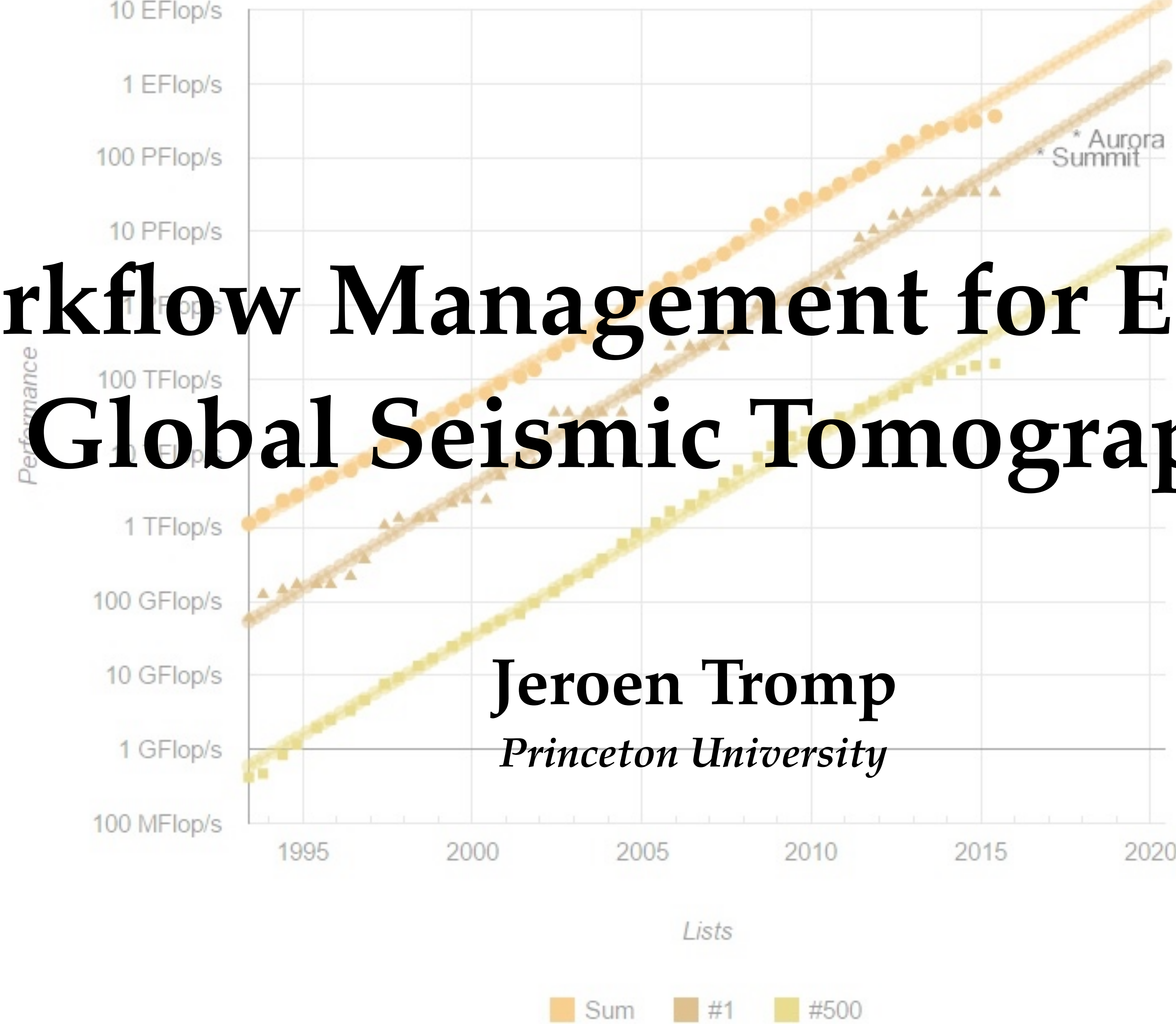
Workflow Management for Exascale Global Seismic Tomography



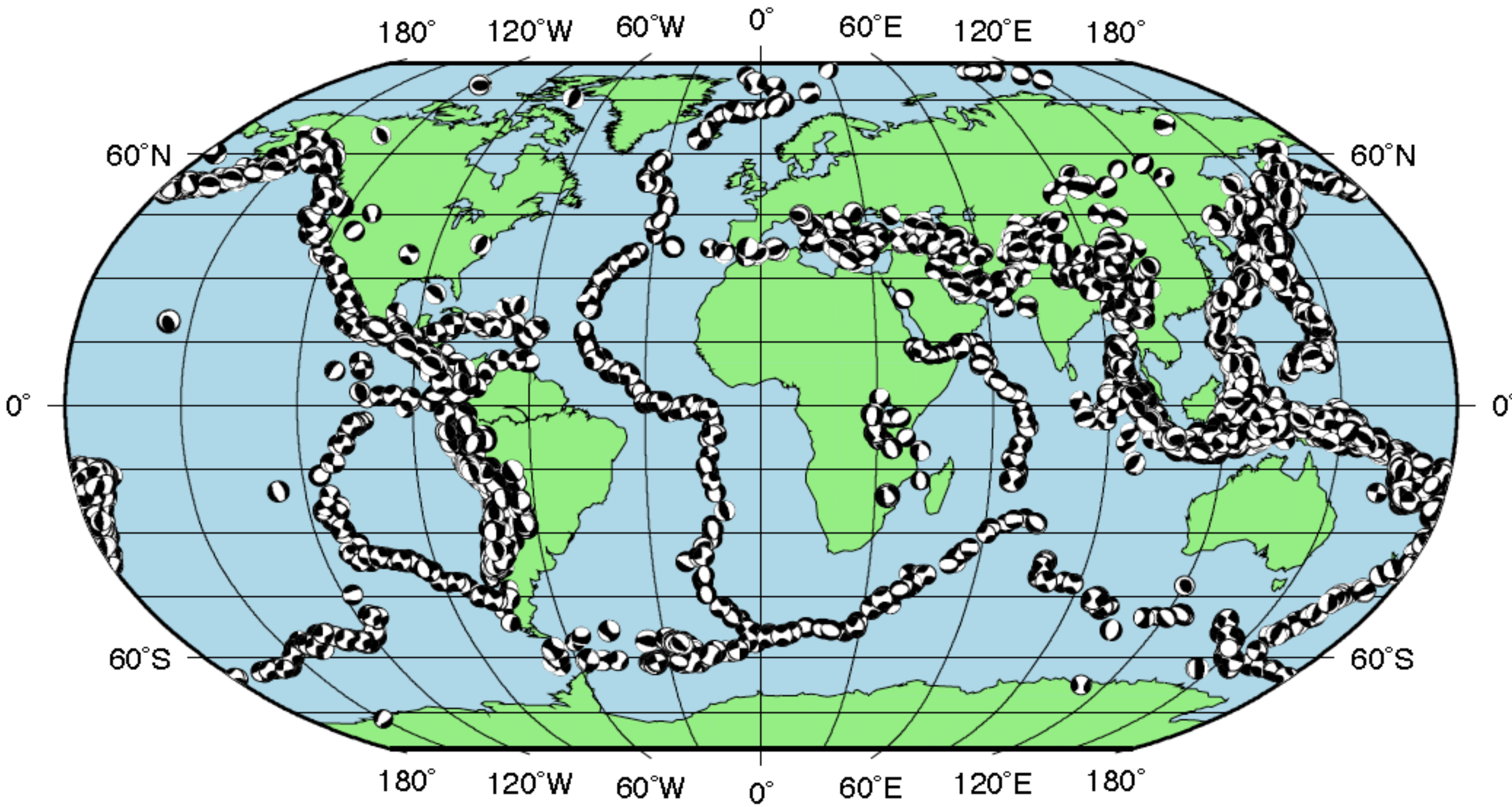
Jeroen Tromp

Princeton University

Ebru Bozdağ, Judith Hill, Shantenu Jha, Dimitri Komatitsch, Matthieu Lefebvre, Wenjie Lei, Daniel Peter, Norbert Podhorszki, David Pugmire, Youyi Ruan, James Smith & Matteo Turilli

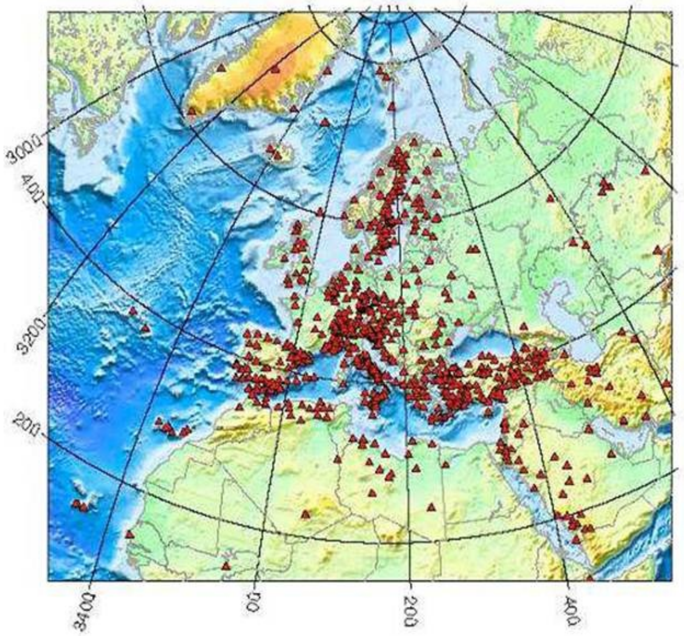
Colorado School of Mines, University of Marseille, Princeton University, Rutgers University, KAUST & ORNL

Exascale Goal: Use All Available Data

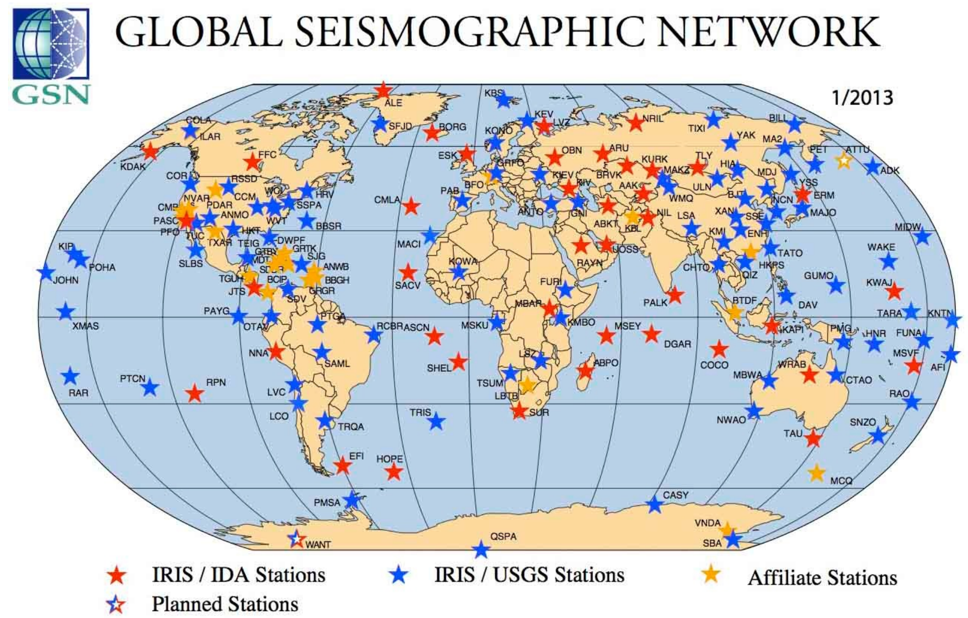


More than 6,000 earthquakes (5.5 ≤ Mw ≤ 7.0) since 1999

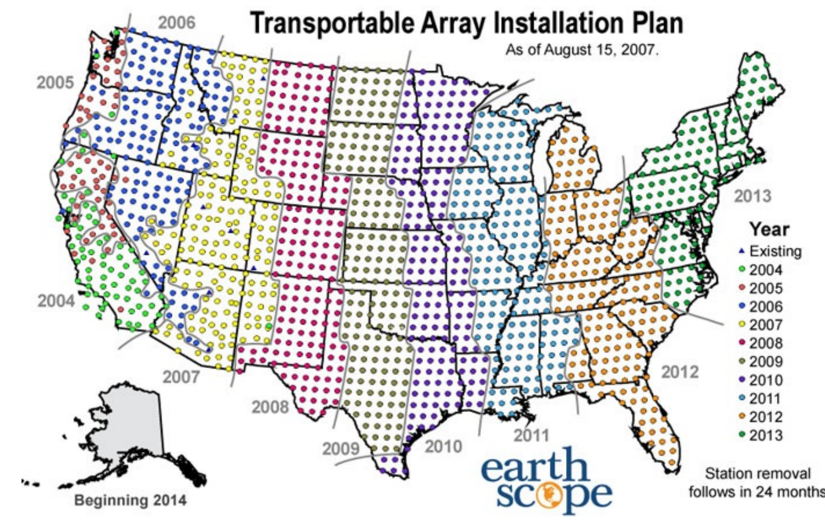
Data Tsunami in Regional & Global Seismology



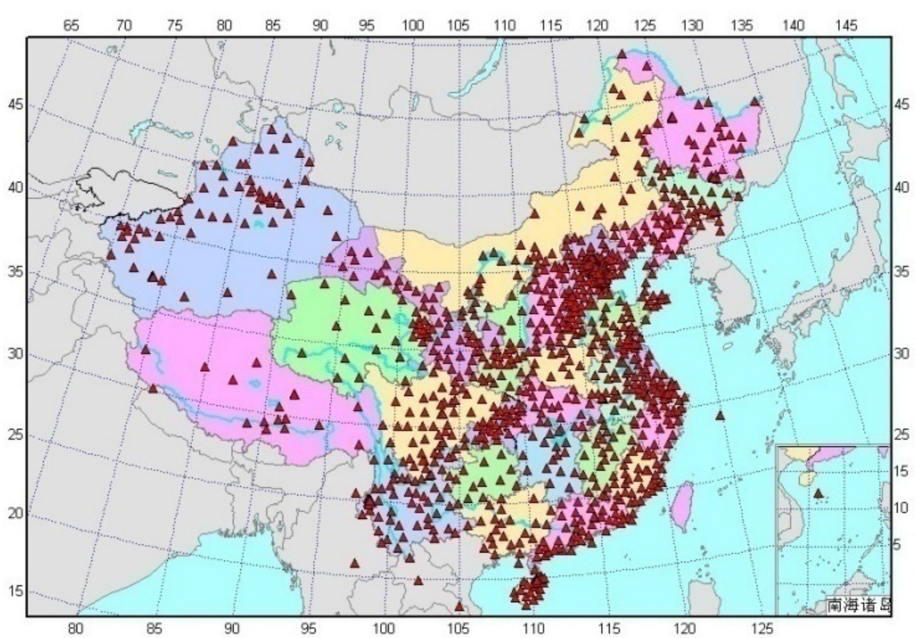
[www.geo.uib.no]



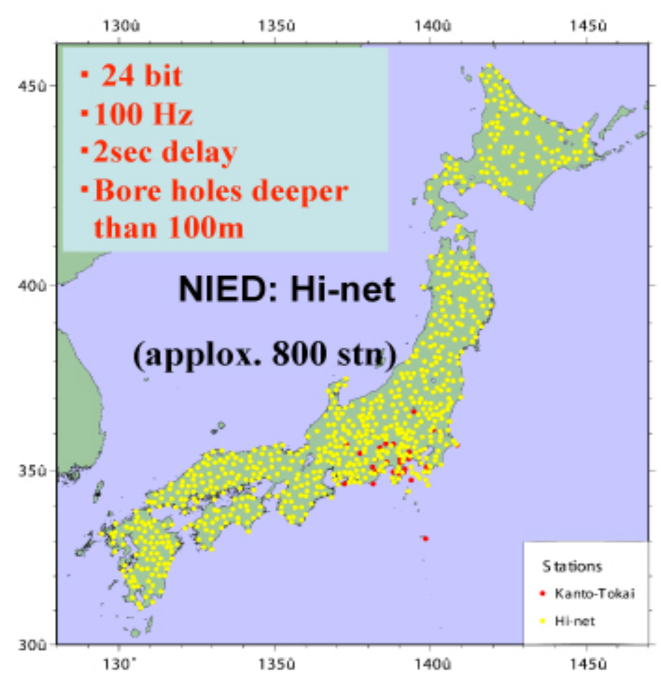
[www.iris.edu]



[web.mst.edu]

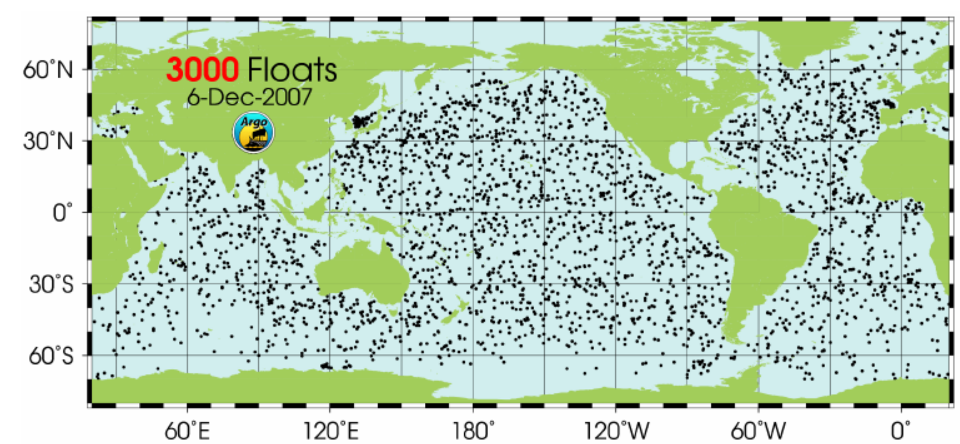


[data.earthquake.cn]



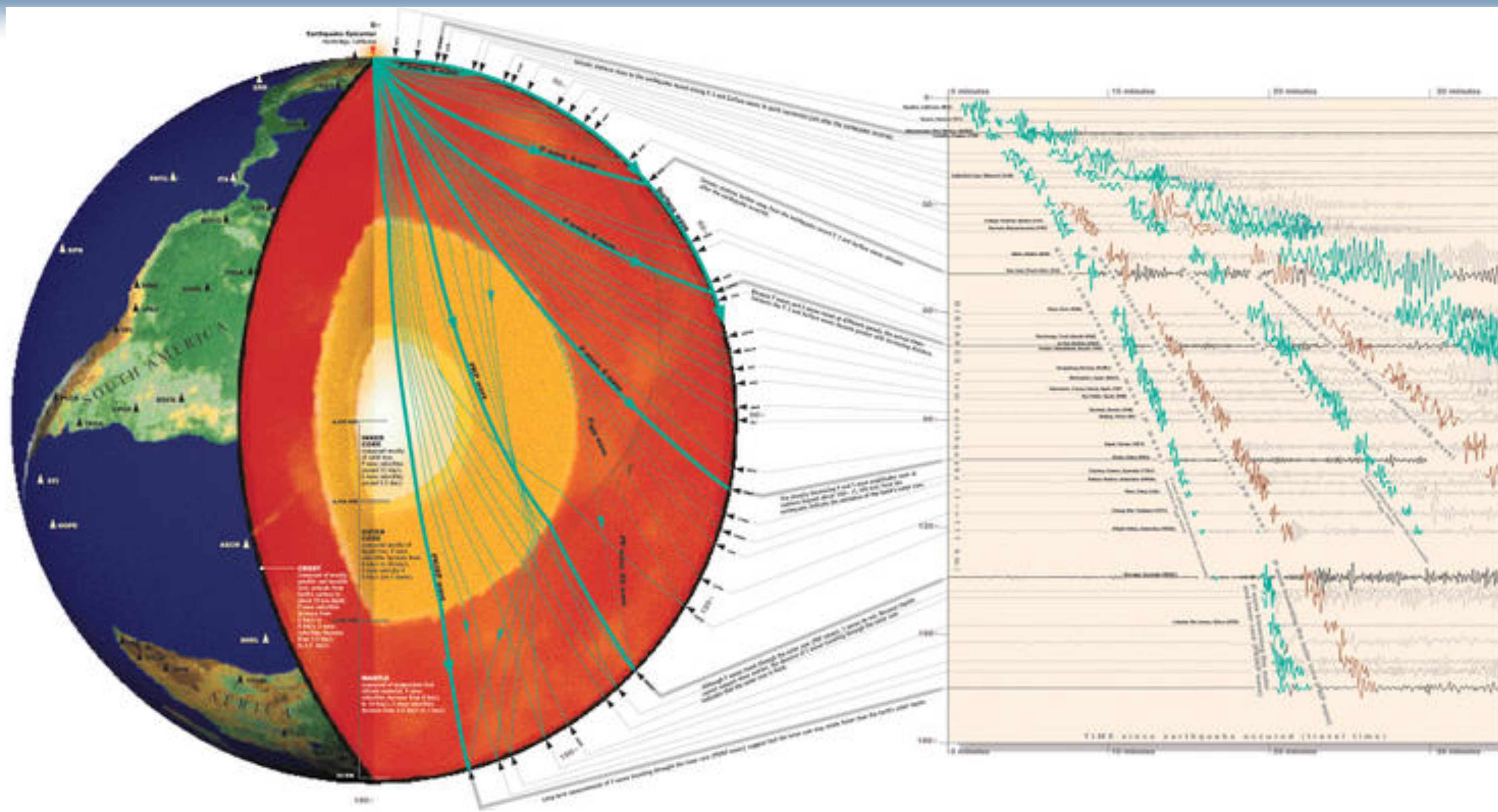
[drh.edm.bosai.go.jp]

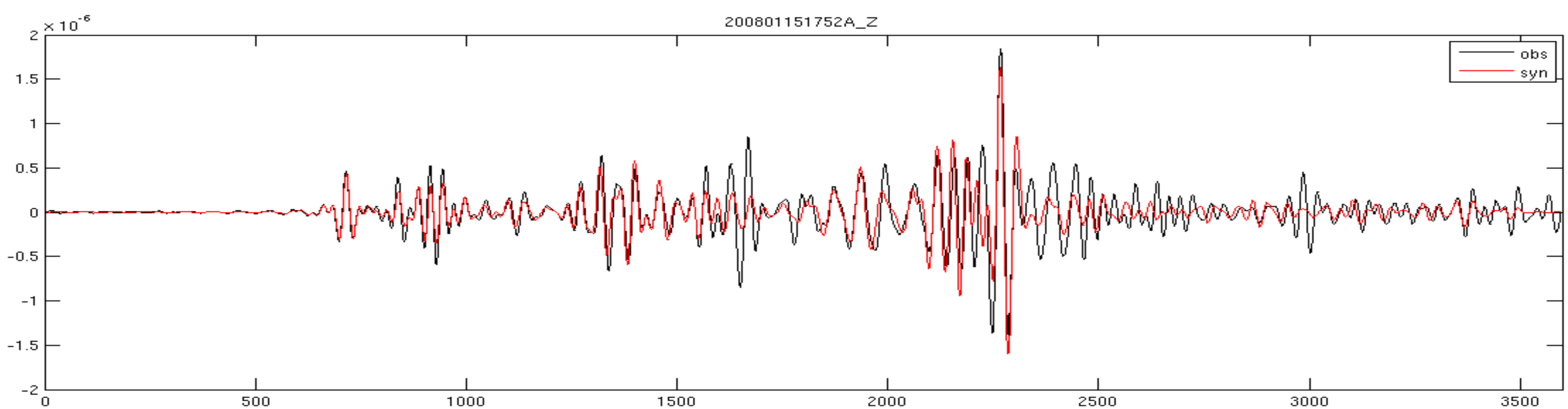
MERMAID/MariScope



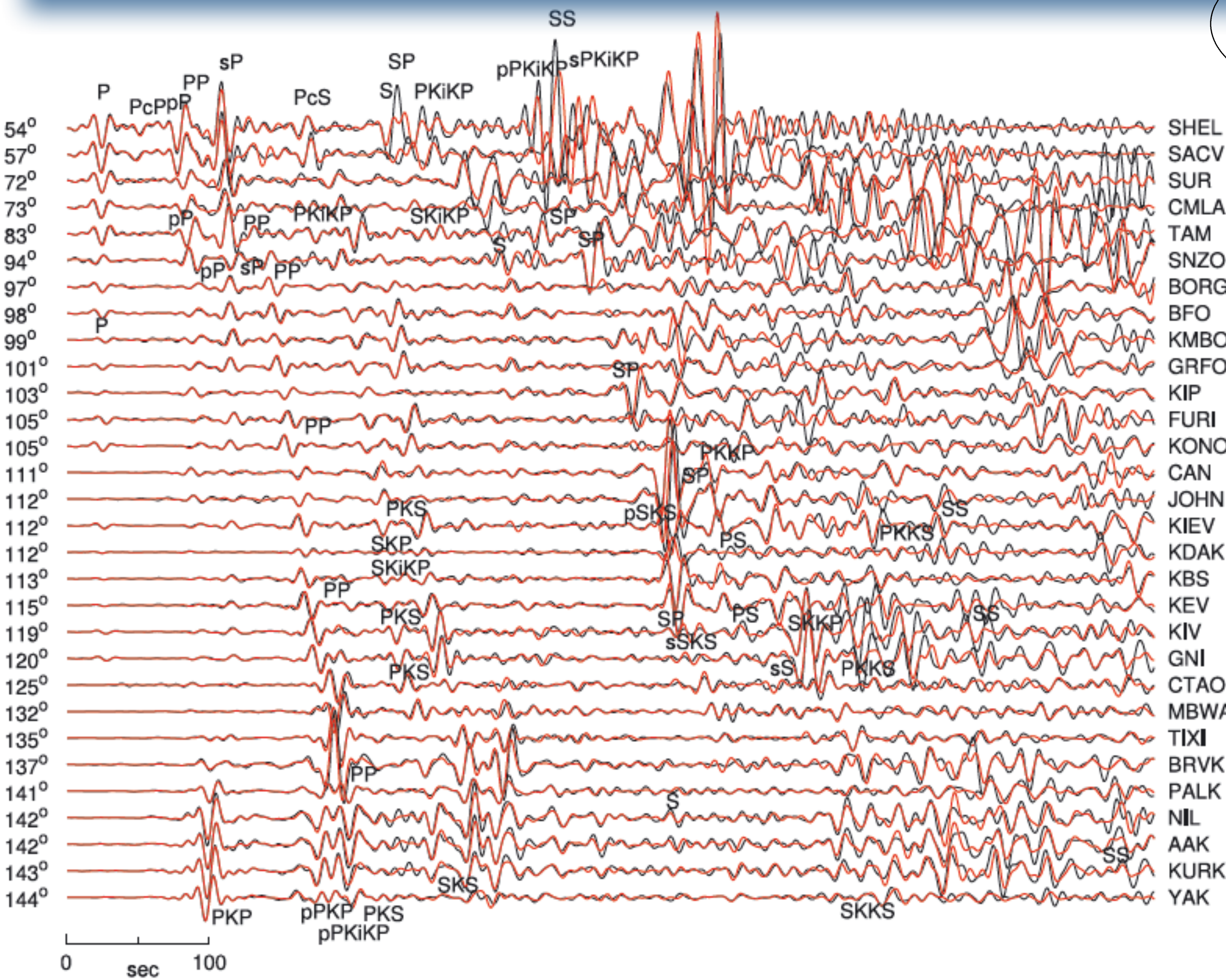
[Simons et al, 2006]

Exascale Goal: Use Complete Seismograms





We Really Use "Full Waveforms"



Sept. 3, 2008,
Argentina
Earthquake
$M_w = 6.3$
d = 571 km

SPECFEM3D_GLOBE

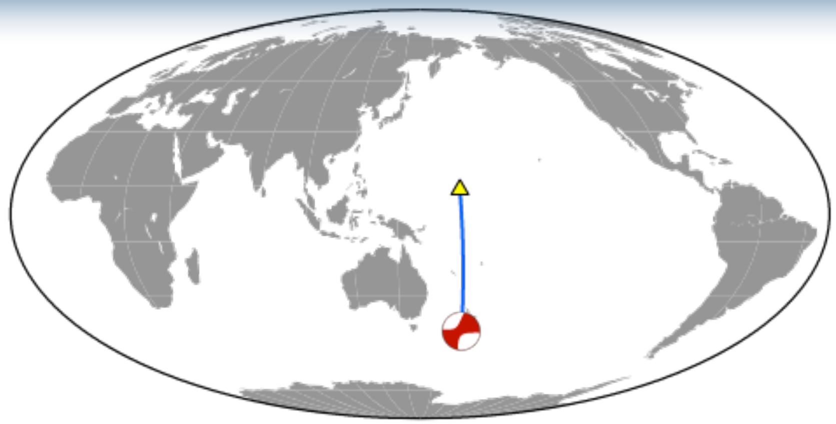
3D Earth model:
S362ANI + Crust2.0

17 – 60 s

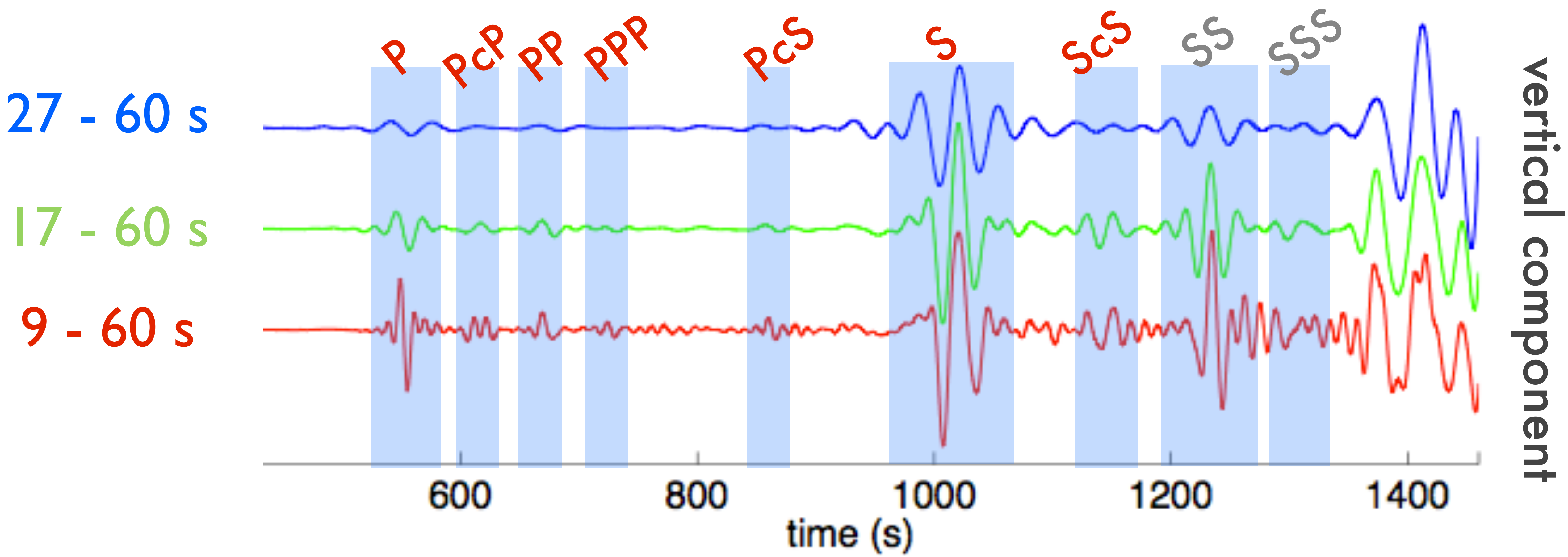
Ebru Bozdağ

Exascale Goal: Higher-Frequency Waves

Short term goal (2019): 9 s (“Summit”)
Long term goal (2021): 1 s (“Frontier”)



Station: KWAJ
Δ= 52 °



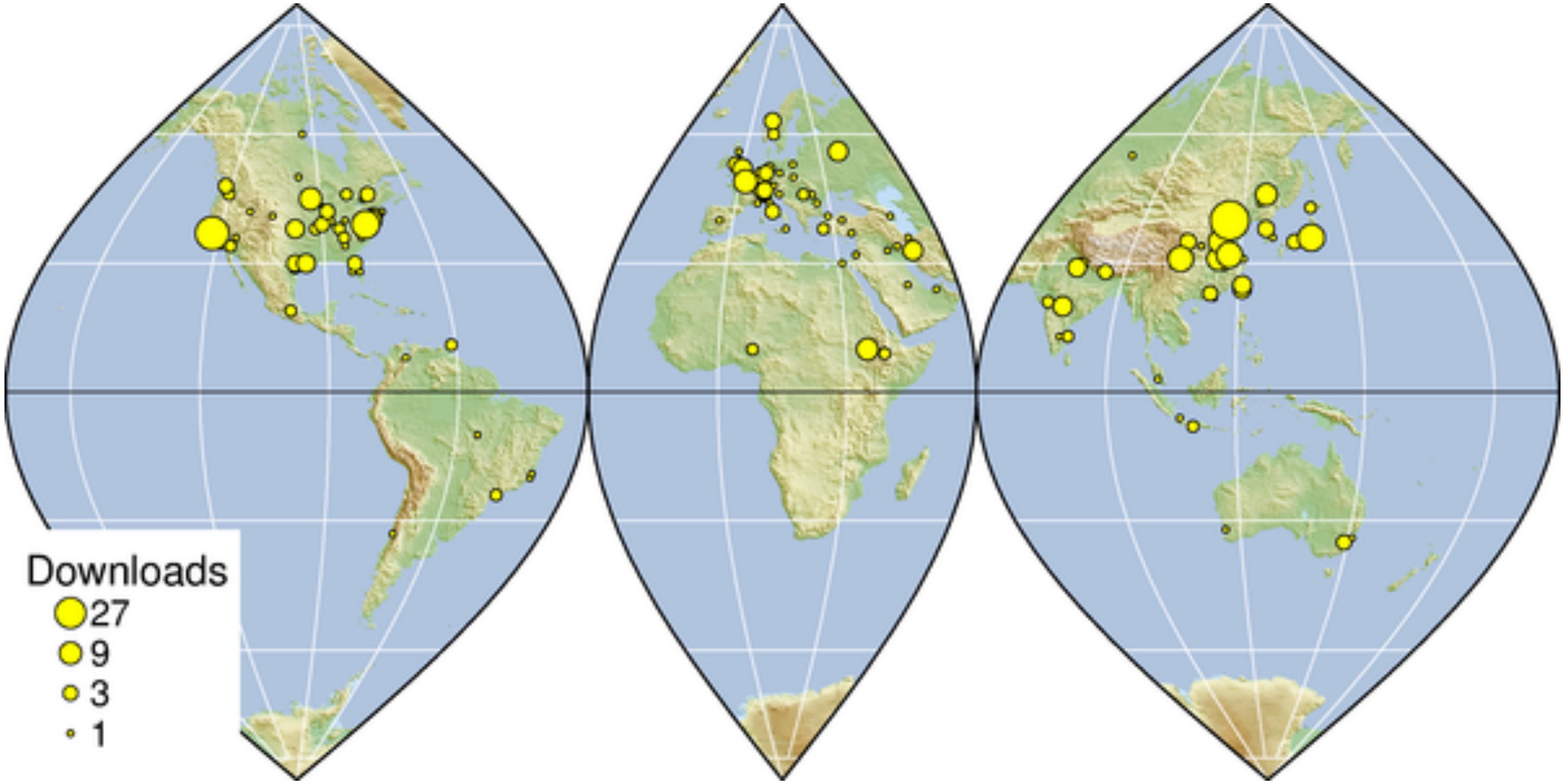
Ebru Bozdağ

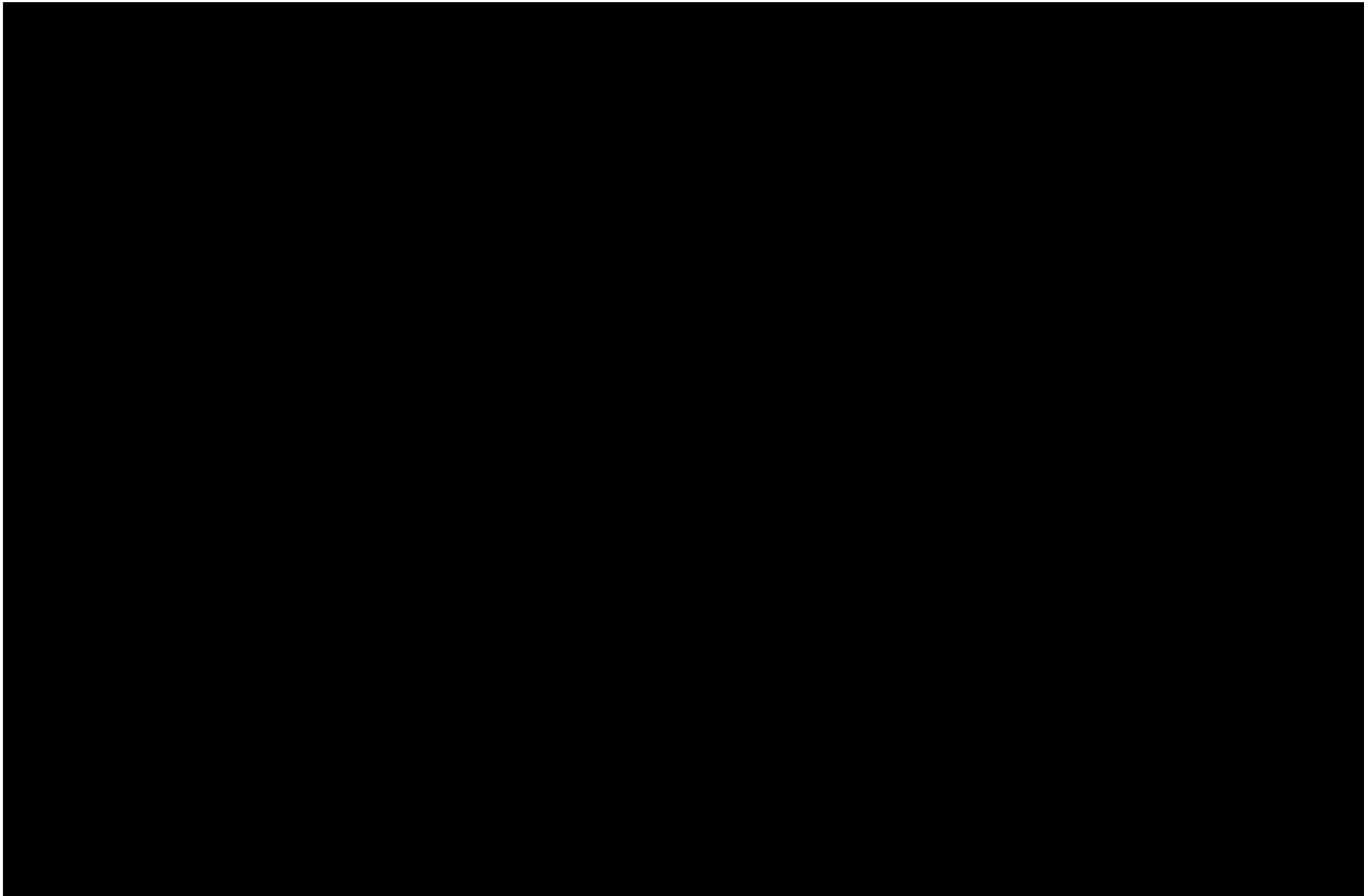
Open Source Forward & Inverse Modeling Software

Spectral-element solvers SPECFEM3D & SPECFEM3D_GLOBE

- 3D crust and mantle models
- Topography & Bathymetry
- Rotation
- Ellipticity
- Gravitation
- Anisotropy
- Attenuation
- Adjoint capabilities
- GPU accelerated

Computational Infrastructure for Geodynamics (CIG)
www.geodynamics.org





Hayden Planetarium (Ben Holtzman, Lamont)

Exascale Seismology Data Science Goals

- Use data with a shortest period of ~1 Hz
- Use all available events with magnitudes greater than ~ 5.5
- Use entire 200 minutes long, 3-component seismograms
- Workflow stabilization & management
- Facilitate uncertainty quantification
- Opportunities for ML/DL/AI in data selection & assimilation
- Data mining, feature extraction & visualization

Seismic Waveform Inversion

Much like a CAT scan or MRI, waveform inversion constructs an image of Earth's interior from waves measured at its surface

Rather than X-rays or magnetic pulses, waveform inversion uses seismic waves, which can be from an earthquake or man-made source

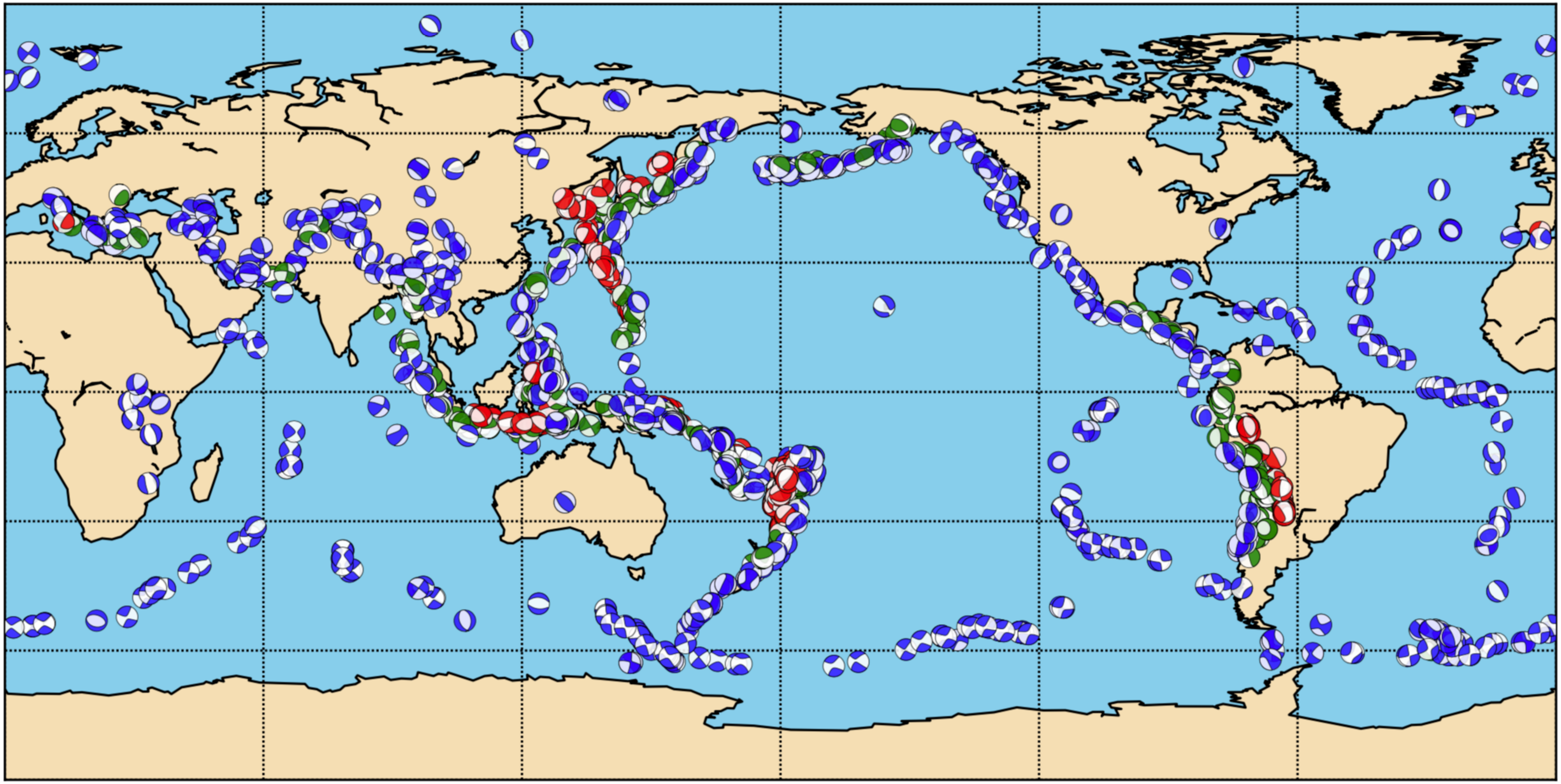
Seismic Waveform Inversion

Widely used in oil and gas exploration, more recently adapted to other applications

Unlike other tomography methods, no simplifying assumptions are made in the solution of governing PDEs

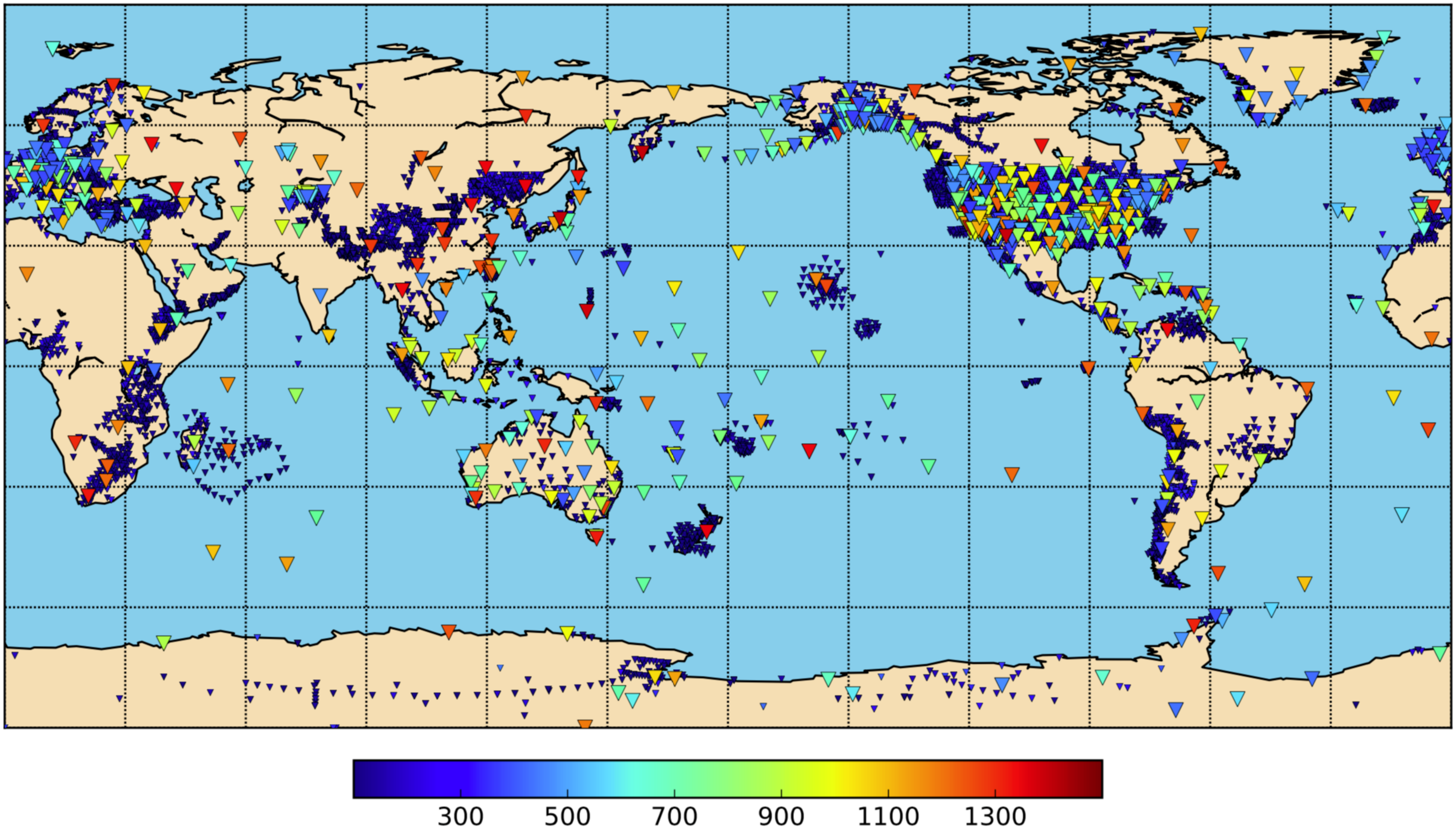
Lack of simplifying assumptions makes waveform inversion very powerful, but also computationally quite challenging

Global Adjoint Tomography: Earthquakes



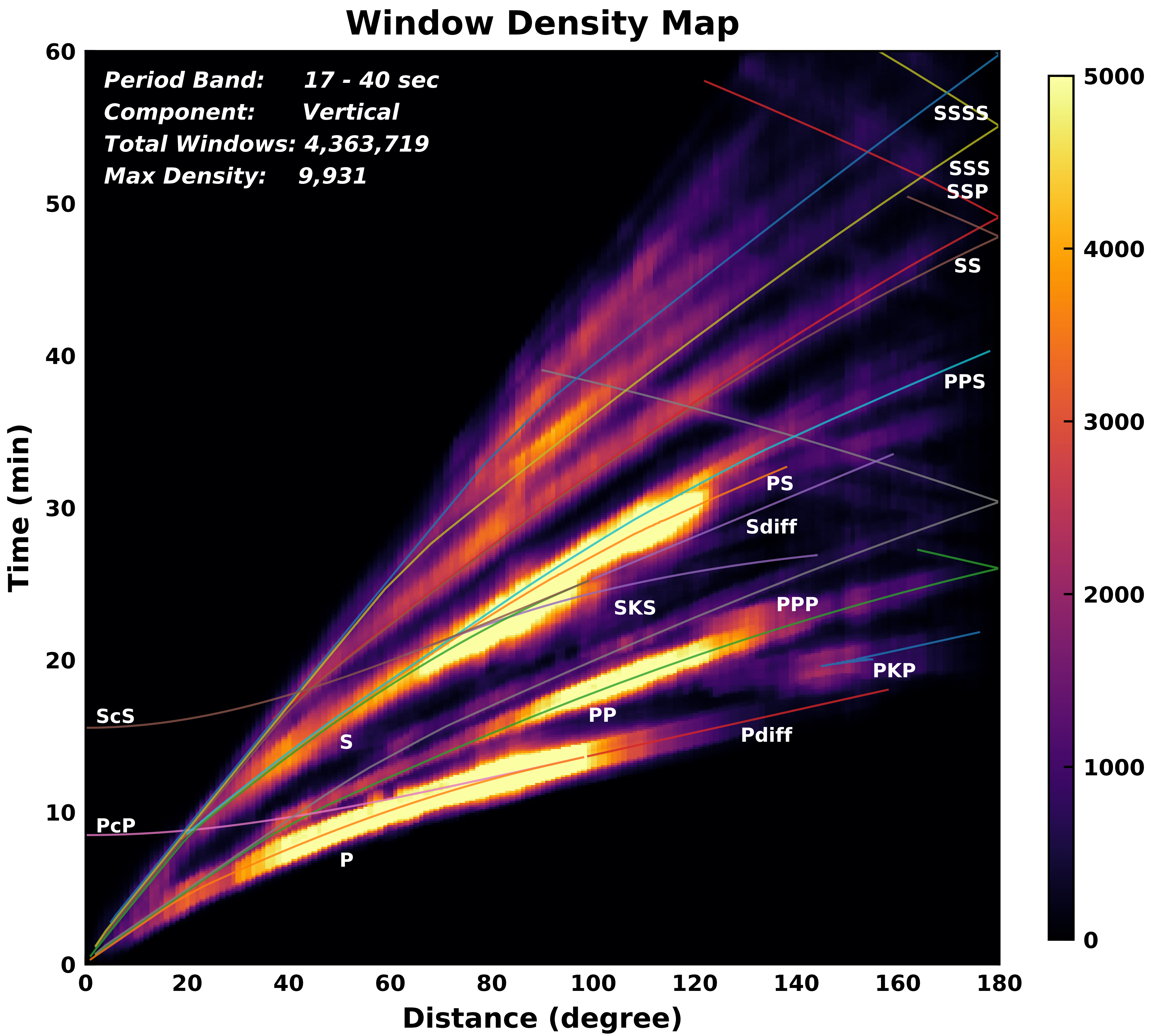
1,480 events

Global Adjoint Tomography: Stations



11,800 permanent and temporary seismographic stations

Window Density Map (Vertical)



South America Subduction Zone

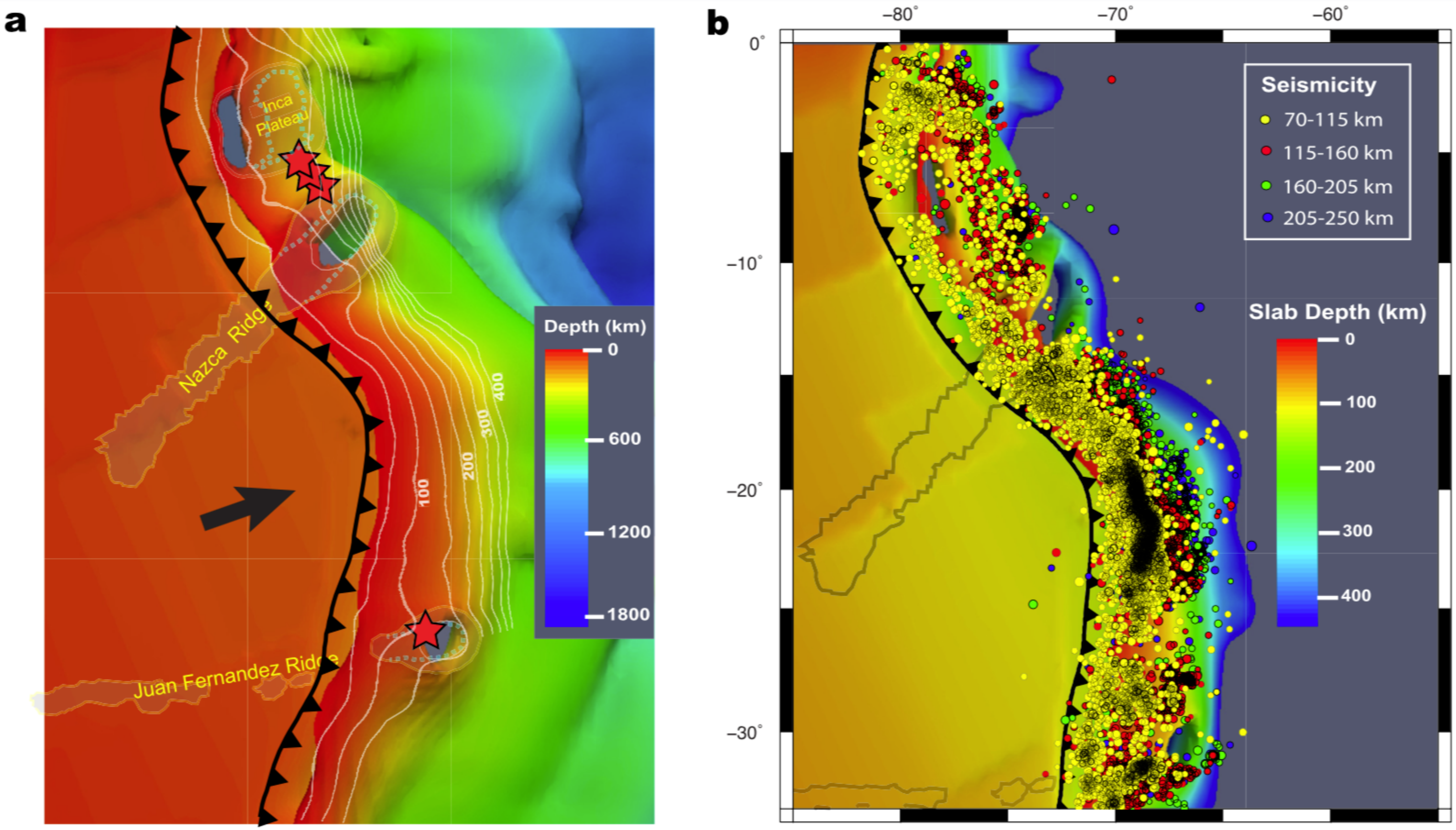
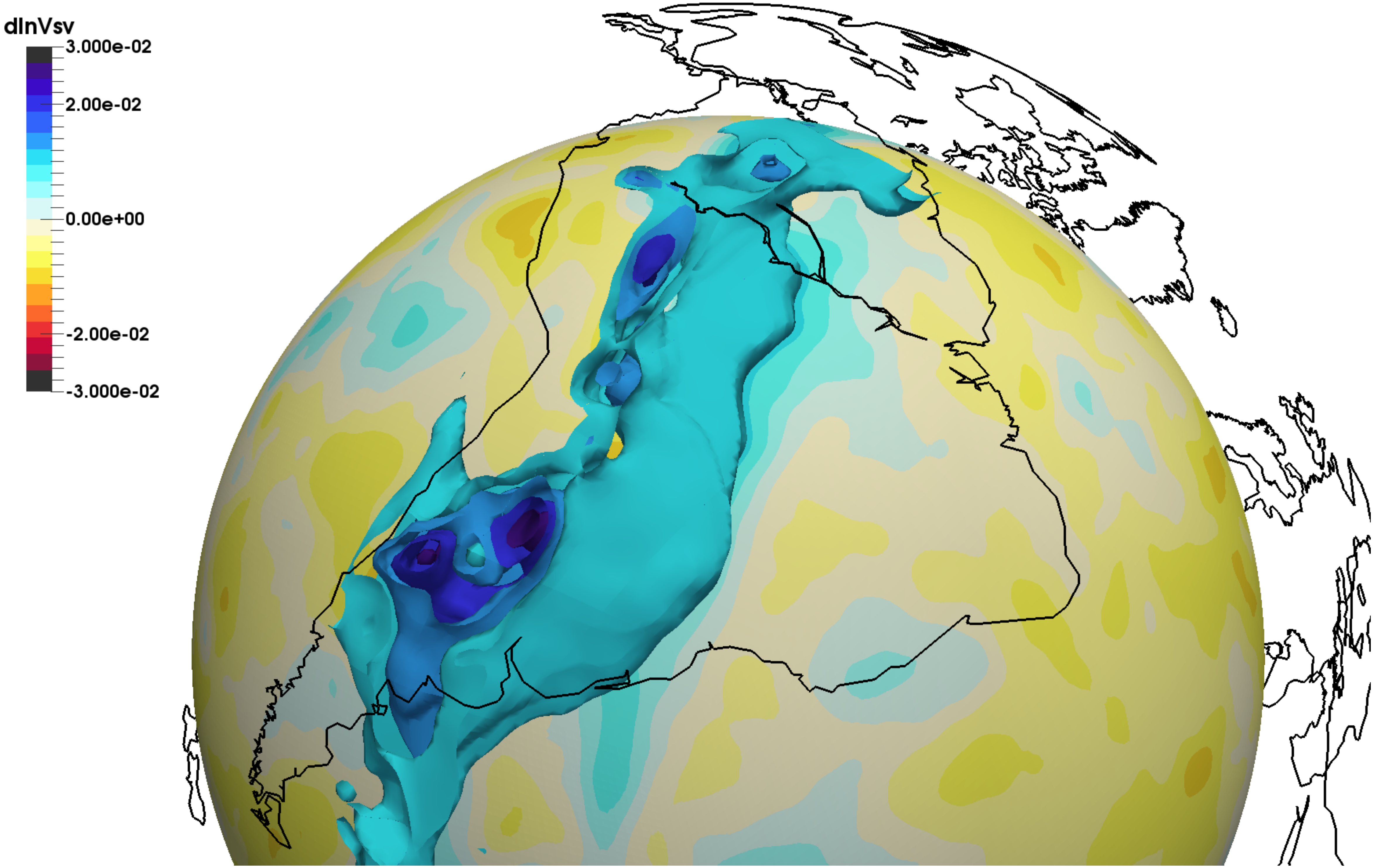


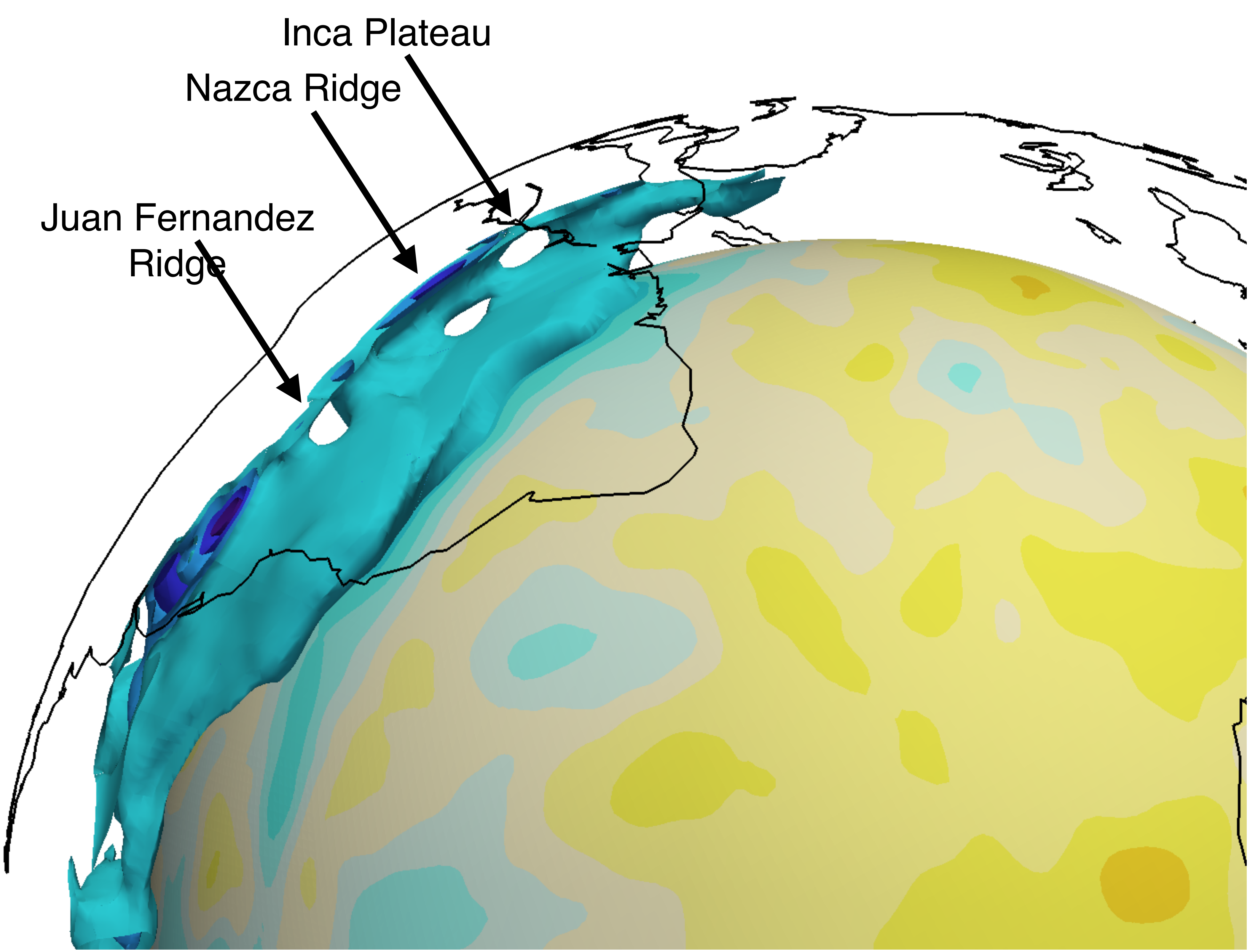
Fig. 5. 3-D geometry of the predicted present-day slab beneath South America and that of the Nazca Plate west of the trench (outlined using an isotherm of 300 °C cooler than the ambient mantle). a) 3D aerial view of the subducting Nazca Plate (temperature isosurface), with colors representing the depth of the slab's upper surface on the right side of the trench and depth of the plate's lower surface on the left side. The slab tears are illustrated with both the isosurface of temperature and the evolution of buoyancy features (translucent gray areas). Thin white lines are the interpolated Benioff zones from Hayes et al. (2012). Dashed lines within the subducting buoyancy features outline their original intact geometry. Red stars indicate the locations of adakitic eruptions. b) Map-view comparison of the slab geometry with the distribution of intermediate-depth seismicity ($M_b > 3.0$ from ISC seismic catalog). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Liu et al, 2016

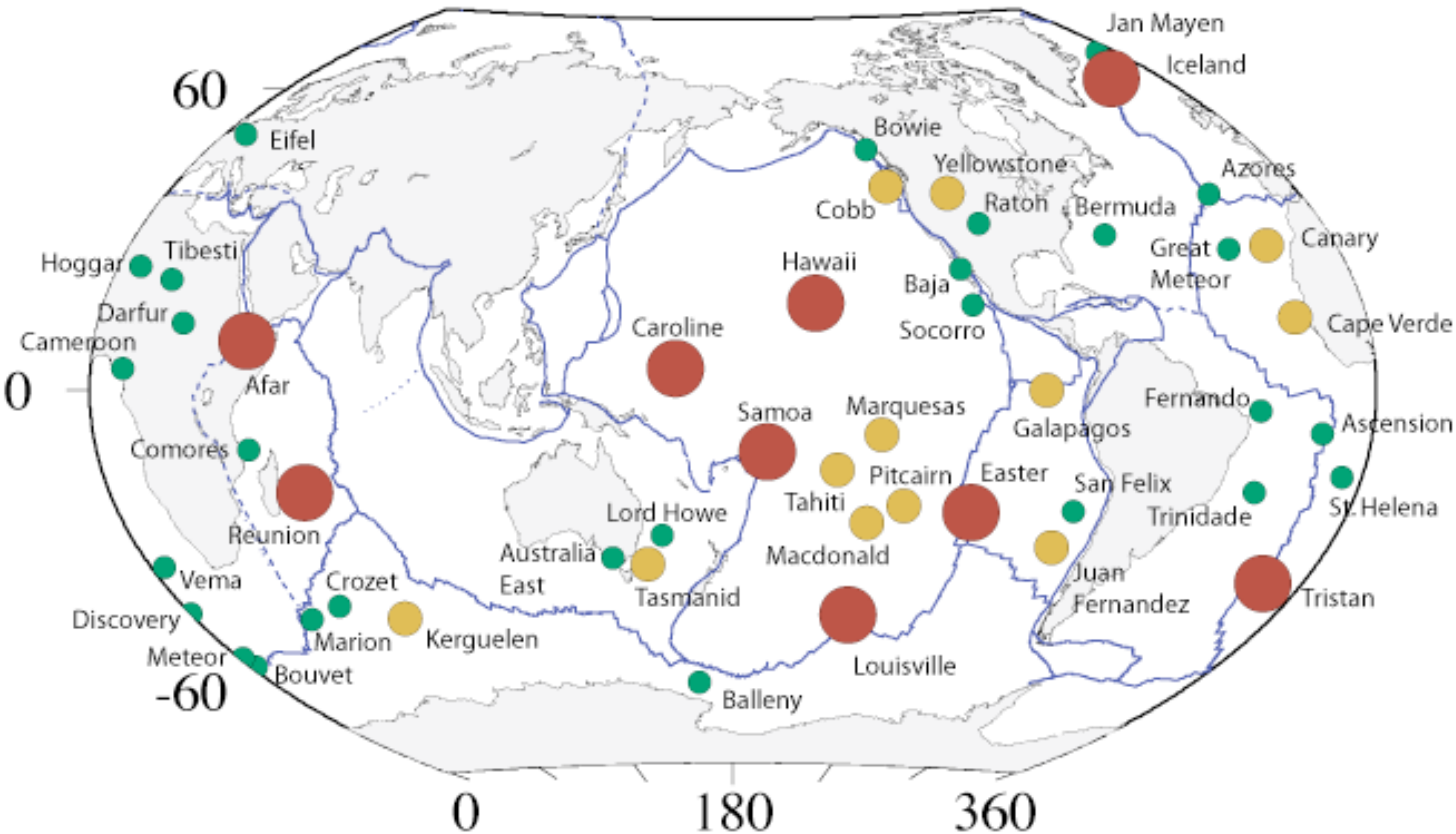
Model GLAD-M25



Model GLAD-M25



Hotspots & Mantle Plumes



Pillars of the Mantle II

Imaging Earth's Interior with Adjoint Tomography

Oak Ridge National Laboratory | Oak Ridge Leadership Computing Facility
Colorado School of Mines — Earth • Energy • Environment
KAUST
Aix*Marseille université
Princeton University

David Pugmire & Ebru Bozdağ

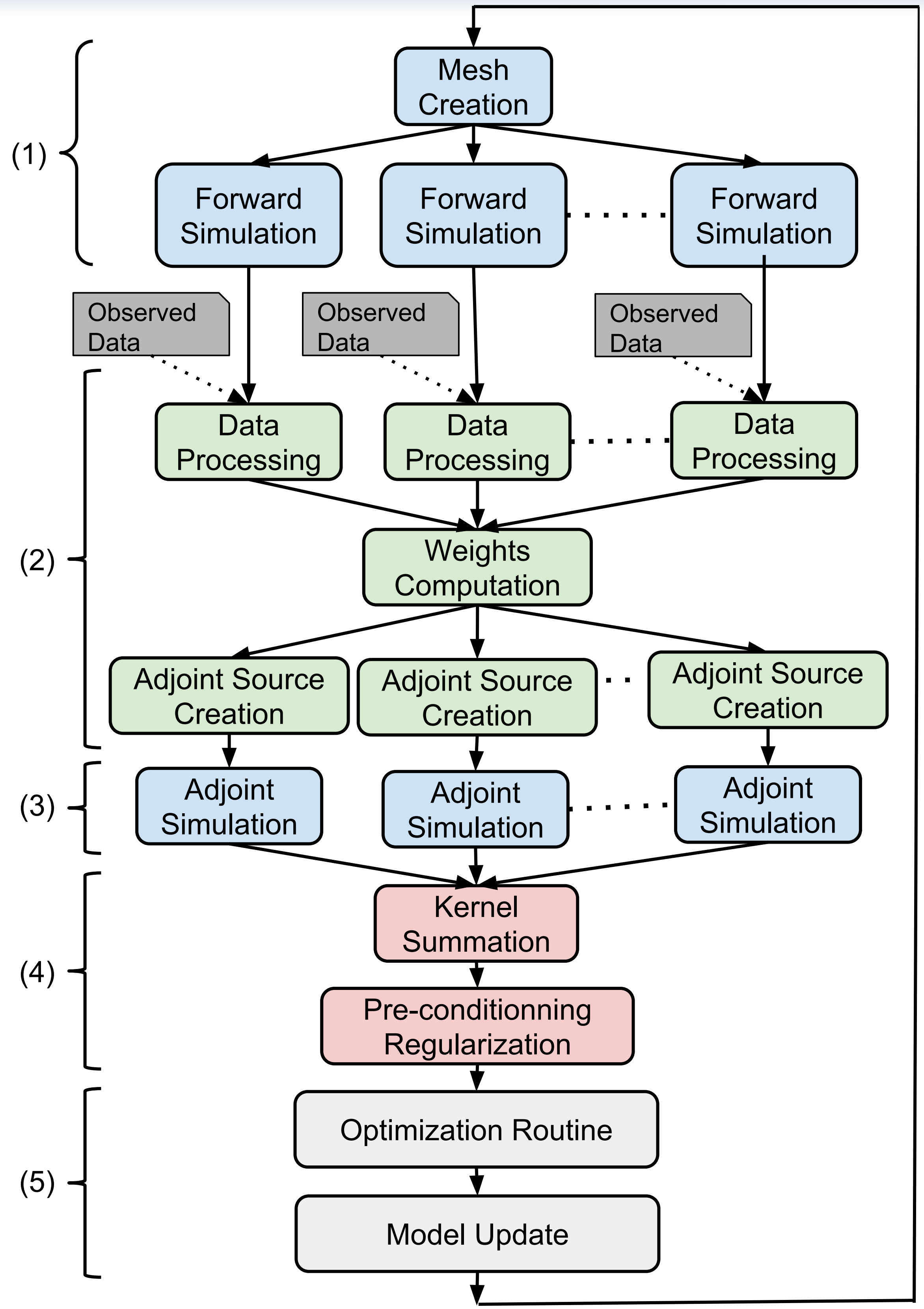
Global Adjoint Tomography Workflow

Challenge#1 Data Volume

8 million 120-min 20 Hz seismograms (6 TB)

10 TB kernels

1 PB wavefield files

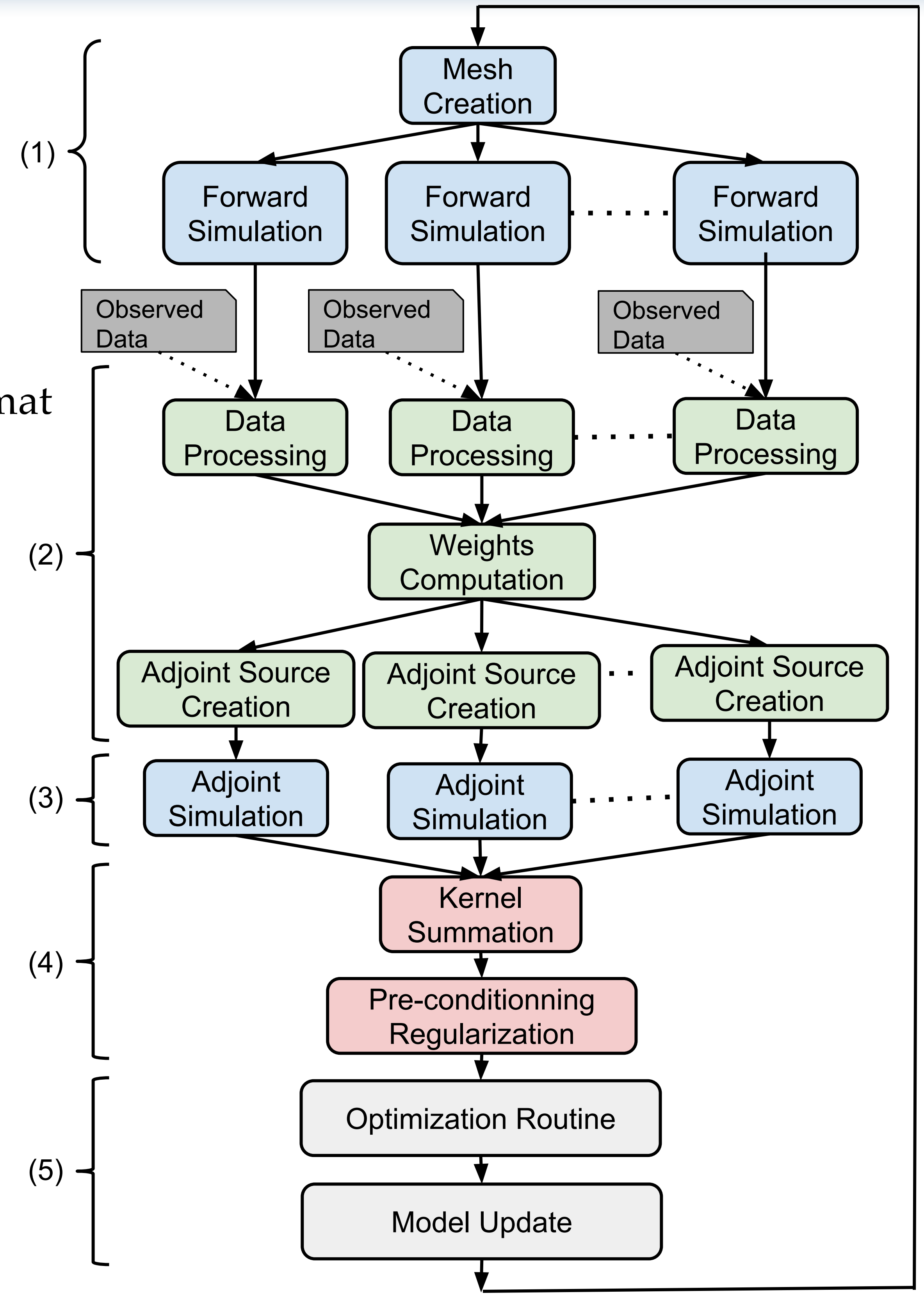


Global Adjoint Tomography Workflow

Challenge#1 Data Volume

Adaptable Seismic Data Format
ASDF (Krischer et al. 2016)

Adaptable I/O System
ADIOS (Liu et al. 2014)



Adaptable Seismic Data Format

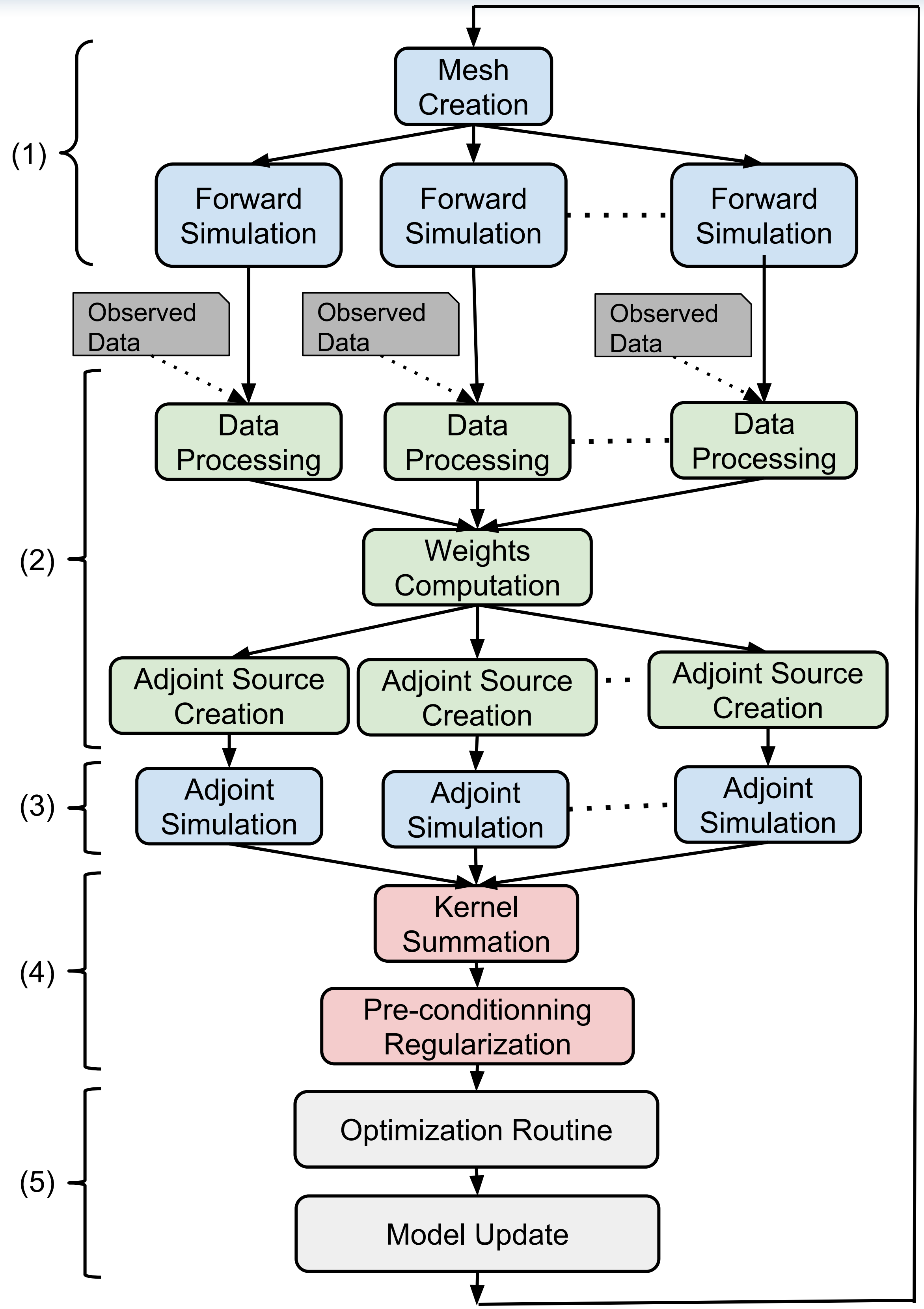
- Collaboration involving Princeton University, Münich University (ObsPy), and Oak Ridge National Laboratory
- Increase I/O performance by combining all the time series for a single shot or earthquake into one file
- Take advantage of parallel processing
- Use modern file format as container (PHDF5)
- Store provenance inside the file for reproducibility
- Use existing standards when possible (e.g., StationXML, QuakeML)
- Open wiki for development (seismic-data.org)

Krischer et al. 2016

Global Adjoint Tomography Workflow

Challenge #2
Expensive Simulations
&
Complex Workflow

0.1 million core hours for data processing

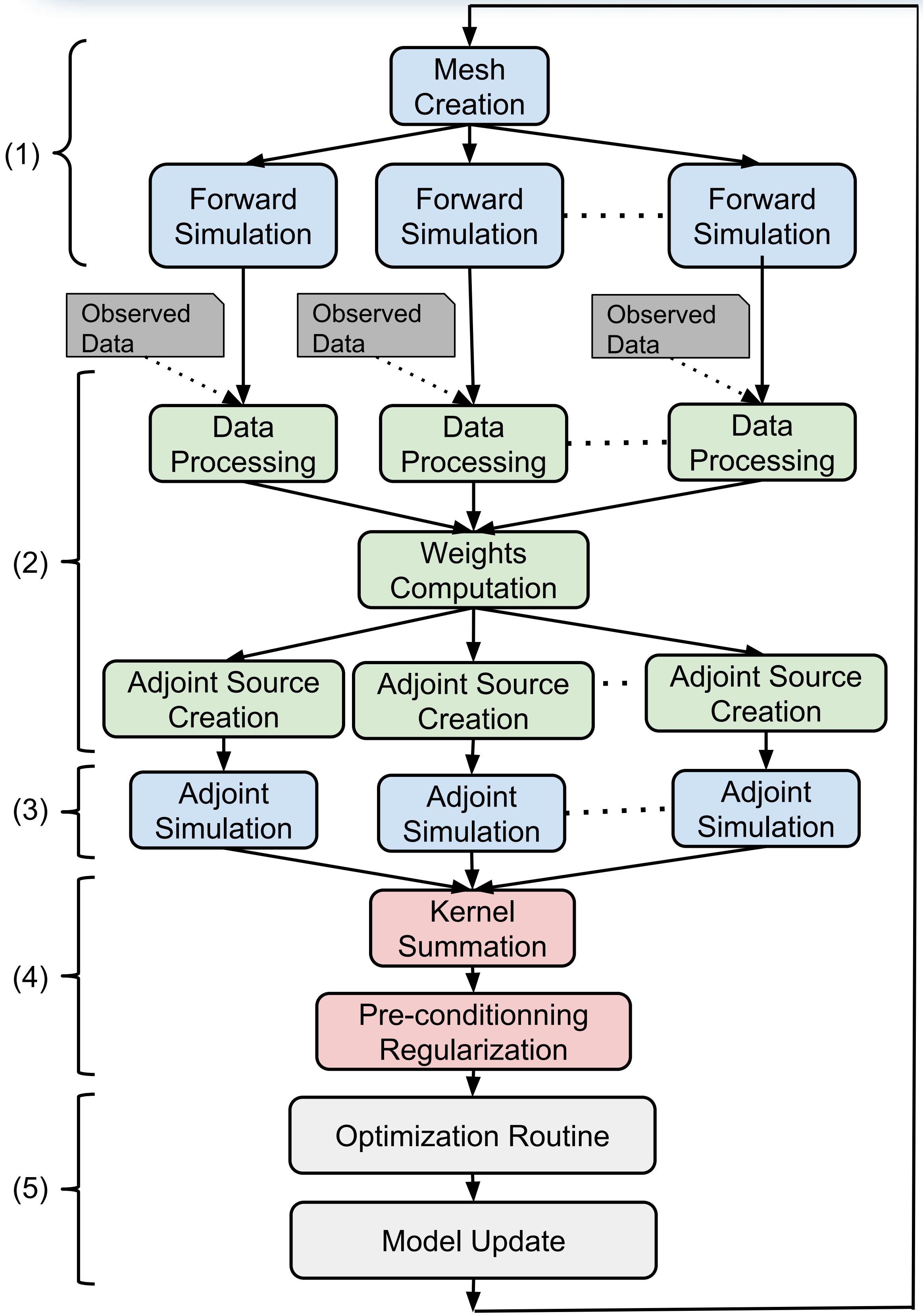


3 million core hours for forward simulation

6 million core hours for adjoint simulation

1 million core hours for line search

Global Adjoint Tomography Workflow Management



Main sources of trouble:

- Hardware failures
- Human errors

We are implementing the RADICAL EnTK workflow management toolkit:

- *Automation*: save time & human effort for repeated tasks
- *Efficiency*: acceleration, taking full advantage of HPC systems
- *Fault tolerance*: automated job failure detection & recovery

Conclusions

Exascale readiness requires the following investments:

- Continual open source software development, e.g., GPUs & Intel Phi
- All software needs to be under source control (e.g., GitHub) and needs to be automatically and continually tested (e.g., BuildBot, Travis, Jenkins)
- Modern seismological data format with full provenance (ASDF)
- Workflow management tools (e.g., RADICAL EnTK)
- Uncertainty quantification