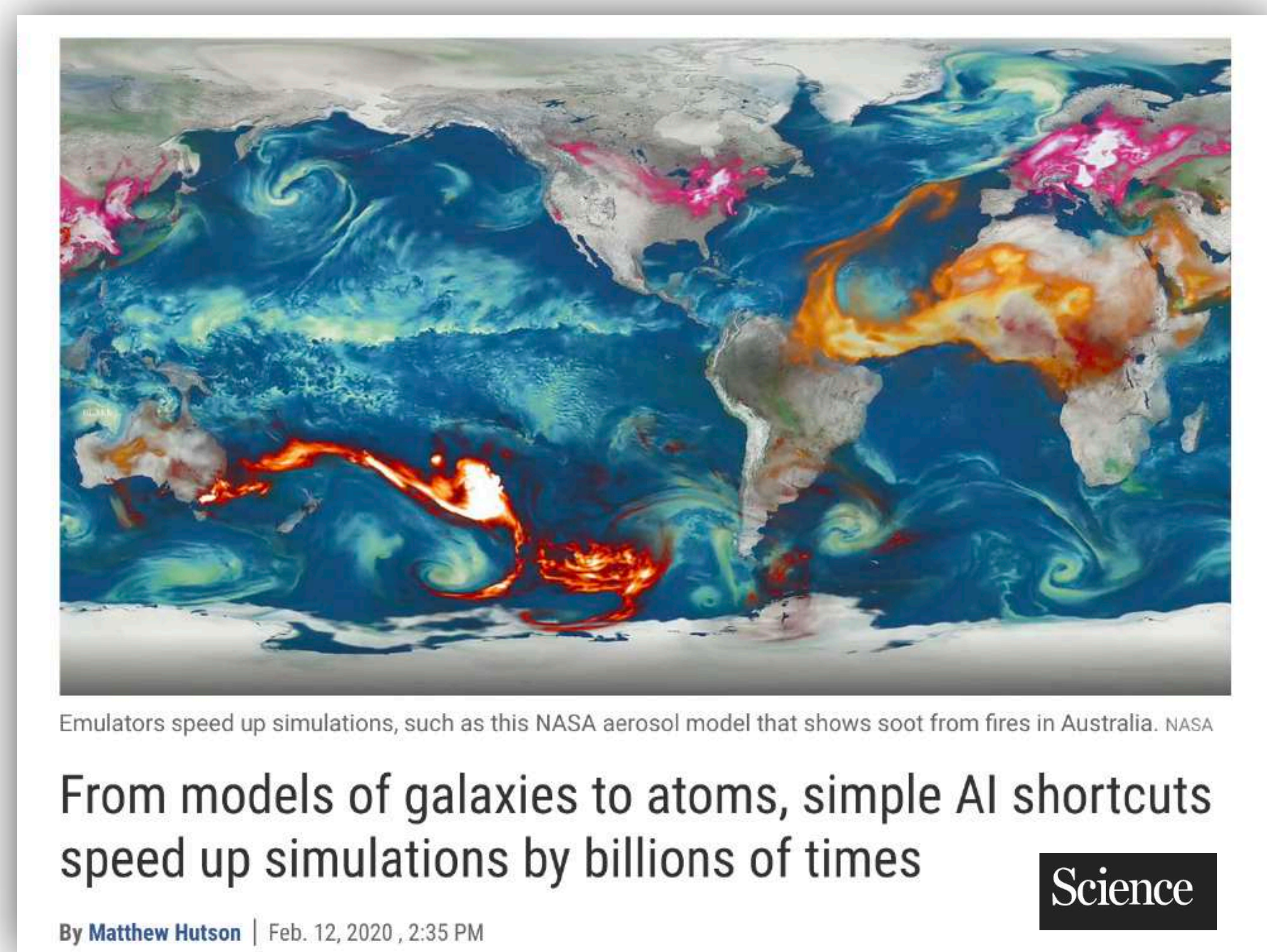




- x100 faster
- 18 million network parameters
- trained with 600,000 simulations

Conditional autoencoder architecture WaveNet

Neural architecture search


Emulators speed up simulations, such as this NASA aerosol model that shows soot from fires in Australia. NASA

From models of galaxies to atoms, simple AI shortcuts speed up simulations by billions of times

Science

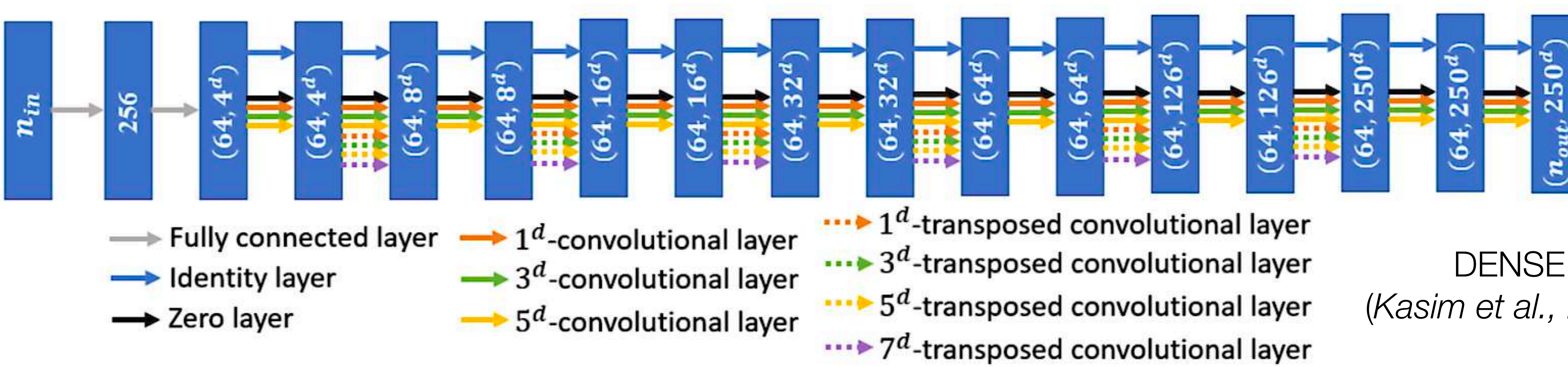
By Matthew Hutson | Feb. 12, 2020 , 2:35 PM

EMULATOR with architecture variability

$$\mathbf{d} \approx F(\mathbf{n}; \mathbf{w}; \mathbf{m})$$

授人以“**魚**”不如授之以“**漁**”

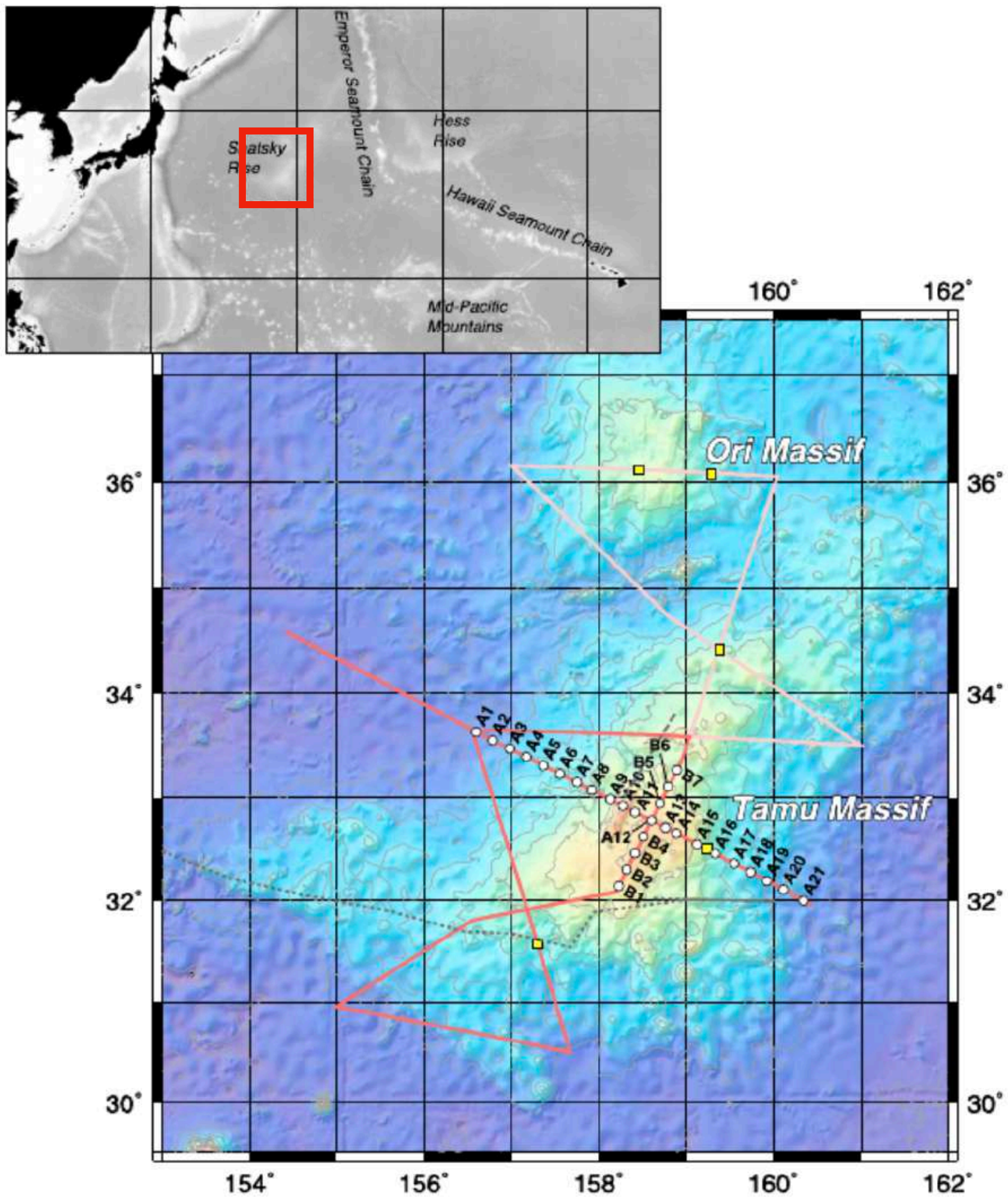
- Fast (10^4~10^9 speedup)
- **Using limited training data** by letting the architecture capture the “**prior**” in the physics

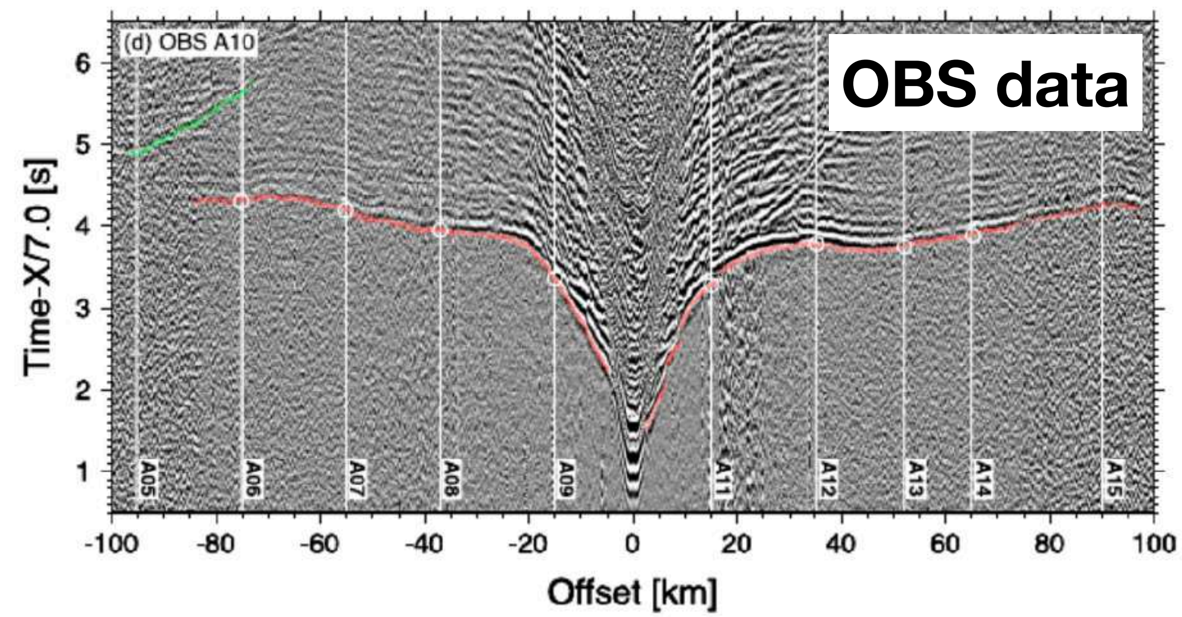


DENSE
(*Kasim et al., 2020*)

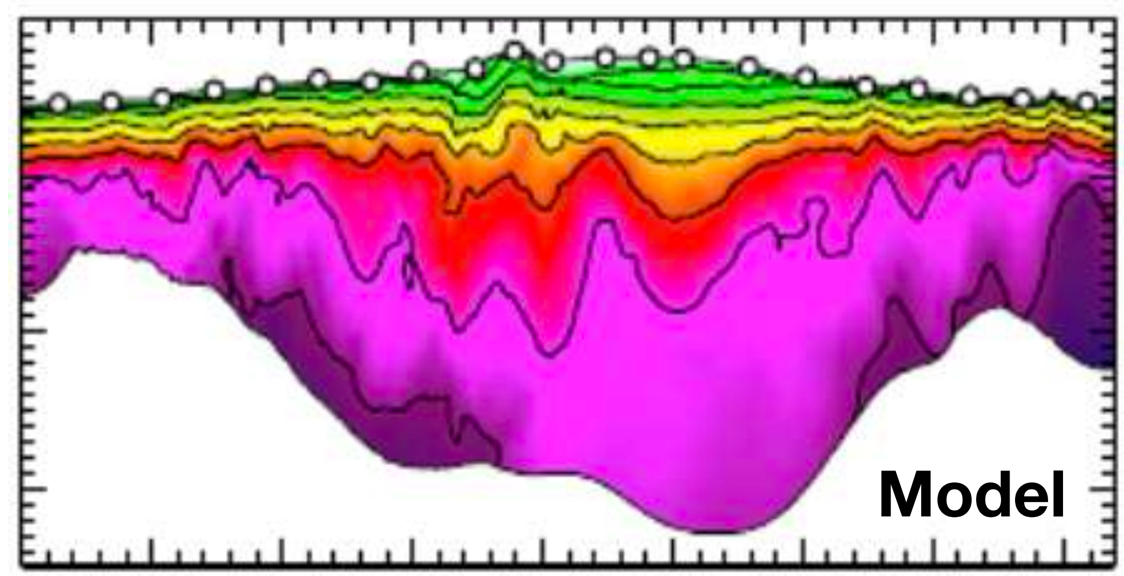
Neural architecture search

- Seismic tomographic of Shatsky Rise (*Korenaga & Sager, 2012*)



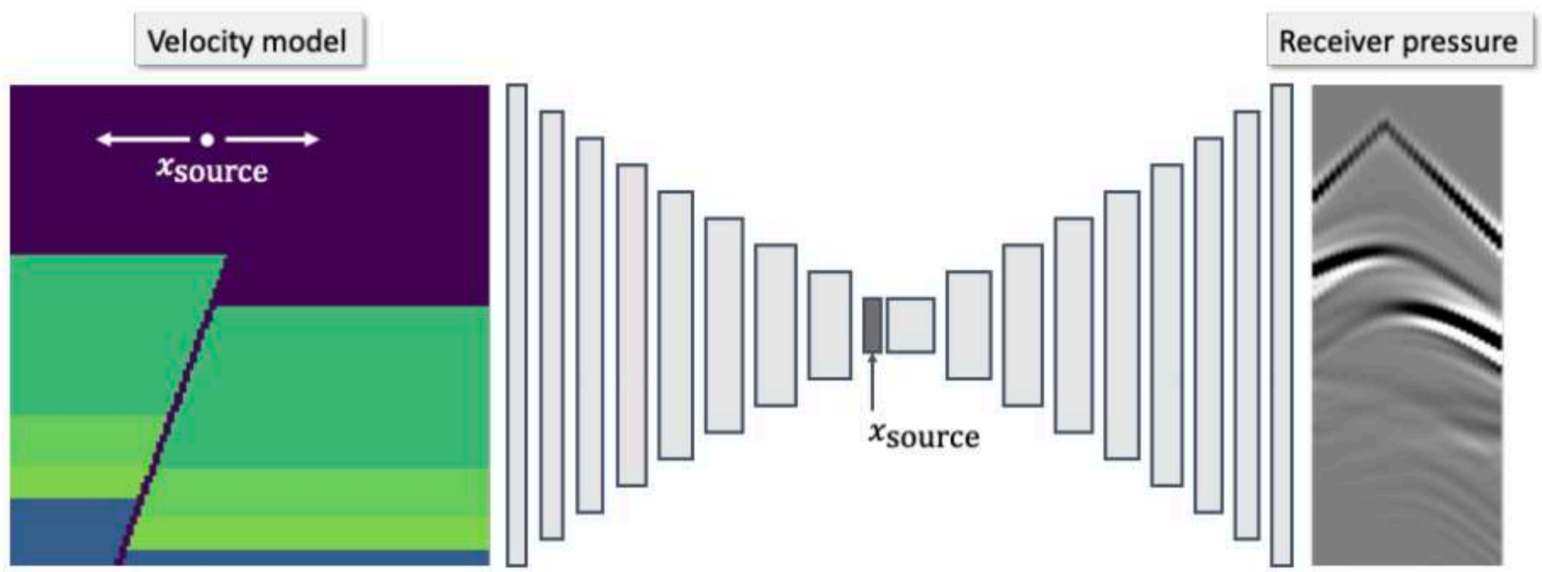


DENSE with x10^8 ?



Problems with end-to-end emulators

- Interpretability (out-of-sample data)



- Accuracy (Probably Approximately Correct, PAC)

$$m \geq \frac{1}{\epsilon}\left(\ln \mathscr{H} + \ln \frac{1}{\delta}\right)$$

m: minimum number of data

ϵ: upper bound of true error

$\mathscr{H}$: size of hypothesis space

$1 - \delta$: confidence

- Complete solution of PDE

Prior in physics

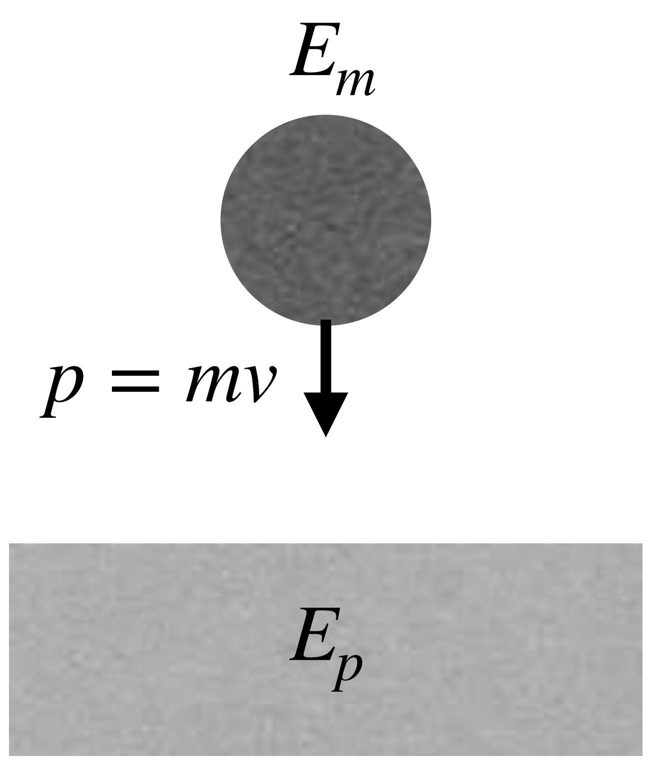
- How ML works

$$m \geq \frac{1}{\epsilon}\left(\ln \mathscr{H} + \ln \frac{1}{\delta}\right)$$

Prior
=
Sparsity

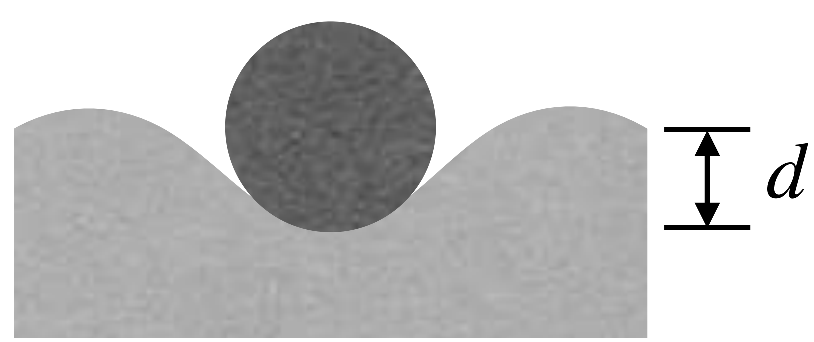
Understanding sparsity in physics is the key to build interpretable and accurate emulators.

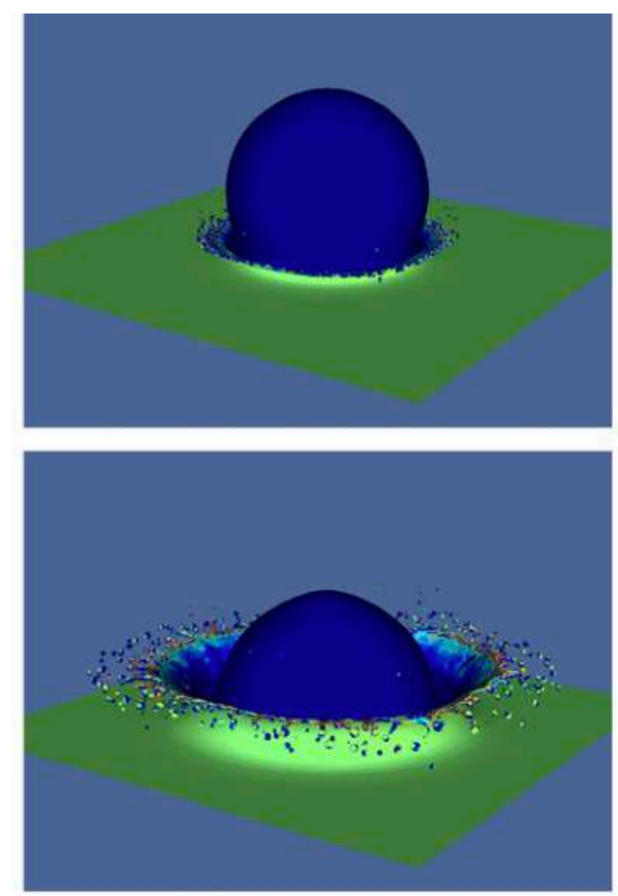
- Planetary impact: $d = d(p, E_m, E_p)$



- Difficult to simulate
- Easy to emulate
- Prior:

$$d \uparrow \Leftrightarrow p \uparrow, E_m \uparrow, E_p \downarrow$$



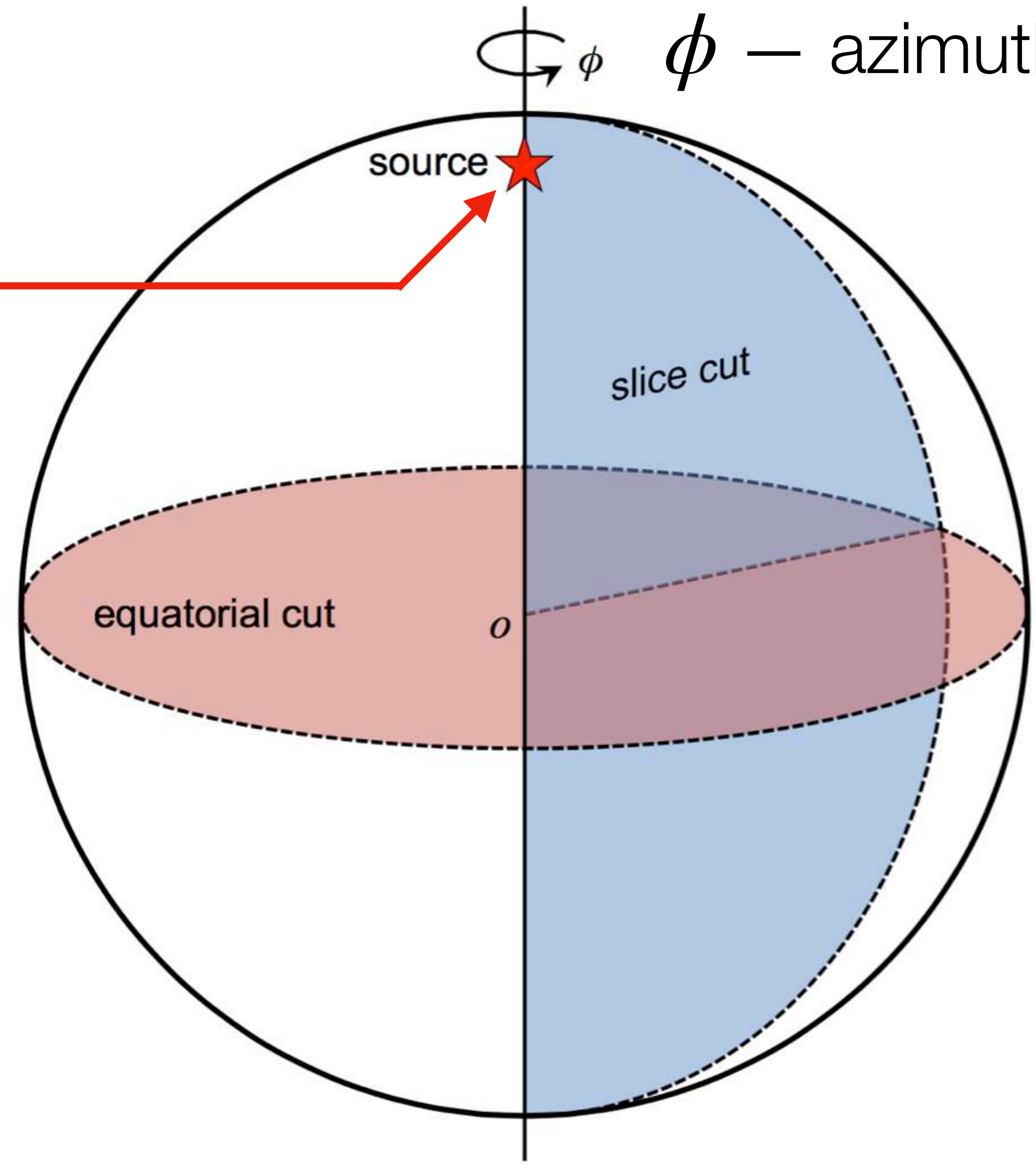


Agbaglah et al., 2011

AxiSEM3D: azimuthal sparsity

A vertical impact

ϕ — azimuth



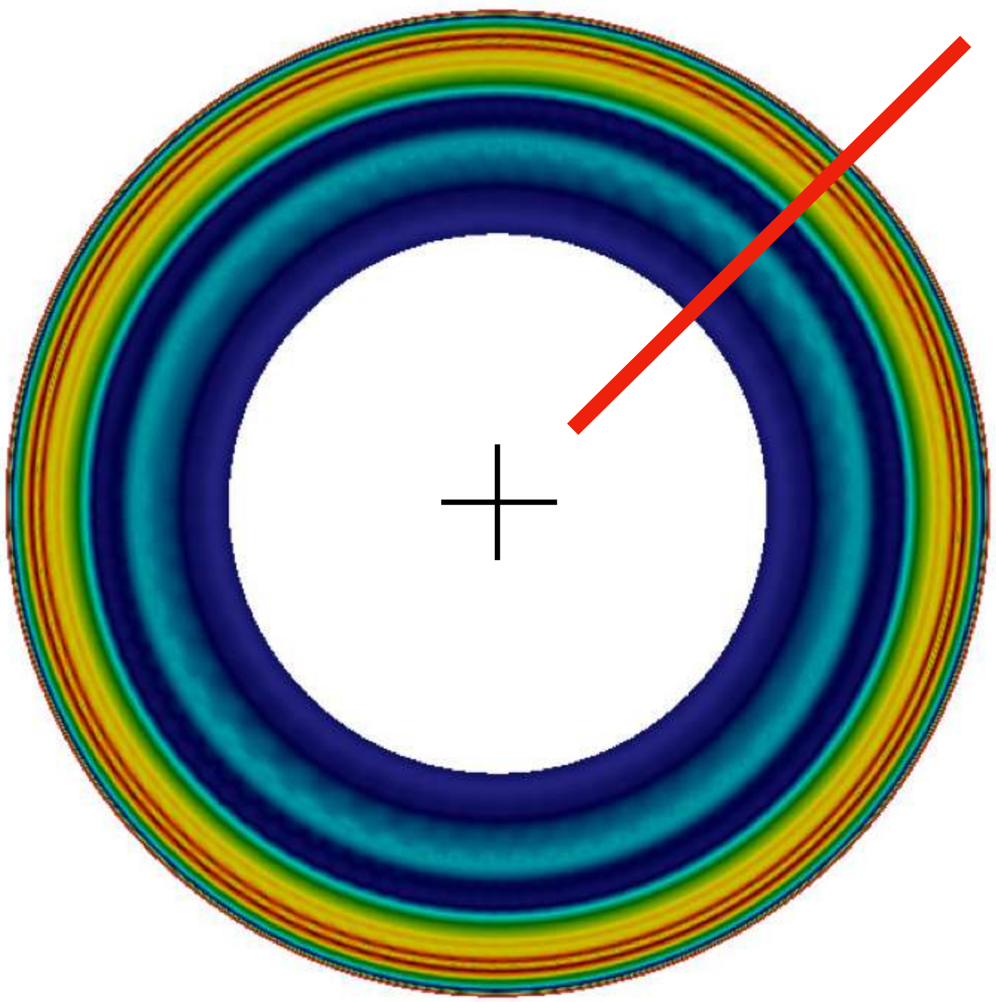
Source-centred system

Leng et al. 2016

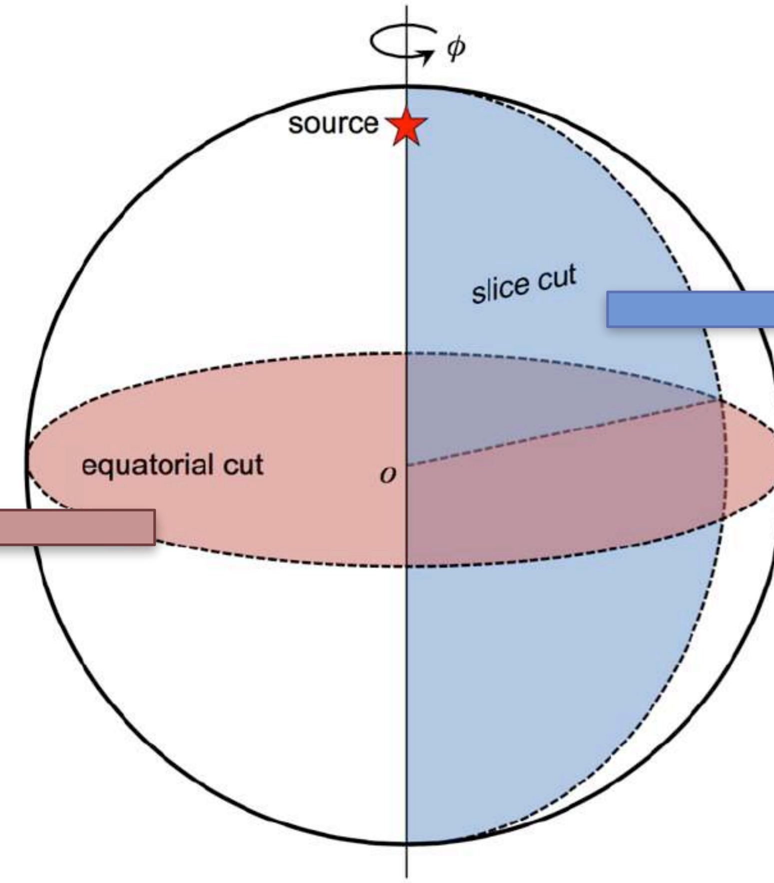
Wavefield in **1D** Earth

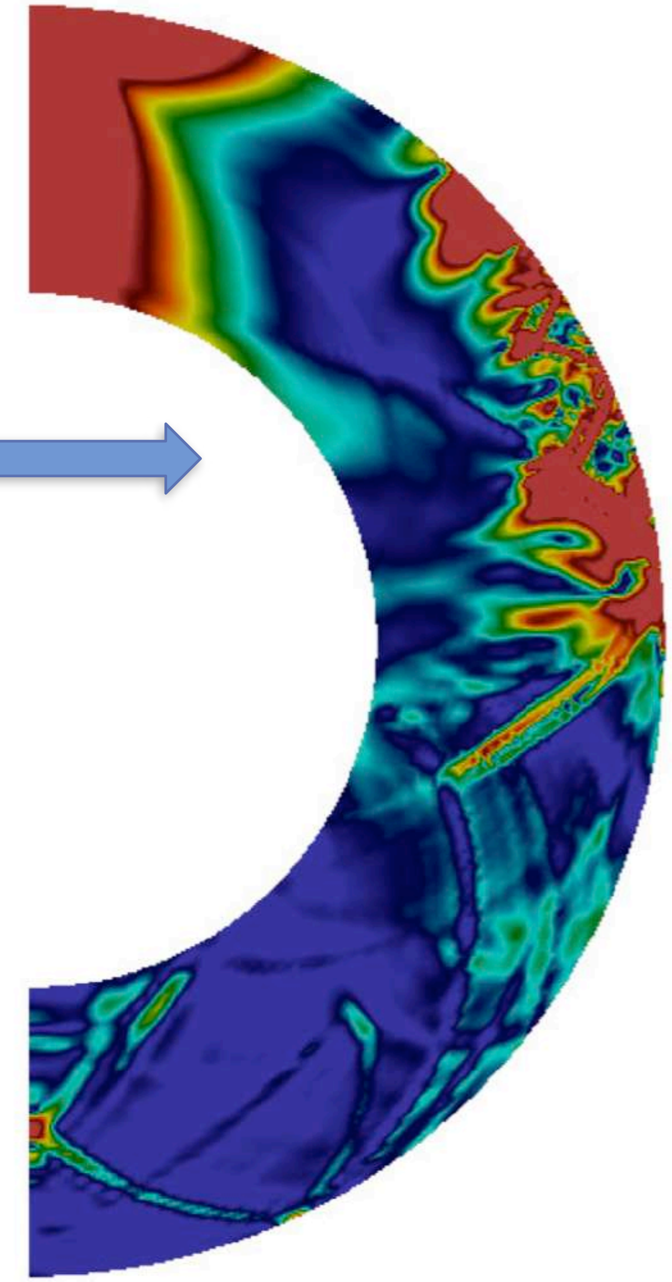
We can solve it on **ONE slice**:
3D problem reduced to 2D

Model: PREM



Equatorial cut



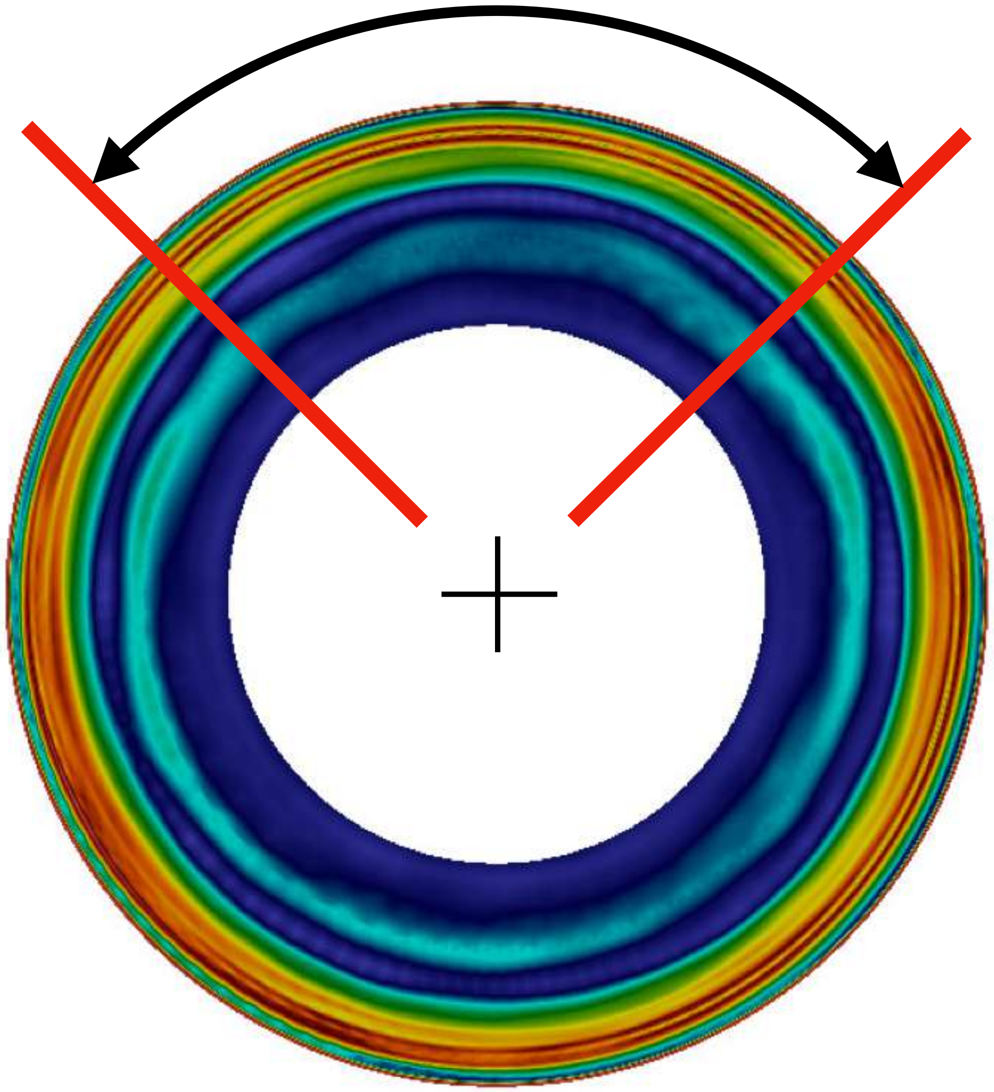


Slice cut

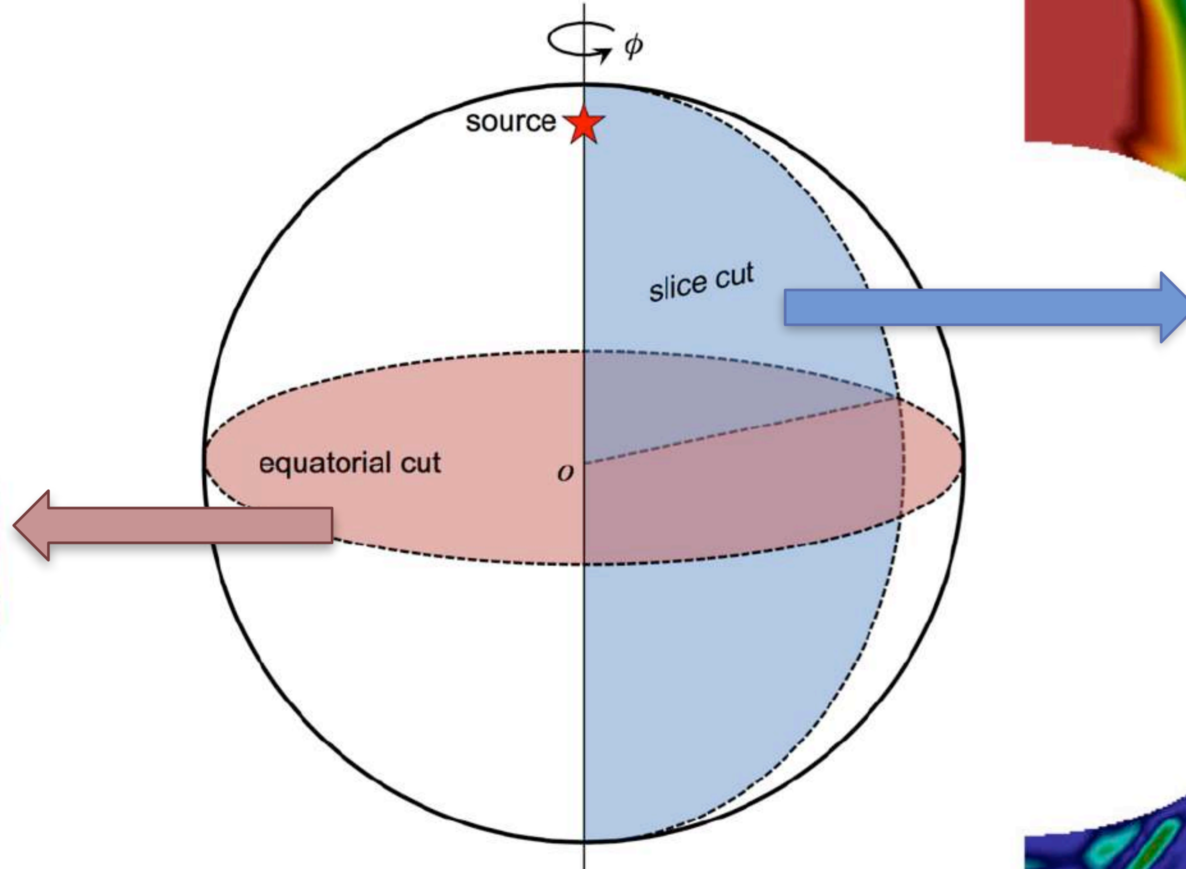
Wavefield in **3D** Earth

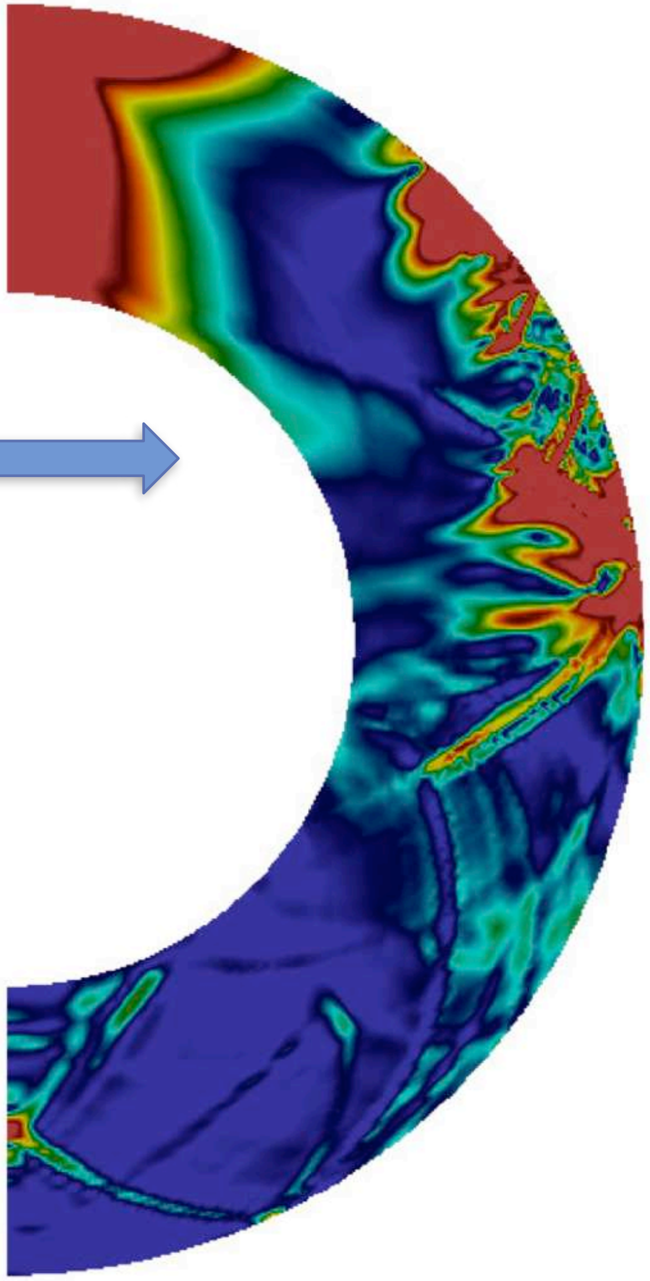
Wavefield becomes non-axisymmetric,
but remains **smooth** along azimuth

Model: S40RTS



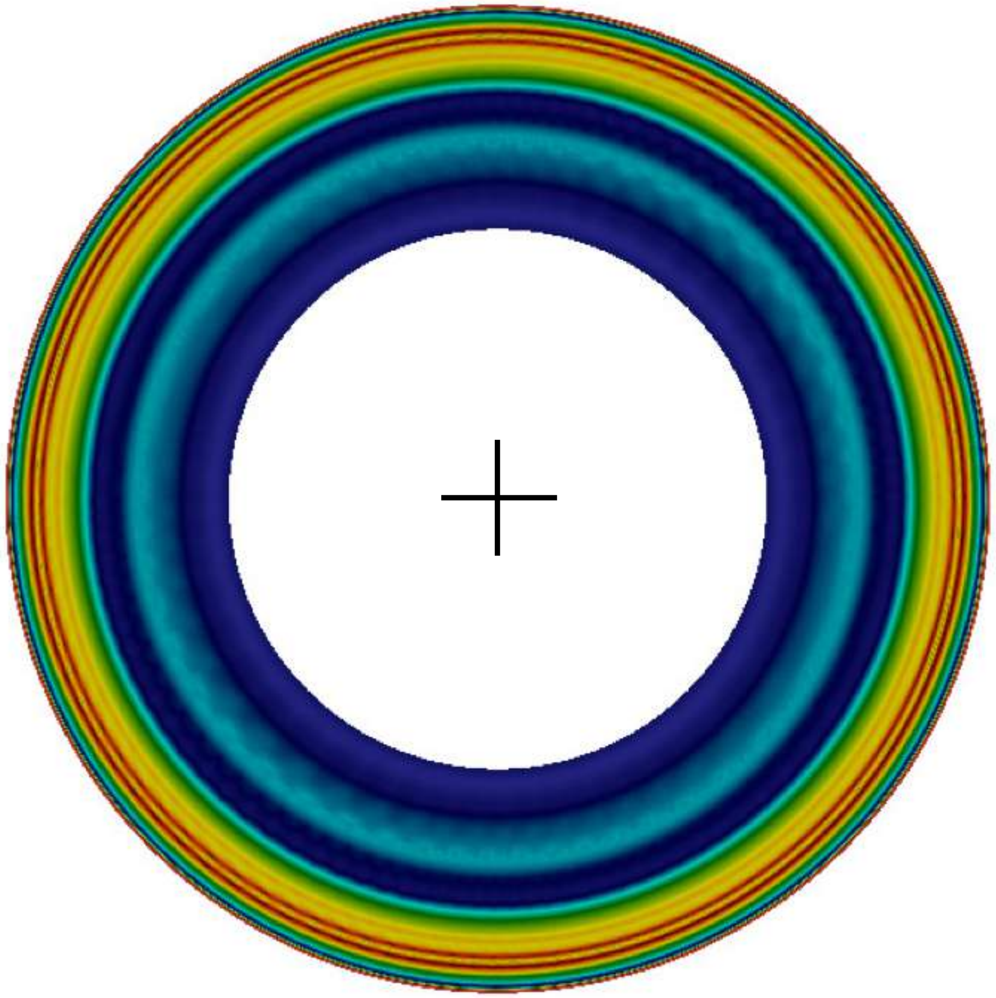
Equatorial cut





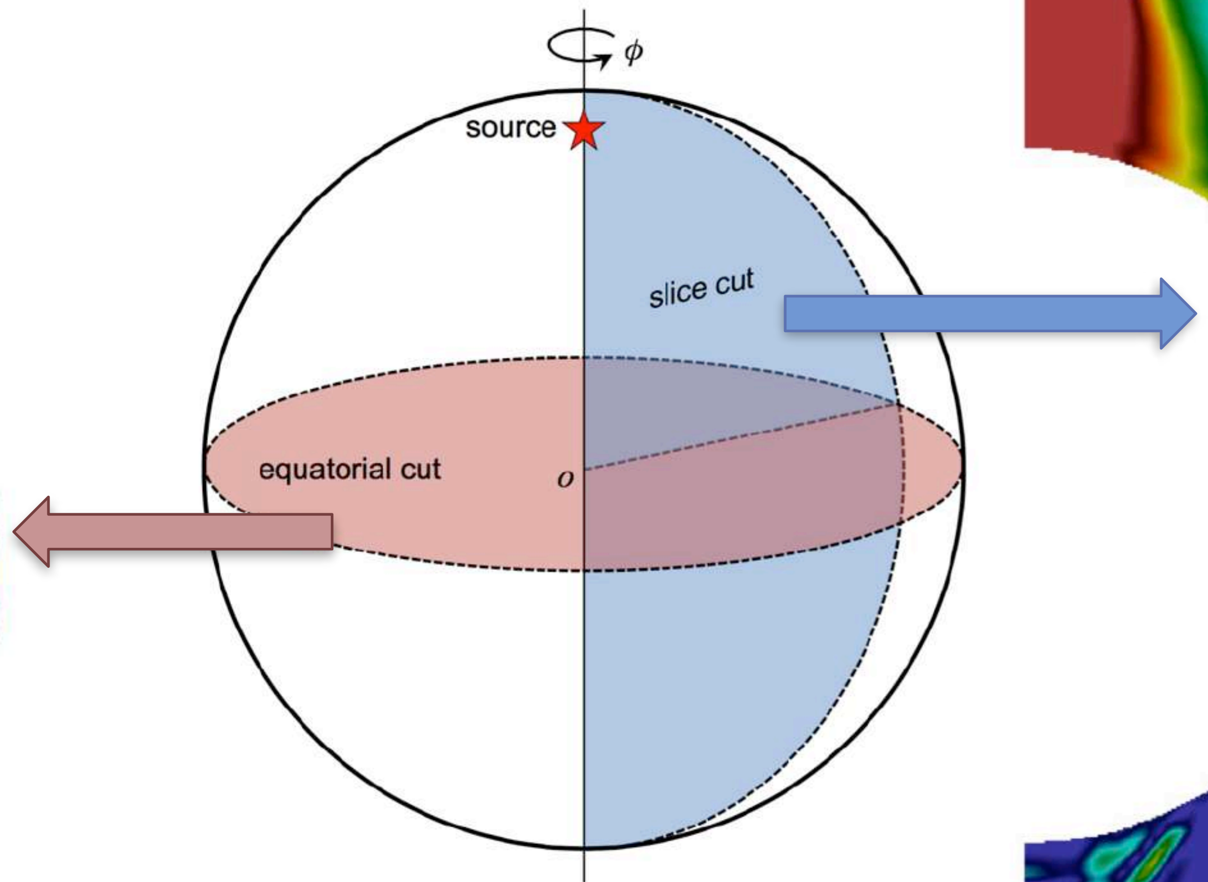
Slice cut

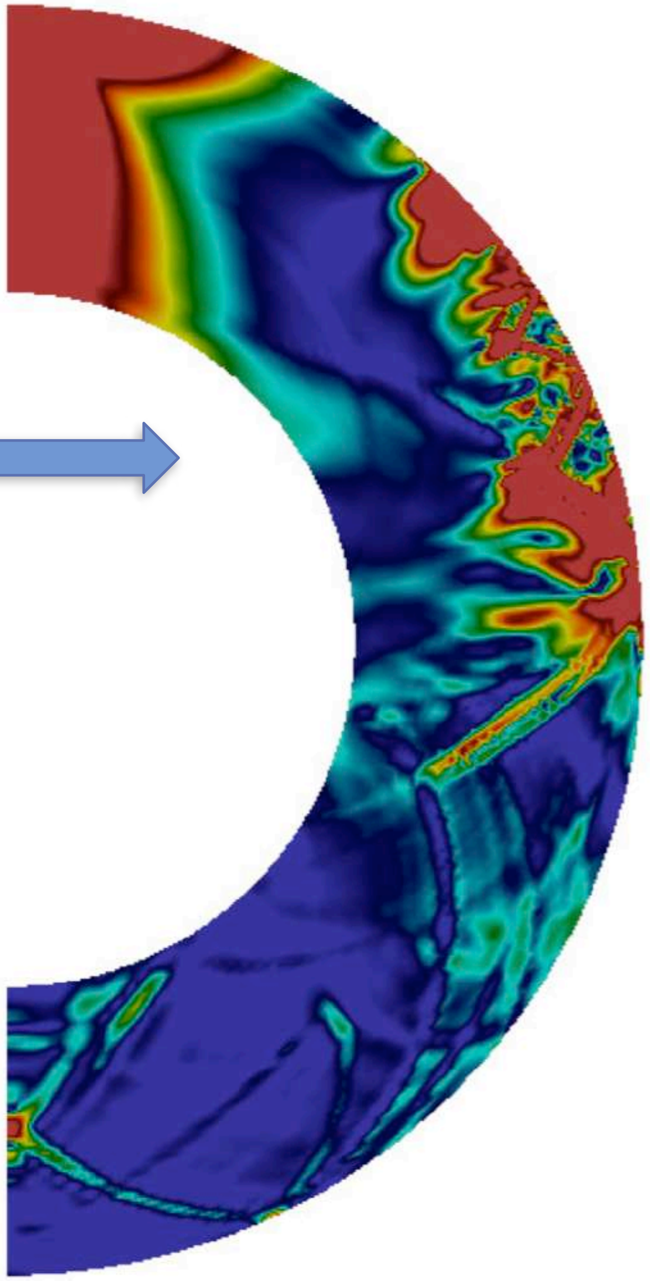
Wavefield in **1D** Earth



Equatorial cut

Model: PREM

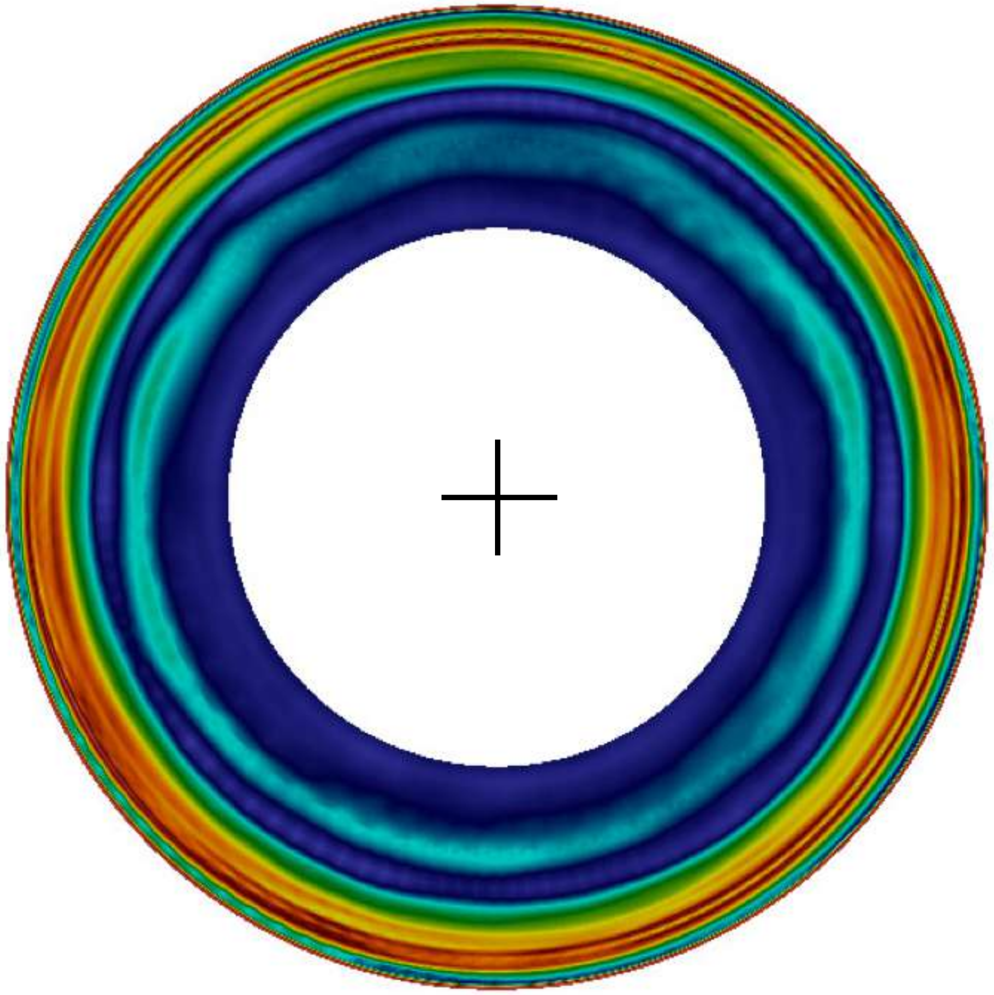




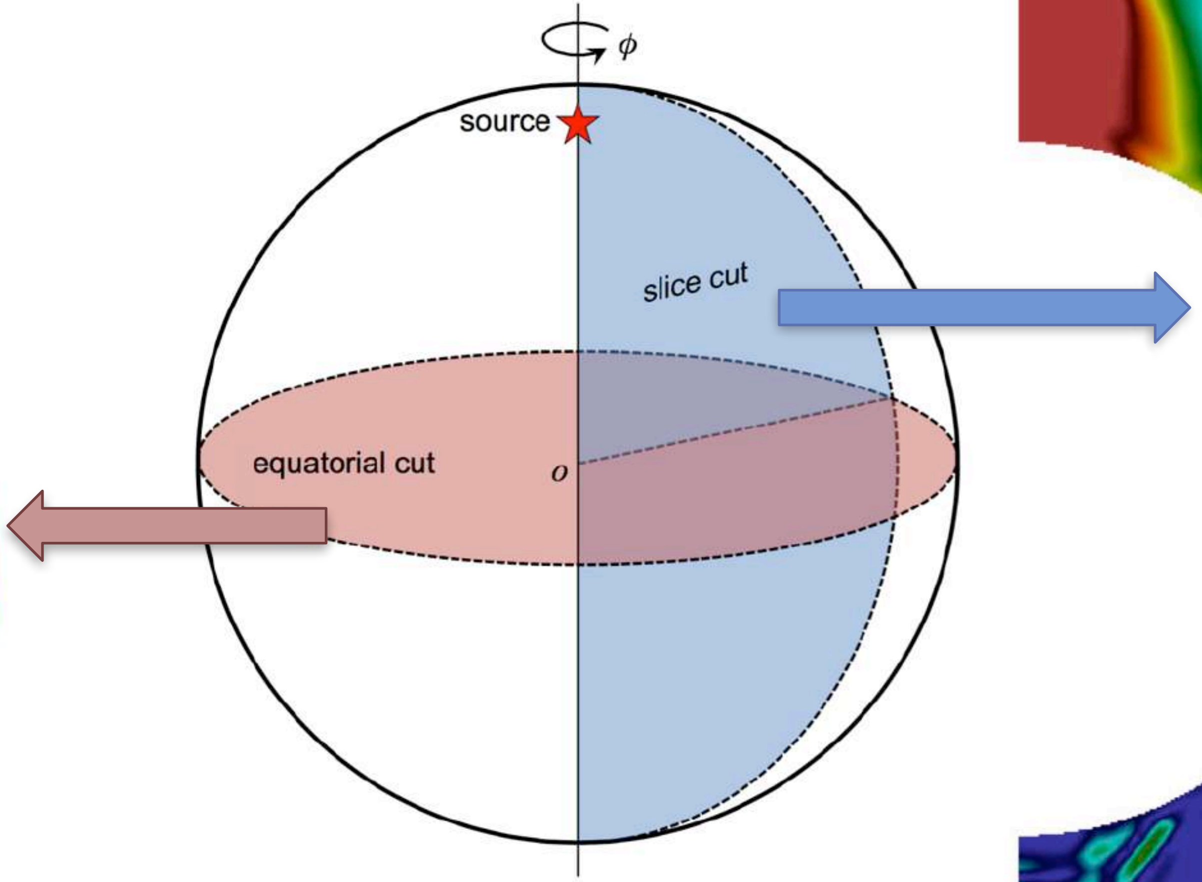
Slice cut

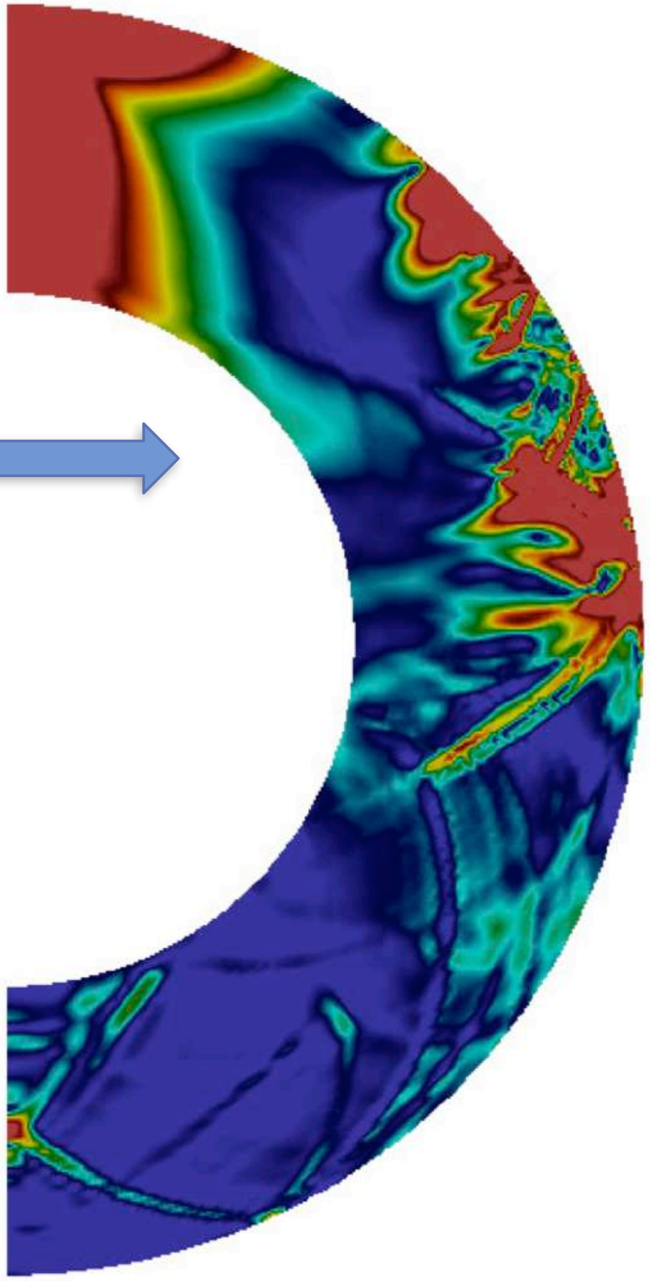
Wavefield in **3D** Earth

Model: S40RTS



Equatorial cut

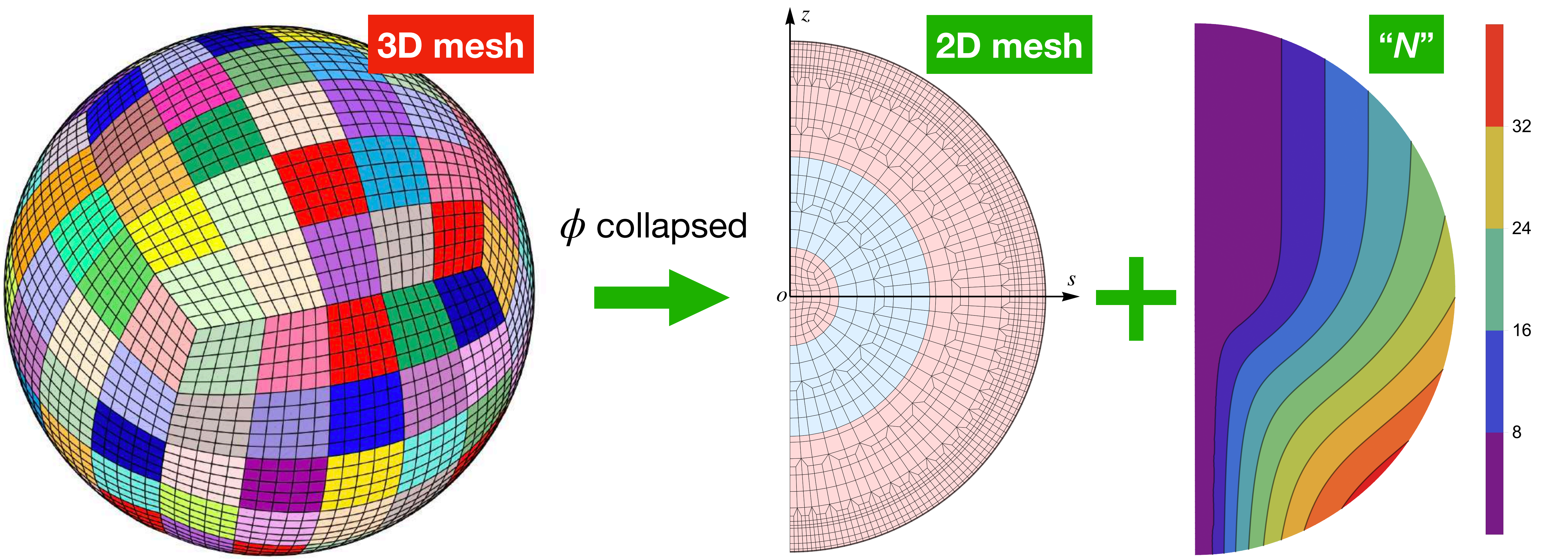




Slice cut

AxiSEM3D

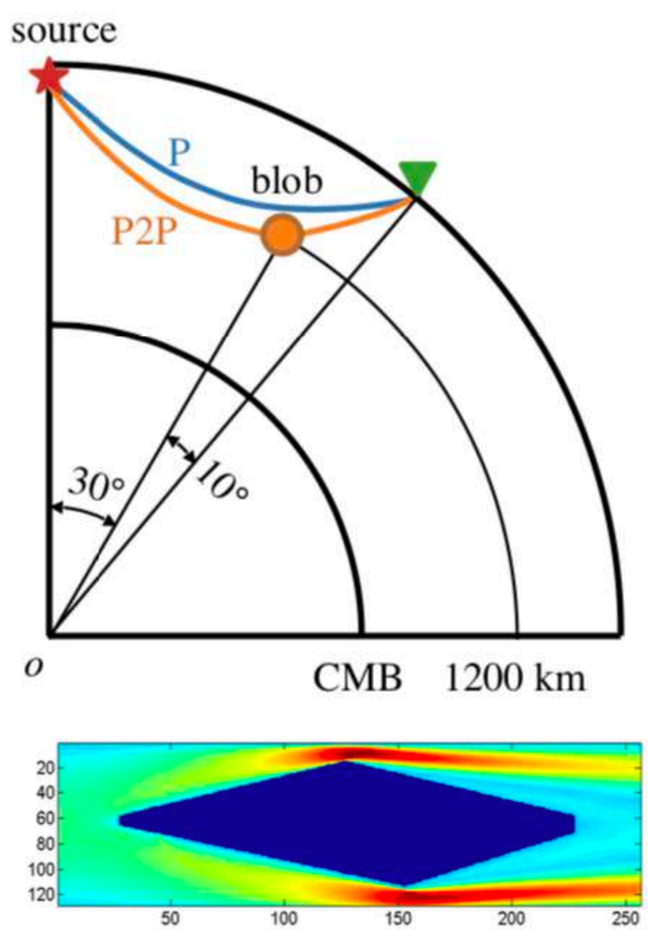
Fourier characterisation of solution: $$u(s,\theta,z) = \sum_{|\alpha|<N(s,z)} u^{\alpha}(s,z)\; e^{i\alpha\phi}$$

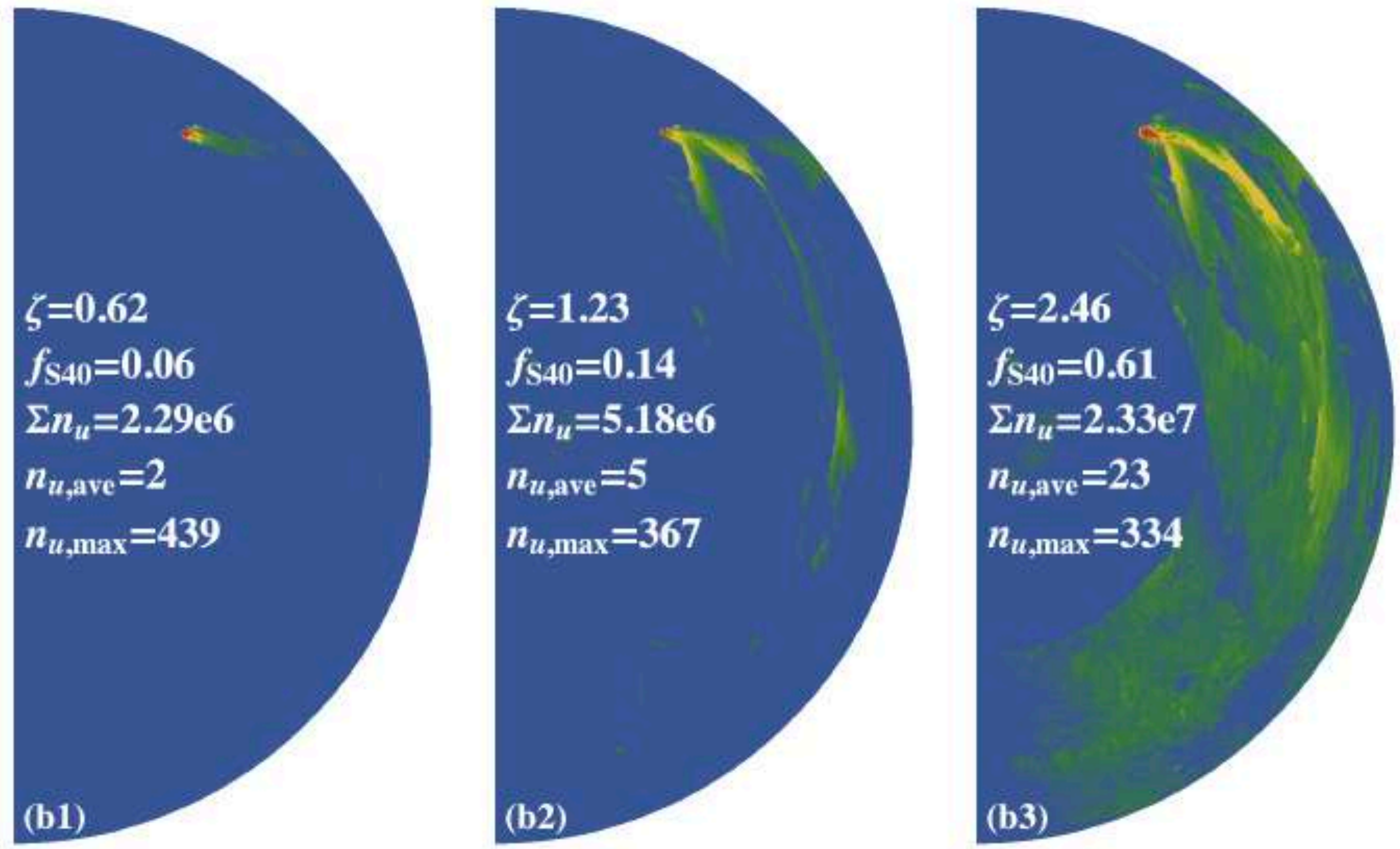


- “N” can be **locally** adapted to wavefield complexity
- Substantially decreases n-DOFs based on **a sparsity in wave physics**
- x100~1,000 speedup for Earth, Mars (Insight mission), Moon and asteroids

Wavefield learning

- *N*-field is the most elusive part of the wave scattering physics
- Difficult (impossible) to derive from the theory





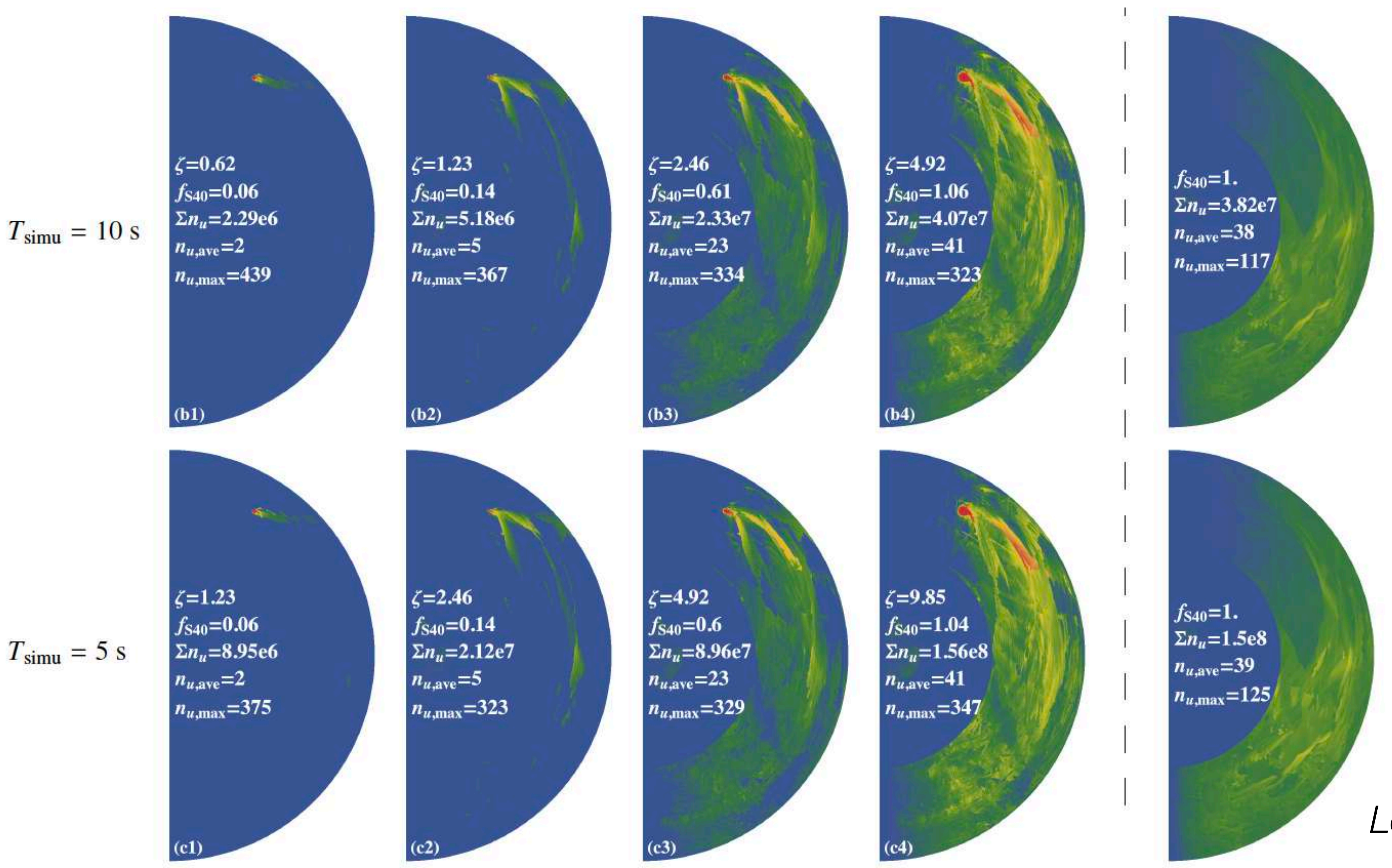
→ Size

- We leave **only** this part to ML

Model End —**FFT**→ *N*-field Model —**ML**→ *N*-field Wave —**SEM**→ Wave End

Wavefield learning

- Interpretable: *N*-field has a clear physical meaning
- Accuracy
  - *N*-field by ML dose not have to be accurate or optimal
  - Error can be evaluated at low-cost end
- Full spatial-temporal solution (*N*-field is static)



Leng et al. 2019

2D inference in model-frequency space

- Extrapolation is via the *N*-fields rather than wavefields
- Maximally keeping the wave physics (interpretability)

✔ Simple model
✔ Low frequency
(laptop)

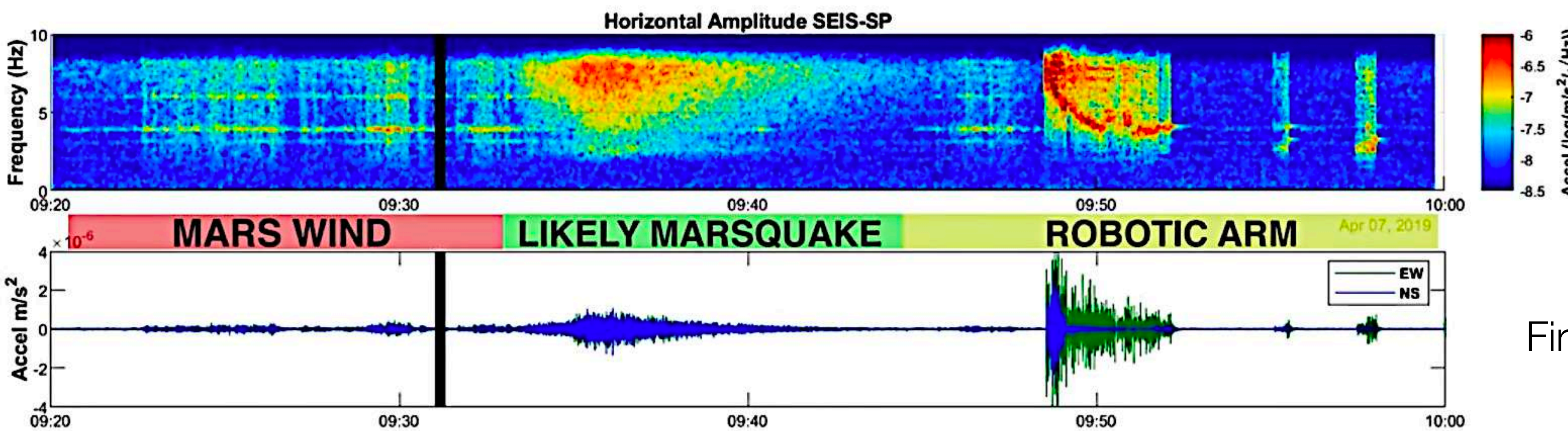
✔ Simple model
✖ High frequency
(small HPC)

✖ Complex model
✔ Low frequency
(small HPC)

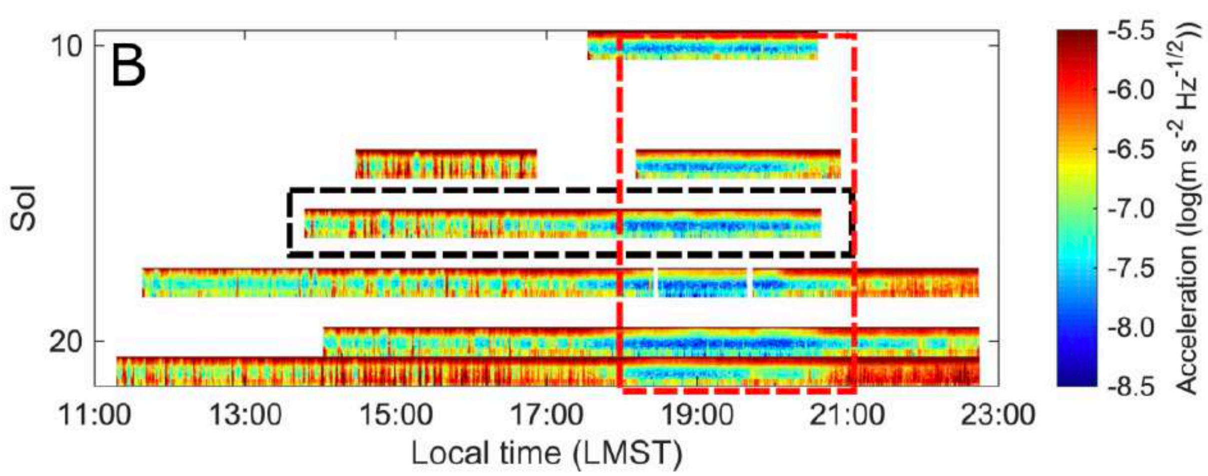
✖ Complex model
✖ High frequency
(large HPC)

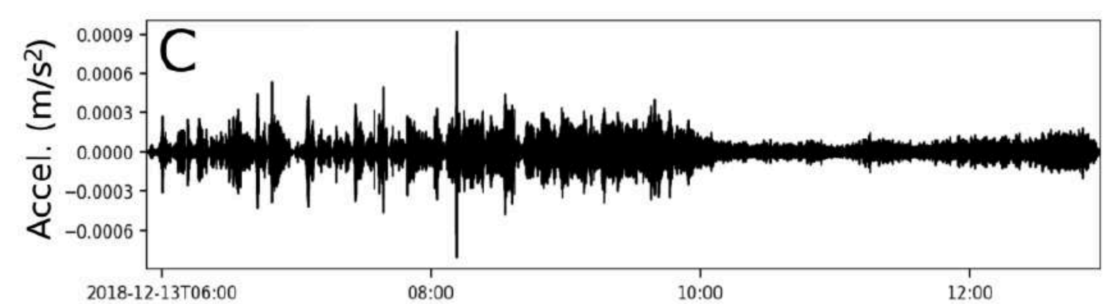
NASA's InSight Mission

- The first mission to explore Mars' deep interior
- Extremely limited observations
- Poorly constrained source and structure
- Seismograms are dissimilar to Earth's



First event detected

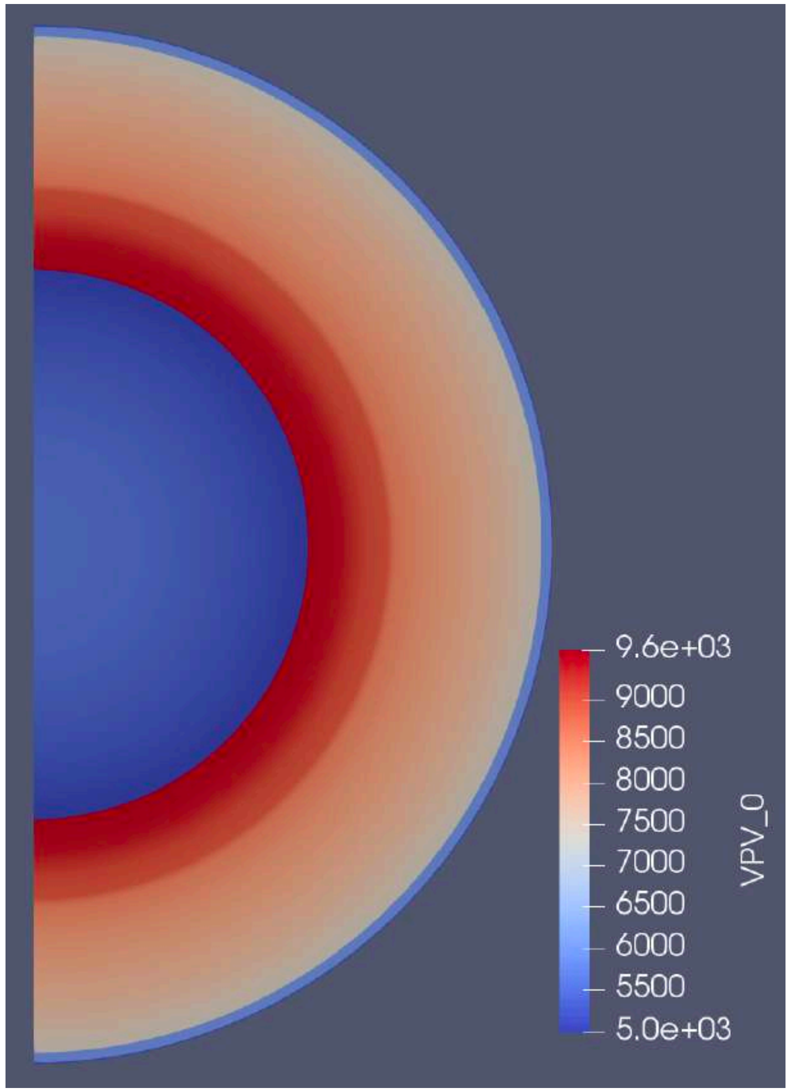




Panning et al. 2020

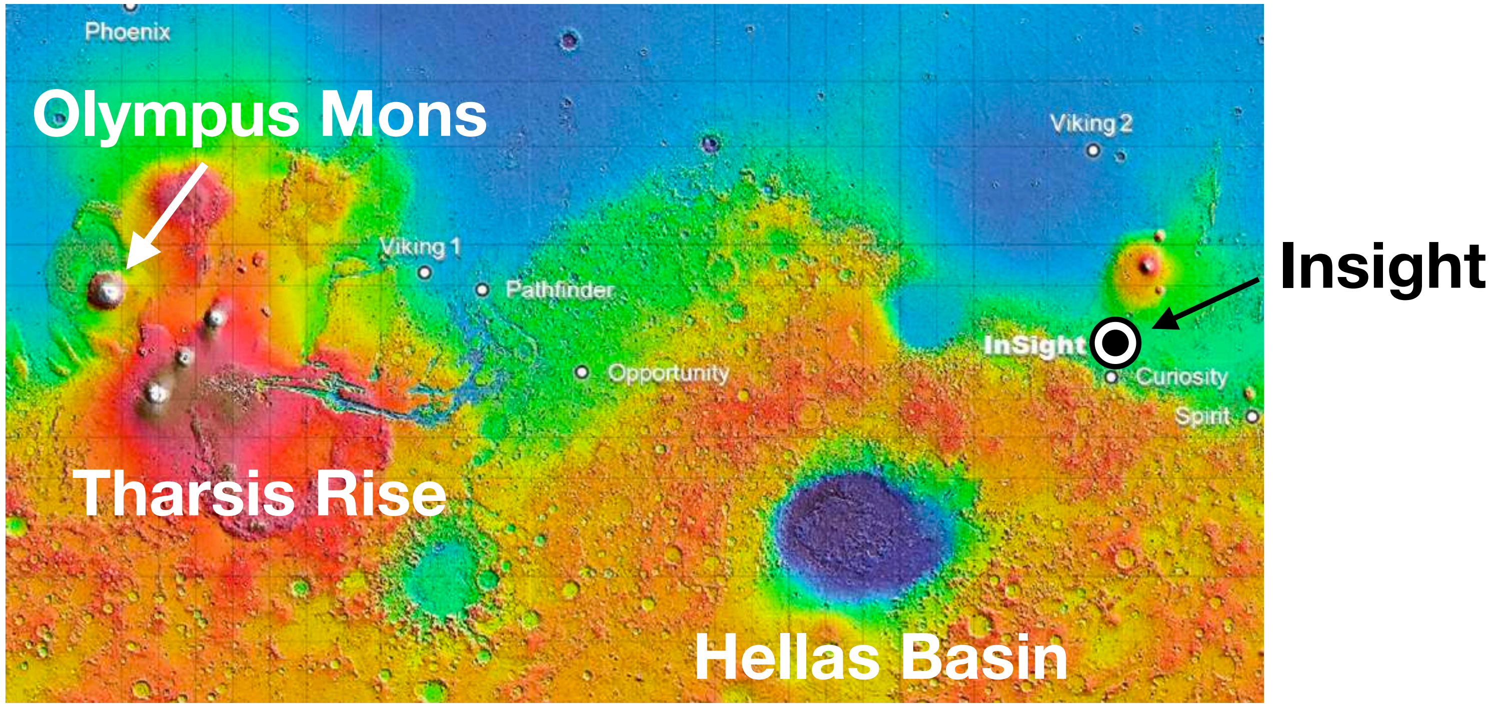
NASA's InSight Mission

- Mars Model

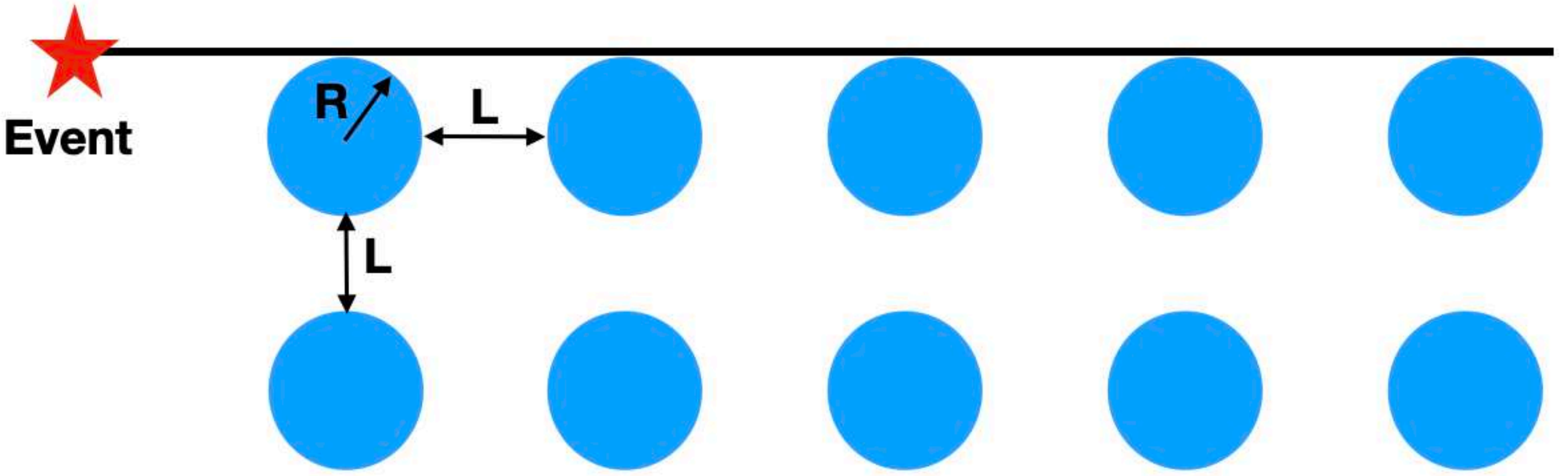


Radial model

Topography



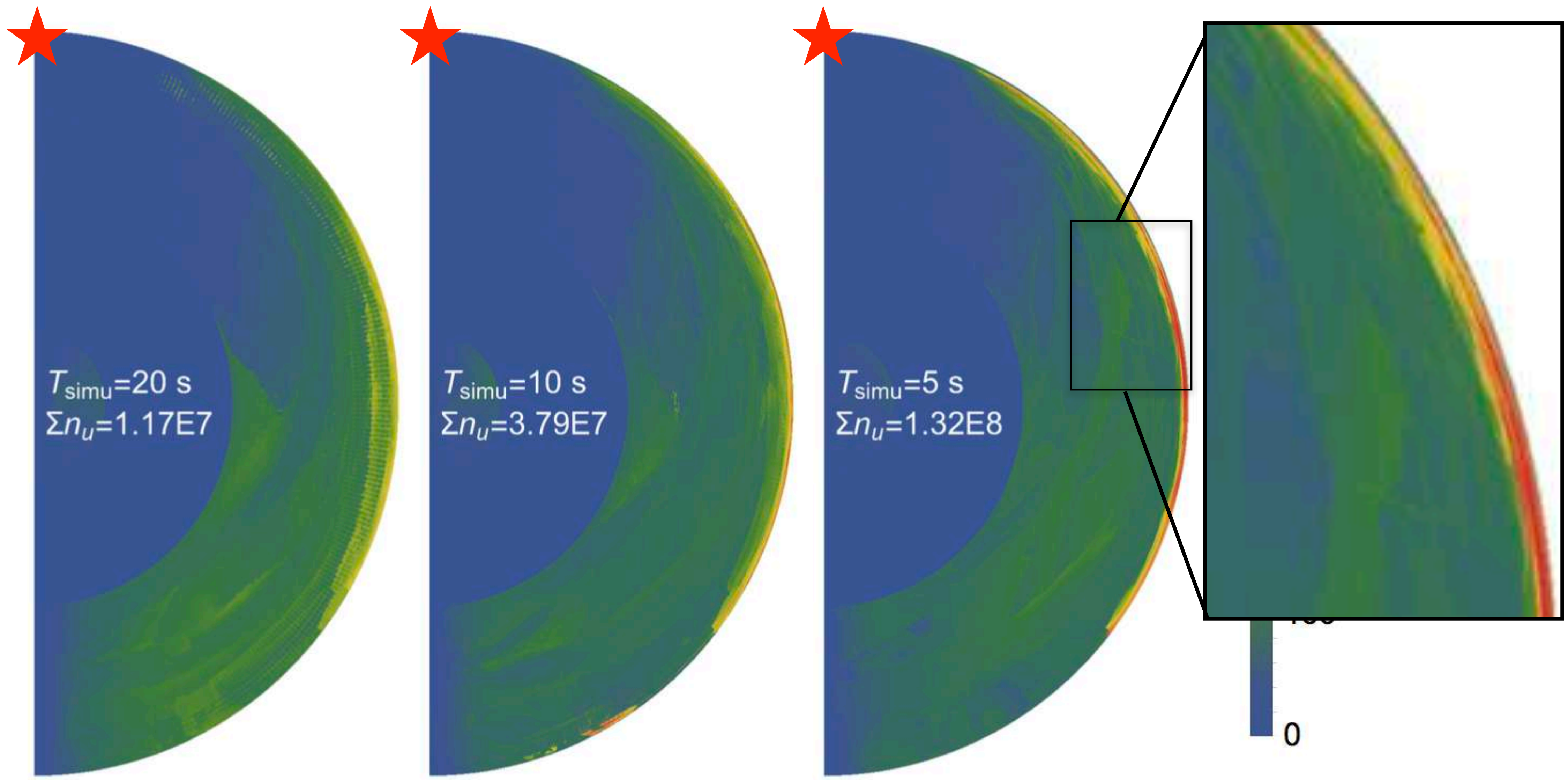
Scatterer model



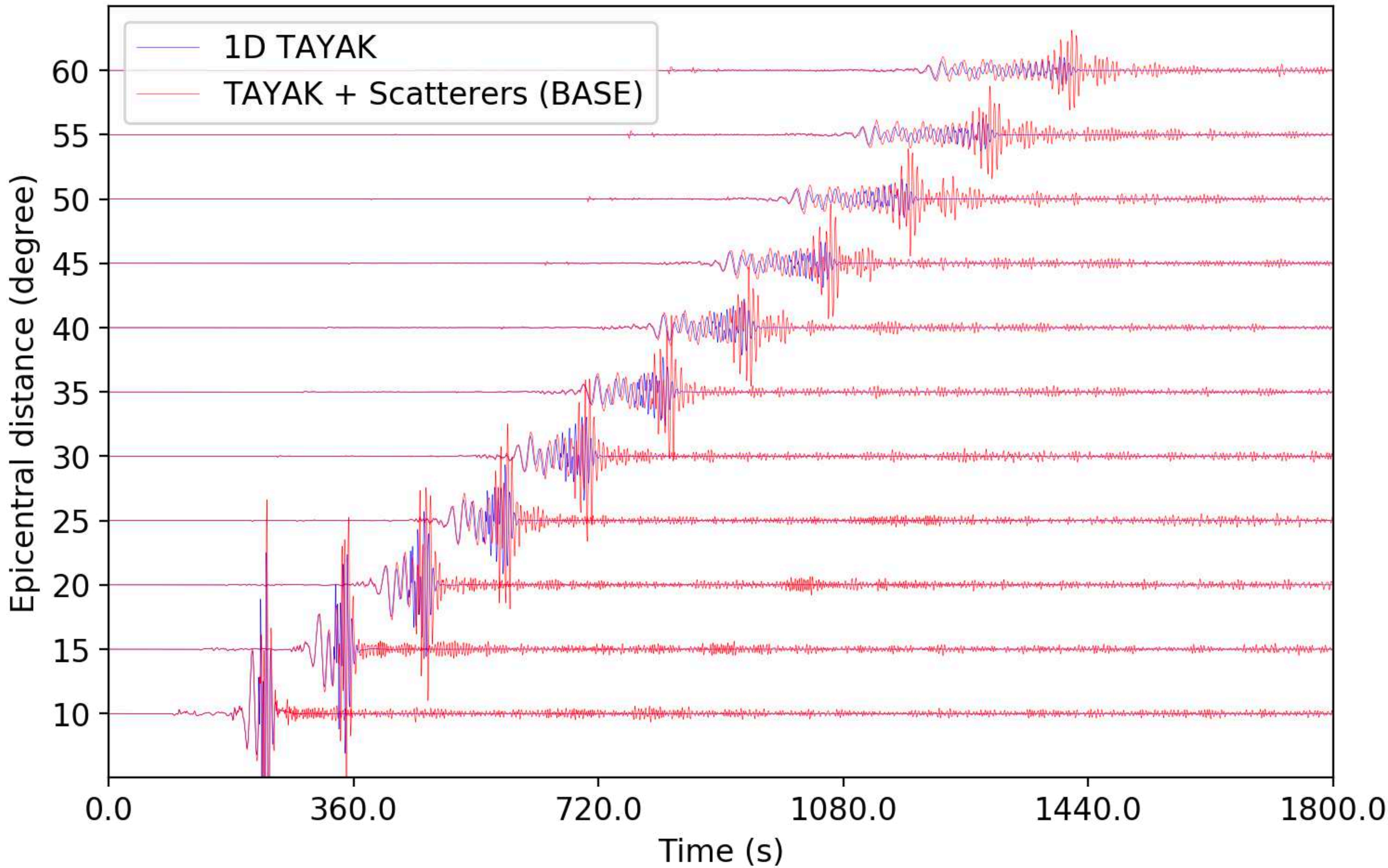
Here N=2, i.e., two scatterer layers along depth

NASA's InSight Mission

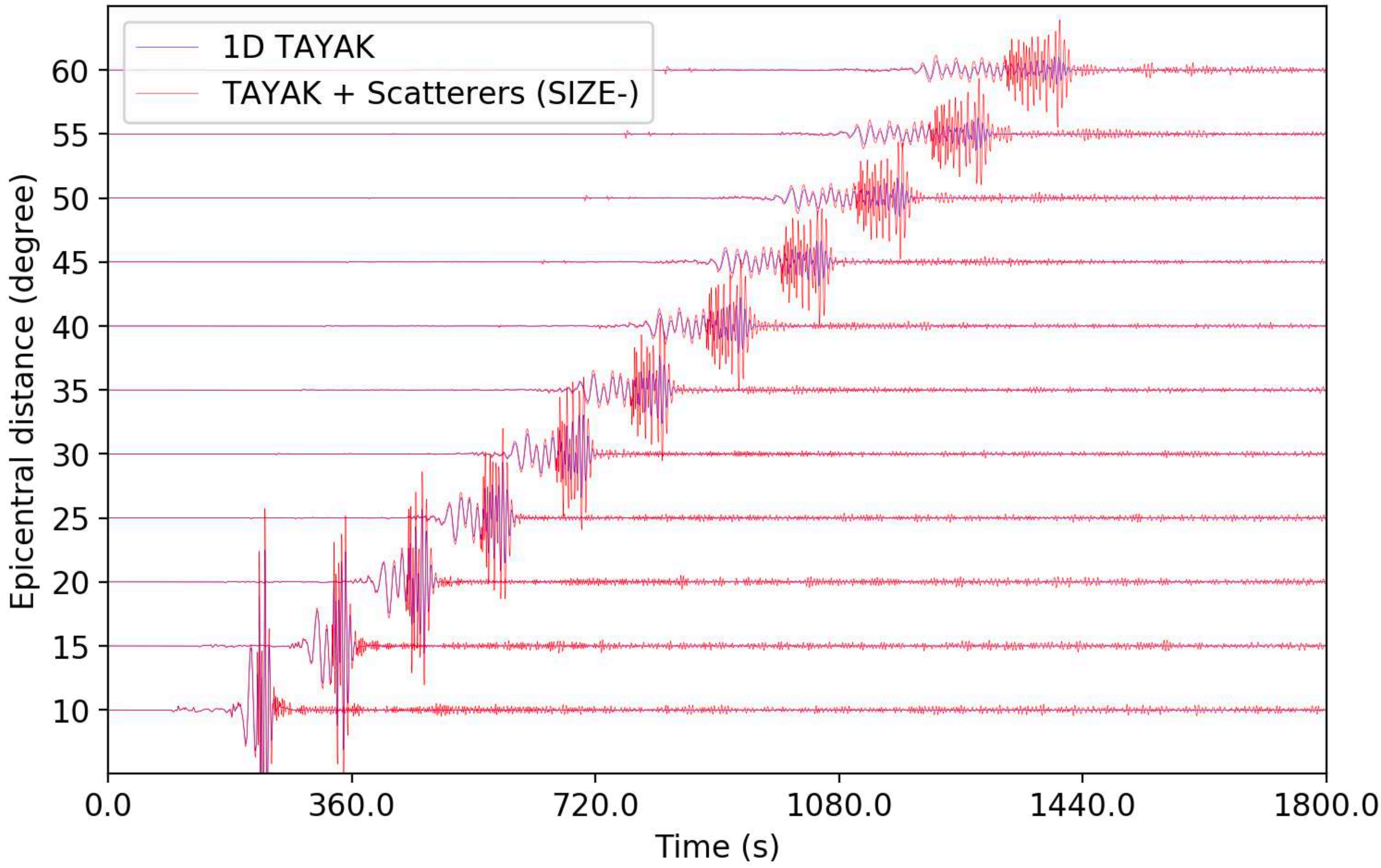
- *N*-field for surface waves
  - Increased near the surface because of strong scattering
  - The large interior is much less affected



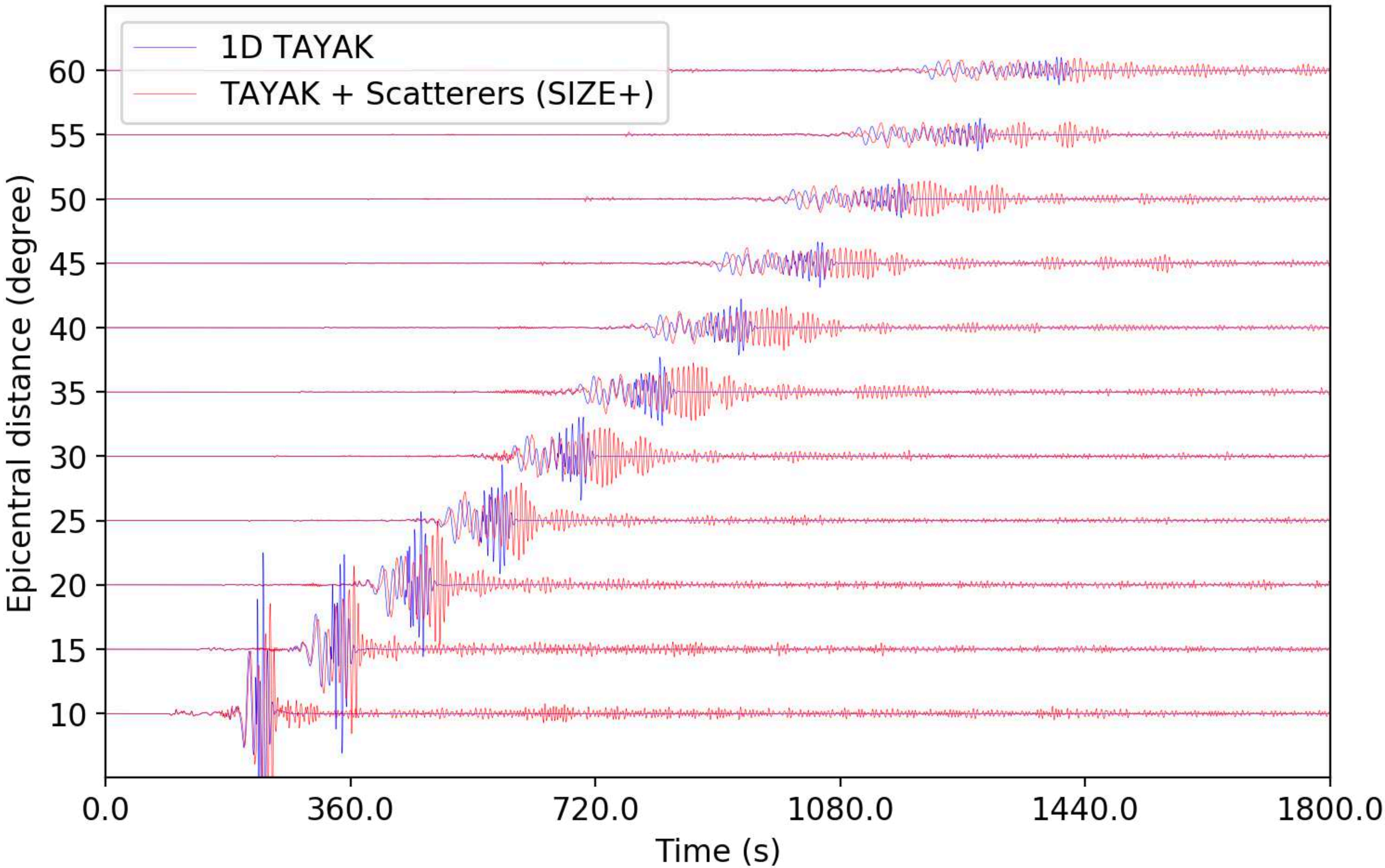
A reference model



Smaller scatterers



Larger scatterers



Summary

- Using ML as an end-to-end blackbox causes low interpretability and difficulty in accuracy evaluation
- Better to identify any sparsity in the physics and built it into the network as a prior

$$m \geq \frac{1}{\epsilon}\left(\ln \mathscr{H} + \ln \frac{1}{\delta}\right)$$

Prior
=
Sparsity

Topics

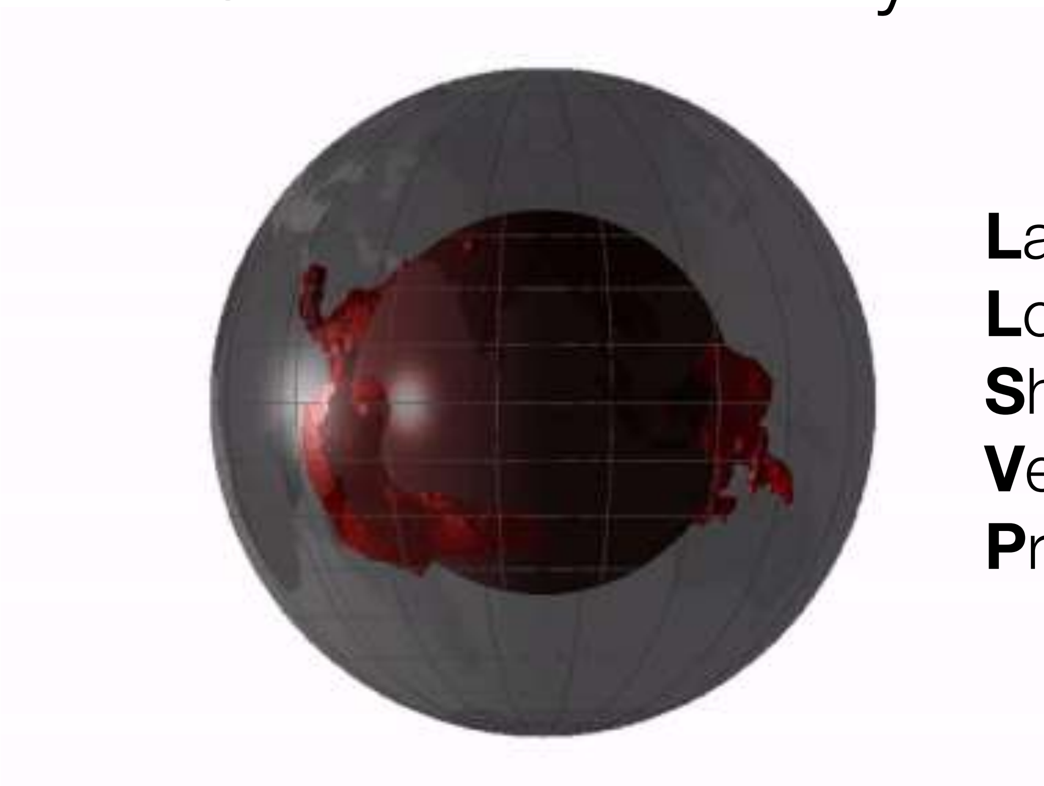
- Global seismology and Earth structure
- “Solving” a PDE-informed system by ML
- Qualifying seismic data by ML

Seismic data

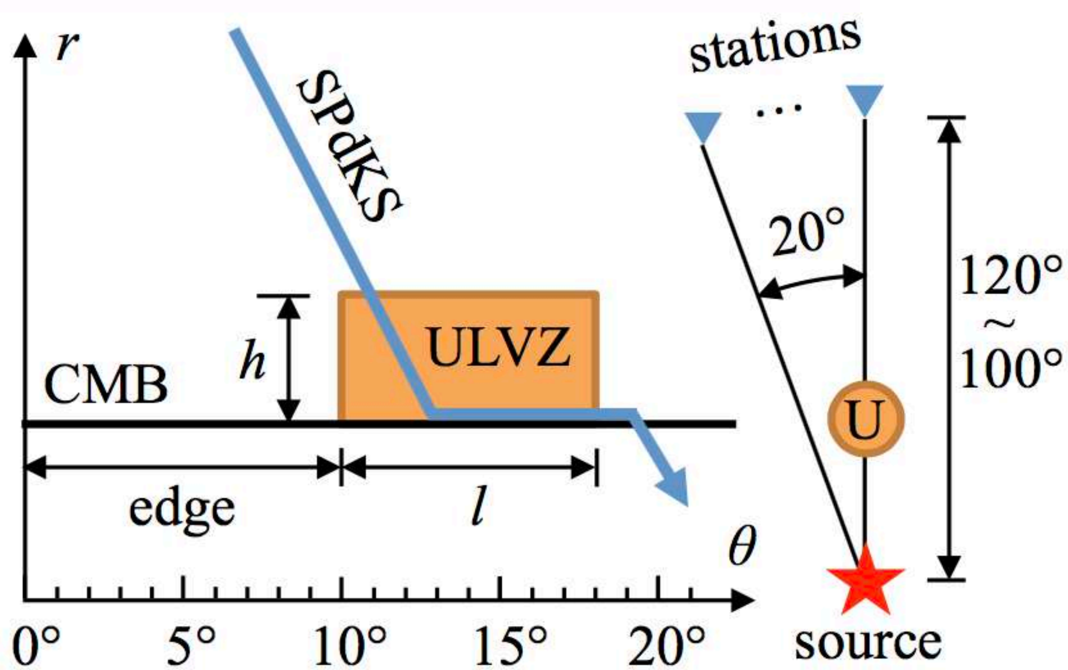
- Seismology is a physics-driven data science
- Teleseismic data is noisy with diverse noise patterns (scattered wave is one of them, careful with denoising)
- Data qualification: $\mathbf{d} = \underset{\mathbf{d}}{\operatorname{argmax}} P(\mathbf{d} \,|\, \mathbf{m}_0)$
- Previous ML-based approaches (SVM, GAN…) for seismic data is based on hand-labelled training data
  - Limited number of training data (3 months for 10,000 data)
  - Subjective, probably biased
  - Undefined uncertainty

Our data

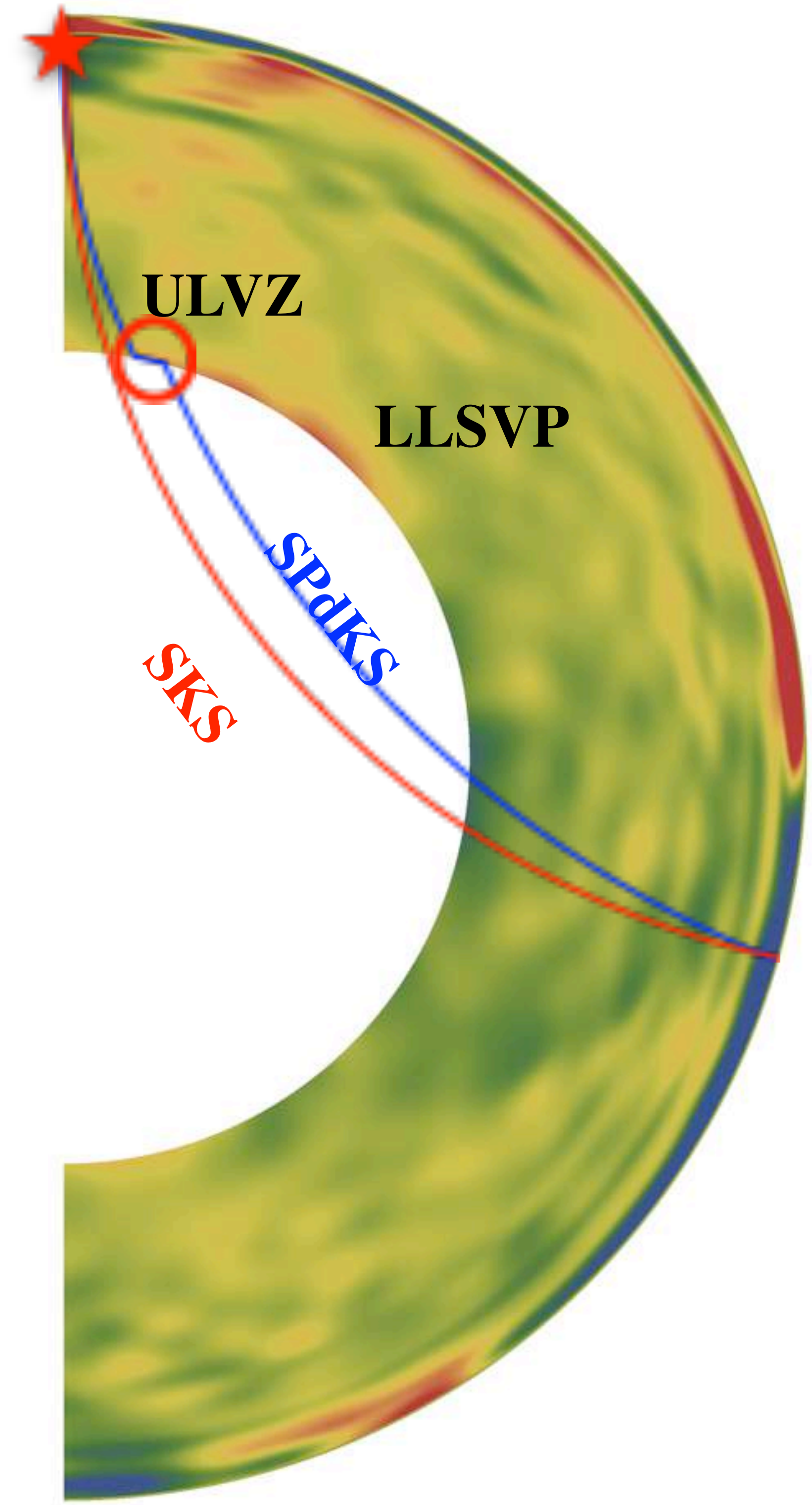
- **U**ltra-**L**ow **V**elocity **Z**one (ULVZ) and SKS-SPdKS



Large
Low
Shear
Velocity
Province

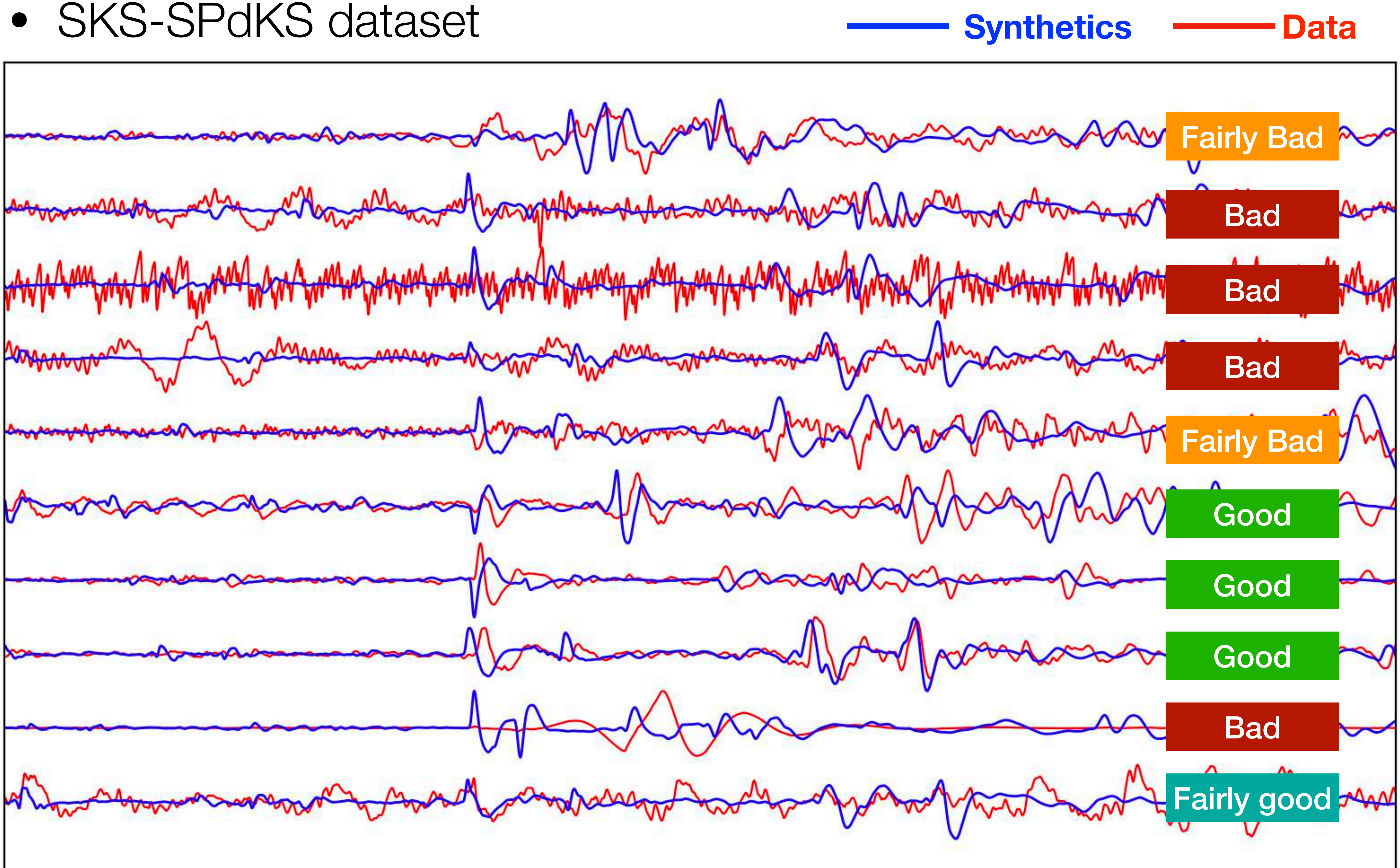


Thorne et al, 2013



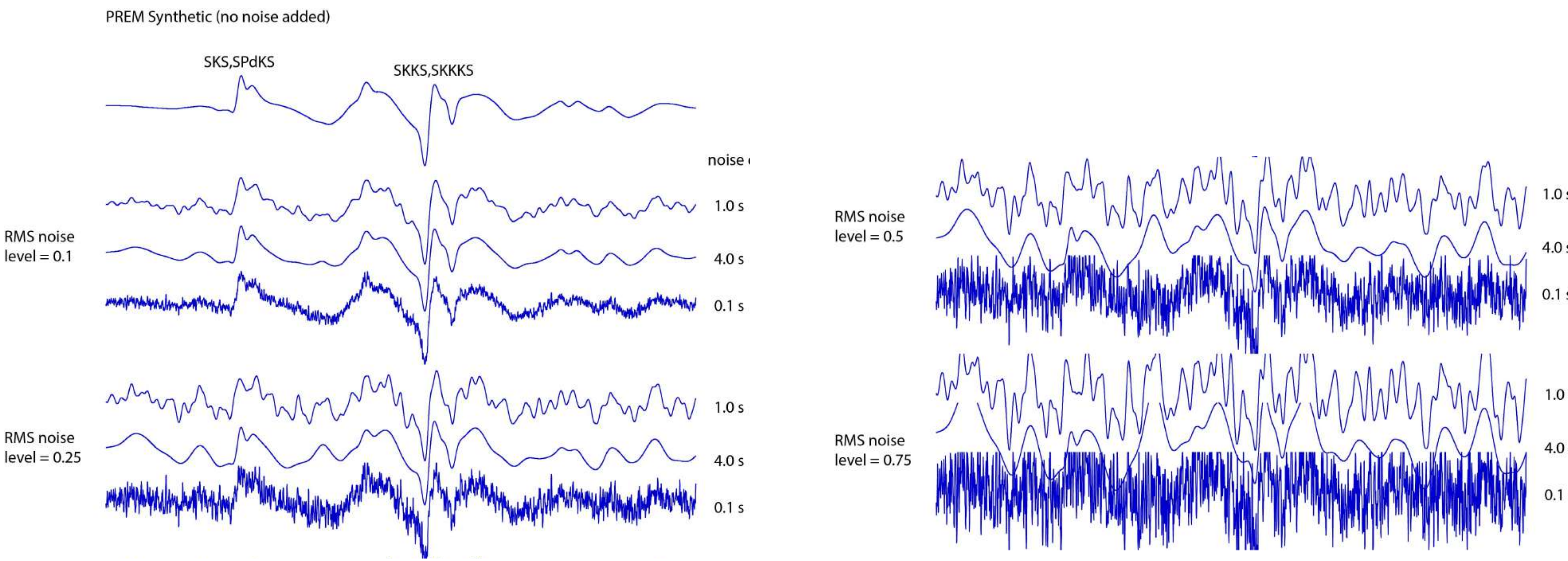
Our data

- SKS-SPdKS dataset

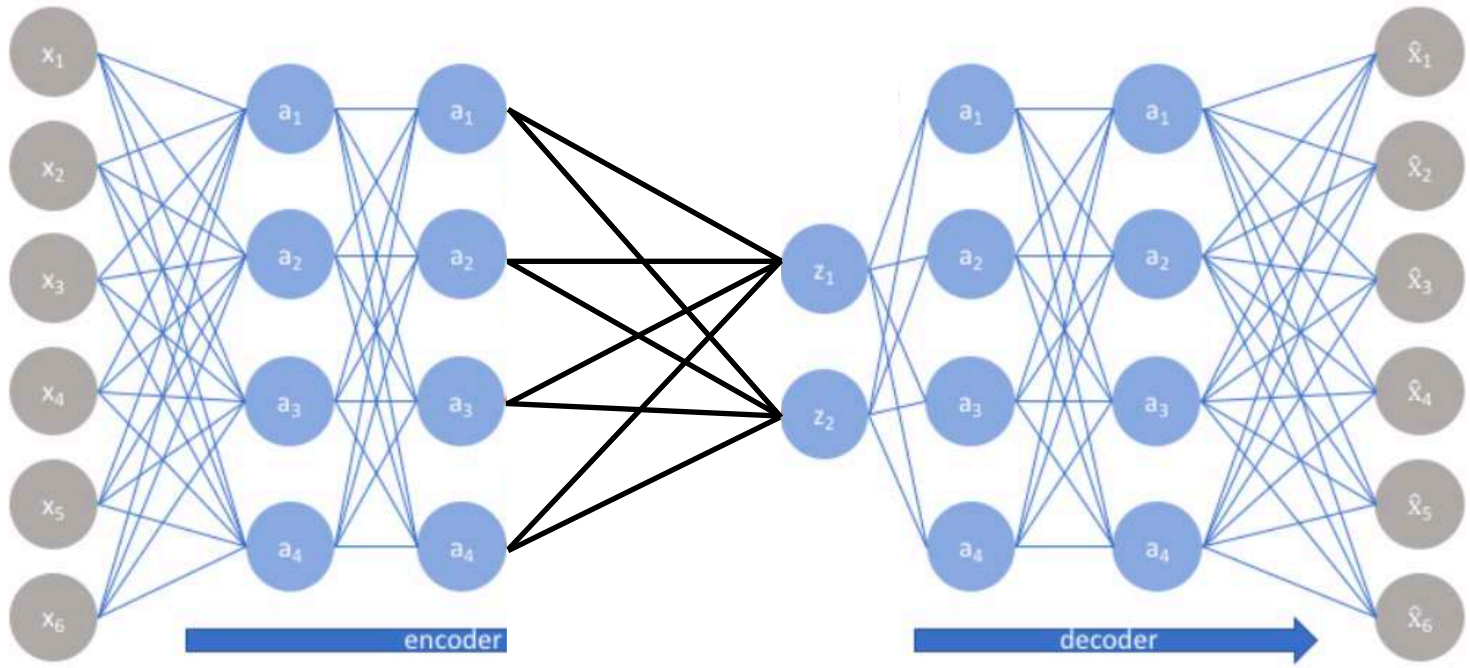


Methods

- Traditional, non-ML (still subjective, satisfaction ~70%)
- Active noise (satisfaction ~80%, strong overfitting)
    - The machine quickly learns the synthetic noise patterns (satisfaction 96% for synthetic noise data)
- Variational autoencoder (ongoing, satisfaction ~100% for “good”)



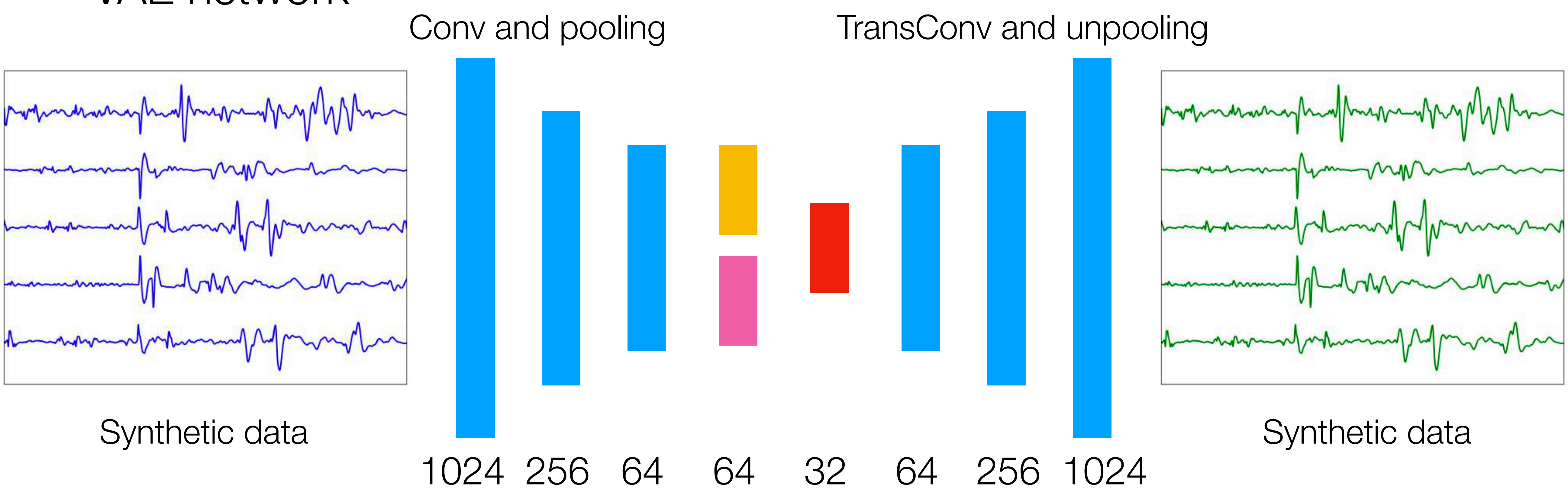
Variational autoencoder (VAE)



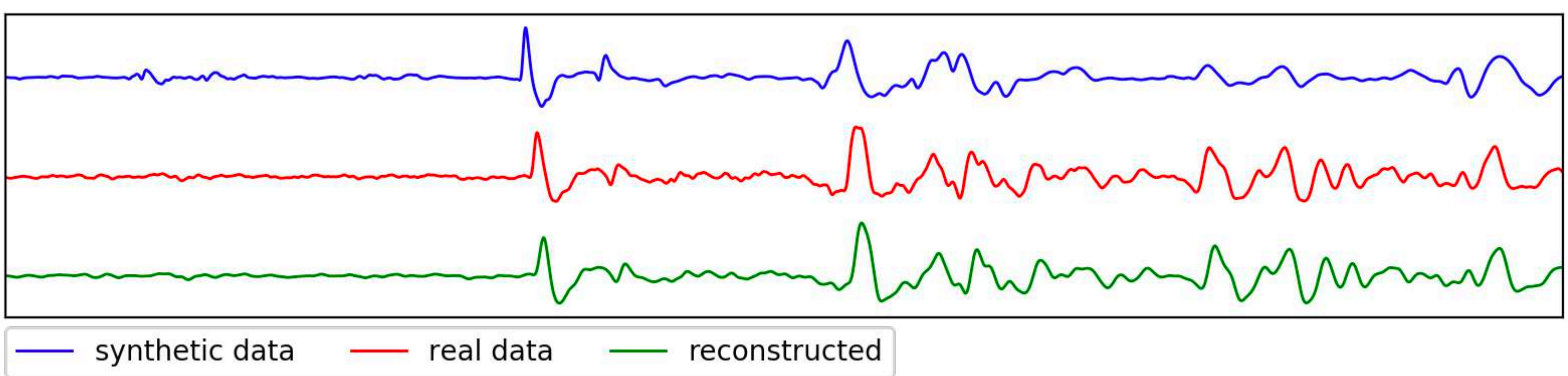
- Use the same data for input and output during training
- VAE can find the sparsity in data (principal components)
- Can be used for image compression and denoising
- If a test datum resembles the training dataset, it should be highly reconstructable: we may use **pure synthetic data** for training

Variational autoencoder (VAE)

- VAE network

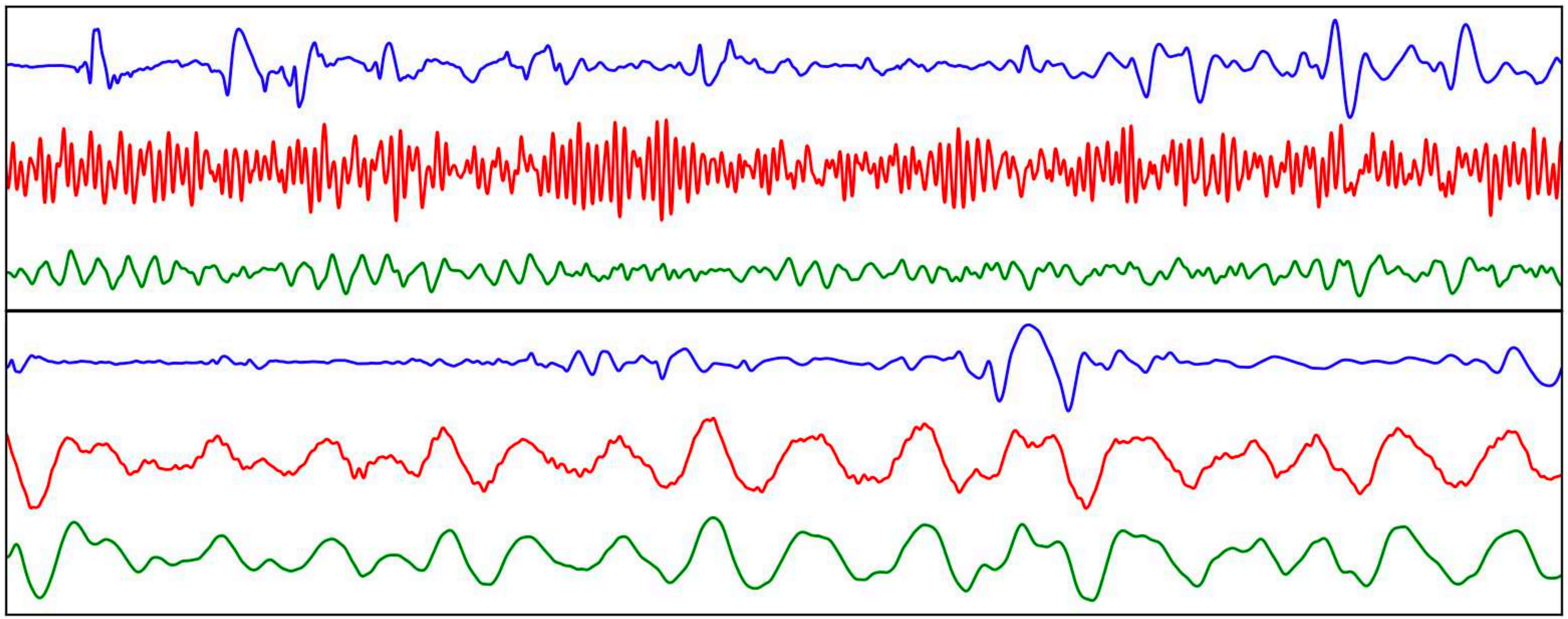


- A GOOD one



Variational autoencoder (VAE)

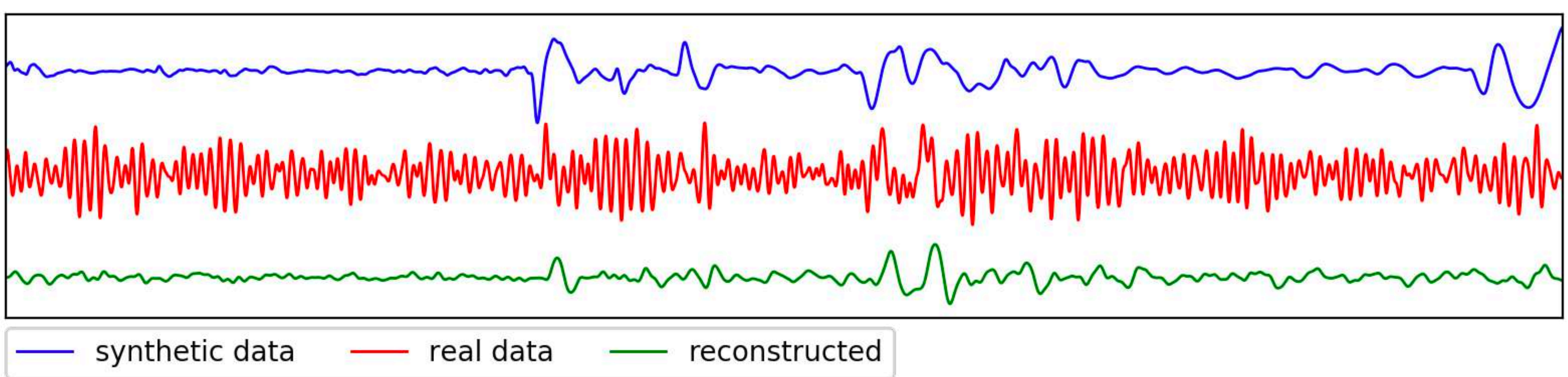
- BAD ones



✔

✗

- One from BAD to GOOD



✔?

Thank you!