

# Global adjoint tomography – Model GLAD-M25

Wenjie Lei<sup>1</sup>, Youyi Ruan<sup>1,2</sup>, Ebru Bozdaḡ<sup>3</sup>, Daniel Peter<sup>4</sup>, Matthieu Lefebvre<sup>1</sup>, Dimitri Komatitsch<sup>5</sup>, Jeroen Tromp<sup>1,6</sup>, Judith Hill<sup>7</sup>, Norbert Podhorszki<sup>7</sup>, and David Pugmire<sup>7</sup>

<sup>1</sup> Department of Geosciences, Princeton University, Princeton, NJ 08544, USA

<sup>2</sup> Nanjing University, Nanjing, China

<sup>3</sup> Department of Geophysics, Colorado School of Mines, Golden, CO 80401, USA

<sup>4</sup> Extreme Computing Research Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia

<sup>5</sup> LMA, CNRS UPR 7051, Aix-Marseille University, Centrale Marseille, 13453 Marseille Cedex 13, France

<sup>6</sup> Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA

<sup>7</sup> Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

## 1 INTRODUCTION

Wenjie: I want to leave this part to the end, when I finished reading all the papers.

With the development of high performance computing and numerical methods, it is now possible to use adjoint method to image the interior of the globe using thousands of earthquakes.

There are few agreements about the earth model and yet many unknown to be resolved. There are quite a lot of global models exist however there is no such consensus.

The GLAD-M25 is the first earth model that incorporate the P and S wave speed using full waveform inversion. Also, people are using specific set of phases in their inversion. For P models, people are using P phase and other body wave phases and for S models, people are using mostly S wave phases and surface waves to constrain the structure. Those two kinds of

models usually reveals very different parts of the earth. For example, for P wave, people used it to image the subduction slabs and it is not very good at imaging lower mantle structures, such as plumes and lower mantle convection. For S waves, it is the verse verso.

There are many challenges coming along the way. First, the data volume is huge. There are in total of 1480 earthquakes used in the inversion stage, added into the database step by step. Second, workflow management is crucial dealing with such a large project. One factor is coming from the complex workflow. There are 4 major building blocks, forward simulation, seismic data processing, adjoint simulation and post-processing. Each block contains several small blocks. The other factor is coming from the number of earthquakes and high demands of computation. In each iteration, we are doing 1480 forward and adjoint simulation, generating a few Petabytes of wavefield files and consuming 16 million of CPU hours on the supercomputer at Oak Ridge National Lab. There are yet more in data processing stages. How to deal with hardware failures is very important and prevent contamination into our inversion results.

Here, we present our GLAD-M25 earth model, with significantly improvements on the plumes and subductions.

test1 (Zhu et al. (2012), Zhu et al. (2012)) test2 (Zhu et al. 2015; Ekström et al. 2012)  
 test3 Ekström et al. (2012)

## 2 STARTING MODEL GLAD-M15

As a continuation of our previous work (Bozdağ et al. 2016), we used “first generation” model GLAD-M15 as our starting model. GLAD-M15 is a 3D transversely isotropic earth model, which combined 3D mantle model S362ANI (Kustowski et al. 2008) with 3D crustal model CRUST2.0 (Bassin et al. 2000) as its starting model. In this study, we continue to use the same transversely isotropic model parametrization. Instead of relying on ubiquitous “crustal corrections”, the mesh implementation of the crust in the spectral-element solver SPECFEM3D\_GLOBE (Komatitsch & Tromp 2002a,b; Peter et al. 2011) enables us accurately accommodate topography and bathymetry as well as variations in the Moho. GLAD-M15 was constructed using a global data base of 256 earthquakes. The first 12 iterations of the GLAD-M15 inversion used three-component seismograms with a shortest period of 27 s, and the final 3 iterations reduced this further to 17 s. In this study we also used three-component seismograms with a shortest period of 17 s.

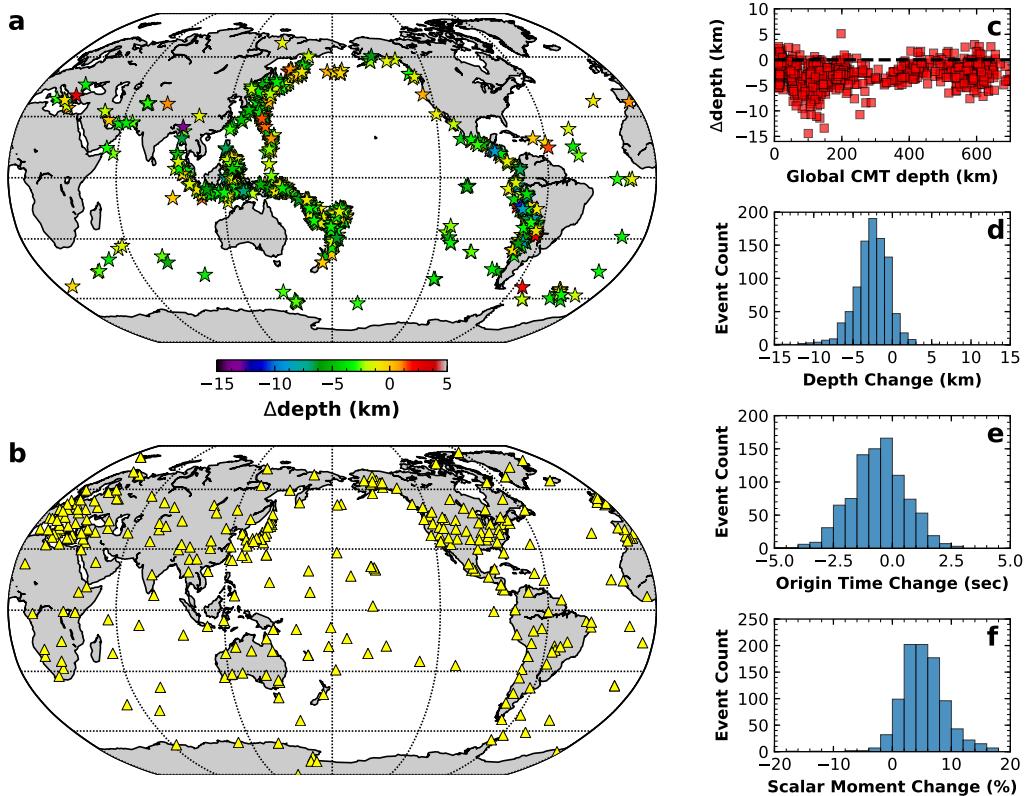


Figure 1: Source corrections for 1,040 out of 1,480 events used in the structural inversion. (a) Map view of depth changes. (b) Distribution of seismographic stations used for source inversions. (c) Depth change relative to the original global CMT Catalog depth. (d)–(f) Histograms of depth, origin time, and scalar moment changes relative to the CMT Catalog.

### 3 EARTHQUAKES

In addition to the 256 earthquakes used for the construction of first generation model GLAD-M15, we initially selected 784 carefully chosen additional earthquakes from the global Centroid-Moment Tensor (CMT) catalog (e.g., Ekström et al. 2012), and added them to our database, resulting in a total number of 1,040 earthquakes. To ensure a good signal-to-noise ratio on a global scale, the smallest moment magnitude in the database is set to 5.5, and to avoid complications associated with source size, the largest moment magnitude is set to 7.2.

Before using the 1,040 earthquakes in our structural inversion, we performed CMT inversions in model GLAD-M15, the results of which are summarized in Fig. 1. To ensure even global coverage, we used seismographic stations from the Global Seismic Network (II, IU, IC, US, CU, and GT), GEOFON (GE), GEOSCOPE (G), and several regional networks, such as

MedNet (MN), the Brazilian Lithospheric Seismic Project (BL), the Chilean National Seismic Network (C), and the Japan Meteorological Agency Seismic Network (JP) (Fig. 1b). The number of stations used for CMT inversion usually ranges from 150 to 500.

To determine the source parameters in the starting model, we used the CMT inversion algorithm of Liu et al. (2004). This algorithm combines a normalized waveform difference misfit with an envelope difference misfit. Specifically, the algorithm minimizes the misfit function

$$\Phi = \sum_{c=1}^C \omega_c \sum_{r=1}^{R_c} \omega_{cr} \sum_{w=1}^{W_{cr}} \omega_{crw} \left\{ \lambda \frac{\int [d_w(t) - s_w(t - \Delta t)]^2 dt}{\int [d_w(t)]^2 dt} + (1 - \lambda) \frac{\int [e(d_w(t)) - e(s_w(t - \Delta t))]^2 dt}{\int [e(d_w(t))]^2 dt} \right\}, \quad (1)$$

where  $d_w(t)$  denotes data in time window  $w$ , and  $s_w(t - \Delta t)$  the corresponding synthetic for model GLAD-M15. We allow the synthetics to shift relative to the data by an amount  $\Delta t$  — effectively a station correction — determined by cross correlation. The envelope function is denoted by  $e(\cdot)$ , and the parameter  $\lambda$  determines the balance between fitting waveforms versus fitting envelopes. The number of measurement categories is denoted by  $C$ . Following Ekström et al. (2012), seismograms were filtered between 50 s and 100 s to select three-component body-wave windows, and between 60 s and 100 s to select three-component surface-wave windows, resulting in six measurement categories,  $c = 1, \dots, C$ . To balance their contributions, each category is weighted by the reciprocal of the number of measurements in that category,  $\omega_c$ . The number of windows for a given receiver  $r$  in a given measurement category  $c$  is  $W_{cr}$ . Each such window is weighted equally, i.e.,  $\omega_{crw} = 1$ . The number of receivers that records data in category  $c$  is denoted by  $R_c$ . As shown in Fig. 1b, the receivers are unevenly distributed across the globe, with several dense arrays in the Northern Hemisphere and poor coverage in the Southern Hemisphere. To balance the coverage, we assign receiver weights  $\omega_{cr}$  based on the expression

$$\omega_{cr}^{-1} = N_c \sum_{r'=1}^{R_c} \exp \left[ - \left( \frac{\Delta_{rr'}}{\Delta_c} \right)^2 \right], \quad (2)$$

where  $N_c$  is a normalization factor for category  $c$ , and where  $\Delta_{rr'}$  denotes the angular distance between receivers  $r$  and  $r'$ . The reference angular distance  $\Delta_c$  for each category  $c$  needs to be chosen such that the condition number of the diagonal weighting matrix defined by Eqn. (2) is not too large. The calculation of the weights may be abstracted as: given a spatial distribution of points on the unit sphere, determine a weighting associated with each point.

The algorithm inverts for the six elements of the moment tensor and the centroid location (latitude, longitude, and depth). The calculation of each associated Fréchet derivative

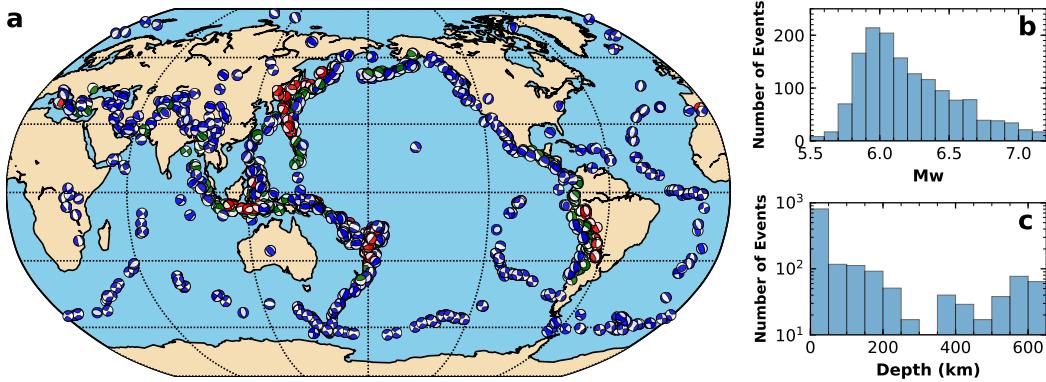


Figure 2: 1,480 earthquakes used in this study. (a) Distribution of earthquakes. The color of each beach ball reflects its depth range, where blue designate events shallower than 50 km, green events between 50 km and 300 km, and red events deeper than 300 km. (b) & (c) Histograms of earthquake moment magnitudes and depths.

requires one full 3D forward simulation, so, considering we have more than one thousand events, these source inversions are computationally very expensive.

Because we allow the synthetics to shift by an amount  $\Delta t$  relative to the data, the CMT inversion has unreliable sensitivity to the centroid time. To alleviate this problem, following Zhu et al. (2012), we perform a subsequent grid search for the centroid time and the scalar moment. The grid-search calibration involves simple shift and multiply operations on seismograms and, unlike the CMT inversion, requires minimal simulation time.

Most earthquakes show a shallower depth after inversion (Fig. 1c), consistent with our previous experiences (e.g., Zhu et al. 2015; Chen et al. 2015; Bozdağ et al. 2016) and with experiments by Hjörleifsdóttir & Ekström (2010). We observed an average depth change of  $-2.62 \pm 2.49$  km relative to the global CMT solutions, an average scalar moment change of  $5.31 \pm 3.91\%$ , and an average centroid time shift of  $-0.60 \pm 1.17$  s. These are relatively minor corrections, especially in view of the significant expense of the source inversions. For this reason, when we added another 440 earthquakes during iteration 22, thereby bringing the total to 1,480 events, we only performed a grid search to calibrate the centroid times and scalar moments. Fig. 2 summarizes the characteristics of all 1,480 earthquakes used in this study.

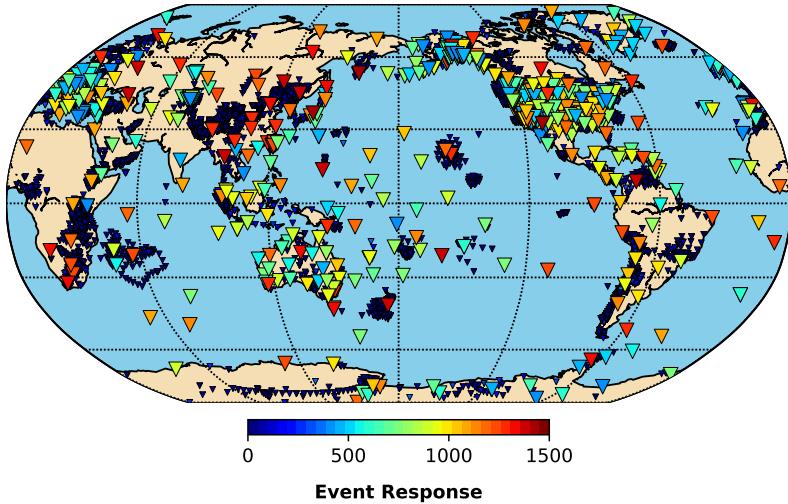


Figure 3: Distribution of 11,800 seismographic stations used in this study. Colors denote the number of events for which a given station contributes waveforms to the structural inversion. Stations with a number of event responses  $< 400$  are plotted in smaller size; these are usually temporary arrays deployed over a short period of time, frequently ocean bottom seismometers. The maximum number of event responses comes from ANMO, in Albuquerque, New Mexico, which contributed to 1,442 out of 1,480 earthquakes in the data set.

## 4 SEISMIC DATA

The seismographic stations used in this study were carefully selected to ensure global coverage and high data quality (Fig. 3). In addition to the seismic stations used for the source inversions, we included all available data for our 1,480 event earthquake database from many data centers, including IRIS, ORFEUS, INGV, IPGP, ETH, and GEONET. Regional and temporary networks, such as US Array (TA), Africa Array (AF), the Canadian National Seismograph Network (CN), Geoscience Australia (AU), the Antarctic Seismographic Argentinean Italian Network (AI), and the New Zealand National Seismograph Network (NZ), constitute a significant part of our database and greatly improve coverage in certain regions.

### 4.1 Adaptable Seismic Data Format

Conventional seismic data formats, such as SAC, involve one file per time series plus related files with instrument response and event information. Since every earthquake in the database is typically recorded by thousands of instruments, I/O during the preprocessing stage of the adjoint tomography workflow quickly cripples the file system. The Adaptable Seismic Data

Format (ASDF) (Krischer et al. 2016, <https://seismic-data.org/>) was developed with complete reproducibility and fast parallel processing in mind.

In this context, there are four key issues ASDF resolves.

- Robustness and stability: the data container is developed and maintained to ensure accuracy of scientific results.
- Data organization: the container is self-describing. Data, including waveform, source, and station information, are organized into well-defined structures.
- Reproducibility: the container enables scientists to keep track of what has been done to data so that results can be reproduced.
- Efficiency: the format provides easy mechanisms for parallel processing.

ASDF serves as a self-contained-and-explained data container while taking full advantages of parallel computing. In our workflow, one ASDF file contains all the information needed for data processing, including the seismic traces, event information files, and station response files. The related APIs are carefully designed for easy data extraction and parallel implementation.

## 5 MISFIT FUNCTION

The misfit function to be minimized during the iterative inversion process must be carefully constructed. It will typically involve cross-correlation traveltimes for body waves and frequency-dependent multitaper phase measurements for surface waves. Measurements are made in several passbands of three-component seismograms rotated into vertical, radial, and transverse components, resulting in a number of measurement categories. In this study, we consider four passbands, namely, a 17–40 s passband targeting body waves, two 40–100 s passbands separately targeting body and surface waves, and a 90–250 s passband targeting longer-period surface waves. Each passband involves measurements on all three components, which results in a total of twelve measurement categories. Although the inversion is designed to minimize the overall misfit, we will be tracking the misfit reduction in each of the categories to ensure that the model is improving the fit to the data roughly equally across the board. One of the biggest challenges in the construction of the misfit function is the highly uneven distribution of earthquakes and seismographic stations, which must be counterbalanced by geographically weighting the data. This issue is discussed in detail by ?); in this section we present a brief synopsis.

With these considerations in mind, the overall misfit,  $\Phi$ , is defined as follows:

$$\Phi = \sum_s^S \omega_s \sum_c^C \omega_c \sum_r^{R_{sc}} \omega_{scr} \sum_w^{W_{scr}} \omega_{scrw} \chi_{scrw} , \quad (3)$$

where  $S$  denotes the number of sources,  $C$  the number of categories,  $R_{sc}$  the number of receivers recording source  $s$  in category  $c$ , and  $W_{scr}$  the number of measurement windows for source  $s$ , category  $c$  and receiver  $r$ . The misfit for a specific source  $s$ , category  $c$ , receiver  $r$ , and window  $w$  is

$$\chi_{scrw} = \frac{1}{\Delta\omega} \int_{\omega_1}^{\omega_2} \left( \frac{\Delta\tau_{scrw}}{\sigma_{scrw}} \right)^2 d\omega , \quad (4)$$

where  $\Delta\tau_{scrw}$  denotes a multi-taper frequency-dependent phase measurement over the frequency interval  $\Delta\omega = \omega_2 - \omega_1$ , with associated standard deviation  $\sigma_{scrw}$ . For body waves, the window misfit is simply the cross-correlation traveltime anomaly,  $\Delta T_{scrw}$ , weighted by its standard deviation, i.e.,  $\chi_{scrw} = (\Delta T_{scrw}/\sigma_{scrw})^2$ . As in the source inversions, the window weights are equal,  $\omega_{scrw} = 1$ , and the category weight,  $\omega_c$ , is the reciprocal of the number of measurements in that category. Earthquakes are mainly confined to plate boundaries, and most seismic stations are confined to the continents, which leads to a very uneven distribution of sources and stations, as illustrated in Figs. 2 and 3. Therefore, weighting is crucial in global tomography to balance this uneven sampling of Earth's interior. Our source and receiver weighting strategy, encapsulated by the weights  $\omega_s$  and  $\omega_{scr}$ , is developed to compensate for uneven spatial sampling. Following the same strategy as for the receiver weights in the source inversions, we define the source and receiver weights as

$$\omega_s^{-1} = N \sum_{s'=1}^S \exp \left[ - \left( \frac{\Delta_{ss'}}{\Delta} \right)^2 \right] , \quad (5)$$

and

$$\omega_{scr}^{-1} = N_{sc} \sum_{r'=1}^{R_{sc}} \exp \left[ - \left( \frac{\Delta_{rr'}}{\Delta_{sc}} \right)^2 \right] , \quad (6)$$

respectively. Here  $N$  and  $N_{sc}$  are a normalization factors, and  $\Delta_{ss'}$  and  $\Delta_{rr'}$  denote angular distances between source and receiver pairs  $\{s, s'\}$  and  $\{r, r'\}$ . The reference angular distances  $\Delta$  and  $\Delta_{sc}$  need to be chosen based on the condition numbers of the diagonal weighting matrices defined by Eqns. (5) and (6).

## 6 MODEL PARAMETRIZATION

We use the same transversely isotropic model parametrization as starting model GLAD-M15. Such a model is described by the five Love parameters  $A$ ,  $C$ ,  $L$ ,  $N$ , and  $F$  (Love 1927), or,

alternatively, using the mass density  $\rho$ , in terms of the wavespeeds  $\alpha_v = \sqrt{C/\rho}$ ,  $\alpha_h = \sqrt{A/\rho}$ ,  $\beta_v = \sqrt{L/\rho}$ ,  $\beta_h = \sqrt{N/\rho}$  and the dimensionless parameter  $\eta = F/(A - 2L)$  (Dziewonski & Anderson 1981; Dahlen & Tromp 1998). Assuming the radial anisotropy is due to shear anisotropy, these five parameters may be further reduced to four by introducing the bulk sound speed,  $c = \sqrt{\kappa/\rho}$ . Therefore, the final four parameters are  $c$ ,  $\beta_v$ ,  $\beta_h$ , and  $\eta$ .

Since density is difficult to constrain with seismic data, density perturbations are scaled to isotropic (Voigt averaged) shear wavespeed perturbations based on the relationship  $\delta \ln \rho = 0.33 \delta \ln \beta$  (Montagner & Anderson 1989).

Based on this parametrization, the variation in the misfit function (3) may be expressed as (Zhu et al. 2015; Bozdağ et al. 2016)

$$\delta\Phi = \int_V (\delta \ln c K_c + \delta \ln \beta_v K_{\beta_v} + \delta \ln \beta_h K_{\beta_h} + \delta \ln \eta K_\eta) dV , \quad (7)$$

where  $K_c$ ,  $K_{\beta_v}$ ,  $K_{\beta_h}$ , and  $K_\eta$  denote the four Fréchet derivatives, which are calculated based on an adjoint-state method (e.g., Plessix 2006; Tromp et al. 2005).

## 7 ADJOINT TOMOGRAPHY WORKFLOW

The adjoint tomography workflow is shown in Fig. 4. It starts with the selection of earthquakes, as discussed in Section 3. Given the earthquake selection, observed seismographic data and related response information are acquired, as discussed in Section 4. For a given earthquake, the corresponding seismograms and related response information are combined in a single ASDF file, described in Section 4.1. For a given earthquake dataset, the conversion to ASDF needs to be performed once and for all. In the following sections we highlight several important aspects of the workflow.

### 7.1 Forward simulations

The observed seismograms need to be compared to corresponding synthetic seismograms for the purpose of making measurements. The forward simulations are based on the global spectral-element solver SPECFEM3D\_GLOBE (Komatitsch & Tromp 2002a,a), and take a CMT solution, a relevant list of stations, and the current earth model as input. The solver is GPU accelerated, and the simulations are performed on the Cray ‘Titan’ at the Oak Ridge Leadership Computing Facility. To simulate 120 min seismograms with a shortest period of 17 s on 384 NVIDIA Tesla K20X GPUs takes approximately 10 min of wall clock time. The calculation of Fréchet derivatives at a later stage of the workflow requires access to snapshots

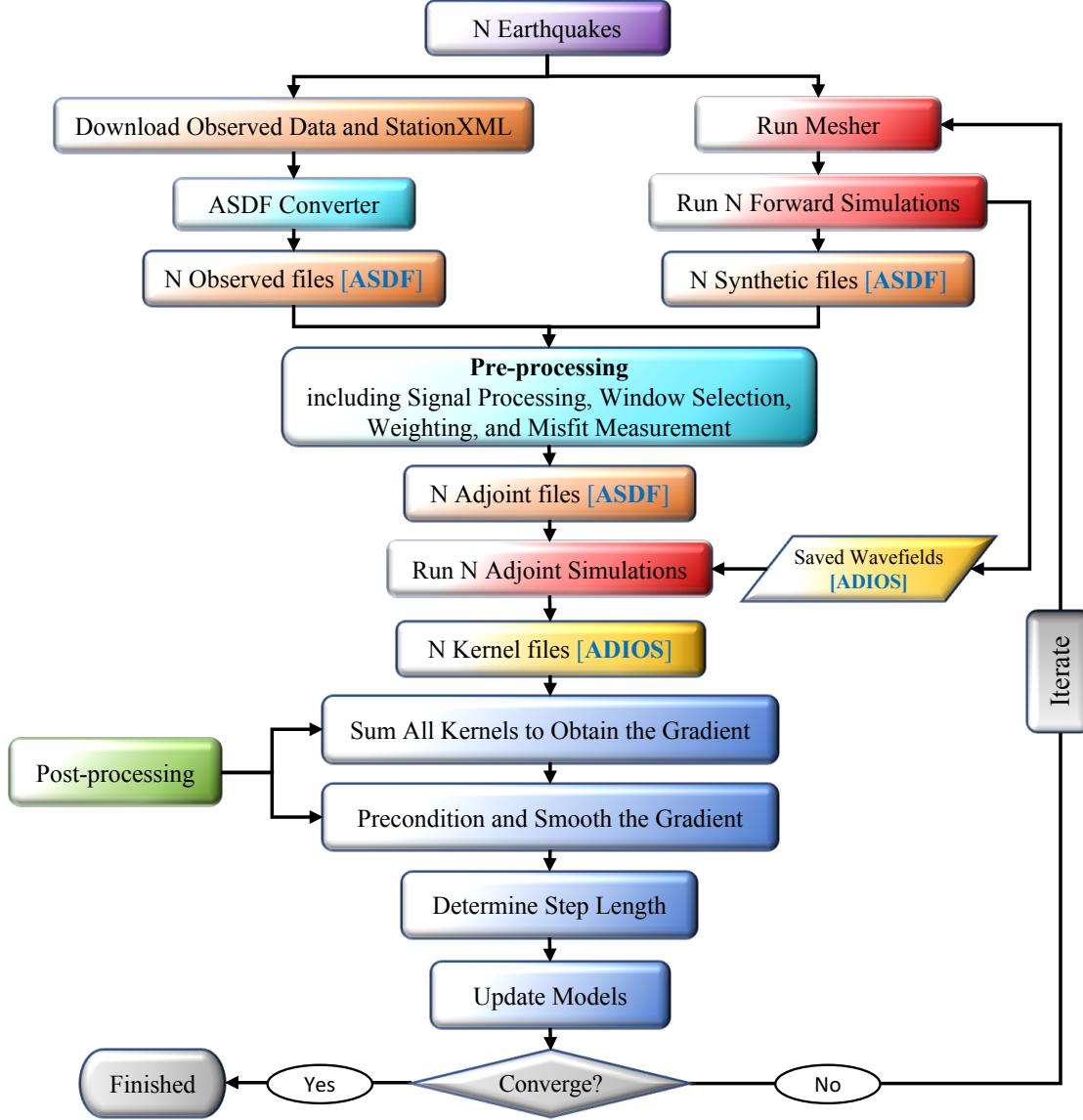


Figure 4: Adjoint tomography workflow.

of the forward wavefield as part of the parsimonious storage algorithm developed by Komatitsch et al. (2016), with rapid I/O facilitated by ADIOS (Liu et al. 2014). For the current simulation setup, this requires about 1 TB of storage per earthquake, so for 1,480 earthquakes this amounts to 1.5 PB of storage.

## 7.2 Seismic data processing

The seismic data preprocessing workflow takes raw observed and synthetic data as input, and generates adjoint sources for data assimilation as output. From a computational perspective, the preprocessing workflow consumes only 1% of the overall computational requirements. But

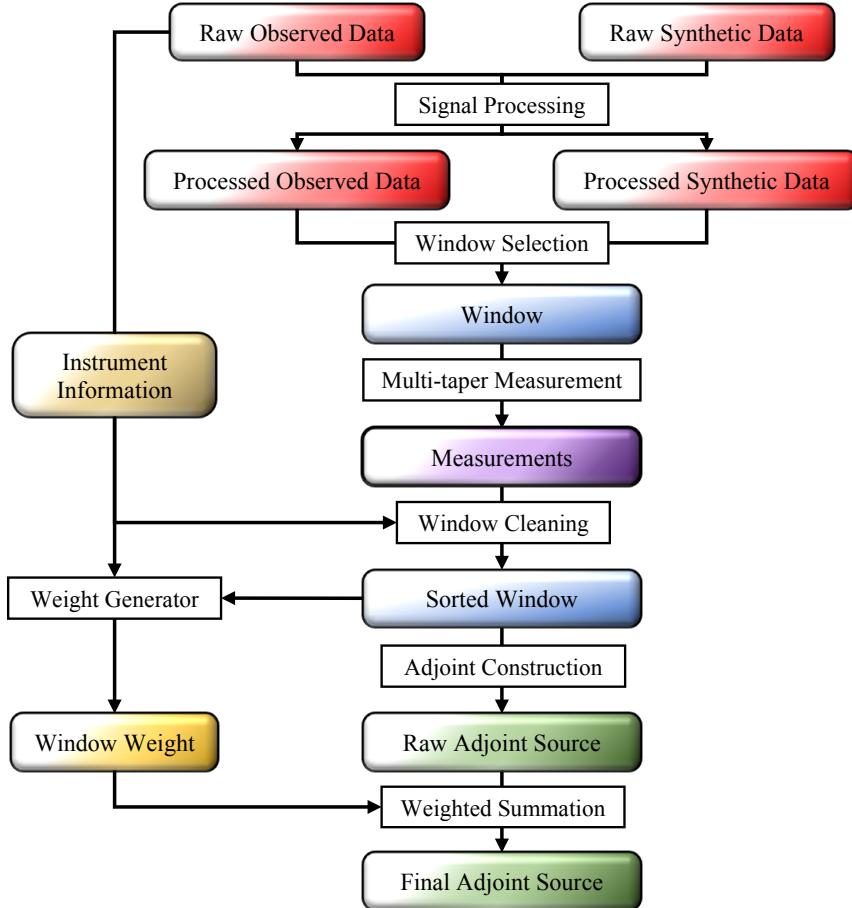


Figure 5: Preprocessing workflow for seismic data.

from the perspective of the tomographic inversion, this is by far the most important stage of the adjoint tomography workflow, directly and fundamentally impacting the outcome. Any “bad data” assimilated at this stage may contaminate the gradient and ultimately the final model.

We use millions of seismograms generating tens of millions of measurement windows, making data selection through human interaction impossible. The preprocessing workflow automates this task, most importantly via the time window selection tool FLEXWIN (Maggi et al. 2009). Every seismogram and potential measurement goes through multiple checks and balances before acceptance or rejection.

As illustrated in Fig. 5, the preprocessing workflow includes the following phases.

- (i) Signal processing to remove the instrument response from observed data to recover ground displacement. Both observed and synthetic data are bandpass filtered, and the horizontal components are rotated to obtain the radial and transverse components of motion.

- (ii) Window selection on a pair of observed and synthetic seismograms. Pyflex (a Python version of FLEXWIN) is used to automatically generate windows where observed and synthetic data are sufficiently close to make measurements based on user defined criteria.
- (iii) Cross-correlation traveltimes or multi-taper phase measurements in selected windows.
- (iv) Window sorting and cleaning based on statistical analyses of all measurements to eliminate outliers.
- (v) Preliminary adjoint source construction based on the sorted windows.
- (vi) Calculation and assignment of weights, as discussed in Section 5.
- (vii) Construction of the final, properly weighted, adjoint sources.

### 7.3 Adjoint Simulations

Collectively, the adjoint simulations are the most expensive stage of the adjoint tomography workflow. Adjoint simulations take the adjoint sources as input, and generate Fréchet derivatives of the four model parameters as output. The computational cost of one adjoint simulation is roughly twice ( $\sim 25$  min) that of a forward simulation, since the forward wavefield is recalculated for convolution with the adjoint wavefield during the adjoint simulation, as discussed in Komatitsch et al. (2016). Using this procedure, both the forward and adjoint wavefield are calculated in forward time, and attenuation is accurately taken into account in both simulations.

### 7.4 Postprocessing

The postprocessing stage of the workflow takes Fréchet derivatives as input and generates a model update as output. It involves the follow steps.

- (i) Summation of the individual Fréchet derivatives for each earthquake to obtain the overall gradient of the misfit function. The contribution of each source is weighted to balance the uneven distribution of earthquakes based on the weighting strategy discussed in Section 5 .
- (ii) Smoothing of the raw misfit gradient using a 3D Gaussian, which serves as a regularization procedure. Instead of using a changing smoothing radius based upon the “ray density” (Bozdağ et al. 2016), we used a fixed value at a given iteration, following Zhu et al. (2012).
- (iii) Preconditioning of the smoothed gradient based on a technique proposed by Luo et al. (2013).
- (iv) Line search to determine the magnitude of the model update. The search direction is

determined using a nonlinear conjugate gradient method (Wright & Nocedal 1999) (iterations 16–21) or an L-BFGS quasi-Newton method (iterations 22–25). We used a subset of 120 earthquakes to conduct the line search and determine the step length.

## 7.5 Workflow management

There are more than ten stages in the adjoint tomography workflow, and each stage involves thousands of small tasks. For every model iteration, we perform 1,480 forward and adjoint simulations, each involving hundreds of compute cores and graphics cards, as well as heavy I/O. The workflow is prone to human error and hardware failure, making it fragile. For these reasons, we wish to harden the workflow by taking advantage of modern workflow management software. With this goal in mind, we selected RADICAL-SAGA and RADICAL-EnTK as our workflow management engines, and developed complementary seismic tomography workflow tools (Balasubramanian et al. 2017). The workflow engine can automatically detect job failures both from the HPC system and via user-defined functions. This enables us to keep track of all tasks and semi-automatically resubmit jobs if necessary. Given that most HPC system time is spent waiting in the job queue, automatic failure detection and relaunching greatly shortens the overall time to solution.

## 8 MISFIT EVOLUTION

The inversion went through ten iterations in a number of stages, as documented in this section. The evolution of the overall misfit function (3) as well as its behavior in the various measurement categories is summarized in Fig. 6.

### 8.0.1 Stage I: Iterations 15–17

Starting model GLAD-M15 was constructed using a database of 256 earthquakes. For iterations 15–17 of the inversion we used an expanded database of 520 events. Our focus was to validate and test our software and workflow before scaling up to a larger database of 1,040 events. For these iterations we used three period bands, namely, 17–40 s, 40–100 s, and 90–250 s, and we used a nonlinear conjugate gradient method to update the model.

### 8.0.2 Stage II: Iterations 18–21

We see an abrupt change in the value of the misfit function at iteration 18 in Fig. 6, reflecting the addition of 520 events to the inversion database, and a change in the weighting strategy

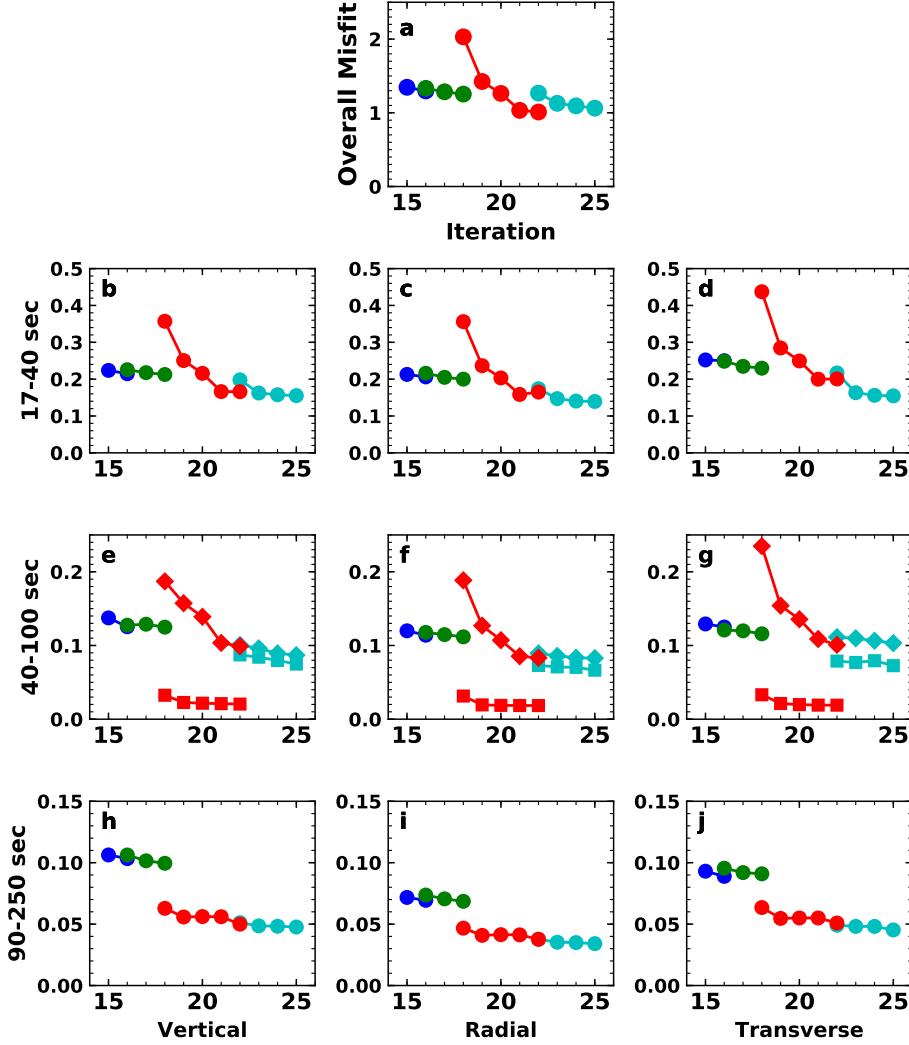


Figure 6: Evolution of phase misfits from GLAD-M15 to GLAD-M25. Each color denotes the different stages of inversions, between which we make changes to the inversion by either tuning the weighting strategies or adding more data. (a) Evolution of the total phase misfit. (b) - (j) Misfit in each period band and component. (e) - (g) From GLAD-M18, we future split the 40-100 sec period band into two measurement categories: body(red diamond) waves and surface waves (red square). **Wenjie, why do the graphs overlap at iteration 22, but not at iteration 18?** Shouldn't 22 look like 18, i.e., no overlapping points? **To Jeroen:** overlap means I calculate the misfit twice using the same mode. For example, in iteration M22, I caculated the misfit using the 1040 earthquakes and 1480 earthquakes, respectively. In M18, I haven't done so. If you want, I can calculate the misfit values and adding the overlap misfit point. **Wenjie:** OK, then let's be consistent in what we do. I would prefer no overlap.

for 40–100 s surface waves. For iterations 18–21 we split the 40–100 s period band into two, separating the body and surface waves (Figs. 6e–g). Since body and surface waves sample different parts of the mantle, this split facilitated more control over the spatial distribution of the model update. Because the 17th iteration model explains 40–100 s surface wave data relatively well, we reduced their weight on all three components for iterations 18–21, as indicated by the lower red curves in Figs. 6e–g.

The misfit reduction for 17–40 s body waves (Figs. 6b–d) tapers off by iteration 21, prompting us to add more earthquakes and change the weighting strategy again by reintroducing the 17–40 s surface waves.

### 8.0.3 Stage III: Iterations 22–25

At iteration 22 we added another 440 earthquakes to the database, bringing the total to 1,480 event, and we changed the 40–100 s surface wave weights back to their setting in Stage I. During this stage we switched from a nonlinear conjugate gradient optimization algorithm to an L-BFGS quasi-Newton method.

## 8.1 Misfit assessment

In the previous sections we discussed the evolution of the misfit through three key stages of the inversion. It is important to note that “the misfit function” is, in fact, a continually moving target, because at every iteration the number of measurements increases as the model improves, and the number of earthquakes increases at iterations 18 and 22.

In this section we calculate specific changes in misfit using all 1,480 earthquakes and an identical set of weightings and windows for three models, namely, GLAD-M25, GLAD-M15, and S362ANI combined with CRUST2.0. Table 1 summarizes the changes in fit in the twelve measurement categories between the new model, GLAD-M25, and its starting model, GLAD-M15, and between GLAD-M25 and model S362ANI combined with CRUST2.0, which was the starting model for the GLAD-M15 inversion. In all categories we observe significant misfit reductions, with dispersive 40–100 s surface waves exhibiting the largest improvements. The improvements in fit per component are comparable in all period bands, indicating that the various categories are reasonably well balanced in the assessment of misfit.

	Vertical (%)	Radial (%)	Transverse (%)
17–40 s body waves	19.1 (37.3)	23.7 (46.3)	28.2 (54.8)
40–100 s body waves	18.3 (37.5)	24.0 (51.4)	24.5 (54.5)
40–100 s surface waves	33.6 (69.6)	33.3 (71.5)	33.5 (77.2)
90–250 s all waves	15.7 (44.9)	14.4 (49.1)	29.6 (48.7)

Table 1: Changes in fit in the twelve measurement categories between the new model, GLAD-M25, and its starting model, GLAD-M15, and, in parentheses, between GLAD-M25 and model S362ANI combined with CRUST2.0, which was the starting model for the GLAD-M15 inversion.

## 8.2 Histogram comparisons

Another way to evaluate model performance is by assessing the distribution of measurements in the various categories. Again we use all 1,480 events and a set of identical windows on all three components to assess GLAD-M25, GLAD-M15, and S362ANI combined with CRUST2.0. The total number of selected windows exceeds 18 million. Fig 7 shows histograms of the resulting phase measurements in all twelve measurement categories. We observe that the distributions generally become better centered on zero, and that all standard deviations are steadily reduced. Again we note that the inversion is aimed at reducing the overall misfit, so there are small tradeoffs between different measurement categories.

Fig 8 shows histograms of the amplitude measurements for all twelve measurement categories, which were not used in the current inversion. Despite this, we observe modest reductions in all standard deviations of the histograms. The histograms are nicely centered, indicating that the moment magnitudes are suitably selected, because the scalar moment affects all measurement categories equally. In the next phase of our ongoing inversion we plan to begin assimilating amplitude measurements, while simultaneously adding shear attenuation as a new model parameter.

Wenjie: Also include the P arrival measurements here.

## 9 MODEL EVALUATION

In this section we evaluate model GLAD-M25 further based on a point-spread function resolution analysis as well as a “held-out” dataset, that is, a set of data not used in the actual structural inversion.

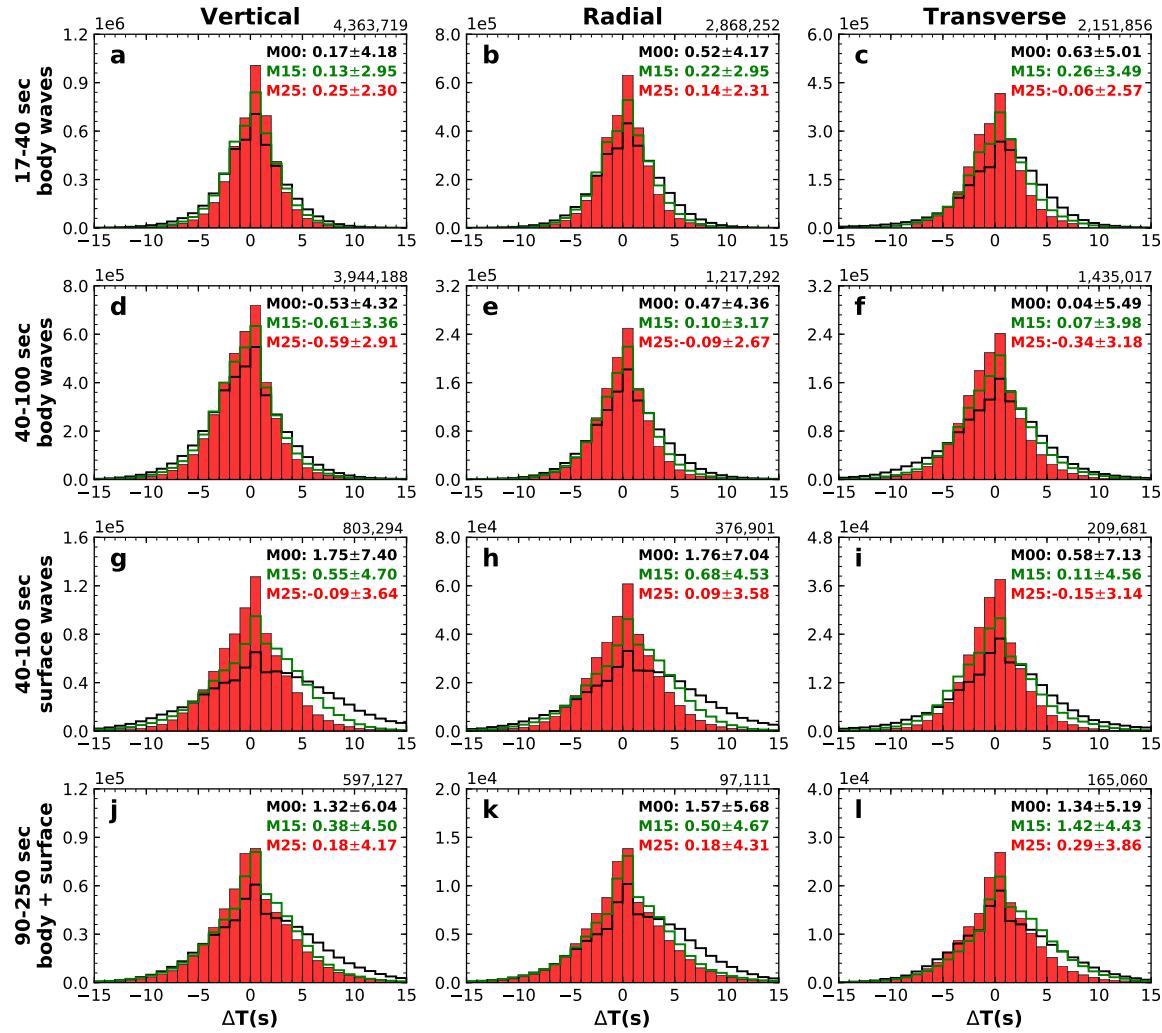


Figure 7: Histograms of phase measurements in all twelve measurement categories for S362ANI combined with CRUST2.0 (M00, Black), GLAD-M15 (M15, Green) and GLAD-M25 (M25, Red). Each column represents one component, and each row corresponds to a period band. The numbers above the top right of each panel denote the number of measurements in the corresponding category. The total number of measurements is 18.2 million. The mean and standard deviations of the phase measurements for the three models are displayed in the top right corner of each panel.

## 9.1 Point-spread function analysis

Checkboard tests or tests involving a chosen known structure, e.g., a plume or slab, are frequently used to evaluate resolution in a tomographic inversion. Such tests are extremely expensive in FWI and adjoint tomography, because they require roughly the same computa-

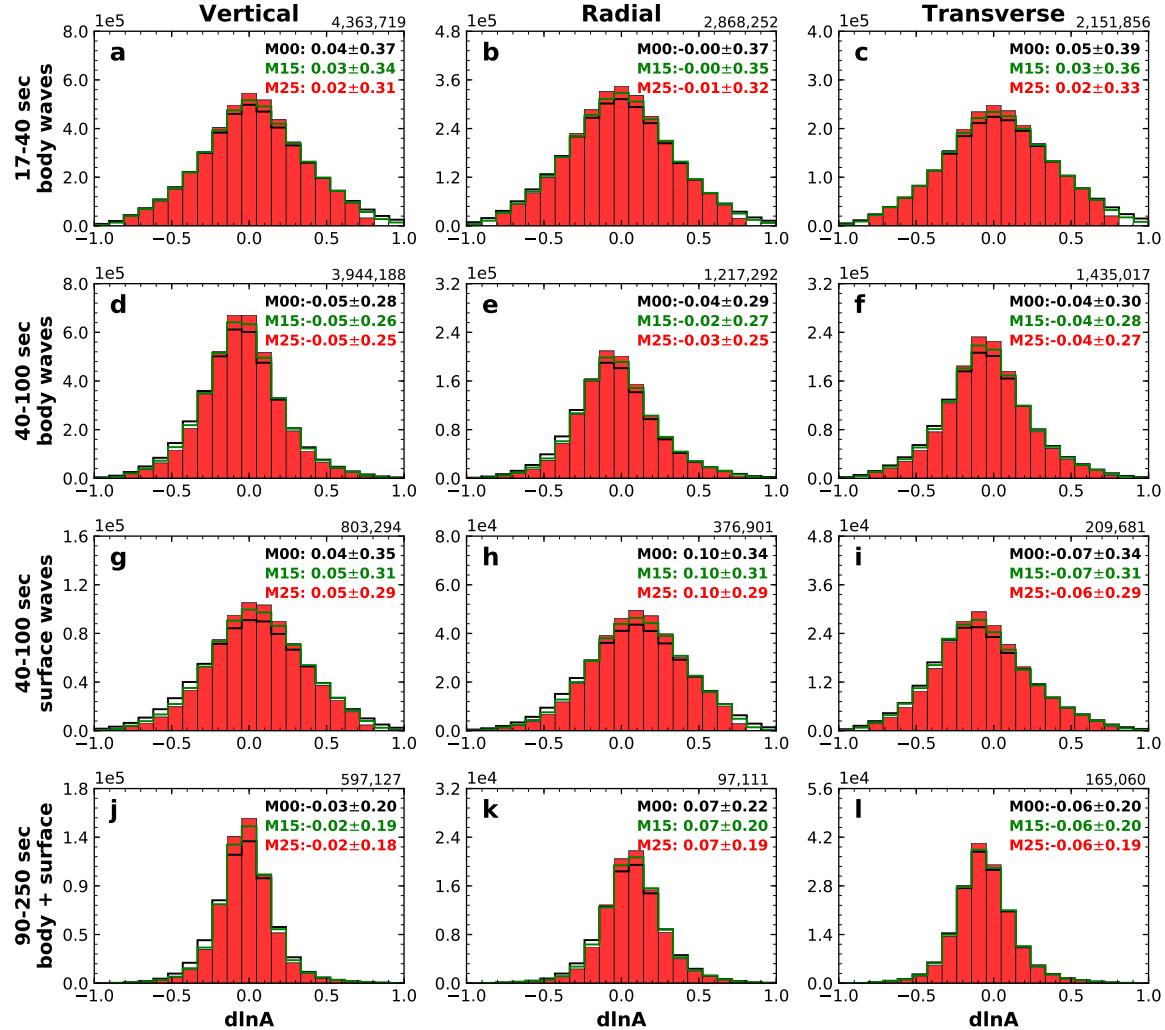


Figure 8: Same as Fig. 7 except for amplitude measurements. These measurements are currently not used in the inversion.

tional resources as the actual inversion. In lieu of such tests, we performed two point-spread function analyses (Fichtner & Trampert 2011; Zhu et al. 2015; Bozdağ et al. 2016).

## 9.2 Held-out dataset

Following the approach of Tape et al. (2009), Chen et al. (2015), and Bozdağ et al. (2016), we selected 360 earthquakes that were not used in the actual inversion. This held-out dataset consists of all magnitude 6.3–7.0 earthquakes in the global CMT catalog that were not used in the inversion. We chose larger events to generate as many measurements as possible. We calibrated the centroid time and scalar moment using a grid search based on 3D synthetics calculated in GLAD-M25. Fig. 9 summarizes the properties of the held-out database.

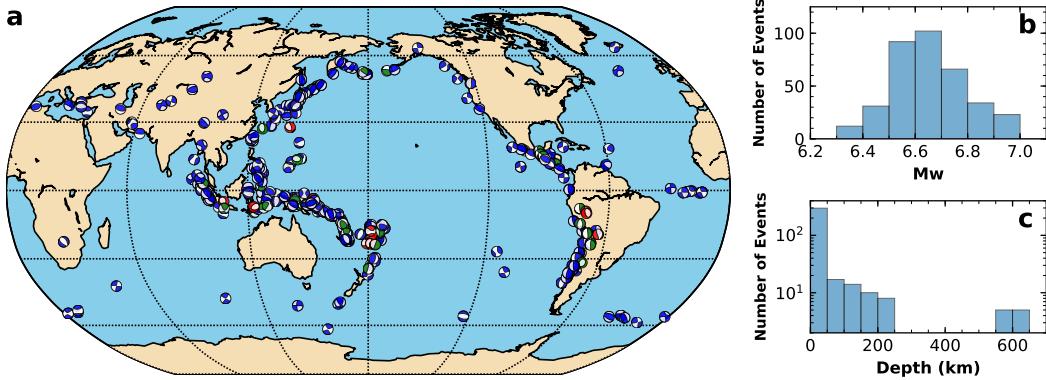


Figure 9: Summary of the held-out database containing 360 earthquakes. These events were not used in the structural inversion. (a) Distribution of events, color coded by depth range as in Fig. 2. (b) & (c) Histograms of moment magnitude and depth.

First, we repeated the misfit assessment discussed in Section 8.1 for the held-out database. Table 2 summarizes the reductions in misfit for the held-out database for model GLAD-M25 compared to GLAD-M15 as well as S362ANI combined with CRUST2.0. It is comforting to see that the reductions are comparable to the values tabulated in Table 1 for the actual inversion. Overall, the reductions are slightly smaller, partly due to the fact that we did not conduct full CMT inversions for the held-out database.

	Vertical (%)	Radial (%)	Transverse (%)
17–40 s	11.8 (29.2)	14.1 (33.7)	21.5 (43.4)
40–100 s body waves	14.5 (26.2)	12.9 (32.2)	16.9 (39.0)
40–100 s surface waves	29.3 (49.1)	28.1 (49.4)	23.7 (51.2)
90–250 s	10.9 (30.9)	14.3 (34.8)	24.0 (34.9)

Table 2: Changes in fit for 360 earthquakes not used in the inversion in the twelve measurement categories between the new model, GLAD-M25, and its starting model, GLAD-M15, and, in parentheses, between GLAD-M25 and model S362ANI combined with CRUST2.0, which was the starting model for the GLAD-M15 inversion.

Second, we repeated the histogram comparisons discussed in Section 8.2 for the held-out database. We used a set of identical windows on all three components to assess phase anomalies in models GLAD-M25, GLAD-M15, and S362ANI combined with CRUST2.0. Fig. 10 shows histograms of the resulting phase measurements in all twelve measurement categories.

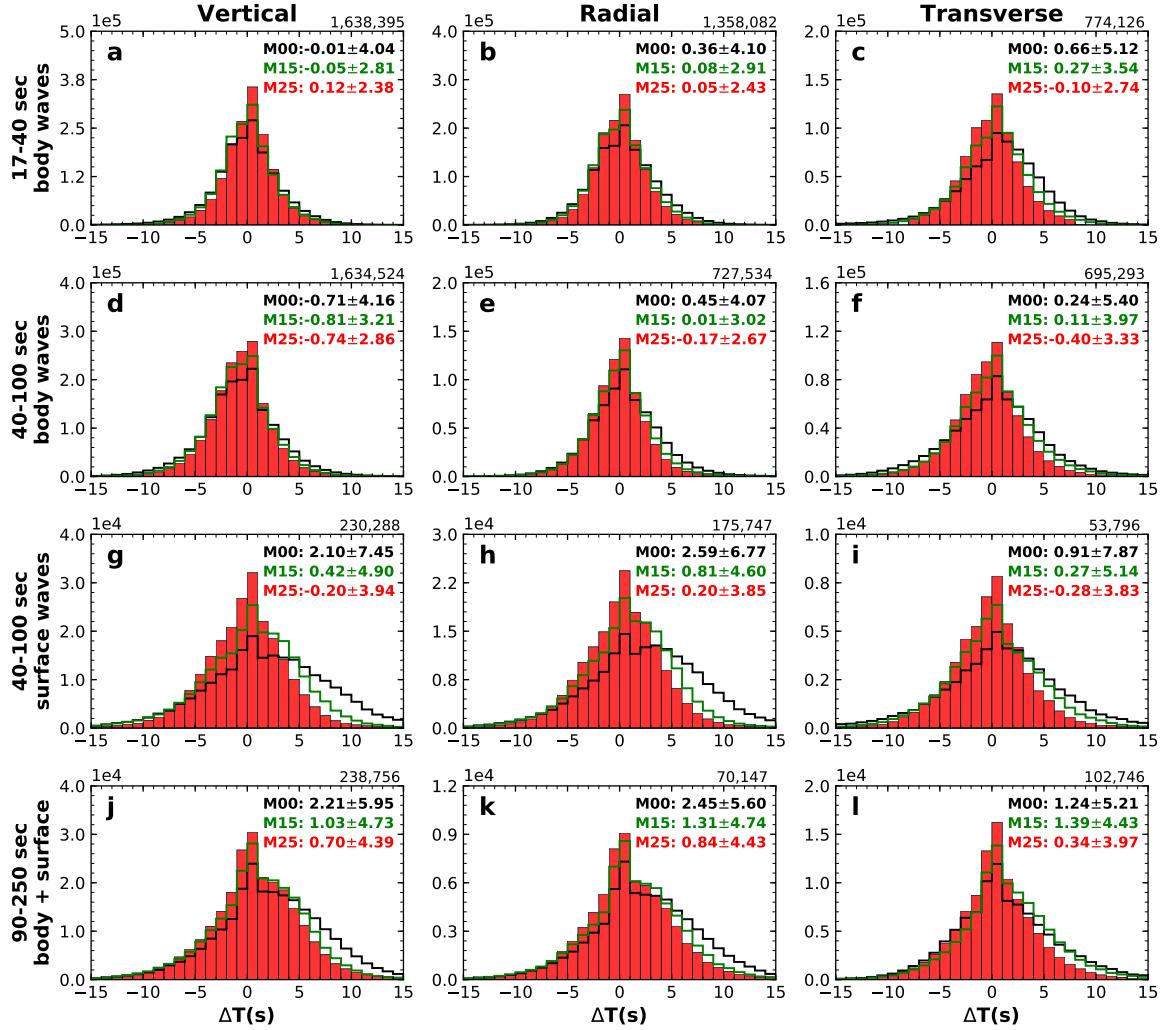


Figure 10: Histograms of phase measurements for the 360 events in the held-out database.

The total number of measurements exceeds 7.8 million. As in the actual inversion, we observe systematic decreases in all standard deviations in all twelve categories, and we note a sharpening of the distributions centered around zero.

## 10 MODEL GLAD-M25

In this section we discuss model GLAD-M25 in some detail. We begin with a global overview of the model before concentrating on specific geographical regions. We conclude by highlighting various slabs and plumes in the model.

## 10.1 Global structure

We first examine model GLAD-M25 in the context of existing global tomographic models. In Fig. 11 we compare global maps of the isotropic part of our shear wavespeed model with model S362ANI+M (Moulik & Ekström 2014), which is an updated version of the starting model for the GLAD-M15 inversion, and model S40RTS (Ritsema et al. 2011). The range of perturbations in shear wavespeed varies from map to map, as indicated in the top left of each panel. Overall, these models are in good agreement at the longest wavelengths, especially Glad-M25 and S362ANI+M. The perturbations in GLAD-M15 tend to be larger than in the other models, consistent with observations by French & Romanowicz (2014), whose model is also based on a form of waveform inversion.

An important aspect of our inversion is that it constrains shear and compressional waves simultaneously. In Fig. 11 we compare global maps of our compressional wavespeed model with P models LLNL-G3Dv3 (Simmons et al. 2012) and GAP-P4 (Fukao & Obayashi 2013). At a depth of 100 km, GLAD-M25 shows prominent mid-oceanic ridges, cooling of the Pacific lithosphere, and the roots of numerous continental cratons. In the upper mantle, LLNL-G3Dv3 exhibits the largest lateral variations, e.g., very slow compressional wavespeeds at 500 km depth below Hawaii, Samoa, and Tahiti. In the midmantle we can see remnants of subduction beneath South America, Indonesia, and Tonga-Kermadec. In the deep mantle, LLNL-G3Dv3 and GAP-P4 begin to fade out, but the long-wavelength patterns, in particular the lowerparts of the Africa and Pacific superplumes, are very similar.

Wenjie: Note: add spherical harmonic analysis later.

### 10.1.1 North America

The deployment of USArray has dramatically increased data coverage across the continental United States. In this section, we compare GLAD-M25 in this region with two global models, S362ANI+M and S40RTS, and two regional models, US22 (Zhu et al. 2017) and SL2013NA (Schaeffer & Lebedev 2014). US22 is a radially anisotropic model based on adjoint tomography using USArray data from 180 regional earthquakes using 15–50 s body waves and 25–150 s surface waves. SL2013NA is a global upper-mantle shear wavespeed model based on multimode surface waveform tomography primarily focused on North America.

North America is characterized by a large, high-wavespeed lithospheric craton, bounded by the Rocky Mountain Front to the West and continental margin to the East. On both sides the lithosphere has been deformed by low wavespeed structures, thus forming sharp tectonic boundaries.

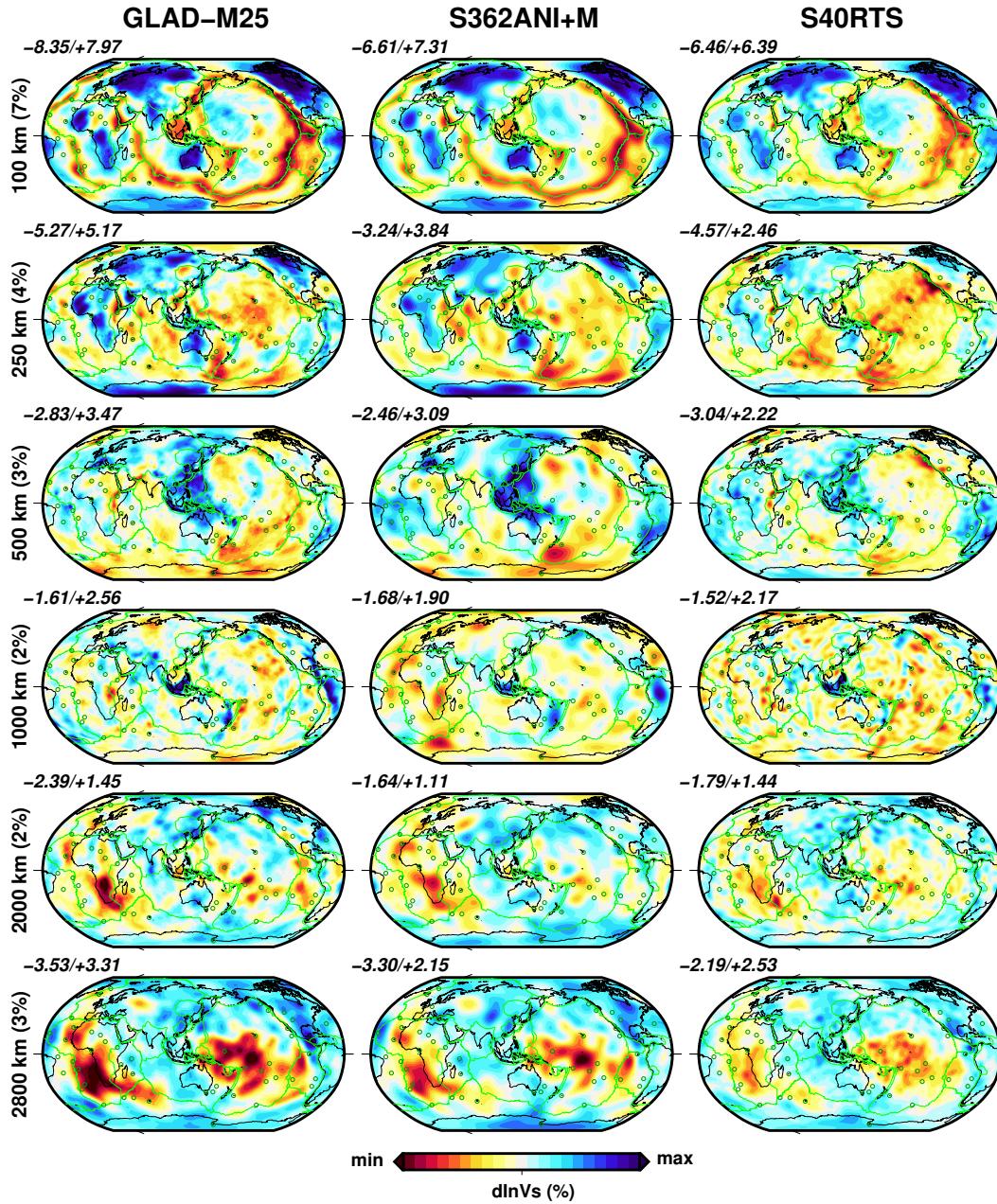


Figure 11: Map views of global shear wavespeed perturbations at various depths for model GLAD-M25 (left column), S362ANI+M (middle column; Moulik & Ekström 2014) and S40RTS (right column; Ritsema et al. 2011). At a given radius, perturbations are calculated relative to each model's average. The range of perturbations in shear wavespeed varies from map to map, as indicated in the top left of each panel. The green circles denote locations of various hotspots (Montelli et al. 2006). The range of the colorbar is the same for each row, and its maximum value is indicated in parentheses on the left, after the depth.

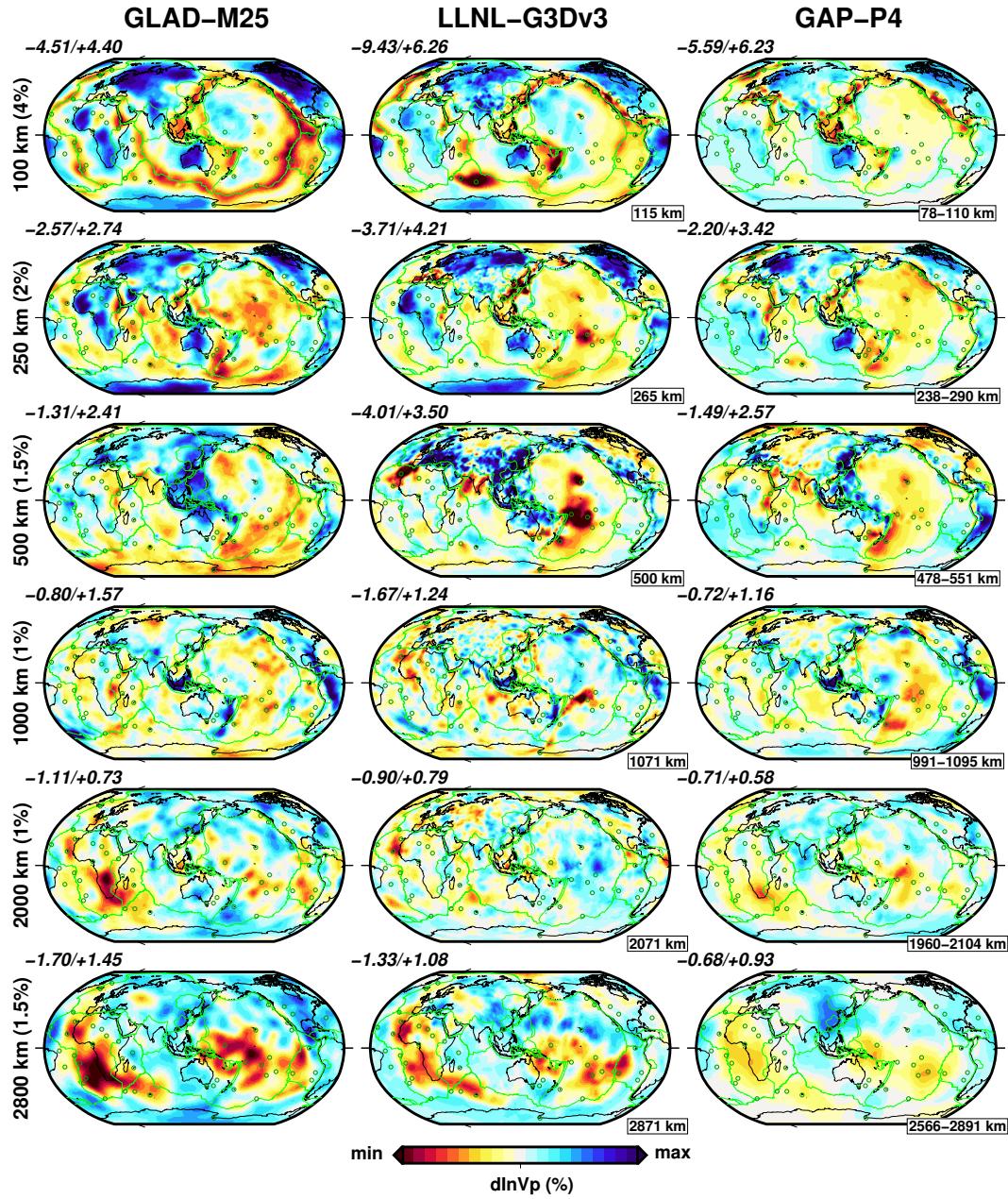


Figure 12: Map views of global compressional wavespeed variations at various depths for model GLAD-M25 (left column), LLNL-G3Dv3 (middle column; Simmons et al. 2012) and GAP-P4 (right column; Fukao & Obayashi 2013). For LLNL-G3Dv3, depth ranges are labeled in the bottom right bottom of each panel. Other plotting conventions are as in Fig. 11.

At a depth of 100 km, we observe that the edges of the craton are sharper in GLAD-M15 than in the other two global models, and quite similar in shape and sharpness to the two regional models. We can clearly see impressions of the Snake River Plain (with Yellowstone) and the Raton hotspot, which also have distinct slow expressions at depths of 200 km. The

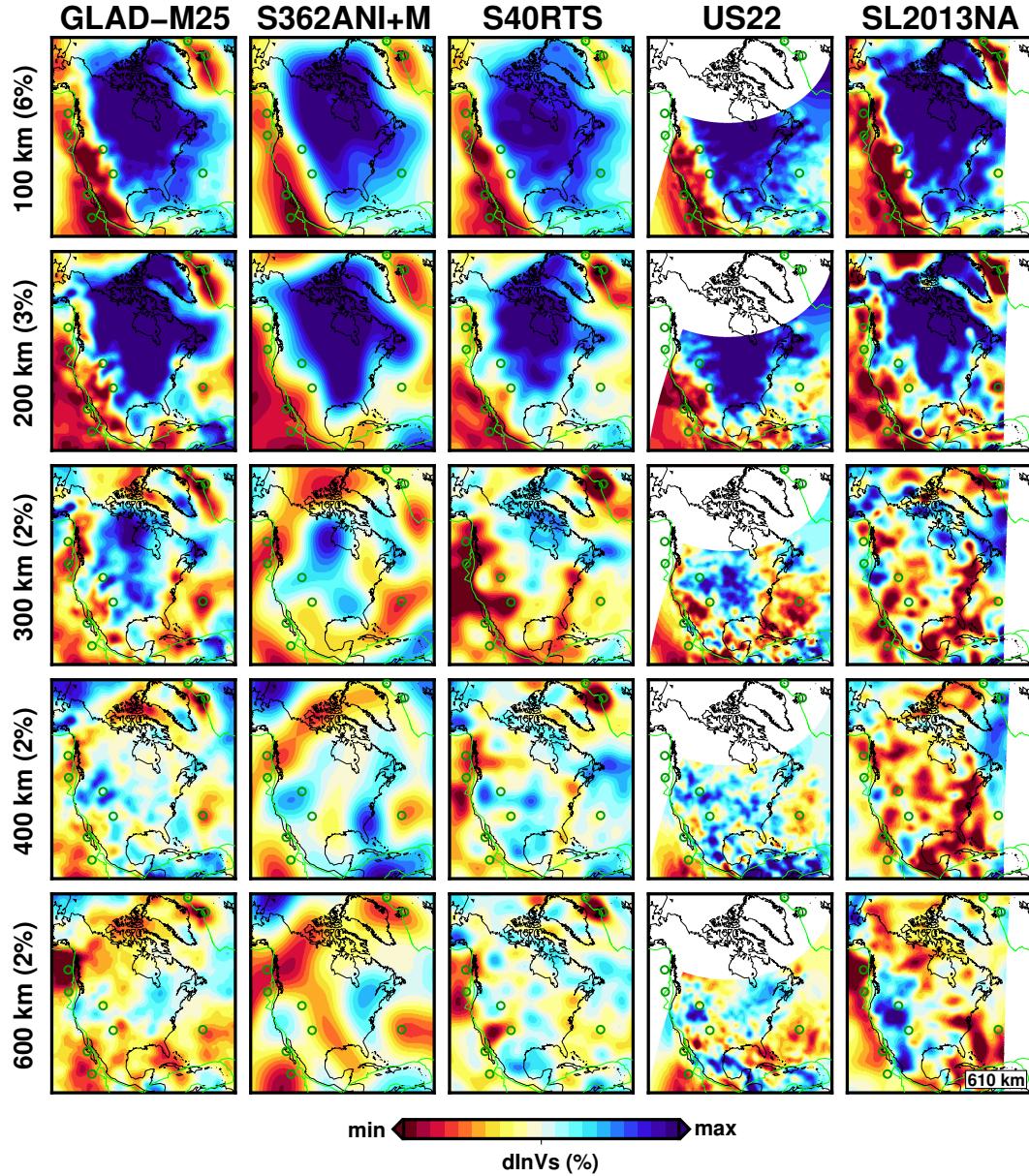


Figure 13: Map views of isotropic shear wavespeed variations beneath North America for model GLAD-M25 (first column) and several global (S362ANI+M and S40RTS) and regional (US22; Zhu et al. 2017) and (SL2013NA; Schaeffer & Lebedev 2014) models. Green circles denote the locations of hotspots. For SL2013NA, we show a depth of 610 km instead of 600 km, reflecting its model parametrization.

core of the craton exhibits fast wavespeeds at depths reaching 300 km Northwest of Hudson Bay. In the regional models the craton has basically faded away at these depths. The Wyoming and Medicine Hat Blocks show a very strong and thick lithospheric root from shallow depths to  $\sim 400$  km, similar to US22 but difficult to discern in SL2013NA. Similar observations can

be made for the Yavapai and Mazatzal Blocks, which define the southern boundary of the large craton.

In the 200–400 km depth range, we observe a high wavespeed anomaly within the Gulf of Mexico, which coincides spatially with the deepest bathymetry and corresponds to a portion of ancient oceanic Lithosphere (?). A remnant of the Juan de Fuca plate, with a high wavespeed footprint, is visible at depths of 300 km and 400 km, where SL2013NA shows a much stronger anomaly.

In the North, we see a high wavespeed craton beneath Greenland which is partly separated from the North American craton by low wavespeed structures beneath Baffin Bay and the Labrador Sea, where the North American craton nicely follows the coastline. In the Northeast corner of the maps we see that the Iceland hotspot is very prominent in GLAD-M25 in all depth ranges, similar to S40RTS but largely missing in S362ANI+M. In the Northwest we see a clear imprint of the Aleutian subduction zone in the 200–400 km depth range, comparable to SL2013NA but absent in the other global models.

In the East, at 100 km depth, we see ageing of the Atlantic lithosphere, comparable to the other global models (?Schaeffer & Lebedev 2014). All models show fast anomalies beneath Newfoundland and Nova Scotia extending to depths of 300 km. The Bermuda hotspot shows a very nice slow wavespeed impression, getting sharper and stronger at depths of 200 km and 300 km (note differences in the scales of the maps). The Caribbean slab, a relatively young subduction zone, is clearly visible in our model in the 200–400 km depth range. At greater depths, GLAD-M25 shows a very distinct image of the Farallon slab, but since it goes well beyond the depth range of this comparison, we return to this feature in Section 10.3.

### 10.1.2 Europe

Fig. 14 shows isotropic shear wavespeed perturbations beneath Europe. In addition to models GLAD-M25, S362ANI+M, and S40RTS, we included regional model EU60 (Zhu et al. 2015) for comparison. EU60 is the result of adjoint tomography of the crust and upper mantle beneath the European continent and the North Atlantic based on the assimilation of data from 190 earthquakes recorded by 745 seismographic station.

A defining tectonic feature of Europe is the Trans-European Suture Zone (TESZ), also known as the Tornquist-Tesisyre Zone, separating the Precambrian East European Craton (EEC) in the Northeast from Phanerozoic Europe in the Southwest. The TESZ is very consistent across all models, even though is much sharper in GLAD-M25 and EU60.

At depths of 100–250 km we observe high wavespeeds beneath the Central Graben in the

North Sea, which is clearly separated from the EEC. The tectonic history of the Mediterranean is complex, involving several regional plates and multiple past and present subduction zones. We see slab ponding in the transition zone beneath the western Mediterranean associated with slab roll-back of the Appennines-Maghrebides subduction zone (?). At shallower depths, we see low wavespeeds along its western border, which are interpreted as back-arc extension associated with the roll-back. Compared to S362ANI+M and S40RTS, our model shows much sharper features for these tectonic structures.

In the Eastern Mediterranean, there is a distinct fast anomaly beneath the Dinarides Mountains, separating the Adria plate from the Pannonian Basin, indicating subduction of the Adria plate beneath the Eurasian plate. This anomaly gets sharper at 600 km depth. Its southern part interacts with the Hellenic slab, which penetrates the mantle transition zone and extends below 1,000 km in our model. We discuss the Hellenic Arc further in Section 10.3. The Pannonian Basin is characterized by a low wavespeed anomaly at depths from 100 km to 250 km. In this depth range we also see slow wavespeeds beneath Anatolia, stronger than in the two other global models.

In Central Europe we note the Cenozoic Ridge System, characterized by low wavespeeds from the Massif Central to the Eifel. Both EU60 and GLAD-M25 exhibit a slow anomaly beneath the northern part of the Rhine Graben, which is interpreted as the reservoir that fuels the Eifel hotspot.

Beneath the Atlantic Ocean, our model strengthens slow anomalies beneath the ridges, especially underneath Iceland and the Azores. Compared to S362ANI+M, Iceland becomes much stronger in the upper mantle at depths from 250 km to 400 km, extending all the way to the mantle-transition zone. The Azores represent a much shallower hotspot, with a maximum depth of about 250 km.

### 10.1.3 Asia

Asia, like Europe, has a complicated tectonic history, involving interactions among the Indian, Eurasian, Australian, Pacific, and Philippine plates. Map views of GLAD-M15 centered on Asia are plotted in Fig. 15, together with S362ANI+M, S40RTS, and regional model EARA2014 (Chen et al. 2015). EARA2014 is a transversely isotropic model based on adjoint tomography, assimilating 1.7 million frequency-dependent traveltimes from 227 earthquakes recorded by 1,869 seismic stations. It has superior data coverage, mainly by incorporating data from the China Array.

The Southwestern part of China is dominated by the Himalayas, expressing the collision

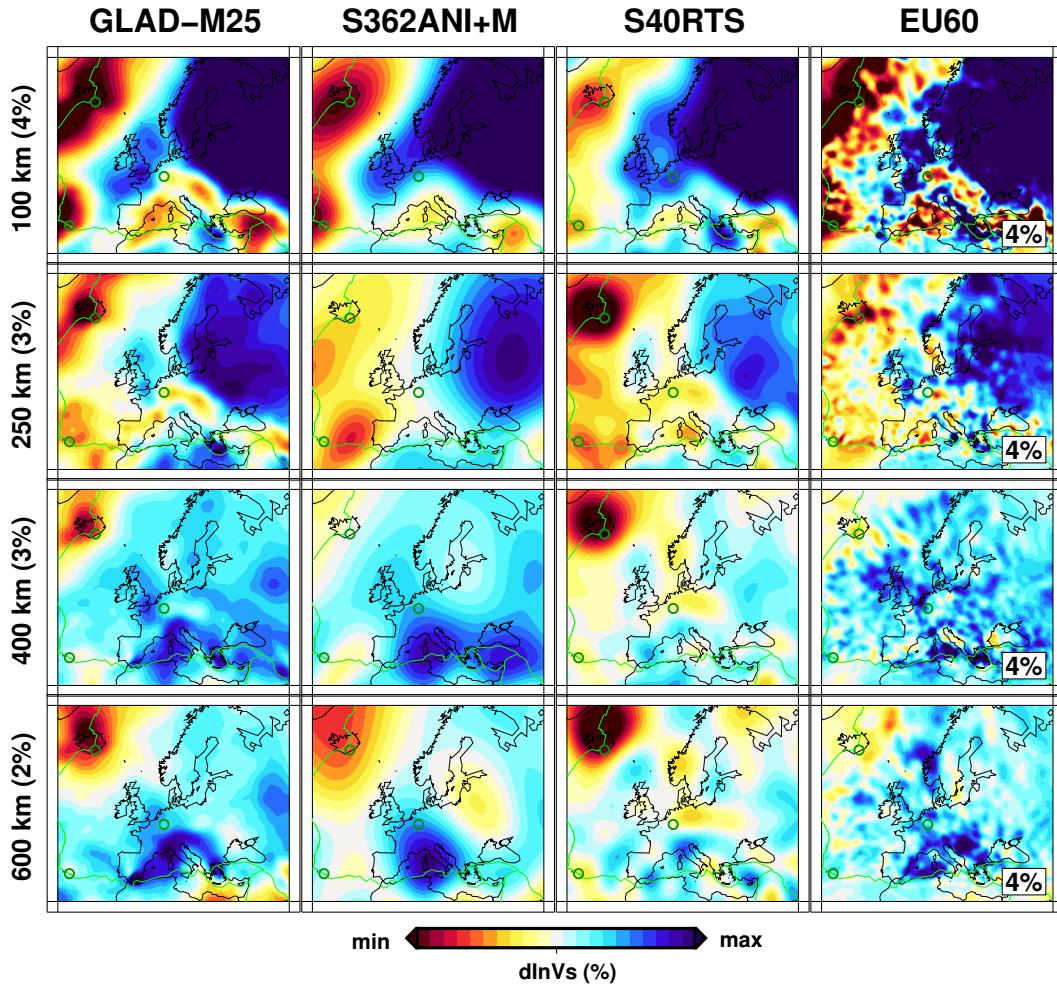


Figure 14: Map views of isotropic shear wavespeed variations beneath Europe for model GLAD-M25 (first column), two global model (S362ANI+M and S40RTS), and one regional model (EU60; Zhu et al. 2015). For EU60, the range of the color bar is 4% at all depths, as indicated in the bottom right bottom of the panels.

between India and Eurasia. GLAD-M25 exhibits strong high wavespeed anomalies in this area, which extend all the way from 100 km to 600 km, narrowing and sharpening with depth, in good agreement with EARA2014. The Tibetan plateau shows relatively low wavespeeds compared to its surroundings, including the Himalayas to its South and the Tianshan Mountains to its North, consistent with the other global and regional studies.

The southern part of Asia shows major tectonic activity between the Philippine, Eurasian, and Australia plates. The Sunda trench is visible as a distinct high wavespeed anomaly below 200 km, nicely following the plate boundary. This feature remains very sharp at 600 km and extends down to 1000 km. We discuss this subducting slab further in Section 10.3. The Manila and Philippine trenches are also much sharper in our model between 200 km and

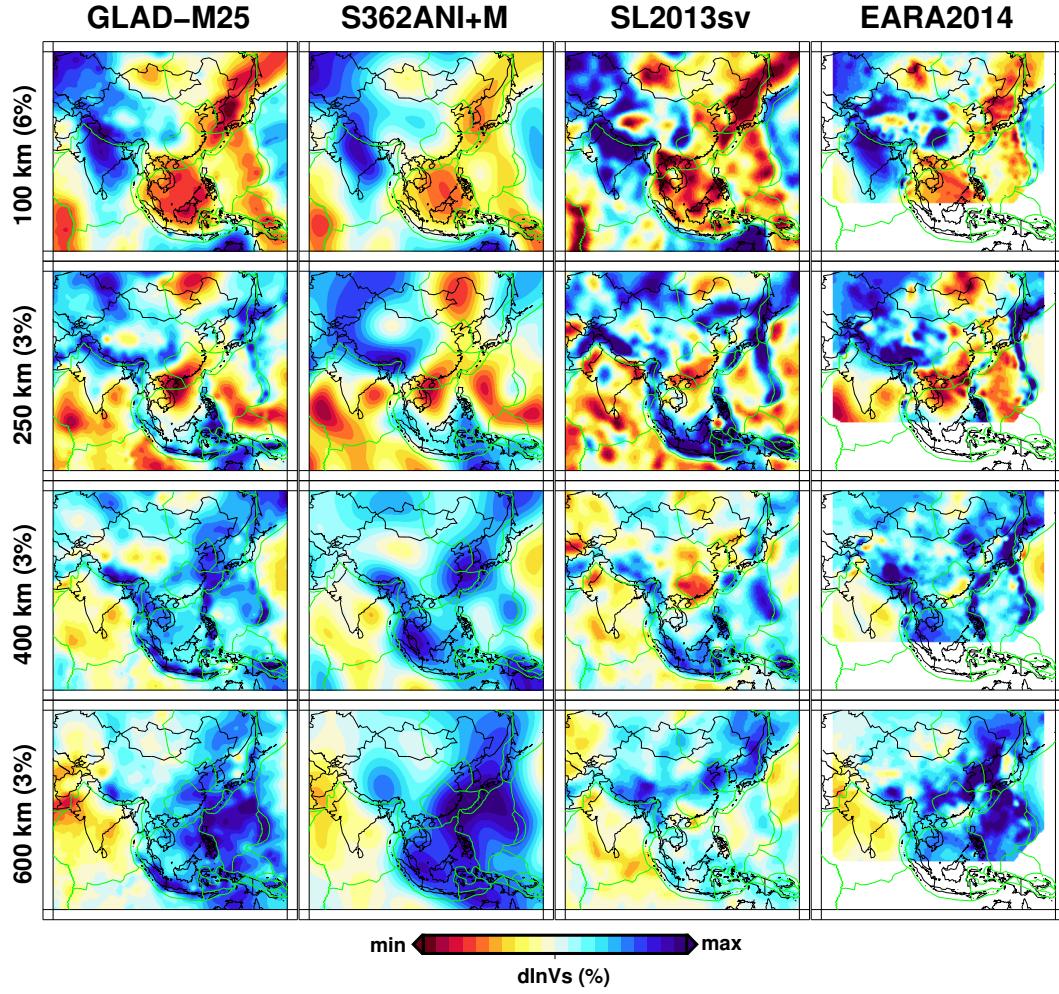


Figure 15: Map views of shear wavespeed variations beneath Asia for GLAD-M25 (first column), global models S362ANI+M (second column) and S40RTS (third column), and regional model EARA2014 (last column; Chen et al. 2015).

600 km than in the other global models, and comparable to EARA2014. The eastern edge of the map features a series of trenches following plate boundaries, including Japan, Izu-Bonin, and Mariana. These slabs are absent in the other global model, but feature more prominently in EARA2014. Along the northern edge we see low wavespeeds associated with the Altai-Sayan and Baikal rift systems. In terms of intra-plate activities, at shallow depths we see clear low wavespeed impressions of the Hainan and Changbai volcanoes, suggesting a lithospheric rather than deeper mantle origin. The Sichuan basin, a very localized high-wavespeed structure, extends below 250 km depth. At 600 km, most models show a large region of high wavespeed beneath the Philippine plate, indicating ponding of ancient subducted plates.

#### 10.1.4 South America

Fig.16 shows the  $V_p$  velocity in our model, together with 3 global models, S362ANI+M, LLNL-G3Dv3(), GAP-P4(Fukao & Obayashi 2013), while the latter two are global  $V_p$  models.

At 100km depth, the northern part of South America is covered by large area of lithosphere craton, showing fast wave anomaly except GAP-P4. The fast anomaly stays strong in LLNL-G3Dv3 while for GLAD-M25 and S362ANI+M it disappear at such depth. At 300km, we are see a sharpened fast anomaly in the south subduction area in GLAD-M25 compared to S372ANI+M. However, the shape and strength are very close between GLAD-M25 and LLNL-G3Dv3 models. Moving to even larger depth, the features of subductions becomes divergent between the north Peruvian slab and south Chilean slab. At 600km, we observed a large chunk of fast anomaly in the Chilean slab, indicating the slab is trapped and stagnant in the mantle transition zone. However, the Peruvian slabs has penetrated the mantle transition zone and extend till 1000, reaching the uppermost part of lower mantle, which is consistent with GAP-P4 (Fukao & Obayashi 2013).

Besides, in the Pacific and Nazca plate, our models is sharpening the Galapagos, San Felix and Juan Fernandez hotspot. In the Scotia plate, our models shows a very strong expression of Scotia Arc subduction till 1000km, which is barely observable in other models. In the Caribbean plate, we also observed fast anomaly, indicating the Caribbean subduction.

## 10.2 Plumes

In this section, we will visit the plumes systems on the earth.

## 10.3 Subduction Slabs

In this sections, we will visit some of the well-known subduction zones.

## 11 CONCLUSION

In summary, the new model performs well for the held-out dataset, both in terms of misfit reductions and measurement distributions in all twelve measurement categories. We expect to see similar behavior for any future earthquakes.

Seismology is driven by data. Back into time, seismologists would look at single seismograms and pick valuable information out the them. Now, we have ten thousands of seismograms deployed globally and it will be a pity not to take all the information we have. Even though GLAD-M25 is following-up of GLAD-M15. It has many superior advantages and improvements. We increase the dataset from 253 to 1480 earthquakes, with carefully inverted sources in the 3D earth structures. We have 360 earthquakes used in the test dataset, which is already larger than the most of the dataset other people used for their inversion. We culled every seismograms from various data centers, thanks to the open environment and data sharing spirit of our seismology community. With very careful checks by assessing the integrity of instrument response and deployment, evaluating the signal-to-noise ratios, picking the time windows using over 20 criteria, cleaning windows by their statistical behavior, and calculating weights to balance the uneven distribution of earthquakes and seismic stations, we really have done a lot to ensure the cleanliness of the data assimilated into our inversion and make sure every piece of value information will be used but rejecting the crap data.

There are also quite a lot of technique challenges on the way but we are seeking the most cutting-edge technologies to solve them. ADIOS was introduced to save mesh and kernel related data, since we are saving petabytes of wavefiled snapshots in a few hours of simulation and it is critical even for the most advanced supercomputers in the world. ASDF is developed by us to store the seismograms since we have ten millions of time series data from thousands of earthquakes. We have to get access and process them in a very efficient manner that most of the existing data format just won't meet our requirements. To handle the complex inversion workflow, we are working RADICAL group to develop the workflow management tools that will help manage every piece of task and make the user aware if there is anything bad happened during the computation. Those tools that help the automation of the workflow, is the "guard" that stand there and monitor things to make sure our scientific results away from contamination. Besides, the developer of SPECFEM packages, they dedicate more than ten years of work, to develop the software and make it fast, efficient and robust for global scale of simulation. They are working closely with hardware vendors such as Cray, IBM and NVidia to adapt the code to architecture of the most advanced supercomputers. We are soon

moving our simulation to Summit, the world No.1 supercomputer at the Oak Ridge National Lab.

## ACKNOWLEDGMENTS

This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under contract DE-AC05-00OR22725. Additional computational resources were provided by the Princeton Institute for Computational Science & Engineering (PICSciE). We acknowledge IRIS ([iris.edu](http://iris.edu)) and ORFEUS ([orfeus-eu.org](http://orfeus-eu.org)) for providing the data used in this study. **Wenjie: Please add all relevant data centers.** We thank Ryan Modrak, Ridvan Orsvuran, Frederik Simons, and James Smith for fruitful discussions. The open source spectral-element software package SPECFEM3D\_GLOBE and the seismic measurement software package FLEXWIN used for this article are freely available via the Computational Infrastructure for Geodynamics (CIG; [geodynamics.org](http://geodynamics.org)). This research was supported by NSF grant 1644826.

## APPENDIX A: SPHERICALLY SYMMETRIC AVERAGE MODEL

In this appendix we determine the spherically symmetric (radial) average of model GLAD-M25. To accomplish this, we first need to transform the 3D spectral-element mesh — which includes ellipticity, topography & bathymetry, and undulations on internal boundaries — into a spherical volume. After these adjustments, we obtain a spectral-element mesh for a perfect cubed sphere (Komatitsch & Tromp 2002a).

In the spherical mesh, the model may be expressed in the form

$$m(\mathbf{x}) = \sum_{\text{elem}} \sum_{\alpha, \beta, \gamma} m^{\alpha\beta\gamma} h_\alpha(\xi) h_\beta(\eta) h_\gamma(\zeta) , \quad (\text{A.1})$$

where  $h_\alpha$  denotes a Lagrange polynomial, and where we have used the invertible mapping

$$\mathbf{x} = \mathbf{x}(\xi, \eta, \zeta) \quad (\text{A.2})$$

between spatial points  $\mathbf{x} = \{x, y, z\}$  and Gauss-Lobatto-Legendre (GLL) points in the reference element  $\{\xi, \eta, \zeta\}$  (Komatitsch & Tromp 1999).

The spherically symmetric part of the model is defined as

$$\bar{m}(r) = \frac{1}{4\pi} \int_{\Omega} m(\mathbf{x}) d\Omega , \quad (\text{A.3})$$

where  $r$  denotes the radius and  $\Omega$  the unit sphere.

Using 2D GLL quadrature (Komatitsch & Tromp 1999) in the unit sphere, the radial average may be determined numerically via GLL quadrature:

$$\bar{m}(r) = \frac{1}{4\pi r^2} \sum_{\alpha, \beta} \omega_\alpha \omega_\beta m^{\alpha\beta}(r) J^{\alpha\beta}(r) . \quad (\text{A.4})$$

Here  $\alpha$  and  $\beta$  denote GLL points in the sphere with radius  $r$ ,  $\omega_\alpha$  denotes a GLL quadrature weight, and  $J^{\alpha\beta}(r)$  is the 2D Jacobian of the mapping to the sphere. It is important to carry out the summations in (A.4) at the elemental level to ensure that first-order discontinuities are honored in the process.

Figure A1 shows the resulting radial averages of the isotropic shear and compressional wavespeeds for model GLAD-M25 compared to 1D starting model STW105 (Kustowski et al. 2008) and PREM (Dziewonski & Anderson 1981). The radial average of GLAD-M25 remains very close to the radial starting model STW105. The main difference between these two models and PREM is the absence of the 220 km discontinuity and a lower  $V_p/V_s$  ratio in the top 200 km.

## APPENDIX B: SPHERICAL HARMONIC MODEL EXPANSION

Wenjie, to save space, this appendix could also go in the supplementary online material.

In this appendix we express model GLAD-M25 in a spherical harmonic basis to facilitate easy plotting, analysis, and comparisons with other models. The resulting spherical harmonic model should only be used for these purposes, not for numerical simulations, which should always be based on the fully 3D spectral-element mesh.

To accomplish the transformation, we first need to transform the 3D mesh — which includes ellipticity, topography & bathymetry, and undulations on internal boundaries — into a spherical volume. After these adjustments, we obtain a spectral element mesh for a perfect sphere, that is, the sort of mesh used for spherically symmetric earth models, such as PREM (Dziewonski & Anderson 1981).

In the spherical mesh, the model may be expressed in the form

$$m(\mathbf{x}) = \sum_{\text{elem } \alpha, \beta, \gamma} \sum_{\alpha, \beta, \gamma} m^{\alpha\beta\gamma} h_\alpha(\xi) h_\beta(\eta) h_\gamma(\zeta) , \quad (\text{B.1})$$

where  $h_\alpha$  denotes a Lagrange polynomial, and where we have used the invertible mapping

$$\mathbf{x} = \mathbf{x}(\xi, \eta, \zeta) \quad (\text{B.2})$$

between spatial points  $\mathbf{x} = \{x, y, z\}$  and Gauss-Lobatto-Legendre (GLL) points in the reference element  $\{\xi, \eta, \zeta\}$  (Komatitsch & Tromp 1999).

Our goal is to expand our spectral-element model in a spherical harmonic basis, i.e.,

$$m(\mathbf{x}) = \sum_{n=0}^N \sum_{\ell=0}^L \sum_{m=-\ell}^{\ell} {}_n C_{\ell m} R_n(r) Y_{\ell m}(\theta, \phi) , \quad (\text{B.3})$$

where  $r$  denotes the radius,  $\theta$  colatitude, and  $\phi$  longitude. We choose a radial basis of the form  $R_n(r)$ ,  $n = 0, \dots, N$ , e.g., layers or B-splines. These radial basis functions need to be chosen sufficiently dense to mimic the density of the radial spectral element mesh. The radial basis may or may not be orthogonal, i.e.,

$$\int_b^a R_{n'}(r) R_n(r) r^2 dr = A_{n'n} , \quad (\text{B.4})$$

where  $b$  denotes the radius of the CMB and  $a$  the free surface. The matrix elements  $A_{n'n}$  define a positive definite matrix which is invertible. As lateral basis functions we use fully normalized spherical harmonics  $Y_{\ell m}(\theta, \phi)$ ,  $\ell = 0, \dots, L$  and  $m = -\ell, \dots, \ell$ , i.e., (Dahlen & Tromp 1998)

$$\int_0^{2\pi} \int_0^\pi Y_{\ell' m'}^*(\theta, \phi) Y_{\ell m}(\theta, \phi) \sin \theta d\theta d\phi = \delta_{\ell' \ell} \delta_{m' m} , \quad (\text{B.5})$$

where an asterisk denotes complex conjugation. The maximum degree  $L$  needs to be chosen to resolve the spectral-element mesh laterally.

To obtain the expansion coefficients  ${}_nC_{\ell m}$ , we multiply Eqn. (B.3) by  $R_{n'}(r) Y_{\ell' m'}^*(\theta, \phi)$  and integrate over the volume of the mantle and crust,  $V$ , using Eqns. (B.4 and (B.5)):

$$\int_V m(\mathbf{x}) R_{n'}(r) Y_{\ell' m'}^*(\theta, \phi) d^3\mathbf{x} = \sum_{n=0}^N {}_nC_{\ell' m'} A_{n'n} , \quad (\text{B.6})$$

We evaluate the integral on the left using GLL quadrature (Komatitsch & Tromp 1999):

$$\sum_{\text{elem}} \sum_{\alpha, \beta, \gamma} \omega_\alpha \omega_\beta \omega_\gamma m^{\alpha\beta\gamma} J^{\alpha\beta\gamma} R_{n'}^{\alpha\beta\gamma} Y_{\ell' m'}^{*\alpha\beta\gamma} = \sum_{n=0}^N {}_nC_{\ell' m'} A_{n'n} . \quad (\text{B.7})$$

Here  $\omega_\alpha$  denotes a GLL quadrature weight,  $J_{\alpha\beta\gamma}$  the Jacobian of the mapping (B.2) evaluated on the GLL points, and  $R_{n'}^{\alpha\beta\gamma}$  and  $Y_{\ell' m'}^{*\alpha\beta\gamma}$  the values of the radial and spherical harmonic basis functions at a GLL point.

Finally, the desired model coefficients may be obtained —one combination of  $n$ ,  $\ell$ , and  $m$  at a time— via

$${}_nC_{\ell m} = \sum_{n'} A_{nn'}^{-1} \sum_{\text{elem}} \sum_{\alpha, \beta, \gamma} \omega_\alpha \omega_\beta \omega_\gamma m^{\alpha\beta\gamma} J^{\alpha\beta\gamma} R_{n'}^{\alpha\beta\gamma} Y_{\ell m}^{*\alpha\beta\gamma} , \quad (\text{B.8})$$

where  $A^{-1}$  denotes the inverse of the  $N \times N$  matrix  $A$ . Expressions of the form (B.8) are commonplace in spectral-element simulations, and thus easily calculated. The normalized power per degree  $\ell$  may be calculated via

$$\sigma_\ell^2 = \frac{1}{2\ell+1} \sum_{n'=0}^N \sum_{n=0}^N \sum_{m=-\ell}^{\ell} {}_{n'}C_{\ell m}^* A_{n'n} {}_nC_{\ell m} . \quad (\text{B.9})$$

Instead of the complex spherical harmonic model expansion (B.3), we may wish to use a real spherical harmonic expansion of the form (Dahlen & Tromp 1998, Section B.8, Eqn. B.99)

$$m(\mathbf{x}) = \sum_{n=0}^N \sum_{\ell=0}^L \left[ {}_n a_{\ell 0} R_n(r) X_{\ell 0}(\theta) + \sqrt{2} \sum_{m=1}^{\ell} R_n(r) ({}_n a_{\ell m} \cos m\phi + {}_n b_{\ell m} \sin m\phi) X_{\ell m}(\theta) \right] , \quad (\text{B.10})$$

where (Dahlen & Tromp 1998, Eqn. B.30)

$$Y_{\ell m}(\theta, \phi) = X_{\ell m}(\theta) \exp(im\phi) , \quad (\text{B.11})$$

and where

$${}_n a_{\ell 0} = \sum_{n'} A_{nn'}^{-1} \sum_{\text{elem}} \sum_{\alpha, \beta, \gamma} \omega_\alpha \omega_\beta \omega_\gamma m^{\alpha\beta\gamma} J^{\alpha\beta\gamma} R_{n'}^{\alpha\beta\gamma} X_{\ell m}^{\alpha\beta\gamma} , \quad (\text{B.12})$$

and for  $1 \leq m \leq \ell$

$${}_n a_{\ell m} = \sqrt{2} \sum_{n'} A_{nn'}^{-1} \sum_{\text{elem}} \sum_{\alpha, \beta, \gamma} \omega_\alpha \omega_\beta \omega_\gamma m^{\alpha\beta\gamma} J^{\alpha\beta\gamma} R_{n'}^{\alpha\beta\gamma} X_{\ell m}^{\alpha\beta\gamma} (\cos m\phi)^{\alpha\beta\gamma} , \quad (\text{B.13})$$

$${}_n b_{\ell m} = \sqrt{2} \sum_{n'} A_{nn'}^{-1} \sum_{\text{elem}} \sum_{\alpha, \beta, \gamma} \omega_\alpha \omega_\beta \omega_\gamma m^{\alpha\beta\gamma} J^{\alpha\beta\gamma} R_{n'}^{\alpha\beta\gamma} X_{\ell m}^{\alpha\beta\gamma} (\sin m\phi)^{\alpha\beta\gamma} . \quad (\text{B.14})$$

The normalized power per degree  $\ell$  may then be calculated via

$$\sigma_\ell^2 = \frac{1}{2\ell+1} \sum_{n'=0}^N \sum_{n=0}^N A_{n'n} \left[ {}_{n'} a_{\ell 0} {}_n a_{\ell 0} + \sum_{m=1}^\ell ({}_{n'} a_{\ell m} {}_n a_{\ell m} + {}_{n'} b_{\ell m} {}_n b_{\ell m}) \right] . \quad (\text{B.15})$$

To assess the accuracy of the expansion, we calculate the model norm in the GLL basis,

$$\|m_{\text{GLL}}\|^2 = \int_V m_{\text{GLL}}^2(\mathbf{x}) d^3\mathbf{x} = \sum_{\text{elem}} \sum_{\alpha, \beta, \gamma} \omega_\alpha \omega_\beta \omega_\gamma (m^{\alpha\beta\gamma})^2 J^{\alpha\beta\gamma} , \quad (\text{B.16})$$

and compare it to the model norm in the spherical harmonic basis,

$$\|m_{\text{SH}}\|^2 = \int_V m_{\text{SH}}^2(\mathbf{x}) d^3\mathbf{x} = \sum_{\ell=0}^L (2\ell+1) \sigma_\ell^2 , \quad (\text{B.17})$$

by calculating the error

$$\epsilon = \log \frac{\|m_{\text{SH}}\|^2}{\|m_{\text{GLL}}\|^2} . \quad (\text{B.18})$$

## REFERENCES

- Balasubramanian, V., Turilli, M., Huy, W., Lefebvre, M., Lei, W., Cervoney, G., Tromp, J., & Jha, S., 2017. Harnessing the power of many: Extensible toolkit for scalable ensemble applications, International Parallel and Distributed Processing Symposium.
- Bassin, C., Laske, G., & Masters, G., 2000. The current limits of resolution for surface wave tomography in {North America}, *Eos*, **81**.
- Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N., & Pugmire, D., 2016. Global adjoint tomography: first-generation model, *Geophysical Journal International*, **207**(3), 1739–1766.
- Chen, M., Niu, F., Liu, Q., Tromp, J., & Zheng, X., 2015. Multiparameter adjoint tomography of the crust and upper mantle beneath east asia: 1. model construction and comparisons, *Journal of Geophysical Research: Solid Earth*, **120**(3), 1762–1786.
- Dahlen, F. A. & Tromp, J., 1998. *Theoretical Global Seismology*, Princeton U. Press, New Jersey.
- Dziewonski, A. & Anderson, D., 1981. Preliminary reference Earth model, *Phys. Earth Planet. Inter.*, **25**, 297–356.
- Ekström, G., Nettles, M., & Dziewoński, A., 2012. The global cmt project 2004–2010: Centroid-moment tensors for 13,017 earthquakes, *Physics of the Earth and Planetary Interiors*, **200**, 1–9.
- Ekström, G., Nettles, M., & Dziewoński, A. M., 2012. The global CMT project 2004-2010: centroid-moment tensors for 13,017 earthquakes, *Phys. Earth Planet. Inter.*, **200-201**, 1–9.
- Fichtner, A. & Trampert, J., 2011. Resolution analysis in full waveform inversion, *Geophysical Journal International*, **187**(3), 1604–1624.
- French, S. & Romanowicz, B., 2014. Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography, *Geophysical Journal International*, **199**(3), 1303–1327.
- Fukao, Y. & Obayashi, M., 2013. Subducted slabs stagnant above, penetrating through, and trapped below the 660 km discontinuity, *Journal of Geophysical Research: Solid Earth*, **118**(11), 5920–5938.
- Hjörleifsdóttir, V. & Ekström, G., 2010. Effects of three-dimensional earth structure on cmt earthquake parameters, *Physics of the Earth and Planetary Interiors*, **179**(3-4), 178–190.
- Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral-element method for 3-D seismic wave propagation, **139**(3), 806–822.
- Komatitsch, D. & Tromp, J., 2002a. Spectral-element simulations of global seismic wave propagation—I. Validation, **149**(2), 390–412.
- Komatitsch, D. & Tromp, J., 2002b. Spectral-element simulations of global seismic wave propagation—II. 3-D models, oceans, rotation, and self-gravitation, **150**(1), 303–318.
- Komatitsch, D., Xie, Z., Bozdağ, E., Sales de Andrade, E., Peter, D., Liu, Q., & Tromp, J., 2016. Anelastic sensitivity kernels with parsimonious storage for adjoint tomography and full waveform inversion, **206**(3), 1467–1478.
- Krischer, L., Smith, J., Lei, W., Lefebvre, M., Ruan, Y., de Andrade, E. S., Podhorszki, N., Bozdağ,

- E., & Tromp, J., 2016. An adaptable seismic data format, *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, **207**(2), 1003–1011.
- Kustowski, B., Ekström, G., & Dziewoński, A., 2008. Anisotropic shear-wave velocity structure of the earth's mantle: A global model, *Journal of Geophysical Research: Solid Earth*, **113**(B6).
- Liu, Q., Polet, J., Komatitsch, D., & Tromp, J., 2004. Spectral-element moment tensor inversions for earthquakes in southern California, *Bulletin of the Seismological Society of America*, **94**(5), 1748–1761.
- Liu, Q., Logan, J., Tian, Y., Abbasi, H., Podhorszki, N., Choi, J. Y., Klasky, S., Tchoua, R., Lofstead, J., Oldfield, R., et al., 2014. Hello adios: the challenges and lessons of developing leadership class i/o frameworks, *Concurrency and Computation: Practice and Experience*, **26**(7), 1453–1473.
- Love, A. E. H., 1927. *A Treatise on the Mathematical Theory of Elasticity*, Cambridge U. Press, Cambridge, 4th edn.
- Luo, Y., Modrak, R., & Tromp, J., 2013. Strategies in adjoint tomography, *Handbook of Geomathematics*, pp. 1–52.
- Maggi, A., Tape, C., Chen, M., Chao, D., & Tromp, J., 2009. An automated time-window selection algorithm for seismic tomography, *Geophysical Journal International*, **178**(1), 257–281.
- Montagner, J.-P. & Anderson, D. L., 1989. Petrological constraints on seismic anisotropy, *Physics of the earth and planetary interiors*, **54**(1-2), 82–105.
- Montelli, R., Nolet, G., Dahlen, F., & Masters, G., 2006. A catalogue of deep mantle plumes: New results from finite-frequency tomography, *Geochemistry, Geophysics, Geosystems*, **7**(11).
- Moulik, P. & Ekström, G., 2014. An anisotropic shear velocity model of the earth's mantle using normal modes, body waves, surface waves and long-period waveforms, *Geophysical Journal International*, **199**(3), 1713–1738.
- Peter, D., Komatitsch, D., Luo, Y., Martin, R., Le Goff, N., Casarotti, E., Le Loher, P., Magnoni, F., Liu, Q., Blitz, C., Nissen-Meyer, T., Basini, P., & Tromp, J., 2011. Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *186*(2), 721–739.
- Plessix, R. E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *167*(2), 495–503.
- Ritsema, J., Deuss, a. A., Van Heijst, H., & Woodhouse, J., 2011. S40rts: a degree-40 shear-velocity model for the mantle from new rayleigh wave dispersion, teleseismic traveltimes and normal-mode splitting function measurements, *Geophysical Journal International*, **184**(3), 1223–1236.
- Schaeffer, A. & Lebedev, S., 2014. Imaging the north american continent using waveform inversion of global and usarray data, *Earth and Planetary Science Letters*, **402**, 26–41.
- Simmons, N. A., Myers, S. C., Johannesson, G., & Matzel, E., 2012. Llnl-g3dv3: Global p wave tomography model for improved regional and teleseismic travel time prediction, *Journal of Geophysical Research: Solid Earth*, **117**(B10).
- Tape, C., Liu, Q., Maggi, A., & Tromp, J., 2009. Adjoint tomography of the southern California crust,

- Science*, **325**(5943), 988–992.
- Tromp, J., Tape, C., & Liu, Q. Y., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J Int.*, **160**, 195–216.
- Wright, S. & Nocedal, J., 1999. Numerical optimization, *Springer Science*, **35**(67-68), 7.
- Zhu, H., Bozdağ, E., Peter, D., & Tromp, J., 2012. Structure of the european upper mantle revealed by adjoint tomography, *Nature Geoscience*, **5**(7), 493.
- Zhu, H., Bozdağ, E., & Tromp, J., 2015. Seismic structure of the european upper mantle based on adjoint tomography, *Geophysical Journal International*, **201**(1), 18–52.
- Zhu, H., Komatitsch, D., & Tromp, J., 2017. Radial anisotropy of the north american upper mantle based on adjoint tomography with usarray, *Geophysical Journal International*, **211**(1), 349–377.

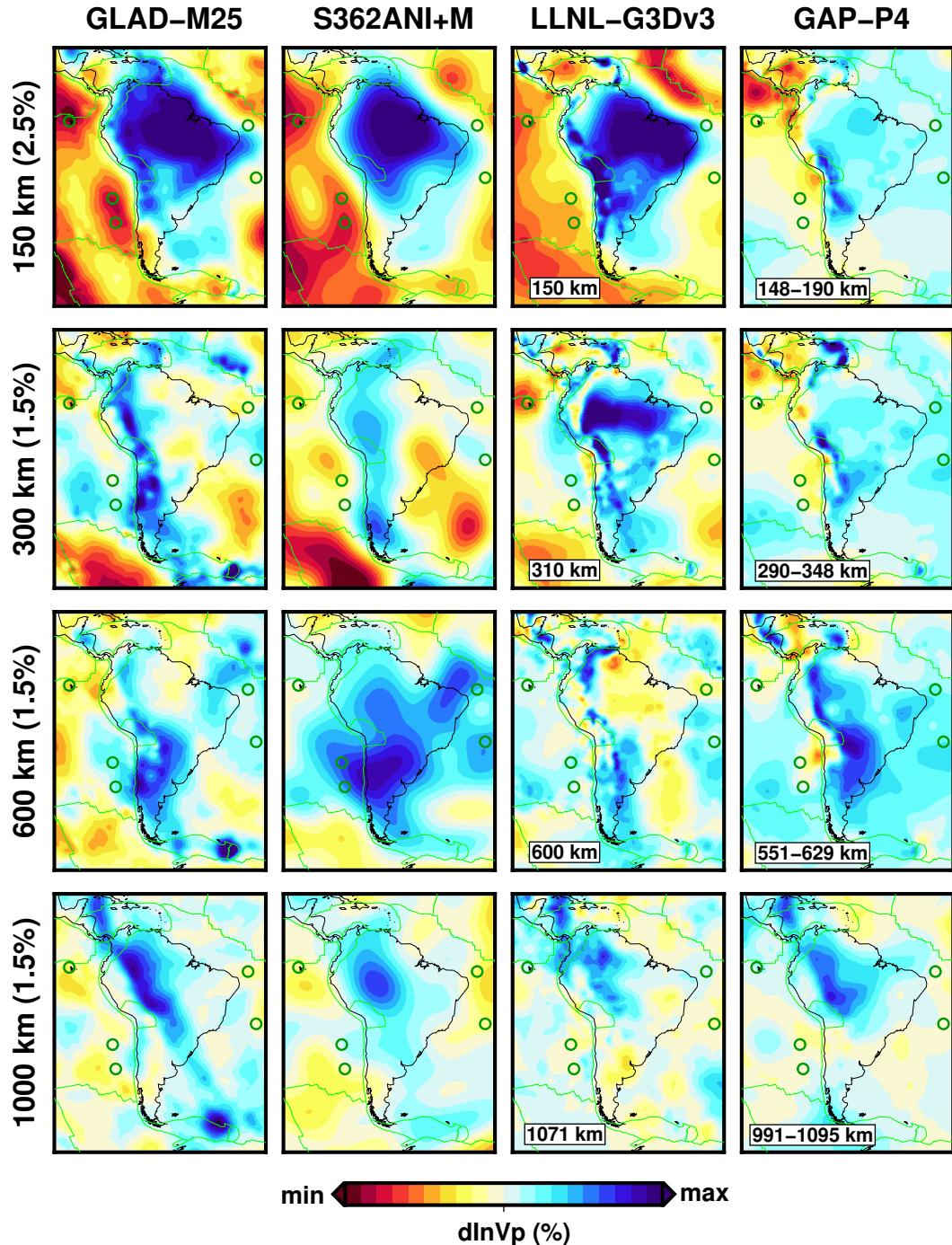


Figure 16: Map views of compressional wavespeed variations beneath South America for GLAD-M25 (first column) and global models S362ANI+M, LLNL-G3Dv3 (), and GAP-P4 (Fukao & Obayashi 2013).

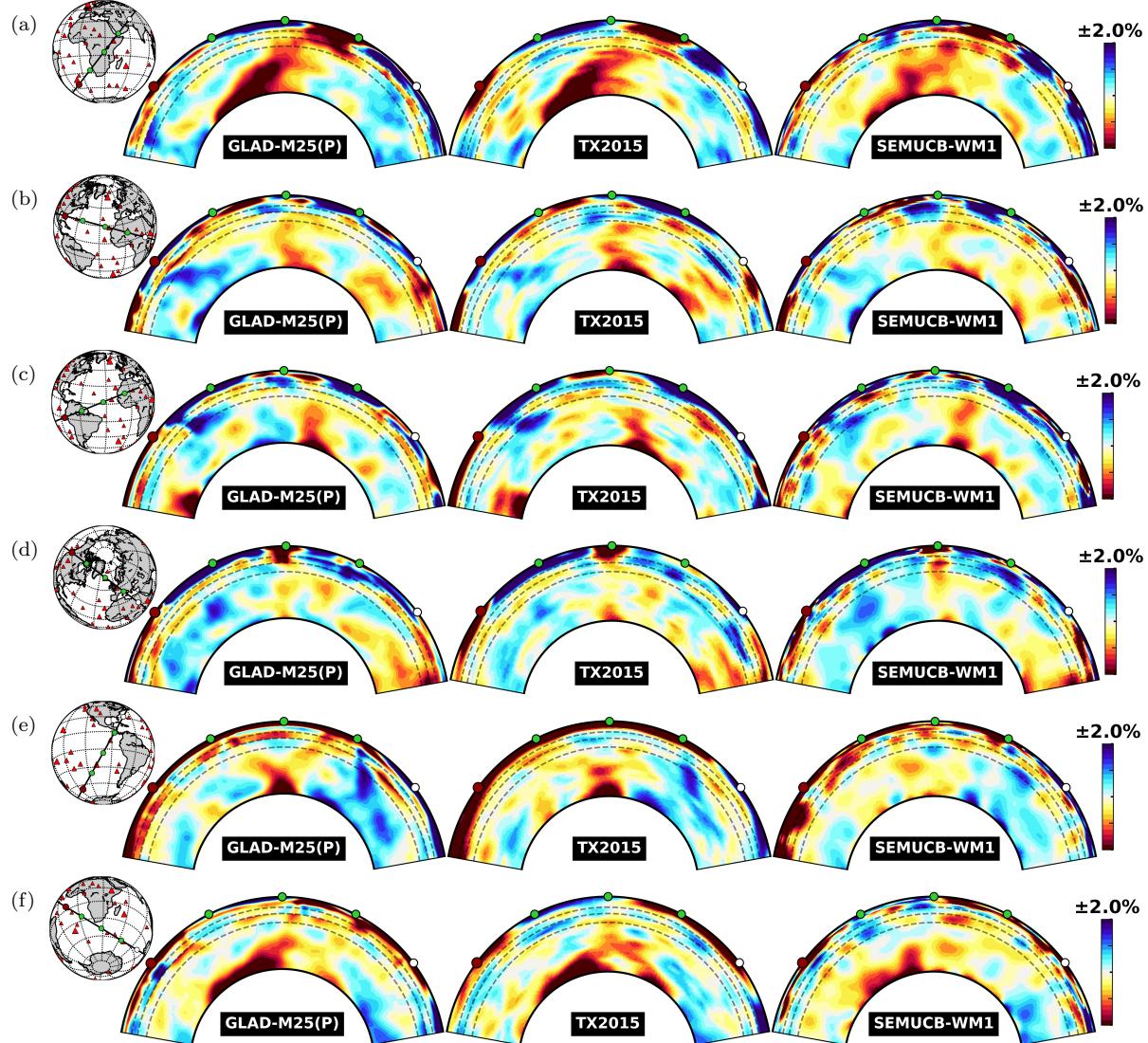


Figure 17: Vertical cross sections of perturbations of shear wave velocity  $V_s$  of plumes.

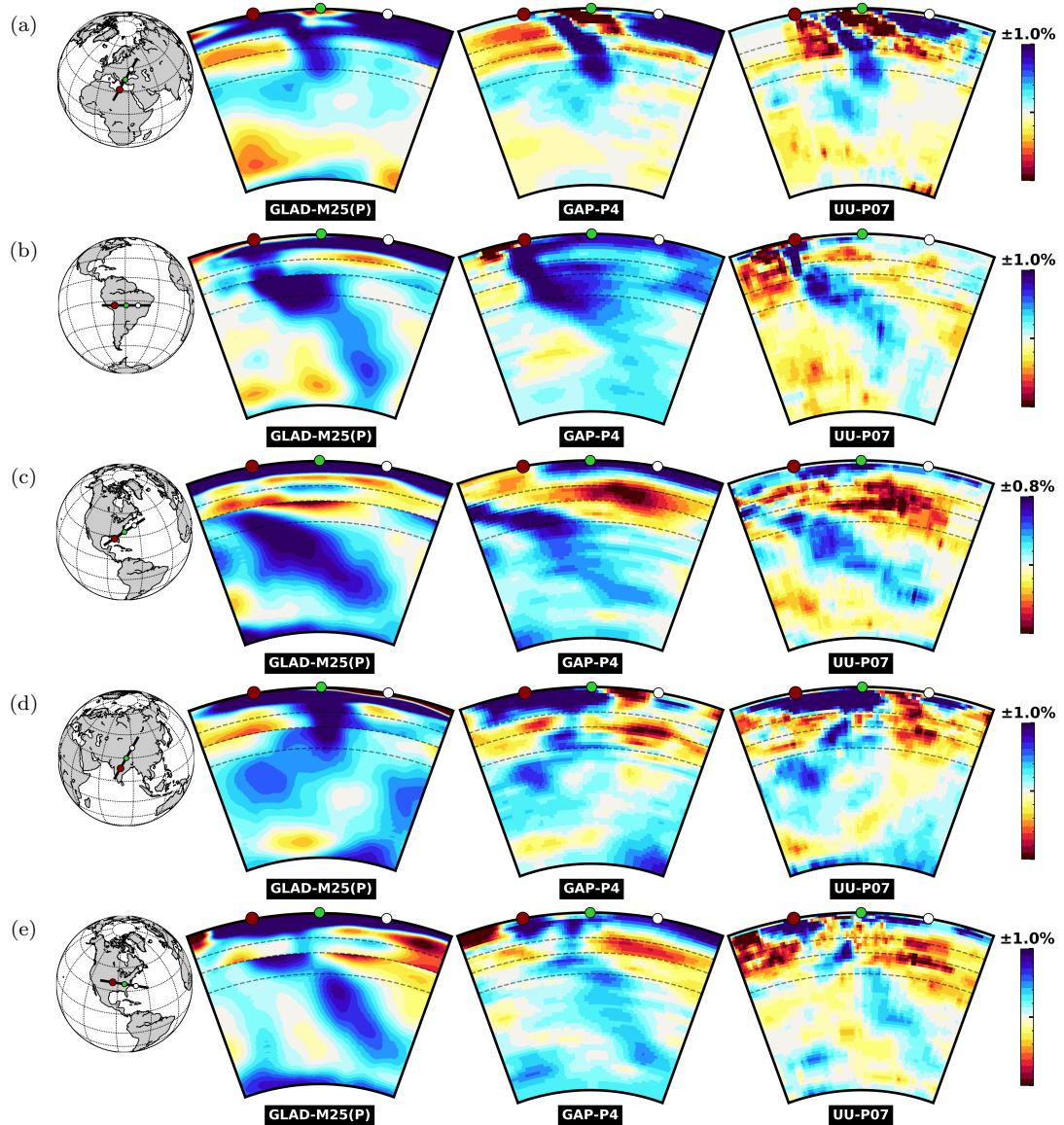


Figure 18: Vertical cross sections of perturbations of shear wave velocity  $V_s$  of plumes.

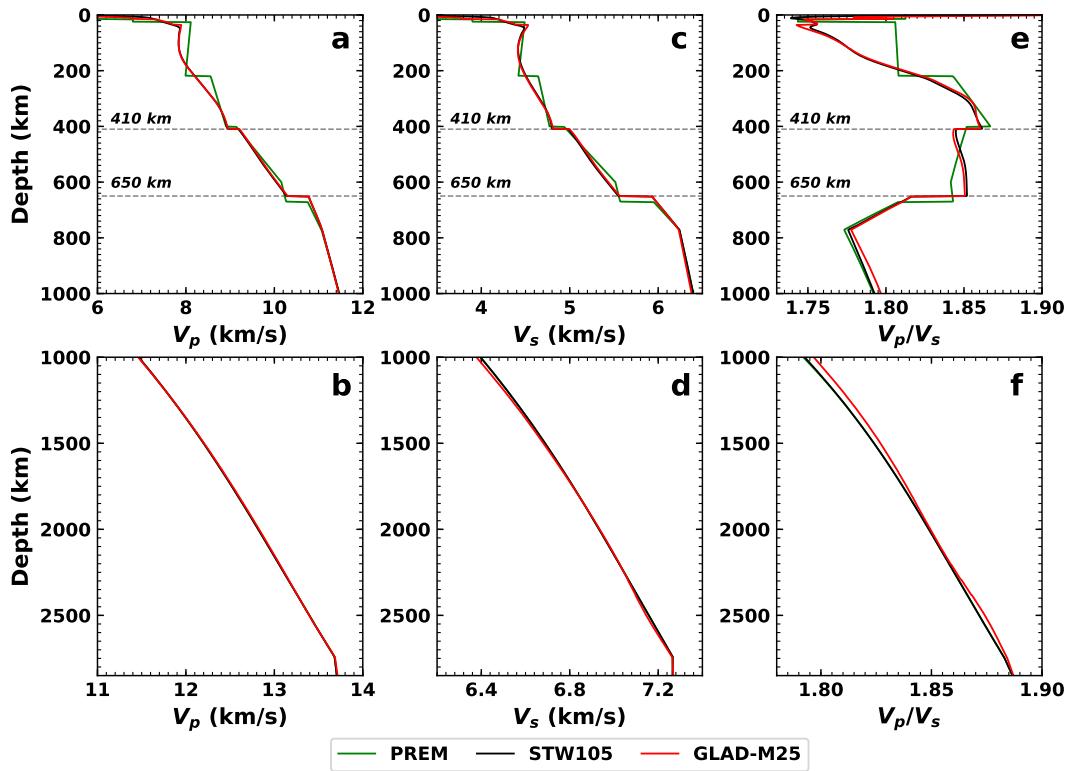


Figure A1: 1D radial shear and compressional wavespeed profiles for GLAD-M25, STW105, and PREM. The reference frequency for physical dispersion is ?? Hz.