Global adjoint tomography: first-generation model

Ebru Bozdağ, 1 Daniel Peter, 2 Matthieu Lefebvre, 3 Dimitri Komatitsch, 4 Jeroen Tromp, 3, 5 Judith Hill, 6 Norbert Podhorszki 6 and David Pugmire 6

1 Laboratory Géoazur, University of Nice Sophia Antipolis, F-06560 Valbonne, France. E-mail: bozdag@geoazur.unice.fr
2 Extreme Computing Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
3 Department of Geosciences, Princeton University, Princeton, NJ 08544, USA
4 LMA, CNRS UPR 7051, Aix-Marseille University, Centrale Marseille, F-13453 Marseille Cedex 13, France
5 Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA
6 Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Accepted 2016 September 21. Received 2016 September 19; in original form 2016 June 9

SUMMARY

We present the first-generation global tomographic model constructed based on adjoint tomography, an iterative full-waveform inversion technique. Synthetic seismograms were calculated using GPU-accelerated spectral-element simulations of global seismic wave propagation, accommodating effects due to 3-D anelastic crust & mantle structure, topography & bathymetry, the ocean load, ellipticity, rotation, and self-gravitation. Fréchet derivatives were calculated in 3-D anelastic models based on an adjoint-state method. The simulations were performed on the Cray XK7 named ‘Titan’, a computer with 18 688 GPU accelerators housed at Oak Ridge National Laboratory. The transversely isotropic global model is the result of 15 tomographic iterations, which systematically reduced differences between observed and simulated three-component seismograms. Our starting model combined 3-D mantle model S362ANI with 3-D crustal model Crust2.0. We simultaneously inverted for structure in the crust and mantle, thereby eliminating the need for widely used ‘crustal corrections’. We used data from 253 earthquakes in the magnitude range 5.8 ≤ Mw ≤ 7.0. We started inversions by combining ∼30 s body-wave data with ∼60 s surface-wave data. The shortest period of the surface waves was gradually decreased, and in the last three iterations we combined ∼17 s body waves with ∼45 s surface waves. We started using 180 min long seismograms after the 12th iteration and assimilated minor- and major-arc body and surface waves. The 15th iteration model features enhancements of well-known slabs, an enhanced image of the Samoa/Tahiti plume, as well as various other plumes and hotspots, such as Caroline, Galapagos, Yellowstone and Erebus. Furthermore, we see clear improvements in slab resolution along the Hellenic and Japan Arcs, as well as subduction along the East of Scotia Plate, which does not exist in the starting model. Point-spread function tests demonstrate that we are approaching the resolution of continental-scale studies in some areas, for example, underneath Yellowstone. This is a consequence of our multiscale smoothing strategy in which we define our smoothing operator as a function of the approximate Hessian kernel, thereby smoothing gradients less wherever we have good ray coverage, such as underneath North America.

Key words: Body waves; Surface waves and free oscillations; Seismic anisotropy; Seismic tomography; Computational seismology; Wave propagation; Waveform inversion.

1 INTRODUCTION

Since the inception of global seismic imaging (Aki et al. 1977; Dziewoński et al. 1977; Sengupta & Toksöz 1977), many models of the mantle have been published based on various types of data, such as body-wave arrival times (e.g. Dziewoński 1984; Bijwaard & Spakman 2000; Boschi & Dziewoński 2000; Zhou et al. 2006), surface-wave dispersion (e.g. Trampert & Woodhouse 1995; Ekström et al. 1997; Shapiro & Ritzwoller 2002; Trampert & Woodhouse 2003; Ekström 2011), shear and surface waveforms (e.g. Woodhouse & Dziewoński 1984; Li & Romanowicz 1996; Lebedev & van der Hilst 2008; Schaeffer & Lebedev 2013) and the Earth’s free oscillations (e.g. He & Tromp 1996; Koelman & et al. 2016). The steady increase in the number of worldwide seismographic stations combined with improvements in data quality have substantially grown the amount of usable data for the construction
of global Earth models, and, at long wavelengths, global shear-wave-speed models are now in general agreement (e.g. Ritzwoller & Lavelle 1995; Trampert & Woodhouse 2001; Becker & Boschi 2002). Several recent global studies have capitalized on this wealth of data (e.g. Ritsema et al. 1999; Mégnin & Romanowicz 2000; Gu et al. 2001; Houser et al. 2008; Kustowski et al. 2008; Ritsema et al. 2011; Schaeffer & Lebedev 2013; Chang et al. 2014), using a broad range of body-wave, surface-wave, and normal-mode observations. Ray-based (infinite-frequency) tomographic inversions have reached their theoretical limits (Wang & Dahlen 1995; Spetzler et al. 2001). Finite-frequency effects for surface waves were recognized much earlier (Woodhouse & Girmuz 1982; Snieder 1993) than for body waves (Marquering et al. 1999). Finite-frequency theory is now widely applied and has been used in global surface-wave (e.g. Zhou et al. 2006) and body-wave (e.g. Montelli et al. 2004) tomography. All these studies are based on tomographic methods rooted in perturbation theory of one form or another.

Current global inversions are severely limited by ‘crustal corrections’, which involve first-order corrections to accommodate the effects of Earth’s 3-D crust on seismic waves. The crust varies in thickness by an order of magnitude, from ~7 km below the oceans to ~70 km beneath the Andes and Tibet. The highly nonlinear effects of the crust on seismic wave propagation, even at long periods (Montagner & Jobert 1988), make crustal corrections questionable because they likely contaminate inferred mantle structure (e.g. Bozdag & Trampert 2008; Lekic et al. 2010; Ferreira et al. 2010).

Despite readily available vast amounts of data, the number of measurements used in classical tomography is limited to arrivals which are easily identified and isolated in seismograms. It is common to use traveltimes of major body-wave arrivals (e.g. P, PP, S, SS, ScS, etc.), Love & Rayleigh surface-wave dispersion measurements, or very long period free oscillations. Since different parts of a seismogram are sensitive to different parts of Earth’s interior, it is also common to integrate complementary data sets. One of the major challenges in global tomography is data coverage due to the uneven distribution of earthquakes and stations. Without permanent ocean-bottom seismographic instruments, it is difficult to change this distribution. However, extracting more information from seismograms will enhance global coverage, for example, by using more exotic—but often prominent—arrivals, such as FS, SP, PKKP and ScS reverberations. Ideally, complete three-component seismograms should be used in global inversions, without worrying about identifying which specific waveforms we are dealing with. Basically, any wiggle in a seismogram should make a suitable measurement, not just the ones we can readily identify with a known phase.

Recent advances in numerical methods combined with developments in high-performance computing have enabled unprecedented simulations of seismic wave propagation in realistic 3-D global Earth models (Komatitsch & Tromp 2002a,b; Capdeville et al. 2003; Chaljub et al. 2003; Chaljub & Valette 2004; Peter et al. 2011). In a complementary development, adjoint-state methods efficiently incorporate the full nonlinearity of 3-D wave propagation in iterative seismic inversions (Akcüelik et al. 2002, 2003; Tromp et al. 2005; Fichtner et al. 2006a,b; Tromp et al. 2008; Plessix 2009; Virieux & Operto 2009; Monteiller et al. 2015; Komatitsch et al. 2016).

‘Adjoint tomography’ provides new opportunities for improving images of Earth’s interior for the following reasons: (1) the full nonlinearity of 3-D seismic wave propagation is taken into account; (2) 3-D background models are used to compute Fréchet derivatives, thereby accommodating nonlinearities due to structure; (3) data may be assimilated based on automated measurement window-selection algorithms (Maggi et al. 2009; Lee & Chen 2013); (4) as a result of (1)–(3), the amount of usable data steadily increases from iteration to iteration, thus enabling the extraction of more information from seismograms, ultimately culminating in global ‘full-waveform inversion’ (FWI), that is, the use of entire three-component seismograms; and (5) the crust and mantle are inverted jointly, thereby eliminating the need for crustal corrections. The goal of this study is to harness 3-D simulations of seismic wave propagation in combination with adjoint-state methods to image the crust and mantle.

Although the basic theory of adjoint methods (Chavent 1974) for seismic inversions was introduced in the 80s (Bamberger et al. 1977; Laïli 1983; Tarantola 1984a,b; Gauthier et al. 1986; Tarantola 1988; Talagrand & Courtier 1987), their application has only recently become possible with the availability of 3-D wave propagation solvers and high-performance computing resources. Currently, there are successful applications of adjoint tomography both on regional and continental scales (Tape et al. 2009; Fichtner et al. 2009, 2013; Zhu et al. 2012, 2013; Zhu & Tromp 2013; Lee et al. 2014; Chen et al. 2015), however, so far it has remained a challenge in global tomography.

At the scale of the globe, the most advanced inversions to date combine 3-D spectral-element simulations of wave propagation in the mantle coupled with a normal-mode solution in a spherically symmetric core (Capdeville et al. 2003; Lekic & Romanowicz 2011; French et al. 2013; French & Romanowicz 2014). This compromise reduces the computational burden, but such coupled simulations do not accommodate Earth’s ellipticity and rotation. Additionally, meshing the Earth’s crust is avoided by replacing it with a smooth anisotropic spherical shell which mimics the behaviour of the actual crust, which is iteratively updated. Furthermore, Fréchet derivatives in the inverse problem are calculated based on the perturbation theory developed by Li & Romanowicz (1996). This hybrid approach has resulted in remarkable images of numerous mantle plumes (French & Romanowicz 2015). We note, however, that Valentine & Trampert (2016) recently reported that hybrid methods may be more error-prone than classical approximate methods. In this paper, no approximations—other than the use of a numerical method for simulating seismic wave propagation—are made in either forward or adjoint simulations and the entire globe is accommodated within a single framework, in which the crust, mantle, and core are all treated equally. A similar approach at the global scale has recently been demonstrated in a multiscale framework by Afanasiev et al. (2015), who performed two iterations with a smaller set of long-period data.

Success of the inversion strategy is closely tied to the choice of misfit function (e.g. Modrak & Tromp 2016). Common measures of misfit include cross-correlation traveltine measurements (e.g. Luo & Schuster 1991; Marquering et al. 1999; Dahlen et al. 2000; Zhao et al. 2000), multitaper phase measurements (e.g. Zhou et al. 2004), relative amplitude variations (e.g. Dahlen & Baig 2002; Ritsema et al. 2002), waveform differences (e.g. Tarantola 1984a,b, 1988; Nolet 1987), generalized seismological data functions (GSDF) (Gee & Jordan 1992), or more recently proposed time-frequency analysis (e.g. Kristekova et al. 2006; Fichtner et al. 2008) and instantaneous phase & envelope misfits (e.g. Bozdag et al. 2011; Rickers et al. 2012); the latter allow separation of phase and amplitude and use of long wave trains. In this study, we use frequency-dependent cross-correlation traveltimes—also called multitaper traveltine measurements—whenever we have dispersive signals, and classical cross-correlation traveltimes for non-dispersive body-wave arrivals. This facilitates an inversion for transversely isotropic lateral heterogeneity. In future studies, we will also
include cross-correlation and multitaper amplitude measurements, thereby enabling inversions that accommodate attenuation.

This paper is organized as follows. We begin by discussing the choice of the starting model, followed by a description of the data set. We then describe the inversion strategy and workflow in some detail, before discussing the first-generation global model based on adjoint tomography. We conclude by discussing our results in the broader context of the current status of global seismic tomography, and highlight a number of future research directions.

2 STARTING MODEL

It is well known that FWIs depend on the starting model. This issue has been addressed in many studies by selecting appropriate initial models and making suitable measurements to avoid getting stuck in a local minimum (e.g. Pratt & Shipp 1999; Brossier et al. 2009; Priex et al. 2013; Yuan & Simons 2014; Yuan et al. 2015). Alternatively, taking advantage of broad-band seismic signals in earthquake seismology, nonlinearities may be avoided by starting with smooth models and long-period signals, and gradually increasing the frequency content in successive iterations (e.g. Nolet et al. 1986; Zhu et al. 2012; Pageot et al. 2013). Unfortunately, a paucity of low-frequency data makes this strategy more difficult in exploration seismology. The 1-D radial structure of the Earth is quite well known, and there is also a basic consensus on the long-wavelength shear-wave-speed structure of the mantle (e.g. Ritzwoller & Lavelle 1995; Becker & Boschi 2002). Recent iterative inversions starting from radially symmetric models confirm said consensus (e.g. Lekić & Romanowicz 2011). Furthermore, reasonable global crustal models are now available, such as 2° × 2° Crust2.0 (Bassin et al. 2000) and its successor Crust1.0 (Laske et al. 2013) with 1° × 1° resolution.

For these reasons, we decided to use a starting model that combines 3-D mantle model S362ANI (Kustowski et al. 2008) with 3-D crustal model Crust2.0 (Bassin et al. 2000); we label it S362ANI+Crust2.0 in what follows.

S362ANI was constructed using surface-wave phase speeds, body-wave traveltimes, and long-period body and mantle waveforms. It has transverse isotropy in the upper mantle down to 420 km. Adding Crust2.0 (Bassin et al. 2000) on top of mantle model S362ANI (Kustowski et al. 2008) poses a challenge, because S362ANI is defined in a spherical shell with bounding radii determined by the core–mantle boundary (CMB) and the PREM Moho. Thus, S362ANI needs to be stretched (underneath the oceans) and squished (underneath the continents) to ‘glue’ it onto Crust2.0 (any other global model poses a similar ‘gluing’ challenge). This procedure affects surface wave speeds, and is another motivation for jointly inverting crust and mantle structure.

We have extensive experience with this starting model, which has been used for near-real-time global ShakeMovie simulations since 2010 (Tromp et al. 2010). There are currently more than 4700 earthquakes in the ShakeMovie database, providing 1-D (PREM) and 3-D (S362ANI+Crust2.0) synthetic seismograms for each event. This model already provides a decent fit to long-period body and surface waves (T > 60 s), and is a significant improvement over a 1-D model.

3 EARTHQUAKES AND SOURCE INVERSIONS

We selected waveform data for 253 earthquakes in the moment-magnitude range 5.8 ≤ Mw ≤ 7.0, as shown in Fig. 1(A). The events were chosen to provide broad geographical coverage, including shallow (depth ≤ 50 km), intermediate (50 km > depth > 300 km), and deep (depth ≥ 300 km) events. Because we used relatively long-period data (> 17 s) and events with magnitudes less than 7, we chose a Centroid Moment-Tensor (CMT) point-source earthquake representation. We used four Mw = 5.8 earthquakes from the East African Rift Valley and the Eastern US (the 2011 Virginia earthquake) to improve coverage, since higher-magnitude events are not observed in these regions. Initial CMT solutions were selected from the global CMT catalogue.

We inverted all source mechanisms in our 3-D starting model using the approach introduced by Liu et al. (2004). Source Fréchet derivatives and 100 min seismograms are calculated based on the spectral-element solver SPECFEM3D_GLOBE (Komatitsch & Tromp 2002a,b), and waveform measurements of body and surface waves are tailored to FLEXWIN (Maggi et al. 2009) window selections. We computed Green’s functions for nine source parameters (six moment-tensor components, depth, latitude, and longitude) in the starting model. When the structural model has changed significantly, this source inversion process may be repeated. Alternatively, source and structural parameters may be determined jointly in iterative adjoint inversions (e.g. Kim et al. 2011), but since the computational requirements are more-or-less the same, we preferred inverting for source and structural parameters separately.

The results of the source inversions are summarized in Figs 1(B)–(D). The scalar moment, Mw, typically changes by less than 30% per cent with an overall tendency for a reduction compared to the initial CMT solution. Hypocentres generally change by less than 10 km, with a typical shallowing of ridge events. These changes are most likely due to the inclusion of a 3-D crustal model in our source inversions, and are consistent with experiments conducted by Hjorleifsdóttir & Ekström (2010).

4 INVERSION STRATEGY AND WORKFLOW

This study is a first attempt at global FWI. The nomenclature FWI means different things in different areas of seismology. We define FWI as follows:

(i) Forward simulations and Fréchet derivatives are computed in fully 3-D models.
(ii) Anelasticity is fully accommodated in all numerical simulations.
(iii) Phase and amplitude information from three-component seismograms is assimilated.
(iv) Crust and mantle are updated simultaneously, thereby avoiding any ‘crustal corrections’.

With the exception of using amplitude information our global adjoint tomography may be considered global FWI. Although it is straightforward to include amplitude information in the inversion process, amplitude anomalies are affected by a host of factors and notoriously nonlinear, which is why we chose to initially focus on phase information. At a later stage, we plan to revisit amplitude anomalies and consider lateral variations in attenuation, as Zhu et al. (2013) did on a continental scale.

Additionally, rather than blindly assimilating complete seismograms, we use the window selection tool FLEXWIN (Maggi et al. 2009) to identify windows in which observed and simulated seismograms are sufficiently close to make a measurement, and to maximize information from our phase measurements, as discussed in...
Figure 1. Summary of source inversions for the 253 globally distributed earthquakes used in the structural inversions. Moment magnitudes vary between 5.8 and 7. (A) Focal mechanisms of the selected CMT events. Shallow (<50 km), intermediate depth (50–300 km) and deep (>300 km) events are shown by red, green and blue beach balls, respectively. (B) Focal mechanisms and relative change in scalar moment, \( \ln(\frac{M_{0\text{new}}}{M_{0\text{cmt}}}) \). The scalar moment changes generally less than 30 per cent, and tends to decrease. (C) Change in depth, \( \Delta \text{depth} = \text{depth}_{\text{new}} - \text{depth}_{\text{cmt}} \) (in kilometres). Shallow events tend to exhibit the largest depth changes, highlighting the influence of the 3-D crust on source parameters. (D) Change in epicentre, \( \Delta \text{loc} = \text{loc}_{\text{new}} - \text{loc}_{\text{cmt}} \) (in kilometres), which is generally less than 5 km.

detail in Section 4.3.1. As the inversion proceeds and the model improves the fit to the data, the number of windows grows, ultimately resulting in the assimilation of complete seismograms.

The inversion strategy in adjoint tomography and FWI is an active area of research. It involves choices with regards to the model (e.g. basis functions, model parametrization, etc.), the data (e.g. period bands, misfit measures, etc.), and the optimization strategy (e.g. regularization, optimization algorithm, etc.), all of which have a direct impact on the final model (e.g. Modrak & Tromp 2016). Once these choices have been made, adjoint inversions are described by a well-defined iterative workflow in which each step may be independently improved for better performance and resolution by adding new capabilities and options. The adjoint tomography workflow consists of four major stages: (1) forward simulations in the current model, (2) pre-processing and construction of adjoint sources, (3) gradient calculation in the current model, and (4) post-processing and model update (Fig. 2). The ultimate goal is to automate the entire workflow by reducing human interaction as much as possible (e.g. Lefebvre et al. 2014; Krischer et al. 2015a). This has been the approach in industrial FWI problems, where tens to hundreds of iterations are performed, which is possible partly due to relatively better data quality and ray coverage. Our global adjoint tomography workflow is complex and involves a significant number of steps. User interaction is error-prone, especially when performing repetitive tasks. In order to stabilize the entire process, we are currently experimenting with workflow management systems, such as Pegasus (pegasus.isi.edu) (Deelman et al. 2015) and RADICAL-Pilot (Merzyk et al. 2016).

In the following sections, we explain our workflow and FWI strategy in more detail.

4.1 Model basis functions and parametrization

In global tomography it is common to use spherical and cubic splines (e.g. Ritsema et al. 1999; Boschi & Ekström 2002; Lebedev et al. 2005; Kustowski et al. 2008; Ritsema et al. 2011), local cells (e.g. Zhou 1996; van der Hilst et al. 1997; Kennett et al. 1998) or triangular grid points (e.g. Zhou et al. 2006). We prefer to use the numerical integration points used in the spectral-element method, that is, the Gauss–Lobatto–Legendre (GLL) points, and smooth the model at a later stage, if need be, rather than projecting it on a smooth basis at the stage of the kernel calculation to minimize possible effects of parametrization on final models (e.g. Trampert & Snieder 1996).

We use a transversely isotropic model parametrization confined to the upper mantle, starting below the Moho. Transverse isotropy is described by five Love parameters, namely, \( A, C, L, N \) and \( F \) (Love 1927). By introducing the mass density, \( \rho \), transverse isotropy may alternatively be specified in terms of the speeds of vertically and horizontally polarized \( P \) waves, \( \alpha_v \) and \( \alpha_h \), the speeds of horizontally travelling and vertically or horizontally polarized \( S \) waves, \( \beta_v \) (or \( Vsv \)) and \( \beta_h \) (or \( Vsh \)), and the dimensionless parameter \( \eta \). To reduce the dependency of \( P \) and \( S \) wave-speed models on each other through the shear modulus, we use the bulk sound speed, \( c \), which depends on the bulk modulus, \( \kappa \). Thus, we are left with five parameters, namely, density, \( \rho \), bulk sound speed \( c = \sqrt{\kappa/\rho} \), vertically and...
horizontally polarized shear wave speeds \( \beta_v = \sqrt{L/\rho} \) and \( \beta_h = \sqrt{N/\rho} \), and the dimensionless parameter \( \eta = F/(A - 2L) \).

Density is generally difficult to constrain within the period range of this study. Therefore, to further simplify the model parametrization, we follow classical global tomographic studies and scale density to shear wave speed via the relation (Montagner & Anderson 1989)

\[
\delta \ln \rho = 0.33 \delta \ln \beta, \quad (1)
\]

where \( \beta \) is the Voigt average (Babuška & Cara 1991):

\[
\beta = \sqrt{\frac{2 \beta_v^2 + \beta_h^2}{3}}. \quad (2)
\]

This further reduces the number of unknown parameters from five to four, and the gradient of the misfit function, \( \delta \chi \), may be expressed as

\[
\delta \chi = \int_V K_c \delta \ln c + K_{\beta_v} \delta \ln \beta_v + K_{\beta_h} \delta \ln \beta_h + K_{\eta} \delta \ln \eta \, dV, \quad (3)
\]

where \( K_c, K_{\beta_v}, K_{\beta_h}, \) and \( K_{\eta} \) are the Fréchet derivatives with respect to the four dimensionless model parameters \( \delta \ln c, \delta \ln \beta_v, \delta \ln \beta_h \) and \( \delta \ln \eta \). Perturbations may be defined with respect to either 1-D or 3-D models. In our iterative inversion, perturbations are always with respect to the 3-D model from the previous iteration.

4.2 Numerical simulations

Today’s hybrid-architecture high-performance computing (HPC) systems employ graphics cards (GPUs—Graphics Processing Unit)
as hardware accelerators connected to the CPU (Central Processing Unit). We used the spectral-element solver SPECFEM3D_GLOBE (Komatitsch & Tromp 2002a,b) accelerated by graphics cards (Komatitsch et al. 2010; Komatitsch 2011) for all forward and adjoint simulations. The first twelve iterations were performed with a shortest period of ~27 s, and the following three iterations with a shortest period of ~17 s. Synthetic seismograms were calculated for the 253 earthquakes shown in Fig. 1(A) recorded by the stations shown in Fig. 3.

In the following, we describe how we combined observed and simulated data to update models.

4.2.1 Forward simulations

3-D forward simulations incorporate the effects of self-gravitation (in the Cowling approximation) (Cowling 1941), rotation, attenuation, ellipticity, the ocean load, and topography & bathymetry, as discussed in Komatitsch & Tromp (2002b). We currently use the 1-D $Q$ model from PREM (Dziewonski & Anderson 1981), which is fixed during the inversion. In the future, when we also assimilate amplitude measurements, we plan to attempt an inversion for lateral variations in attenuation.

For the first nine iterations, we calculated 100 min-long seismograms, containing minor-arc surface waves (G1 & R1) at all epicentral distances. In subsequent iterations, after incorporating the full effects of attenuation (Komatitsch et al. 2016) during the calculation of Fréchet derivatives (discussed in more detail later), we used 180 min long seismograms, containing full-orbit Love and Rayleigh surface waves as well as body waves. For topography/bathymetry, we used ETOPO4, a 4 min resolution model sub-sampled and smoothed from ETOPO1 (Amante & Eakins 2009).

4.2.2 Implementation of the crust

Earth’s highly heterogeneous crust has a strong influence on seismic waves in general and on surface waves in particular (Montagner & Jobert 1988), but may also significantly affect body-wave travel-times (Ritsema et al. 2009). Joint inversions for the crust and mantle are challenging, and ‘crustal corrections’ of one form or another are ubiquitous. Two commonly used approximations are: (1) crustal effects are smooth enough to be captured by first-order perturbation theory, and (2) Earth’s crust is assumed known and fixed in the inversion. Concerns about the former have been raised by Bozdağ & Trampert (2008) and Lekić et al. (2010), whereas Ferreira et al. (2010) showed that the latter biases inversions for mantle heterogeneity, for example, by introducing transverse isotropy or azimuthal anisotropy when there is none. Consequently, crustal corrections may strongly affect models of the mantle and core.

To accommodate the crust more accurately, Fichtner et al. (2009, 2013) prefer to fit a long-wavelength equivalent of the crustal signal and update the crust separately using a Backus-averaging technique (Backus 1962), and Lekić & Romanowicz (2011) and French & Romanowicz (2014) follow a similar approach (Capdeville & Marigo 2007). The goal of these efforts is to reduce the computational burden of accommodating the effects of the 3-D crust. Our preferred solution is to accept the complications induced by the crust and fully incorporate it in forward simulations and inversions. As described in Tromp et al. (2010), the Moho is honoured by the spectral-element mesh if the crust is less than 15 km thick (mainly oceanic crust) and thicker than 35 km (continental crust), and the Moho runs through mesh elements in ocean-continent transitions. This meshing strategy ensures accurate simulations of global surface-wave propagation.

4.2.3 Adjoint simulations: calculation of Fréchet derivatives

Using the adjoint method, Fréchet derivatives are computed based on two numerical simulations: a forward simulation initiated by a regular source, such as an earthquake, and recorded at a receiver, and an adjoint simulation initiated by placing a fictitious source at the location of a regular station and recorded at the location of the regular source (Tarantola 1984a; Tromp et al. 2005). Since the Green’s functions are the same in both numerical simulations, if one can simulate the regular forward wavefield, the adjoint wavefield can be simulated in the same fashion by simply changing the source term. The adjoint source term is directly dependent on the chosen misfit function (e.g. Tromp et al. 2005; Bozdağ et al. 2011), such that the resulting Fréchet derivative, or sensitivity kernel, reflects the measurement.

The biggest challenge in gradient calculations used to be taking into account full attenuation, because the time-reversed reconstruction of the forward wavefield during the convolution with the adjoint wavefield is numerically unstable in the presence of dissipation, as described in detail in Liu & Tromp (2006). Based on a comparison with normal-mode calculations, Zhou et al. (2011) showed that for body waves and long-period surface waves physical dispersion is the most important aspect of attenuation for kernel construction, and this effect can be readily accommodated. This was our strategy for the first eight iterations, up to which point we only assimilated minor-arc surface waves with periods longer than 50 s. Indeed, this is a valid approximation at long periods and short epicentral distances, which may safely be used in continental- and regional-scale studies (e.g. Tape et al. 2009; Zhu et al. 2012; Chen et al. 2015). However, at the global scale, especially with the use of major-arc waves at longer epicentral distances and shorter periods ($T < 50$ s), the approximation may no longer be valid. After the stable implementation of full attenuation in adjoint simulations (Komatitsch et al. 2016), we switched to exact anelastic kernel calculations after the ninth iteration, and we immediately observed a major benefit for the Love-wave misfit reduction.

In an independent theoretical study, Valentine & Trampert (2016) reported that combining exact wave simulations with approximate kernels may generate larger errors in imaging than a fully asymptotic or approximate approach in both forward and kernel computations. Based on our observations, as we go down to shorter periods, any approximations in wave and kernel simulations should be avoided.
4.3 Pre-processing

We selected data for the 253 earthquakes shown in Fig. 1(A) from the Global Seismographic Network (GSN) and several local continental arrays, such as USArray, and European, Japanese, and Australian networks (Fig. 3). Data are freely available from data centres operated by IRIS (USA) and ORFEUS (Europe). The pre-processing phase of the adjoint tomography workflow involves data culling, time-series analysis, window selection, making measurements, and adjoint source construction (Fig. 2).

4.3.1 Measurement strategy

To avoid nonlinearities, which may occur in FWIs, we used only phase information—targeting elastic structure in the first-generation model—and defined appropriate period bands for measurements at each iteration.

All measurements were made on three-component (vertical, radial, transverse) seismograms, assimilating both body and surface waves. In relatively short time windows, for example, for body waves, we make cross-correlation traveltime measurements, and in sufficiently long time windows, for example, for surface waves, we make frequency-dependent (multitaper) traveltime measurements. The measurements are divided into a number of categories. For example, for our four final period bands on three components we have twelve measurement categories (Fig. 4). The frequency-dependent traveltime misfit in category $c$ may be expressed as

$$
\chi_c = \frac{1}{N_c} \sum_{w=1}^{E} \sum_{i=1}^{N_c} \int w_i(\omega) \left( \frac{\Delta t_c(\omega)}{\sigma_i} \right)^2 d\omega / \int w_i(\omega) d\omega,
$$

where $\Delta t_c$ denotes the traveltime anomaly in frequency window $w_i$, $\sigma_i$ the associated standard deviation, $N_c$ the number of measurements in category $c$ for earthquake $e$, $E$ the total number of earthquakes, and $N_c = \sum_{w=1}^{E} N_c^w$ the total number of measurements in category $c$. If the time window is too short to make a multitaper measurement, we use a cross-correlation measurement instead. The total misfit in all $C$ categories is

$$
\chi_{\text{total}} = \frac{1}{C} \sum_{c=1}^{C} \chi_c. \quad (5)
$$

We selected our period bands as follows:

(i) 1st to 5th iteration: We initiated iterations with 100 min-long seismograms with $\sim27$ s resolution, using two period bands, namely, 30–60 s for body waves and 60–120 s for surface waves and long-period body waves. Our strategy was to decrease the lower corner of the surface-wave pass band gradually, as the overall misfit improved.

(ii) 6th to 8th iteration: We added a 96–250 s long-period surface-wave band. We adjusted the other two period bands to 30–66 s and 56–110 s, respectively, with $\sim10$ per cent overlap between bands.

(iii) 9th to 11th iteration: We incorporated full attenuation in the frequency window $\omega$, and long-period body waves. The period bands were adjusted to 30–59 s, 50–106 s and 90–250 s, respectively.

(iv) 12th to 15th iteration: We increased the resolution of our simulations by interpolating and resampling our 11th-iteration model from 160 surface elements along each side of the cubed sphere (Komatitsch & Tromp 2002a) to 256 surface elements, thereby reducing the shortest period from $\sim27$ s to $\sim17$ s. This allowed us to add one more shorter-period body-wave measurement category. Thus, we performed the last four iterations with four period bands, namely, 17–38 s for shorter-period body waves, 30–56 s for intermediate-period body waves, 45–110 s for surface waves & long-period body waves, and 92–250 s for long-period surface waves. Note that we used any selected phase in the 45–110 s period band, including minor-and major-arc surface and body waves, whereas we used only body waves in the 17–38 s and 30–56 s period bands, and only surface waves in the 92–250 s period band.

In Fig. 4, our last four period bands together with FLEXWIN window selections are illustrated for a path across the Indian Ocean. We initiated our inversion with about $\sim1.2$ million measurements, and gradually increased this number to $\sim2.6$ million after the 9th iteration, culminating in the assimilation of more than 3.8 million measurements during the last four iterations.

4.3.2 Challenges of data pre-processing on large HPC systems

While several different groups have their own data formats, Seismic Analysis Code (SAC; Helffrich et al. 2013) has been the standard data format in earthquake seismology. However, handling data in SAC format during pre-processing involves millions of files, and the related I/O traffic can cripple the file system. This is undesirable on high-performance clusters, and highlights the need for a new seismic data format which satisfies the needs of modern seismology.

For this reason, a new Adaptable Seismic Data Format (ASDF) is being developed (Krischer et al. 2016). ASDF is based on HDF5 and combines all seismic traces for an event in a single file. Thus, one needs only two files per event, one for observed data and one for synthetic data. Additionally and importantly, ASDF enables users to keep track of data provenance, which is stored with the data in the same container. We are in the process of migrating the entire pre-processing phase to a Python-based workflow which seamlessly integrates ASDF with ObsPy (Krischer et al. 2015b), a Python framework for processing seismological data. As part of this migration, Python versions of FLEXWIN (pyflex) and the measurement code (pyadj) are being developed.

4.4 Post-processing

Once the gradient calculations for all earthquakes are completed, the adjoint tomography workflow continues with a post-processing phase leading to a model update (Fig. 2). The post-processing phase uses the Adaptable I/O System (ADIOS) (Liu et al. 2014) developed by Oak Ridge National Laboratory for fast parallel I/O, which also greatly reduces the number of files. The post-processing steps leading to the model update are summarized in the next sections.

4.4.1 Summation of event kernels

Adjoint simulations result in event kernels for each earthquake, which are summed to obtain the full gradient of the misfit function. This summation is performed at the GLL level.

4.4.2 Smoothing the gradient

Smoothing serves the same purpose as damping in classical tomography and is applied for the following reasons: (1) The gradient is a result of numerical simulations and should be smoothed to reflect the numerical resolution. (2) Smoothing should be applied to balance imperfect ray coverage, which is an issue for global
Figure 4. Sample window selections by FLEXWIN (Maggi et al. 2009; blue windows) showing the period bands used during the last three iterations. Shown are vertical, radial, and transverse component records of observed (black) and synthetic (red) seismograms of the 2010 September 3 New Zealand earthquake ($M_w = 7$, depth = 12 km) recorded at station KBL in Kabul, Afghanistan.

4.4.3 Pre-conditioning

Following Luo et al. (2013), we used a pre-conditioner based on the interaction between the forward and adjoint accelerations, namely the pseudo-Hessian

$$P(x) = \sum_{e=1}^{E} \int \partial^2_s s(x, t) \cdot \partial^2_s \tilde{s}(x, T - t) \, dt.$$  (6)

Here $s$ and $\tilde{s}$ denote the forward and adjoint displacements, respectively, and $E$ denotes the number of earthquakes. This pre-conditioner corresponds to the diagonal terms of the Hessian. These diagonal terms mimic ray (kernel) coverage, and thus this pre-conditioner not only suppresses high amplitudes around sources and receivers, but also balances imperfect coverage.

4.4.4 Optimization

We performed all iterations based on a conjugate-gradient method (Fletcher & Reeves 1964). Following Tromp et al. (2005), Tape et al. (2010) and Zhu et al. (2012), we determined the search direction via

$$d_i = -g_i + \beta d_{i-1},$$  (7)

where $g$ and $d$ are the gradient and search direction from the current and previous iterations, respectively, and $\beta$ is given by

$$\beta = \frac{g_i^T (g_i - g_{i-1})}{g_{i-1}^T g_i}.$$  (8)

Although some studies show that conjugate gradient and quasi-Newton methods give similar convergence rates during the first few iterations (e.g. Luo et al. 2013), we are planning to switch to the L-BFGS method (Nocedal 1980) in future iterations, which may help with imperfect ray (kernel) coverage.
Figure 5. Pseudo-Hessian kernel defined by eq. (6) calculated based on the measurements for the final model to illustrate global ray (kernel) coverage. (A) Northern Hemisphere, (B) Southern Hemisphere. Minimum and maximum values denote areas with poor and good coverage, respectively. The pseudo-Hessian is used to determine the amount of smoothing of the gradient, as well as a pre-conditioner.

4.4.5 Determining the step length

Once we establish the search direction, we use a line search to determine the step length for the model update, as described in Tape et al. (2007). Following Zhu et al. (2015), we run forward simulations for a subset of 24 earthquakes for various step lengths. In global inversions, we generally use 0.5–2 per cent perturbations in the search direction. The challenge is to find a step length that satisfies all measurement categories described in Section 4.3.1. Once the step length is determined, the model parameters $m$ may be updated via

$$
\ln \frac{m_{i+1}}{m_i} = \alpha d_i,
$$

where $\alpha$ and $d_i$ are the step length and the search direction from the $i$th iteration, respectively.

4.5 Computational requirements

All numerical simulations were performed in parallel with the spectral-element seismic wave propagation solver SPECFEM3D_GLOBE. The computational cost is independent of the number of seismic stations and scales linearly with the number of earthquakes. The computational requirements are summarized in Table 1. We observed longer simulation times during adjoint calculations due to SAC I/O traffic. We expect better performance with ASDF, which is designed to reduce I/O.

<table>
<thead>
<tr>
<th>Source inversions</th>
<th>1 event</th>
<th>253 events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural Inversions–I</td>
<td>1 event</td>
<td>1 iteration (253 events)</td>
</tr>
<tr>
<td>CPU-h, $T_{\text{min}} \sim 27$ s</td>
<td>$\sim 750$ h (forward)</td>
<td>$\sim 760 000$ h</td>
</tr>
<tr>
<td>100 min seismograms</td>
<td>$\sim 2250$ h (adjoint)</td>
<td></td>
</tr>
<tr>
<td>Structural Inversions–II</td>
<td>1 event</td>
<td>1 iteration (253 events)</td>
</tr>
<tr>
<td>GPU-h, $T_{\text{min}} \sim 27$ s</td>
<td>$\sim 12.5$ h (forward)</td>
<td>$\sim 12 650$ h</td>
</tr>
<tr>
<td>100 min seismograms</td>
<td>$\sim 38$ h (adjoint)</td>
<td></td>
</tr>
<tr>
<td>Structural Inversions–III</td>
<td>1 event</td>
<td>1 iteration (253 events)</td>
</tr>
<tr>
<td>GPU-h, $T_{\text{min}} \sim 17$ s</td>
<td>$\sim 22.5$ h (forward)</td>
<td>$\sim 52 600$ h</td>
</tr>
<tr>
<td>180 min seismograms</td>
<td>$\sim 60$ h (adjoint)</td>
<td></td>
</tr>
</tbody>
</table>
In this section, we present the ‘first generation’ global adjoint tomography model GLAD-M15 (GLobal ADjoint tomography-Model iteration 15), which is the result of 15 tomographic iterations.

### 5.1 Misfit reduction

Fig. 6 summarizes the misfit reduction. The inversion seeks to minimize the total misfit, given by eq. (5), obtained by summing the misfits in each of the sub-categories, given by eq. (4). Thus, we expect the total misfit to be steadily reduced, even though the misfits in each subcategory may not be. Note, however, that the misfit function is a continually moving target, because we seek to increase the number of measurements and gradually broaden the frequency content as the iterations progress. Consequently, when new categories are introduced, the new misfit values are sometimes slightly higher than they were in the previous iteration. The overall misfit reductions in all categories indicate that our gradient is well balanced.

We incorporated the longest-period surface waves (∼90–250 s) and shortest-period body waves (∼17–38 s) during the 6th and 12th iterations, respectively. Slight jumps in misfits are observed at the 6th, 9th and 12th iterations due to changes in the number of windows and period bands. The overall misfit reduction is smooth and gradual, and flattens towards the 15th iteration, which is an indication of convergence with the current data set within data errors. Note that up to the 9th iteration, body- and surface-wave misfits on the transverse component (Fig. 6, third row, third column) decreased significantly slower than on the other components. This signals the introduction of full attenuation in adjoint simulations, as described in Komatitsch et al. (2016).

### 5.2 Traveltime histograms

In Fig. 7, we show multitaper (for dispersive waveforms) and cross-correlation (for non-dispersive waveforms) traveltime anomaly histograms for the final four measurement categories on all three components for starting model S362AN1+Crust2.0 (M00) and final model GLAD-M15 (M15). Note how, unlike the M00 histograms, the M15 histograms are nicely peaked and centred on zero and more Gaussian in shape in all 12 misfit categories.
Figure 7. Multitaper (dispersive waves) and cross-correlation (non-dispersive waves) traveltime histograms for the starting model S362ANI+Crust2.0 (Kustowski et al. 2008; Bassin et al. 2000) (M00) and the 15th iteration model GLAD-M15 (M15) in the 12 measurement categories used during the last four iterations. The numbers in the top-right of each plot denote the number of measurements in each category. The total number of measurements exceeds 3.8 million.

5.3 Map views

In GLAD-M15, we observe well-known plumes, hotspots, and slabs emerging from smooth starting model S362ANI+Crust2.0, particularly in regions with good ray coverage. Figs 8 and 9 show map views centred on the Pacific and Africa at 250 km depth. Major hotspots and plumes, such as Tahiti, Caroline, Hawaii, Bermuda and Kerguelen, are nicely resolved, as are slabs in the Aleutians, Scotia Arc, Hellenic Arc and Tonga, and collision zones, such as the Himalayas. The changes in our model are non-uniform due to our multisMOOTHing strategy, in which we smooth areas with good coverage less to allow the introduction of smaller-scale features, whereas we smooth areas with relatively poor coverage, such as Africa or the Southern Hemisphere, more. The most pronounced changes occur in the upper mantle, where we have the densest ray coverage (Fig. 5). GLAD-M15 naturally resembles S362ANI at long wavelengths, and remains close to it in areas of poorer coverage.

To better depict differences between our final and starting models, we plot vertically polarized shear-wave-speed perturbations in GLAD-M15 with respect to S362ANI+Crust2.0 in Fig. 10. The major absolute changes (>2 per cent) are in the upper mantle, particularly beneath North America and Europe, thanks to dense seismic networks. Perturbations gradually diminish with depth due to reduced data coverage and our multiscale smoothing strategy. Near the CMB, the absolute changes are within ~0.5 per cent, and the largest perturbations are observed beneath the Pacific. These perturbations are generally larger than in model S362ANI+M (Moulik & Ekström 2014) —a recent updated version of S362ANI with a larger data set that includes normal-mode splitting functions— except near the CMB beneath the Pacific. GLAD-M15 also introduces more localized and higher-resolution features, for example, in subduction zones.

5.4 Notable features: plumes, hotspots and slabs

In this section, we present some of the plume, hotspot, and slab features in GLAD-M15. In Fig. 11, three vertical cross-sections are shown, one along the equator and two along meridians. We observe enhancements of Pacific plumes, hotspots, and subduction zones. We also see enhancement of the African plume, as well as the Caroline and Galapagos hotspots in the Pacific. As shown in the bottom row of Fig. 11, the Pacific plume is enhanced near the CMB. Changes underneath Africa are less dramatic than underneath the Pacific due to poorer sampling. We also observe subducted plates and their remnants in the lower mantle, for example, underneath Asia.

One of the most striking features in GLAD-M15 is the Tahiti plume, as shown in Figs 12(A) and (B). The plume originates at the CMB, gets flattened around 1000 km, which may be associated with a viscosity change (e.g. Rudolph et al. 2016), and bends towards Tonga, likely interacting with the slab along the trench (e.g. Chang et al. 2016). The Tahiti and Samoa plumes appear to originate from one superplume in the lower mantle, and their continuation in the upper mantle is most pronounced in Vp/Vs ratios. We see a similar enhancement of the Caroline plume, which also flattens at around 1000 km, as supported by Vp/Vs ratios.

In the horizontal sections shown in Fig. 13, most of the North American low-wave-speed zones appear in GLAD-M15, such as Yellowstone, Raton and Anahim, as well as Bowie and Cobb.
Figure 8. Map views of vertically polarized shear-wave-speed perturbations in starting mantle model S362ANI (left) and GLAD-M15 at 250 km depth. Notable slabs and plumes/hotspots enhanced in GLAD-M15 are marked. Each model is shown with respect to its own mean. Plate boundaries are from Bird (2003).

Yellowstone is currently debated in terms of its size, depth extent and resolution (e.g. Smith & Braile 1994; Pierce & Morgan 2009; Faccenna et al. 2010; Fouch 2012). Thus, it is exciting to observe such a local upper mantle feature in a global tomographic model in an area where we have some of the best ray coverage. Furthermore, the slab along the Aleutians has become clearly visible, both in map view and in vertical cross-sections. Yellowstone, Raton, Anahim, and Bowie extend down to the 660-km discontinuity, as best illustrated in Vp/Vs ratios. Transverse isotropy (TI) underneath Yellowstone and Raton is mainly showing Vsh > Vsv, which is consistent with an interpretation in terms of predominantly horizontal flow in a plume head. Although the resolution of TI may not be perfect, particularly at this scale, we report a clear slab signature in the TI plots with persistent Vsv > Vsh all around the globe, consistent with predominantly vertical flow (e.g. Montagner 1998).

In the lithosphere and asthenosphere, Vsh is typically larger than...
Figure 10. Vertically polarized shear-wave-speed perturbations in GLAD-M15 with respect to S362ANI+Crust2.0, highlighting differences between the 15th iteration model (M15) and the 3-D starting model (M00) ($\ln(M15_{Vsv}/M00_{Vsv})$). Note the changing colour scales, as indicated. Note also that in the rest of this article all shear-wave-speed perturbations are plotted with respect to their own mean.

\[ \ln(M15_{Vsv}/M00_{Vsv}) \] 

Vsv, consistent with flow/strain-induced horizontal alignment of the olivine fast axis. Subduction of the lithosphere gradually tilts this picture, resulting in Vsv being larger than Vsh in steeply subducting slabs (e.g. Song & Kawakatsu 2012).

In Fig. 14 we consider Antarctica, with a focus on the Erebus hotspot, a well-known active Antarctic volcano. As previously mentioned, resolution in this part of the globe is challenging due to a paucity of data. Despite this, we clearly observe an enhanced image of the Erebus hotspot, illustrating the power of the methods and tools that we are currently using for imaging. With the help of temporary Antarctic seismic networks (see Fig. 3), we observe thickening of the low-wave-speed structure underneath Erebus, which goes down to about 1200 km, as supported by Vp/Vs ratios and transverse isotropy characterized by Vsh $>$ Vsv.

Subduction zones are distinctly enhanced in GLAD-M15, for example, in Japan, Izu-Bonin, Marianna, Indonesia and the Aleutians. We also resolve slabs that do not exist in the starting model, such as the Hellenic and Scotia Arcs (Figs 15 and 16). We clearly observe a slab signature in Vp/Vs ratios, with relatively low values, and in Vsv/Vsh ratios, showing significant transverse isotropy with faster Vsv speeds all around the globe. We see a continuation of the Hellenic slab below the 660 km discontinuity, in agreement with previous studies (e.g. Spakman et al. 1993; Zhu et al. 2012). The Scotia Arc is another challenging location for imaging due to poor
ray coverage. Li et al. (2008) obtained a $P$-wave slab signature down to $\sim660$ km, and it has been argued that the slab likely does not penetrate into the lower mantle (e.g. Loiselet et al. 2010). Our images of the Scotia arc are in overall agreement with Li et al. (2008), but we observe stronger perturbations and likely penetration into the lower mantle. Despite being a young slab, lower-mantle penetration is tectonically possible considering its age and the current 69–78 mm yr$^{-1}$ subduction rate (Thomas et al. 2003).
Comparisons of transverse isotropy and Vp/Vs ratios between GLAD-M15 and S362ANI+Crust2.0 at several depths are shown in Fig. 17. Our large-scale transversely isotropic perturbations in the upper mantle are in overall agreement with model S362ANI+M (Moulik & Ekström 2014) which is an updated version of our starting model S362ANI. However, our perturbations diminish more rapidly below ∼250 km, which is more consistent with Panning et al. (2010) and Chang et al. (2014). GLAD-M15 exhibits more localized anomalies around slabs and plumes, and contains features consistent with Chang et al. (2016) in the upper mantle beneath the Samoa-Tonga region, which may indicate a slab-plume interaction. Similarly, our Vp/Vs ratios are also in agreement with values determined by Moulik & Ekström (2016), but again reveal sharper anomalies around slabs and plumes.

5.5 Resolution tests

It is common to use checkerboard tests to estimate resolution in tomographic studies, but this is computationally unfeasible for 3-D FWI, particularly on a global scale. Such tests would require the same number of iterations—and hence the same computational resources—as the actual inversion. To ameliorate this problem, Fichtner & Trampert (2011) introduced the ‘point-spread function’ (PSF) test. To perform such a test, a finite-difference approximation is used to calculate the action of the Hessian on a localized model perturbation:

\[ H \cdot \delta m \approx g(m + \delta m) - g(m), \]
Figure 13. Map views at 250 km depth (top row) and vertical cross-sections (middle row) of vertically polarized shear-wave-speed perturbations in the starting model S362ANI+Crust2.0 and GLAD-M15 underneath North America. Also shown are Vp/Vs ratios (bottom row, left) and transverse isotropy (bottom row, right). Map views in the middle and third row (left) denote the CMB, and the map view in the bottom right panel denotes the 660 km discontinuity, below which transverse isotropy vanishes. Each model is plotted relative to its own radial mean.

where $H$ denotes the Hessian and $\delta m$ refers to a localized model perturbation with respect to the current model $m$. The misfit gradient $g$ is evaluated for both models $m$ and $m + \delta m$. Based on the action of the Hessian on the model perturbation, $H \cdot \delta m$, one is able to assess the curvature of the misfit function at a particular ‘point’ in the model space, reflecting the degree of ‘blurring’ of that point. Since we have to calculate the misfit gradient $g(m + \delta m)$ for the perturbed model and we already have the gradient $g(m)$ for the current model, the computational requirements for a single spot analysis are the same as for one full iteration. Recently, a stochastic extension to this approach has been proposed by Fichtner & van Leeuwen (2015) based on random probing of the Hessian and of the model parameters. In this approach the resolution length of each parameter of interest may be obtained with roughly 5 iterations. We intend to consider such tests in the future.

We selected two specific locations for PSF tests, namely, Yellowstone and Erebus. We perturbed the 14th-iteration model by a spherical Gaussian with a size close to the hotspot of interest, and
Figure 14. Map views at 450 km depth (top row) and vertical cross-sections (middle row) of vertically polarized shear-wave-speed perturbations in the starting model S362ANI+Crust2.0 and GLAD-M15 under Antarctica. Also shown are Vp/Vs ratios (bottom row, left) and transverse isotropy (bottom row, right). Map views in the middle and third row (left) denote the CMB, and the map view in the bottom right panel denotes the 660 km discontinuity, below which transverse isotropy vanishes. Each model is plotted relative to its own radial mean.
computed the difference between the gradients of the perturbed and unperturbed 14th-iteration models, thereby giving the action of the Hessian on the model parameters according to eqn. (10).

In Figs 18 and 19 we show the results for vertically polarized shear-wave-speed perturbations centred on Yellowston and Erebus, respectively. The Gaussians are reasonably well retrieved without much bias or smearing in the upper mantle, which supports the resolution of the observed features. Furthermore, trade off with other model parameters mainly occurs as random noise near the surface and does not generate a significant anomaly at the location of perturbation.

5.6 Independent earthquake database

Following an approach used by Tape et al. (2009, 2010) and Chen et al. (2015), we further investigated the quality of our model with an independent database of 40 randomly selected $6.5 \leq M_w \leq 7.0$ earthquakes, shown in Fig. 20. We chose slightly larger events because these generate more measurements for analysis. An earthquake not used in the tomographic inversion may be used to independently assess the misfit reduction from M00 to M15. In Fig. 21, we show multitaper (for dispersive waveforms) and cross-correlation (for non-dispersive waveforms) traveltime anomaly histograms for the final four measurement categories on all three components for starting model S362ANI+Crust2.0 (M00) and final model GLAD-M15 (M15). Like the histograms for the data used in the actual inversion (shown in Fig. 7), these histograms show a clear reduction in the traveltime anomalies in M15 compared to M00 in the form of more sharply centred distributions in all 12 categories. This result provides validation for our global model and suggests that future earthquakes will see similar misfit reductions.

5.7 Comparisons with S40RTS

It is well known that global models differ significantly from each other at smaller scales. Detailed model comparisons may be found in numerous studies (e.g. Schaeffer & Lebedev 2013; Chang et al. 2014; French & Romanowicz 2015). Here, we present a
comparison with S40RTS (Ritsema et al. 2011), a recent degree-40 global model. In Figs 22 and 23, we show map views at various depths of our model together with starting model S362ANI (Kustowski et al. 2008) and S40RTS (Ritsema et al. 2011). We observe that our 15th-iteration model generally takes the common ground between S362ANI and S40RTS.

6 DISCUSSION

GLAD-M15 is the first global model based on fully 3-D forward and adjoint simulations of seismic wave propagation since the inception of ‘FW’ by Tarantola (1984b). It naturally unifies the crust and mantle by inverting them jointly, using anything and everything in three-component seismograms that passes automated misfit and data-quality selection criteria. Many global models use bigger data sets in terms of the number of earthquakes (e.g. Schaeffer & Lebedev 2013), or merge various complementary secondary data types, such as phase and group wave speeds, traveltimes, and splitting functions, sometimes even including isolated waveforms (e.g. Ritsema et al. 2011; Chang et al. 2014). Our study demonstrates what is feasible with a limited data set of 253 earthquakes and just 15 tomographic iterations. Imagine what more can be done with the thousands of suitable
Figure 17. Map views of transverse isotropy and Vp/Vs ratios at various depths in GLAD-M15 and starting mantle model S362ANI. Each model is shown with respect to its own mean.

earthquakes that have already been recorded by worldwide seismographic networks!

Granted, our approach is currently computationally expensive. However, we are at a stage where such expenses are justified, even necessary. The significance of using full-attenuation in adjoint kernel simulations serves as a case in point: approximate kernels based on physical-dispersion only are inadequate for full-orbit surface waves. If the goal is to assimilate anything and everything,
Figure 18. 3-D contour plot of a point-spread function to assess resolution at Yellowstone which is cut through to view the inside of the anomaly. The 14th iteration $\beta_v$ model was perturbed by a 2 per cent spherical Gaussian located at 125 km depth with a radius of 250 km. Grey spheres denote the size of the spherical Gaussian, and a vertical section is taken on the contour plot to show the values on the inside. $\beta_h$ and $c$ plots in the bottom row show trade-offs with these model parameters. Map views denote 660 km discontinuity.

Figure 19. Same as Fig. 18, but for Erebus. The 14th iteration $\beta_v$ model was perturbed by a 2 per cent spherical Gaussian located at 300 km depth with a radius of 300 km.
Figure 20. Collection of 40 independent global earthquakes (6.5 ≤ Mw ≤ 7.0) used to assess traveltime misfit in model GLAD-M15. These events were not used in the actual structural inversion.

synthetic seismograms must be calculated as accurately as possible to avoid errors in the forward theory from contaminating the model. And the advantage of adjoint-state methods is that they end up solving the fully 3-D nonlinear inverse problem, albeit iteratively and therefore not cheaply.

The impact of the starting model on FWI is well recognized. Since most global models are in agreement at long wavelengths (e.g. Ritzwoller & Lavely 1995; Becker & Boschi 2002), we chose to start with such a model rather than a spherically symmetric model. Broadly speaking, our iterations only modify the starting model where such modifications are warranted by the data, as expressed in the Fréchet derivatives. It is for this reason that we see much more detailed structural variations underneath North America and Europe in GLAD-M15, and the resolution of Erebus clearly benefited from temporary array deployments in Antarctica.

Despite the power of our approach, it remains a challenge to fit every wiggle in 180 min broad-band teleseismic seismograms both in phase and amplitude. More ocean-bottom seismometers or recently proposed floating acoustic sensors (e.g. Simons et al. 2009; Sukhovich et al. 2015) would of course help in terms of global coverage. Moreover there is still scope for improving imbalanced coverage and reducing uncertainties based on new measurement strategies (e.g. Choi & Alkhalifah 2012; Yuan et al. 2016). But the most natural way forward is to use all available data from all earthquakes in the global CMT catalogue. That data is readily available, and we should be using it all. In theory, there is no impediment to assimilating all suitable data in global adjoint tomography. In practice, we need robust workflows and modern data formats to make this possible, in addition to substantial computational resources. Workflow management and stabilization is an active area of research in computational science in general and computational seismology in particular (Lefebvre et al. 2014; Krischer et al. 2015a).

We currently take advantage of GPU computing by having access to more than 18K graphics cards on the Oak Ridge Leadership Computing Facility (OLCF) Cray ‘Titan’, a machine with a peak performance of more than 20 petaflops. Exascale computers are expected to become available in the 2020–2022 time frame, and we

Figure 21. Same as Fig. 7, except for the set of 40 additional earthquakes shown in Fig. 20, which were not used in structural inversions. There are ~938 000 measurements.
Figure 22. Comparison of horizontal isotropic shear-wave-speed cross-sections of GLAD-M15 with starting mantle model S362ANI and recent degree-40 mantle model S40RTS (Ritsema et al. 2011).
Figure 23. Same as Fig. 22, but for greater depths.
want to be ready to harness such systems when they do. Needless to say, this requires continual investments in code development and optimization. With this goal in mind, we are a partner in ORNL’s Center for Accelerated Application Readiness (CAAR). CAAR has established eight partnerships to prepare computational science & engineering applications for use on the OLCF system to be named ‘Summit’, which will become available in 2018. The Summit system, an IBM with Power-9 CPUs and NVIDIA Volta GPU accelerators, will help determine what exascale hardware might look like in the early 2020s. Summit will enable us to reduce the shortest period in our global simulations from 17 to 9 s, and exascale systems will reduce this further to just a few seconds.

Tomographic resolution depends in part on the chosen model parametrization. To make the problem tractable, we currently keep the 1-D model constant in numerical simulations and assume that the Earth is elastic with transverse isotropy confined to the upper mantle, and use (frequency-dependent) phase information only. The PSF tests confirm that such inversions are feasible with the current data set. Building on our experiences in Europe (Zhu et al. 2013; Zhu & Trampert 2013), we plan to invert for global azimuthal anisotropy and attenuation in the future. In the latter case, we will investigate the inclusion of frequency-dependent amplitude measurements.

7 CONCLUSIONS AND FUTURE WORK

We determined the first global tomographic model based on fully 3-D forward and adjoint simulations of anelastic seismic wave propagation. We assimilated 3.8 million measurements in three-component data from 253 earthquakes with a shortest period of 17 s, using 180 min seismograms containing full-orbit surface waves. Our ‘first generation’ model is the result of 15 conjugate-gradient iterations performed on the Cray XK7 ‘Titan’, a supercomputer located at Oak Ridge National Laboratory (USA). We simultaneously inverted for crust and mantle structure, thereby avoiding ‘crustal corrections’; thus ours is the first global model which naturally unifies the crust and mantle. The model is transversely isotropic in the upper mantle, and contains numerous distinct signatures of plumes, hotspots, and slabs. Such anomalies are seen in lateral variations in shear wave speed, but also in the Vp/Vs ratio and in transverse isotropy. Our multiscale smoothing strategy helps bring out smaller-scale features where coverage is good, for example, underneath USArray. Point-spread function tests show that a number of interesting features are well resolved in our models, with limited parameter trade off. Finally, we used a data set of 40 additional earthquakes not used in the construction of our global model to demonstrate that it provides a clear improvement in traveltime fit compared to the starting model.

Looking forward, our goal is to assimilate data from thousands of earthquakes that have already been recorded by global and regional networks. This requires further optimizing and stabilizing the adjoint tomography workflow by taking advantage of workflow management tools, such as Pegasus (pegasus.isi.edu).

ACKNOWLEDGEMENTS

EB dedicates this manuscript to her parents, Zahire & Hilmi Bozdağ. 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) and ORFEUS (orfeus-eu.org) for providing the data used in this study. We gratefully acknowledge editor Gabi Laske, Carl Tape and an anonymous reviewer for constructive feedback which improved the manuscript. We thank Heiner Igel, Suzan van der Lee, Guust Nolet, Jeroen Ritsema, Frederik J. Simons and Jeanneit Trampert for their support and fruitful discussions. We also thank Wenjie Lei and Youyi Ruan for providing 40 reinverted CMT solutions for tests with an independent set of earthquakes, and Yanhua Yuan, Ryan Modrak, Vadim Monteiller, Lion Krischer and James Smith for various discussions on FWI. 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). This research was supported by NSF grant 1112906. EB was partly supported by her UNS–CNRS Chaire d’Excellence grant.

REFERENCES


