Imaging global mantle discontinuities: a test using full-waveforms and adjoint kernels

Maria Koroni [1] and Jeannot Trampert[2]

[1]*Department of Earth Sciences, Institute of Geophysics, ETH Zürich, Sonneggstrasse 5, CH-8092 Zürich, Switzerland. E-mail:* maria.koroni@erdw.ethz.ch
[2]*Department of Earth Sciences, Utrecht University, Princetonlaan 8A, 3584 CB Utrecht, the Netherlands*

SUMMARY

We present a novel approach for imaging global mantle discontinuities based on full-waveform inversion (FWI). Over the past decades, extensive research has been done on imaging mantle discontinuities at approximately 400 and 670 km depth. Accurate knowledge of their topography can put strong constraints on thermal and compositional variations and hence geodynamic modelling. So far, however, there is little consensus on their topography. We present an approach based on adjoint tomography, which has the advantage that Fréchet derivatives for discontinuities and measurements, to be inverted for, are fully consistent. Rather than working with real data, we focus on synthetic tests, where the answer is known in order to be able to evaluate the performance of the developed method. All calculations are based on the community code SPECFEM3D_GLOBE. We generate data in fixed 1-D or 3-D elastic background models of mantle velocity. Our 'data' to be inverted contain topography along the 400 and 670 km mantle discontinuities. To investigate the approach, we perform several tests: (i) In a situation where we know the elastic background model 1-D or 3-D, we recover the target topography fast and accurately; (ii) The exact misfit is not of great importance here, except in terms of convergence speed, similar to a different inverse algorithm and (iii) In a situation where the background model is not known, the convergence is markedly slower, but there is reasonable convergence towards the correct target model of discontinuity topography. It has to be noted that our synthetic test is idealized and in a real data situation, the convergence to and uncertainty of the inferred model is bound to be larger. However, the use of data consistent with Fréchet kernels seems to pay off and might improve our consensus on the nature of mantle discontinuities. Our workflow could be incorporated in future FWI mantle models to adequately infer boundary interface topography.

Key words: Computational seismology; Global seismology; Mantle discontinuities; Fréchet derivatives; Inverse problem.

1 INTRODUCTION

The Earth's mantle has extensively been studied using seismological and mineral physics data as well as geodynamic modelling. Comprehending mantle processes is central to our understanding of planetary evolution. Seismology has clearly played a key role in imaging mantle structure using body and surface waves, normal modes as well as ambient noise (e.g. Dziewoński & Anderson 1981; Ritsema *et al.* 2011). Mantle processes are strongly influenced by global discontinuities (e.g. Tackley *et al.* 1994), which are observed consistently in seismic data. Their importance is paramount as they determine how mantle evolution develops. Many seismological

studies try to infer clear images of these discontinuities, but broad consensus is still lacking (e.g. Meier *et al.* 2009). In this study, we propose a new approach based on full-waveform inversion (FWI), which may improve the consensus on the nature of mantle discontinuities.

Images of seismic wave speeds and density show that the mantle can be separated into two major parts. The boundary between upper and lower mantle is a complex transition zone whose boundaries are identified as global discontinuities at depths around 410 and 660 km (e.g. Shearer 2000; Deuss 2009). The reference depths of the discontinuities are hereinafter denoted as 400 and 670 km, since these correspond to PREM (Dziewoński & Anderson 1981), which

is our chosen reference model. The discontinuities are identified by abrupt (instead of gradual) increases in wave speeds in the radial direction and density.

Mineral physics predicts major phase changes in mantle mineralogy, which roughly correspond to these depths. The transition from olivine to wadsleyite takes place at a depth of approximately 400 km (Akaogi *et al.* 1989; Katsura & Ito 1989), coinciding with the observed seismic discontinuity. The discontinuity around 670 km depth coincides with a transformation from ringwoodite to perovskite and magnesiowüstite (e.g. Shearer 1990; Helffrich 1999; Frost 2008). Ringwoodite results from a transformation of wadsleyite at about 520 km depth. A seismic discontinuity has also been reported at this depth (e.g. Shearer 1990, 1996), although its existence as a global feature is not unequivocal (Bock 1994).

Topographic changes of individual discontinuities is their most interesting feature because it is a manifestation of lateral changes in temperature and composition that forces mineralogical transformations to happen at varying depths (e.g. Helffrich 1999). To access this information, it is necessary to resolve the discontinuity structure with sufficient accuracy in order to establish connections with localized thermal and compositional variations.

Much effort has been put into imaging the topography of mantle discontinuities by using seismic body waves such as *SS*, *PP* and *P'P'* precursors at long periods (usually between 15 and 75 s, e.g. Shearer 1991; Shearer & Masters 1992; Benz & Vidale 1993; Shearer 1993; Estabrook & Kind 1996; Flanagan & Shearer 1998; Xu *et al.* 1998; Gu & Dziewonski 1998; Flanagan & Shearer 1999; Shearer *et al.* 1999; Gu & Dziewonski 2002; Gu *et al.* 2003; Chambers *et al.* 2005; Schmerr & Garnero 2006; Houser *et al.* 2008; Lawrence & Shearer 2008). Near surface conversions of compressional and shear waves have been used to infer the local discontinuity depth using receiver functions. *P*-to-*S* conversions have been used in many studies (e.g. Vinnik 1977; Paulssen 1988; van der Lee *et al.* 1994; Chevrot *et al.* 1999; Lebedev *et al.* 2003; Lawrence & Shearer 2006; Andrews & Deuss 2008; Tauzin *et al.* 2013; Bonatto *et al.* 2013; Gao & Liu 2014; Jenkins *et al.* 2016), as well as *S*-to-*P* conversions (e.g. Lawrence & Shearer 2006; Tauzin *et al.* 2008). *ScS* reverberations, higher order surface waves and ambient noise are other data types used for imaging mantle discontinuities (e.g. Revenaugh & Jordan 1991c; Meier *et al.* 2009; Poli *et al.* 2012; Haugland *et al.* 2019).

The most commonly used and convenient data type for global studies are the precursors to once surface reflected *S* waves. They precede the *SS* wave, travelling similar paths, but instead of reflecting at the earth's surface, *SS* precursors bounce from the underside of the 400 and 670 discontinuities. The bounce points are thought to concentrate most of the sensitivity and cover many areas on the globe, including large oceanic regions, which are otherwise difficult to sample. When it comes to data selection, despite their suitability due to good global data coverage, a major issue with *SS* precursory phases is their low amplitudes, often found to be below noise levels. To enhance signal-to-noise ratio stacking and geographical cap averaging, based on midpoints location, is employed (e.g. Flanagan & Shearer 1998, 1999; Deuss 2009). Bounce points are roughly located half-way between source and receiver along the great circle path. Due to the similarity of the two reflection legs of *SS* and its precursors, it is generally assumed that their traveltime differences are mainly sensitive to the depth variation of the corresponding discontinuity at this particular bounce point. In reality the traveltime difference between *SS* and its precursors also depends on the wave speed along the predicted ray paths. To account for this,

researchers correct for the velocity structure using an existing shear velocity model along the ray paths and the inference of discontinuity variations beneath each bounce point is made using linearized ray theoretical expressions. There have been studies (e.g. Gu *et al.* 2003; Lawrence & Shearer 2008; Kustowski *et al.* 2008) that performed joint inversions for both, namely wave speeds and discontinuities, simultaneously. However, the joint inversions in these studies were based on using ray theoretical sensitivity kernels instead of finite-frequency ones, or by linearly correcting for wave speed structure. This has been shown to decrease resolution of the resulting images, due to complex wave effects which are non-negligible.

A comparison of different global topography maps using this linearized methodology shows considerable discrepancies. Some pictorial examples of observed dissimilarities can be found in Meier *et al.* (2009) and Deuss (2009), which probably are due to different model choices for traveltime corrections and limitations to seismic ray theory itself.

In an effort to explain this, theoretical and modelling studies showed that traveltimes of underside reflections, such as *SS* or *PP* precursors, present considerable sensitivity off the ray path (Neele *et al.* 1997; Chaljub & Tarantola 1997; Zhao & Chevrot 2003; Dahlen 2005; Lawrence & Shearer 2008; Liu & Tromp 2006; Deuss 2009; Bai *et al.* 2012; Koroni *et al.* 2017, 2019). Ray theory is therefore not suitable for mapping discontinuity topography with such data. Additionally, effects of topography and shear wave speed are non-additive for traveltime residuals of *SS* precursors (Koroni & Trampert 2016).

In a different approach, Meier *et al.* (2009) derived maps of the 400 and 670 km topography using a data set of higher mode surface waves together with neural networks in a joint inversion for discontinuity topography and a three-layered, averaged shear velocity model. Their non-linear Bayesian inversion scheme allowed them to properly account for the trade-off between the parameters and obtain error bars for the derived models.

In this study, a new methodology, based on FWI, is investigated for imaging topography. The main objective of the current paper is to establish whether FWI is suitable for imaging global discontinuities, and under which conditions. We focus on the mantle transition zone boundaries and conduct synthetic experiments based on a gradient descent optimization method. To the best of our knowledge, this method has so far not explicitly been developed or applied to mantle discontinuities. The essential ingredient of the presented approach is using Fréchet sensitivity kernels with respect to internal boundaries (e.g. Tromp *et al.* 2005; Dahlen 2005; Liu & Tromp 2008) computed using the adjoint method (e.g. Patera 1984; Talagrand & Courtier 1987; Tromp *et al.* 2005; Fichtner *et al.* 2009; Fichtner 2010). These sensitivity kernels have been used in the past for sensitivity analysis by Liu & Tromp (2008) and Koroni *et al.* (2017, 2019), but, to date and to our knowledge, they have never been used for inversions of discontinuity structure. FWI allows us to extract much more accurate information from the waveforms leading to a better treatment of the inherent non-linearity between observed data and model parameters and, hence, better models.

The methodology aims to improve resolution imaging of mantle transition zone boundaries, using waves interacting with internal interfaces. This approach can be incorporated in future FWI models for mantle and discontinuity structure and/or can be used to improve current high resolution models of earth's mantle on global and/or continental scales (e.g. Fichtner *et al.* 2013, 2018; Bozdağ *et al.* 2016).

The following sections give a description of the synthetic data generated, the methods adopted for FWI for mantle discontinuities,

our results regarding the various tests and, finally, the main conclusions drawn from the experiments. The results are discussed in the context of global FWI and comments on the selected iterative optimization method are made with some suggestions for future work.

2 METHODS

2.1 Forward modelling and synthetic data

We use a spectral-element method for simulating numerical wave propagation to obtain exact, noise-free seismograms for various earth models. The experiments presented require 'observed' and 'synthetic' data sets for comparison. These data sets are created by using an open-source, community developed spectral-element code, that is SPECFEM3D_GLOBE (Komatitsch & Vilotte 1998; Komatitsch & Tromp 1999, 2002a,b,c). We use two earth models: the 1-D velocity model PREM (Dziewoński & Anderson 1981) and the 3-D velocity model S20RTS (Ritsema et al. 1999) plus CRUST2.0 (Bassin et al. 2000) to obtain the starting 'synthetic' data in 1-D and 3-D, respectively. By adding topography models of the 400 and 670 km discontinuities (Meier et al. 2009) to these models, the 'observed' data set with lateral variations along the discontinuities is created. The topographic models have a lateral resolution of spherical harmonic degree eight and are scaled to ±30 km peak to peak. This peak-to-peak amplitude is chosen by reported uncertainty existing in current models. Mostly both discontinuities are varying in depth by ±20 km, however, sometimes the reported variation can reach ±40 km, as mentioned by (Meier et al. 2009).

The experimental setup is equal to that in Koroni & Trampert (2016), consisting of twelve equally spaced events along the equator and 6211 randomly chosen stations in a distance range of 110°−160°, leading to a set of well distributed 6211 *SS* bounce points. It should be kept in mind that this is an unusually even data distribution and a realistic, much sparser data coverage will plague real data applications. The earthquake source is a pure shear strike slip of a moment magnitude equal to 7.9. The source time function has a duration equal to 60 s, which is chosen as pre-smoothing for the sensitivity kernels. The dominant period of the synthetics is between 25 and 150 s, a range chosen in conventional studies using the *SS* precursors for discontinuity topography inference (e.g. Deuss 2009). Both 'observed' and 'synthetic' data are filtered using a bandpass between 4 and 45.4 mHz. The same filter is applied to the adjoint sources. To further idealize forward modelling, certain earth properties that in general affect seismograms, are switched off during the simulations. Attenuation, self-gravitation, earth's ellipticity and rotation, as implemented in the spectral-element code (Komatitsch & Tromp 2002b), are not taken into account here as they will affect 'observed' and 'synthetic' data differently. Therefore, obtained seismograms for models with and without topographic variations differ only due to topography and the observed misfit can directly be interpreted as a result of it.

2.2 Misfit functionals and adjoint sensitivity kernels

To measure differences between 'observed' and 'synthetic' data, two misfit functions are used, namely a classical waveform *L2* norm misfit (WF) and a cross-correlation traveltime least-squares misfit (CC). The misfit functional chosen for calculating differences between real and synthetic data determines the adjoint kernels and
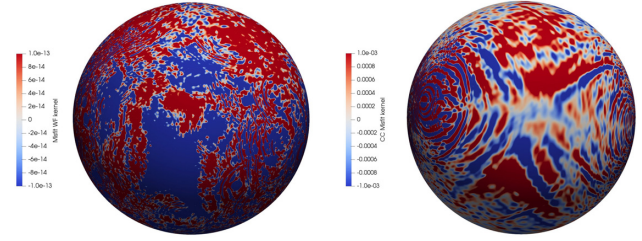
Figure 1. Right-hand panel: CC misfit kernel for the 400 km discontinuity for a given event and all stations. The kernel has units of $s^{-2} \cdot km^{-2}$. Left-hand panel: for the same event and station, now using the WF misfit. The units of this sensitivity kernel are *seconds*. Note the overall difference in pattern as well as the different units.

hence the way in which model parameters are sensed by the data. By calculating the gradient of the misfit functions using adjoint methods with respect to boundary parameters (Dahlen 2005; Tromp et al. 2005; Liu & Tromp 2006), the initial models can be updated within an iterative optimization scheme.

The misfit between 'observed' and 'synthetic' data is done over narrow time windows around the predicted arrivals of *SS* precursors. They are centered at predicted arrival times which are obtained using the TauP software Crotwell et al. (1999) and PREM as a background model. The width of time windows is set to 40 s, 20 s at either side of the predicted precursor arrival. A cosine taper of 20 per cent is utilized to isolate the phases.

The cross-correlation misfit is expressed as (Luo & Schuster 1991; Marquering et al. 1999; Dahlen et al. 2000; Tromp et al. 2005; Fichtner 2010):

$$\chi_{CC}(\mathbf{m}) = \frac{1}{2} \sum_{r=1}^{N} [T_{obs}^r - T_{syn}^r]^2, \tag{1}$$

where 'obs' and 'syn' denote the observed and synthetic data (with and without topography, respectively), r denotes the single receiver where each measurement is made, m stands for the model, N is the total number of stations and CC denotes a cross-correlation traveltime misfit. The gradient of the misfit function due to a traveltime perturbation is given by:

$$\delta\chi_{CC}(\mathbf{m}) = \sum_{r=1}^{N} [T_{obs}^r - T_{syn}^r] \delta T_{syn}^r, \tag{2}$$

The cross-correlation traveltime perturbation measurement is given by the expression:
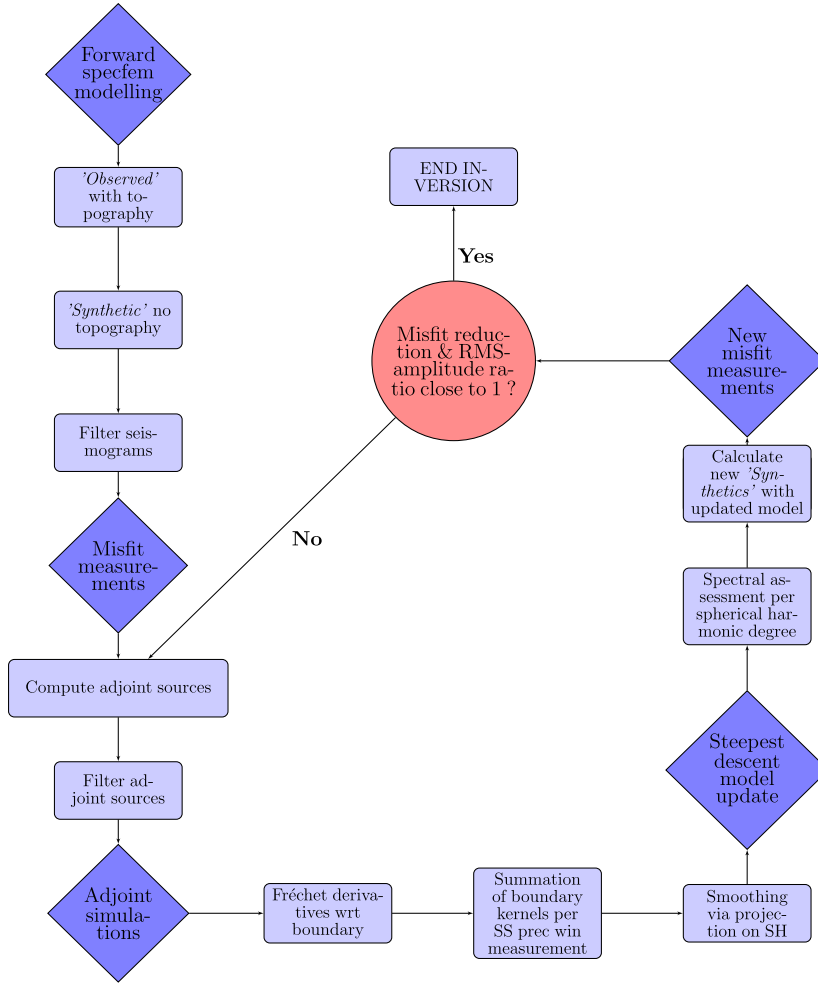
$$\delta T_{syn}^r = \frac{1}{N_r} \int_0^T w_r(t) \partial_t u_i(\mathbf{x_r}, t) \delta u_i(\mathbf{x_r}, t) dt, \tag{3}$$

where $w_r(t)$ denotes the time window around the selected precursor phase and $\mathbf{u}$ represents the displacement field. The normalization factor N_r is given by:

$$N_r = \int_0^T w_r(t) u_i(\mathbf{x_r}, t) \partial_t^2 u_i(\mathbf{x_r}, t) dt. \tag{4}$$

The corresponding CC misfit adjoint source per component i is then given by:

$$f_i^\dagger(\mathbf{x_r}, t) = \frac{1}{N_r} w_r(T - t) \partial_t u_i(\mathbf{x_r}, t) \delta(x - \mathbf{x}_r). \tag{5}$$

Figure 2. This flowchart describes the basic steps taken for each of the iterations shown here, in order to obtain a single model update for the mantle discontinuities. The last step of the workflow where one should decide whether there is sufficient convergence has two options: stop inversion or recalculate adjoint forces based on the new model update and initial model new misfit measurements.

The waveform misfit is given by the expression (Tromp *et al.* 2005):

$$\chi_{WF}(\mathbf{m}) = \frac{1}{2} \sum_{r=1}^{N} \int_0^T ||\mathbf{u}_{syn}(\mathbf{x}_r, t, \mathbf{m}) - \mathbf{d}_{obs}(\mathbf{x}_r, t)||^2 dt. \qquad (6)$$

The gradient of this misfit function is then:

$$\delta\chi_{WF}(\mathbf{m}) = \sum_{r=1}^{N} \int_0^T [\mathbf{u}_{syn}(\mathbf{x}_r, t, m) - \mathbf{d}_{obs}(\mathbf{x}_r, t)]$$
$$\times \delta\mathbf{u}(\mathbf{x}_r, t, \mathbf{m}) dt. \qquad (7)$$

The adjoint source vector for the WF misfit in the time window around the selected *SS* precursor is then:

$$\mathbf{f}^\dagger(t) = \mathbf{u}_{syn}(\mathbf{x}_r, t, m) - \mathbf{d}_{obs}(\mathbf{x}_r, t). \qquad (8)$$

The above expressions relate to a single earthquake and are called event kernels. They are summed over all events to obtain the total misfit kernel (Tape *et al.* 2007).

The expression relating the gradient of the total misfit to boundary changes is given by:

$$\delta\chi = \int_{\Sigma_{SS}} K_h \delta \ln h(\mathbf{x}) d^2 x, \qquad (9)$$

where Σ_{SS} denotes internal solid-solid 400 or 670 km discontinuity surface and h is the radius of the discontinuity.

Various definitions of misfit lead to sensitivity kernels which are sensing the model quite differently. An example of how sensitivity kernels differ due to various misfits is shown in Fig. 1. The example shows a single station event kernel for the *S400S* precursor window. The reader will note that not only are the units different, the overall kernels are in fact incomparable. Comparisons of various misfit functionals and their corresponding kernels have been shown by Bozdağ *et al.* (2011). Despite the overall differences, the final model updates for either misfit should lead to models that can explain the data equally well.

It is essential that the misfit kernels are smoothed prior to using them for a model update. This is done in order to reduce high amplitude contributions at the locations of sources and receivers and to filter small-scale structural changes that are not resolvable by our

data. Not performing smoothing could introduce artefacts in the models. Usually a Gaussian smoothing is applied, but in this study, smoothing is performed by projecting the boundary misfit kernels onto a low degree spherical harmonics expansion up to degree 20. The target model is also expanded to degree 20, but has an effective resolution of degree eight (Meier *et al.* 2009). Since our synthetic data set consists of global and evenly distributed sample points (*SS* precursor bounce points), smoothing by a low degree spherical harmonic expansion should be adequate without introducing spatial bias. It should be kept in mind however that in real data applications uneven data layout may bias this approach (Trampert & Snieder 1996). Another advantage of an expansion based smoothing function is a reduced size of the gradient that is more manageable for numerical optimization and resolution analysis as described in (e.g. Fichtner & Trampert 2011a,b).

2.3 Iterative optimization using a steepest descent algorithm

We have chosen the steepest descent algorithm (referred to as SD hereinafter) to iteratively update the discontinuity structure. The starting models for upper-mantle discontinuity topography are set to zero, meaning that the 400 and 670 km discontinuities are flat at the beginning of the imaging process.

The step length α required at every iteration is chosen based on trial and error. We performed spectral comparisons between potential new models and target models and chose the value of α corresponding to the highest correlation and amplitude ratio, in order to speed up convergence. We note that techniques for estimating the optimal step length are needed when using real data or even more elaborate optimization schemes (Nocedal & Wright 2000). In the Appendix, we provide details on how the model is updated for SPECFEM3D_GLOBE based on the generic algorithm below.

Algorithm 1 Steepest descent
_____
1: k ← 0
   ∘ Initial topography model is:
   $\mathbf{m}_0 = 0$ (flat discontinuities with zero depth variations)
   ∘ Compute the search direction:
   $p_k = -\,\delta_{\mathbf{m}}\,\chi(\mathbf{m}_k)$
   (the search is on the negative gradient direction)
   ∘ Update the model:
   $\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k\, p_k$
   where α_k is a real constant and must ensure that:
   $\chi_{\mathbf{m}_{k+1}} < \chi_{\mathbf{m}_k}$
   k ← k+1
_____

In all experiments, the 1-D (PREM) or 3-D (S20RTS+ CRUST2.0) elastic background models remain unmodified and only the 400 or 670 km topography is updated. Therefore, the are no updates of the model for the velocity or density variations.

Five iterations are performed for experiments using the cross-correlation misfit, while four iterations are performed for the waveform misfit, which seems to converge faster.

The number of adjoint simulations depends on the number of events. Each misfit gradient requires twelve forward and adjoint simulations per misfit function (CC or WF) and per precursor window (*S400S* or *S670S*). The misfit is calculated on transverse components only, as the signal-to-noise ratio of the precursors is too low

on radial and vertical components. The adjoint forces are calculated for all displacement components. All computations are done using 600 CPUs. A brief flowchart of our workflow is shown in Fig. 2.

3 RESULTS

3.1 Retrieval of topography in a fixed elastic 1-D background model

The 'observed' data are calculated in PREM plus a target topography, while 'synthetic' data are calculated in PREM plus the updated topography. The starting model for the iterations is PREM only with no discontinuity topography. We present visual comparisons of the models for a qualitative assessment of the inversion. For a more quantitative comparison, as our models are parametrized on spherical harmonics, we evaluate the correlation between target and update model as well as the RMS amplitude ratio of target over updated model per spherical harmonic degree.

3.1.1 Cross-correlation misfit

The left column of Fig. 3 displays the smoothed gradients using CC misfits measured in *S400S* time windows. The first gradient calculation at iteration number zero (*it #0*) is in the PREM background model and gives the first model update for the 400 km discontinuity, as shown in the same figure, second column (*m1*).

Fig. 4 displays the updates of the 670 km topography model using misfits measured in *S670S* windows. Each iteration provides two separate model updates, one for the 400 km and one for the 670 km discontinuity, which are then used together for the calculation of the new synthetics. The new synthetics are compared once again to the 'observed' data, which of course remain unaltered. This procedure is continued until no significant improvement is observed in the model updates compared to the previous iteration. The misfit reduction for this first experiment, for both time windows is shown in Fig. 5. The total misfit reduction is approximately 71 per cent for the 670 km topography model and 66 per cent for the 400 km discontinuity.

For more quantitative comparisons, we show correlation coefficient and RMS amplitude ratios per spherical harmonic degree. The plots for correlations are shown in Figs 6(a) and (b) and the RMS amplitude ratios are displayed in Figs 7(a) and (b) for the 400 and 670 km discontinuity, respectively. The correlations per spherical harmonic degree increase as the iterations proceed. Similarly, the amplitude ratios converge to the optimal value of one.

We observe that the improvement from model *m4* to *m5* is modest. This indicates that the algorithm has converged towards a model that is close to minimizing the misfit in the selected data window. Another observation is that the inference of 670 km topography model, judging from the spectral comparisons, seems better than that for the 400 km discontinuity topography.

3.1.2 Waveform misfit

We now repeat the previous experiment, but in this case a waveform misfit function is used, instead of a cross-correlation misfit. The smoothed gradients corresponding to the waveform misfit

## S400S 1-D
Cross-correlation
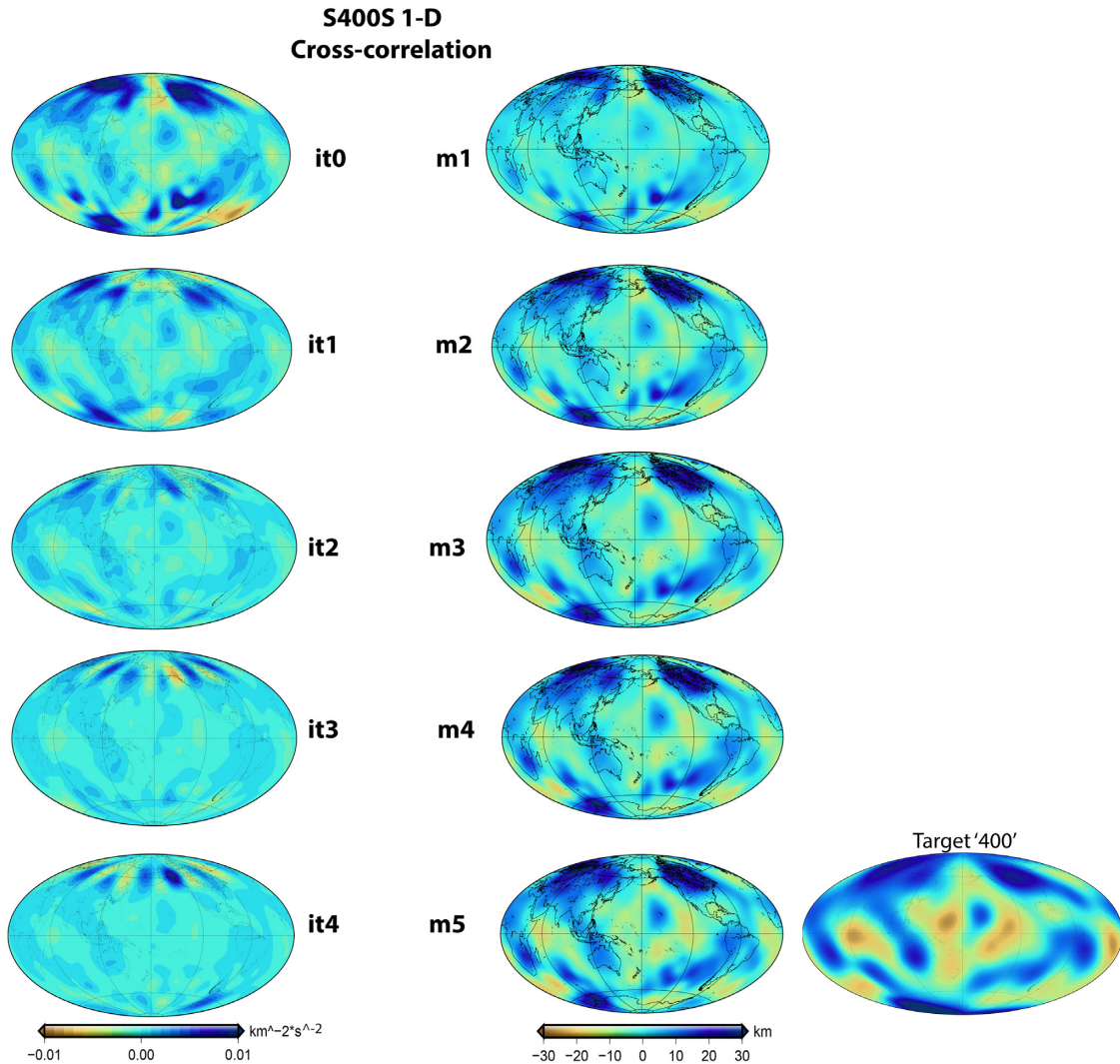
Figure 3. Successive smoothed boundary misfit kernels (left-hand column) per iteration for a cross-correlation misfit in a 1-D elastic background model. The misfit is measured in the *S400S* time windows. The units of the boundary misfit kernels are $s^{-2} \cdot km^{-2}$. In the right-hand column, the five successive model updates for the 400 km discontinuity are shown. The target model is shown to the right of the last update.

measurements are displayed in the left hand side of Figs A1 and A2 for the two data windows, respectively. The model updates derived using these kernels are shown in the second column of the respective figures. While the gradients exhibit large differences compared to those derived using cross-correlation misfits (Figs 3 and 4), the updated models vary much less in the end. This stems from the variable sensitivity of the different misfit functions. The inferred target models, however, are encouragingly good for both misfit functions.

The misfit reduction for the waveform misfit is shown in Fig. 8. We see that the iterations converge, but do not significantly improve during the last two iterations. The waveform misfit seems to converge slightly faster than the cross-correlation misfit. That is why one iteration less was performed. The total misfit reduction for the 400 km topography updates is approximately 81.8 per cent and that for the 670 km discontinuity is around 74.6 per cent. The spectral comparisons are shown in Figs 9 and 10.

It is worth noting that the waveform information leads to a faster reduction of the misfit, as it takes into account amplitude and phase of the precursors. However, this may be challenging in real data applications where the conditions are not as idealized as here, since the non-linearities are larger for waveform misfits. When comparing the misfit reduction for CC and WF misfits, we note that the '400 km' windows present larger misfit reductions than the '670 km' case, however the 670 km discontinuity is better retrieved.

Already after the first update (*m1*), the correlation per spherical harmonic degree is quite high. Until the last iteration, both topography models remain very similar to the target models. The RMS amplitude ratios per spherical harmonic degree also rapidly converge towards one. Once again, the final update for the 670 km topography is closer to the target model than the 400 km discontinuity. This is despite the higher percentage value of the misfit reduction for the 400 km discontinuity. Remember that the synthetics are calculated using both topography models and, thus, the misfits

**S670S 1-D
Cross-correlation**
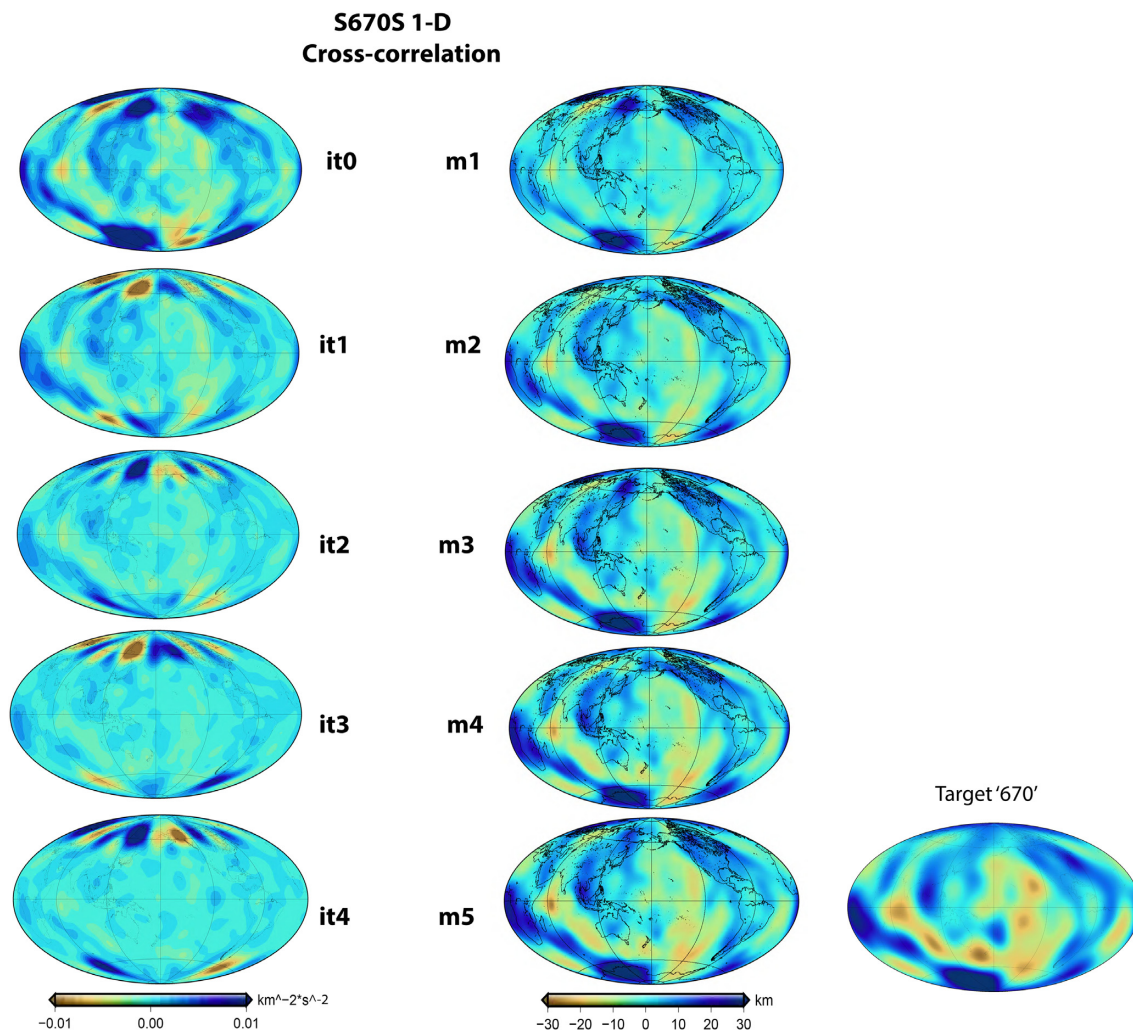
Figure 4. The same as in Fig. (3) but for the 670 km discontinuity.
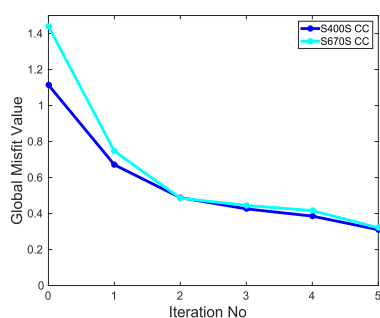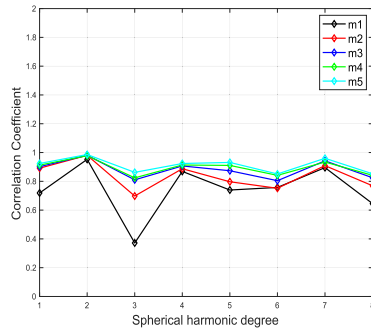

Figure 5. Misfit reduction for cross-correlation measurements in a 1-D elastic background model. The misfit reduction per iteration shows that the inversion converges towards the target models used to create the 'observed' waveforms. The misfit reduction is about 71 per cent for the 670 km and 66 per cent for the 400 km discontinuity. The units of the misfit values are s^2.

are influenced by both as well, since precursory waves interact with both discontinuities on their way to the receivers.

3.2 Retrieval of topography in a fixed 3-D elastic background model

We will now report on the retrieval of topography in a known 3-D elastic background model, which is S20RTS+CRUST2.0. To calculate the 'observed' data, S20RTS+CRUST2.0 plus our target topography models are used. The 'synthetic' data are calculated in S20RTS+CRUST2.0. Again, time shift misfit measurements by cross-correlation are used. The models are updated using the smoothed gradients shown on the left columns of Figs A3 and A4. The model updates are shown on the right columns. The initial gradients (*it #0*) are obtained by measuring the misfit in the same time windows as before, although now the elastic background model is characterized by 3-D lateral variations. Given that the time windows around the precursors are 40 seconds wide, it is assumed that the underside reflections are captured correctly.

After five model updates, the target models of the 400 and 670 km topography are successfully retrieved. The steepest descent algorithm converges to a solution as shown by the misfit reduction curves in Fig. 11. The percentage of misfit reduction is lower than in the 1-D case. For *S400S*, the misfit reduction is 20.7 per cent, while for *S670S* its value is 36.2 per cent. However, by looking at
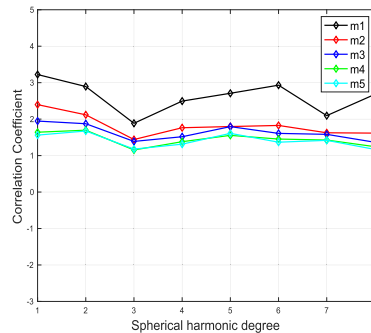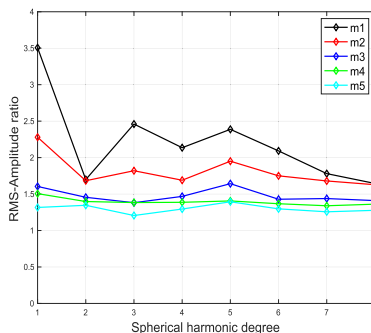
(a) 400 km CC 1-D
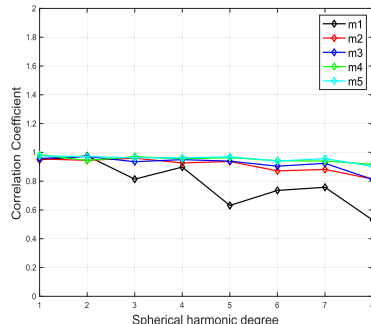
(b) 670 km CC 1-D

Figure 6. Correlation per spherical harmonic degree between the target model and the retrieved models at a given iteration (m1, m2,..m5) using cross-correlation measurements in a 1-D elastic background model for the '400' discontinuity. Bottom plot is for the '670' discontinuity.
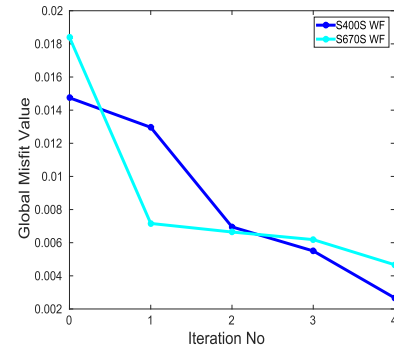

(a) 400 km RMS 1-D

(b) 670 km RMS 1-D

Figure 7. Root-mean-square amplitude per spherical harmonic degree of the target model over the retrieved models at a given iteration using cross-correlation measurements in a 1-D elastic background model.

Figure 8. Misfit reduction for waveform measurements in a 1-D elastic background model. The misfit reduction per iteration shows that the inversion converges towards the target models used to create the 'observed' waveforms. The misfit reduction is 74.6 per cent for the 670 km and 8 1.8 per cent for 400 km discontinuity. The units of the misfit are seconds.
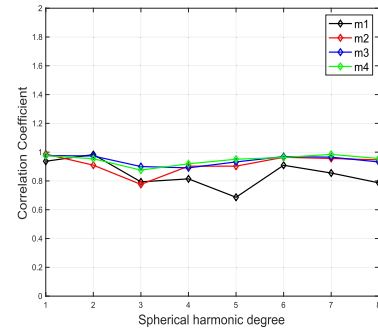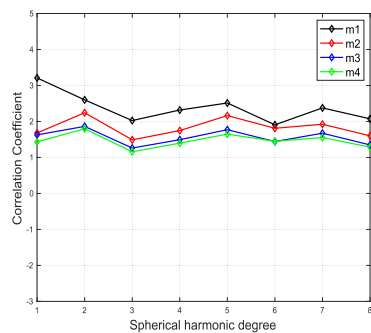

(a) 400 km CC WF 1-D
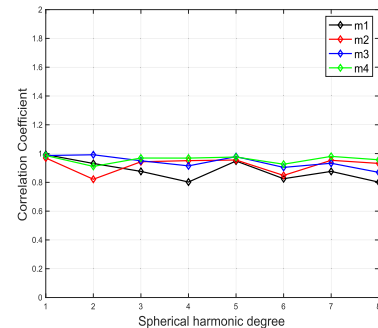
(b) 670 km CC WF 1-D

Figure 9. Correlation per spherical harmonic degree between the target model and the retrieved models at a given iteration using waveform misfit measurements in a 1-D elastic background model.

Fig. 11, we see that the algorithm perhaps needs more iterations to fully converge. This slower convergence compared to the 1-D case, is clearly an imprint of the 3-D elastic structure. More complicated wave scattering has an effect on the kernels, but because the 3-D model is correct and the kernels are compatible with the structure, eventually the non-linear optimization will converge towards the correct target model.

The spectral correlations confirm this and are shown in Figs 12(a) and (b). After each iteration the correlation progresses towards a
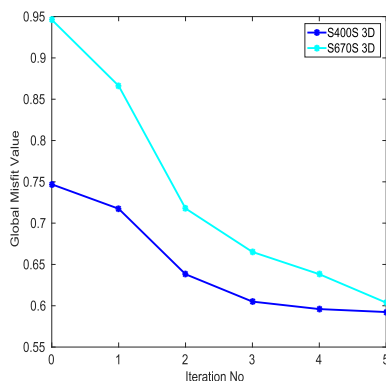
(a) 400 km RMS WF 1-D
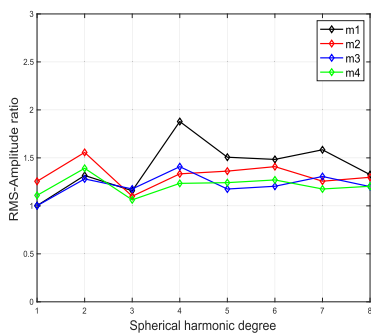
(b) 670 km RMS WF 1-D

Figure 10. Root-mean-square amplitude ratio per spherical harmonic degree of the target model over the retrieved models at a given iteration using waveform misfit measurements in a 1-D elastic background model.
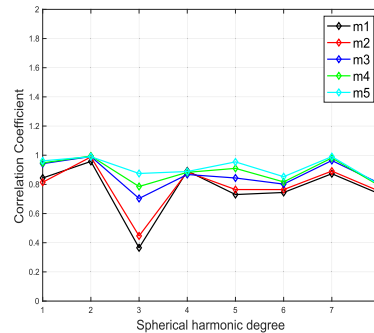

Figure 11. Misfit reduction for cross-correlation measurements in a 3-D elastic background model. The misfit reduction per iteration shows that the inversion converges towards the target models used to create the 'observed' waveforms. The misfit reduction is about 36.2 per cent for the 670 km and 20.7 per cent for the 400 km discontinuity. The units of the misfit values are s^2.

high value close to one for all spherical harmonic degrees. The RMS amplitude ratios are shown in Figs 13(a) and (b). The RMS-amplitude ratios also slowly tend towards one.

3.3 Retrieval of topography in an unknown elastic background model

This is probably the most realistic test: Can the topography models be correctly inferred even though the background elastic structure is basically unknown? Thus far, our results suggest that
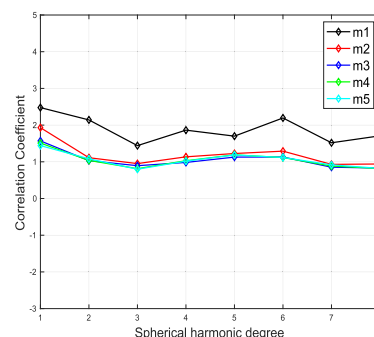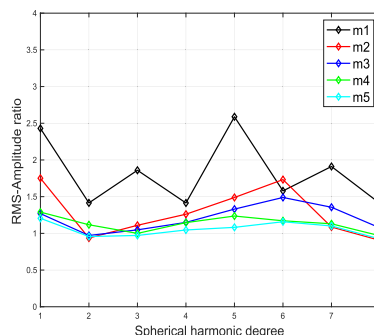

(a) 400 km CC 3-D
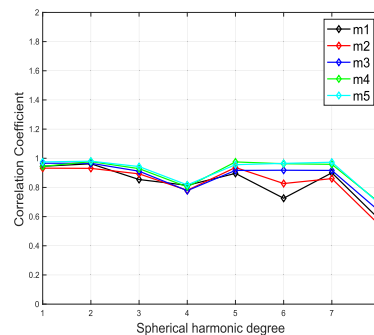
(b) 670 km CC 3-D

Figure 12. Correlation per spherical harmonic degree between the target model and the retrieved models at a given iteration using cross-correlation misfit measurements in a 3-D elastic background model.
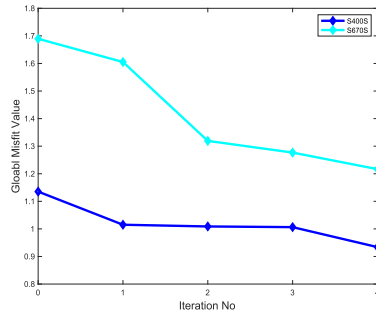

(a) 400 km RMS 3-D

(b) 670 km RMS 3-D

Figure 13. Root-mean-square amplitude ratio per spherical harmonic degree of the target model over the retrieved models at a given iteration using cross-correlation misfit measurements in a 3-D elastic background model.

Figure 14. Misfit reduction for cross-correlation measurements in an un-known elastic background model. The misfit reduction per iteration shows that the inversion slowly converges towards the target models used to create the 'observed' waveforms. The misfit reduction is about 28 per cent for the 670 km and 17 per cent for the 400 km discontinuity. The units of the misfit values are s^2.

(a) 400 km CC Unknown
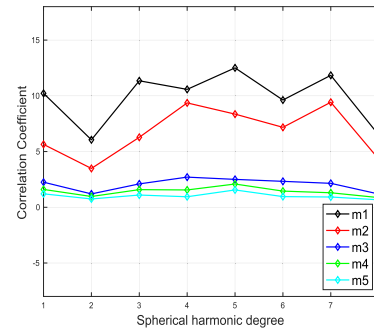

(b) 670 km CC Unknown
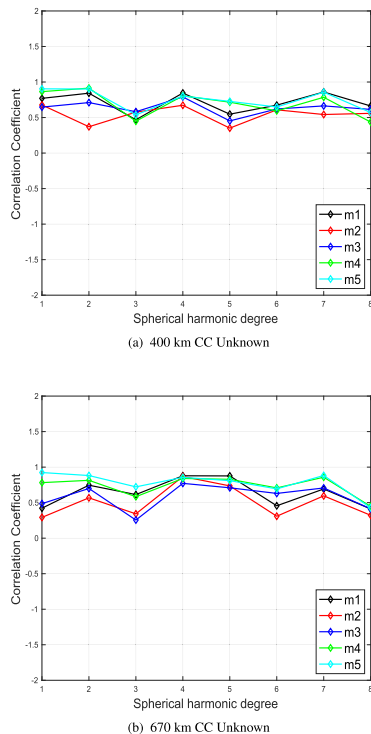
Figure 15. Correlation per spherical harmonic degree between the target model and the retrieved models at a given iteration using cross-correlation measurements in an unknown elastic background model.

in a well-known background elastic model, either 1-D or 3-D, *SS* precursor traveltimes lead to successful retrievals of 400 and 670 km topography models. However, it is usually the case that the velocity model is inaccurate and does not fully represent the data.

The whole experiment is once more performed and again based on cross-correlation misfit measurements within the same time windows around PREM arrivals of *SS* precursors. The 'observed' data are calculated in S20RTS+CRUST2.0 plus our target topography. In the 'synthetic' data, we use PREM and keep it fixed. This is probably quite a severe restriction, as in reality we would
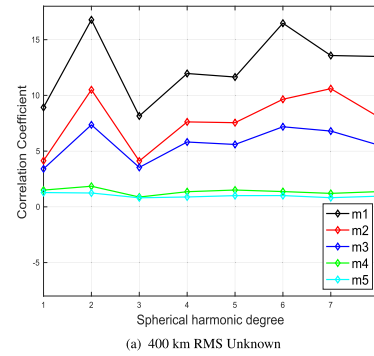
(a) 400 km RMS Unknown

(b) 670 km RMS Unknown

Figure 16. Root-mean-square amplitude ratio per spherical harmonic degree of the target model over the retrieved models at a given iteration using cross-correlation measurements in an unknown elastic background model.

know the earth model better than PREM. In this way, the gradients are based on measurements where the difference is not only due to topography, but also contains the full 3-D elastic variations. Only boundary kernels are used for the updates. Therefore, while topography models are introduced in the new synthetic data at each iteration, the 3-D elastic variations remain unmodelled.

Figs A5 and A6 show the smoothed kernels (left-hand side) and corresponding model updates (right-hand side) at each iteration. We observe that the steepest descent algorithm converges towards the target topography model, albeit slower than in the cases of a known background elastic model. The misfit evolution curves for this test show indeed that there is convergence. However, the values of misfit reduction are much smaller than those in a known 1-D or 3-D elastic background (Fig. 14). The misfit reduction is 17.7 per cent for the 400 and 27.9 per cent for the 670 km discontinuity data. We further see that the misfit values are also not close to zero, despite that the inferred model resembles the target model relatively well. This shows that the unmodelled 3-D elastic structure has a strong impact on the precursors, which is not properly addressed.

The spectral comparison shows that, after five iterations, the inferred models correlate quite well with the target topography (Figs 15a and b). The correlation coefficient per spherical harmonic degree is increasing towards one, however we see that it is slightly lower than in the tests with known 1-D or 3-D background. In the tests with a known background, we also noted that the 670 km discontinuity is better retrieved that the 400 km discontinuity. This is only marginally true here, specifically, the overall correlation

between target and model update *m5* is 0.62 for the 670 km discontinuity, while it is 0.57 for the 400 km. The RMS amplitude ratios shown in Figs 16(a) and (b) are revealing a similar picture. The two discontinuity models perform equally well and the initial updates were far from a good model in terms of amplitude. However, the overall amplitudes of model *m5* are quite close to the target model (RMS amplitude ratio of 1.13 for 670 km and 1.11 for 400 km).

The main result of this experiment is that *SS* precursors contain sufficient information on the discontinuities and the topography models can be obtained even without perfectly knowing the background elastic model, provided we are using Fréchet boundary kernels. These kernels capture essential sensitivity to the boundary interfaces, although the unmodelled 3-D velocity effects have a considerable impact on convergence.

4 CONCLUDING REMARKS

We have shown that adjoint tomography is an appropriate tool for imaging the topography of mantle discontinuities. Boundary sensitivity kernels (e.g. Dahlen 2005; Tromp *et al.* 2005), just as the volumetric ones are consistent with the measurements and therefore will converge towards a less biased model (Valentine & Trampert 2015). Approximate kernels cannot deal with the non-linear trade-off between elastic structure and topography (e.g. Bai *et al.* 2012; Koroni & Trampert 2016). Our main conclusion, therefore, is that *SS* precursors can retrieve depth variations of mantle discontinuities using FWI. The reliability of this inference depends on several factors specified below.

The main strength of adjoint tomography is that the kernels and data are mutually consistent with the full physics of wave propagation. The choice of measurements, therefore, has a profound influence on the nature of the kernel (Bozdağ *et al.* 2011) as we show as an illustration in Fig. 2. In a known background model, the cross-correlation and waveform misfit kernels, both retrieve the target topography model correctly, although the latter seems to converge slightly faster. In the seismological community, in contrast to the exploration one, the waveform misfit is not favoured due to a higher non-linearity between model parameters and waveforms. In a real data case, not as idealized as ours, waveform misfits are therefore probably not ideal. It is very likely that a cross-correlation misfit as a function of frequency or an instantaneous misfit are more beneficial. The optimization scheme also has a big influence (Nocedal & Wright 2000), but we have not explored this here. We only used a basic and simple steepest descent method, where the strength of the update is found by trial and error.

Many adjoint tomography practitioners favour a conjugate gradient method (e.g. Fletcher & Reeves 1964) or a quasi-Newton algorithm, with the most popular among them the limited memory BFGS (Liu & Nocedal 1989). These methods exhibit faster convergence because an optimal pre-conditioner is used. Examples of the use of various numerical optimization techniques can be found in, for example Fichtner *et al.* (2013), Afanasiev *et al.* (2016), Fichtner *et al.* (2018), Bozdağ *et al.* (2016), Tape *et al.* (2010), Blom *et al.* (2017), Simute *et al.* (2016) and Ermert *et al.* (2017).

In the case of a known elastic background model, 1-D or 3-D, the non-linear optimization converges towards the correct target model, although the convergence seems slower in the 3-D background. Because kernels and data are mutually consistent and honour the physics of the problem, the target model will always be retrieved. When approximate kernels are used, due to the non-linearity of the

trade-off between 3-D structure and tomography, complete convergence cannot be achieved (e.g. Bai *et al.* 2012; Koroni & Trampert 2016).

The most realistic case is probably the experiment where we do not fully know the elastic background model. We have chosen to look at the case where we know nothing about the 3-D elastic structure, known to have large focusing and defocusing effects on the waveforms. We observe clear signs of convergence, but the non-linear trade-offs between elastic structure and topography cannot completely be compensated. Because kernels and measurements are compatible, it seems that more information on the topography is extracted than in the case of approximate kernels (e.g. Bai *et al.* 2012; Koroni & Trampert 2016). The inferred models, clearly have a strong resemblance to the target model. A joint approach using data sensitive to elastic structure, together with data more sensitive to topography, where one alternates between 3-D structure and topography model updates, might be worth considering, similar to joint inversions for source and structure.

Although the convergence speed depends on the chosen misfit, the 670 km discontinuity topography seems to be better recovered than the 400 km one, using *SS* precursors. This is understandable when bearing in mind that for SS precursors, near the bounce points, S670S is only strongly affected by the the 670 km discontinuity, while S400S is affected by both. This has already been noted before by just inspecting seismograms from direct calculations in earth models with topography (Koroni & Trampert 2016).

For a real data inversion, where many factors can hinder the experimental set-up, the situation is obviously not as clear-cut as in our synthetic work here. First, the data coverage is bound to be less ideal than assumed in this work, with vast areas remaining unprobed by seismic data. Secondly, real data contain noise and the precursors will probably be even less pronounced in the waveforms. Many Earth properties we neglected will affect the results presented here, such as attenuation and earthquake source uncertainties. Incorporating processing techniques, which could enhance the precursor data, will surely be essential. Another option will be to use data which have a strong sensitivity to the discontinuities and are more easily observable, such as SSS, SSSS as already noted by Koroni *et al.* (2017). A judicial choice of the misfit function will also be essential. On the other hand, our knowledge on the 3-D earth structure is evolving rapidly (e.g. Bozdağ *et al.* 2016; Fichtner *et al.* 2018), meaning that we are probably closer to the situation where the background elastic structure is well known (and seismic sources are usually relocated in those models as well). In that case, our topography modelling experiments indicate that we should be able to extract robust information on the topography from precursor or other data. This is provided that we combine high quality measurements with an optimal numerical optimization algorithm and appropriate kernel smoothing.

REFERENCES

Afanasiev, M., Peter, D., Sager, K., Simute, S., Ermert, L., Krischer, L. & Fichtner, A., 2016. Foundations for a multiscale collaborative earth model, *Geophys. J. Int.,* **204**(1), 39–58.

Akaogi, M., Ito, E. & Navrotsky, A., 1989. Olivine-modified spinel-spinel transitions in the system Mg_2SiO_4-Fe_2SiO_4: calorimetric measurements, thermochemical calculation, and geophysical application, *J. geophys. Res.,* **94**(B11), 15 671–15 685.

Andrews, J. & Deuss, A., 2008. Detailed nature of the 660 km region of the mantle from global receiver function data, *J. geophys. Res.,* **113**(B6), doi:10.1029/2007JB005111.

Bai, L., Zhang, Y. & Ritsema, J., 2012. An analysis of *SS* precursors using spectral-element method seismograms, *Geophys. J. Int.,* **188,** 293–300.

Bassin, C., Laske, G. & Masters, G., 2000. The current limits of resolution for surface wave tomography in North America, *EOS, Trans. Am. geophys. Un.,* **81,** F897.

Benz, H. & Vidale, J., 1993. Sharpness of upper-mantle discontinuities determined from high frequency reflections, *Nature,* **365,** 147–150.

Blom, N., Böhm, C. & Fichtner, A., 2017. Synthetic inversions for density using seismic and gravity data, *Geophys. J. Int.,* **209,** 1204–1220.

Bock, G., 1994. Synthetic seismogram images of upper mantle structure: no evidence for a 520-km discontinuity, *J. geophys. Res.,* **99**(B8), 15 843–15 851.

Bonatto, L., Schimmel, M., Gallart, J. & Morales, J., 2013. Studying the 410-km and 660-km discontinuities beneath Spain and Morocco through detection of P-to-S conversions, *Geophys. J. Int.,* **194,** 920–935.

Bozdağ, E., Tromp, J. & Trampert, J., 2011. Misfit functions for full waveform inversion based on the instantaneous phase and envelope measurements, *Geophys. J. Int.,* **185,** 845–870.

Bozdağ, E., Peter, D., Komatitsch, M.L.D., Tromp, J., Hill, J., Podhorszki, N. & Pugmire, D., 2016. Global adjoint tomography: first generation model, *Geophys. J. Int.,* **207,** 1739–1766.

Chaljub, E. & Tarantola, A., 1997. Sensitivity of *SS* precursors to topography on the upper-mantle 660-km discontinuity, *Geophys. Res. Lett.,* **24**(21), 2613–2616.

Chambers, K., Deuss, A. & Woodhouse, J.H., 2005. Reflectivity of the 410-km discontinuity from PP and SS precursors, *J. geophys. Res.,* **110,** doi:10.1029/2004JB003345

Chevrot, S., Vinnik, L. & Montagner, J.-P., 1999. Global-scale analysis of the mantle *Pds* phases, *J. geophys. Res.,* **104**(B9), 20 203–20 219.

Crotwell, H.P., Owens, T.J. & Ritsema, J., 1999. The TauP ToolKit: flexible seismic travel-time and raypath utilities, *Seismol. Res. Lett.,* **70,** 154–160.

Dahlen, F., 2005. Finite-frequency sensitivity kernels for boundary topography perturbations, *Geophys. J. Int.,* **162,** 525–540.

Dahlen, F.-A., Hung, S.-H. & Nolet, G., 2000. Fréchet kernels for finite frequency traveltimes - I. theory, *Geophys. J. Int.,* **141,** 157–174.

Deuss, A., 2009. Global observations of mantle discontinuities using *SS* and *PP* precursors, *Surv. Geophys.,* **30,** 301–326.

Dziewoński, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.,* **25,** 297–356.

Ermert, L., Sager, K., Afanasiev, M., Boehm, C. & Fichtner, A., 2017. Ambient seismic source inversion in a heterogeneous earth: theory and application to the earth's hum, *J. geophys. Res.,* **122**(11), 9184–9207.

Estabrook, C.H. & Kind, R., 1996. The nature of the 660-kilometer upper-mantle seismic discontinuity from precursors to the *PP* phase, *Science,* **274**(5290), 1179–1182.

Fichtner, A., 2010. *Full Seismic Waveform Modelling and Inversion,* Springer Verlag.

Fichtner, A. & Trampert, J., 2011a. Hessian kernels of seismic data functionals based upon the adjoint techniques, *Geophys. J. Int.,* **185,** 775–798.

Fichtner, A. & Trampert, J., 2011b. Resolution analysis in full waveform inversion, *Geophys. J. Int.,* **185,** 1604–1624.

Fichtner, A., Kennett, B.L., Igel, H. & Bunge, H.-P., 2009. Full waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.,* **179,** 1703–1725.

Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y. & Villasenor, A., 2013. Multiscale full waveform inversion, *Geophys. J. Int.,* **194,** 534–556.

Fichtner, A. *et al.,* 2018. The collaborative seismic earth model: generation 1, *Geophys. Res Lett.,* **45**(9), 4007–4016.

Flanagan, M. & Shearer, P., 1998. Global mapping of topography on transition zone velocity discontinuities by stacking *SS* precursors, *J. geophys. Res.,* **103,** 2673–2692.

Flanagan, M. & Shearer, P., 1999. A map of topography of 410-km discontinuity from *PP* precursors, *Geophys. Res. Lett.,* **26**(5), 549–552.

Fletcher, R. & Reeves, C., 1964. Function minimization by conjugate gradients, *Comput. J.,* **7,** 149–154.

Frost, D.J., 2008. The upper mantle and transition zone, *Elements,* **4,** 171–176.

Gao, S.S. & H.Liu, K., 2014. Imaging mantle discontinuities using multiply-reflected *P-to-S* conversions, *Earth planet. Sci. Lett.,* **402,** 99–106.

Gu, Y.J & Dziewonski, A.M., 1998. Global de-correlation of the topography of transition zone discontinuities, *Earth planet. Sci. Lett.,* **157,** 57–67.

Gu, Y.J. & Dziewonski, A.M., 2002. Global variabilty of transition zone thickness, *J. geophys. Res.,* **107**(B7), 2135.

Gu, Y.J., Dziewonski, A.M. & Ekstrom, G., 2003. Simultaneous inversion for mantle shear velocity and topography of transition zone discontinuities, *Geophys. J. Int.,* **154,** 559–583.

Haugland, S.M., Ritsema, J., Sun, D., Trampert, J. & Koroni, M., 2019. Common reflection point mapping of the mantle transition zone using recorded and 3-D synthetic ScS reverberations, *Geophys. J. Int.,* **220**(1), 724–736.

Helffrich, G., 1999. Thermal variations in the mantle inferred from 660 km discontinuity topography and tomographic wave speed variations, *Geophys. J. Int.,* **151**(3), 935–943.

Houser, C., Masters, G., Flanagan, M. & Shearer, P., 2008. Determination and analysis of long-wavelength transition zone structure using *SS* precursors, *Geophys. J. Int.,* **174**(1), 178–194

Jenkins, J., Cottaar, S., White, R.S. & Deuss, A., 2016. Depressed mantle discontinuities beneath Iceland: evidence of a garnet controlled 660 km discontinuity, *Earth Planet. Sci. Lett.,* **453,** 159–168.

Katsura, T. & Ito, E., 1989. The system mg2sio4-fe2sio4 at high pressures and temperatures: precise determination of stabilities of olivine, modified spinel, and spinel, *J. geophys. Res.,* **94**(B11), 15 663–15 670.

Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral element method for three-dimensional seismic wave propagation, *Geophys. J. Int.,* **139**(3), 806–822.

Komatitsch, D. & Tromp, J., 2002a. Spectral-element simulations of global seismic wave propagation - I. Validation, *Geophys. J. Int.,* **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, *Geophys. J. Int.,* **150**(1), 303–318.

Komatitsch, D. & Tromp, J., 2002c. Propagación de ondas sísmicas en modelos de tierra global 3-D usando elementos espectrales, in *Métodos numéricos en ingeniería y ciencias aplicadas,* pp. 676–686, ed. Oñate, E. *et al.*, CIMNE.

Komatitsch, D. & Vilotte, J.P., 1998. The spectral-element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bull. seism. Soc. Am.,* **88**(2), 368–392.

Koroni, M. & Trampert, J., 2016. The effect of topography of upper mantle discontinuities on *SS* precursors, *Geophys. J. Int.,* **204,** 667–681.

Koroni, M., Bozdağ, E., Paulssen, H. & Trampert, J., 2017. Sensitivity analysis of seismic waveforms to upper-mantle discontinuities using the adjoint method, *Geophys. J. Int.,* **210,** 1965–1980.

Koroni, M., Paulssen, H. & Trampert, J., 2019. Sensitivity kernels of pp precursor traveltimes and their limitations for imaging topography of discontinuities, *Geophys. Res. Lett.,* **46**(2), 698–707.

Kustowski, B., Ekström, G. & Dziewoński, A.M., 2008. Anisotropic shear-wave velocity structure of the earth's mantle: a global model, *J. geophys. Res.,* **113**(B6), doi:10.1029/2007JB005169.

Lawrence, J.F. & Shearer, P.M., 2006. A global study of transition zone thickness using receiver functions, *J. geophys. Res.,* **111**(B06307), doi:10.1029/2005JB003973.

Lawrence, J.F. & Shearer, P.M., 2008. Imaging mantle transition zone thickness with *SdS-SS* finite-frequency sensitivity kernels, *Geophys. J. Int.,* **174,** 143–158.

Lebedev, S., Chevrot, S. & van der Hilst, R., 2003. Correlation between the shear-speed structure and thickness of the mantle transition zone, *Phys. Earth planet. Inter.,* **136,** 25–40.

Liu, D. & Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization, *Math. Program.,* **45,** 503–528.

Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bull. seism. Soc. Am.,* **96**(6), 2383–2397.

Liu, Q. & Tromp, J., 2008. Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, *Geophys. J. Int.,* **174**(1), 265–286.

Luo, Y. & Schuster, G., 1991. Wave-equation traveltime inversion, *Geophysics,* **56**(5), doi:10.1190/1.1443081.

Marquering, H., Nolet, G. & Dahlen, F., 1999. Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, *Geophys. J. Int.,* **137,** 805–815.

Meier, U., Trampert, J. & Curtis, A., 2009. Global variations of temperature and water content in the mantle transition zone from higher mode surface waves, *Earth planet. Sci. Lett.,* **282,** 91–101.

Neele, F., de Regt, H. & VanDecar, J., 1997. Gross errors in upper-mantle discontinuity topography from underside reflection data, *Geophys. J. Int.,* **129,** 194–204.

Nocedal, J. & Wright, S., 2000. *Numerical Optimization,* Springer.

Patera, A.T., 1984. A spectral element method for fluid dynamics: laminar flow in a channel expansion, *J. Comput. Phys.,* **54,** 468–488.

Paulssen, H., 1988. Evidence for a sharp 670-km discontinuity as inferred from *P-to-S* converted waves, *J. geophys. Res.,* **93,** 10 489–10 500.

Poli, P., Campillo, M., Pedersen, H. & Group, L.W., 2012. Body-wave imaging of earth's mantle discontinuities from ambient seismic noise, *Science,* **338,** 1063–1065.

Revenaugh, J. & Jordan, T., 1991c. Mantle layering from ScS reverberations, *J. geophys. Res.,* **96**(B12), 19 781–19 810.

Ritsema, J., Van Heijst, H.J. & Woodhouse, J.H., 1999. Complex shear velocity structure imaged beneath Africa and Iceland, *Science,* **286,** 1925–1928.

Ritsema, J., Deuss, A., Van Heijst, H.J. & Woodhouse, J.H., 2011. S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophys. J. Int.,* **184**(3), 1223–1236.

Schmerr, N. & Garnero, E., 2006. Investigation of upper mantle discontinuity structure beneath the central pacific using *SS* precursors, *J. geophys. Res.,* **111**(B08305), doi:10.1029/2005JB004197.

Shearer, P., 1993. Global mapping of upper mantle reflectors from long-period *SS* precursors, *Geophys. J. Int.,* **115,** 878–904.

Shearer, P. & Masters, G., 1992. Global mapping of topography on 660-km discontinuity, *Nature,* **355,** 791–796.

Shearer, P., Flanagan, M.P. & Hedin, A., 1999. Experiments in migration processing of *SS* precursor data to image upper mantle discontinuity structure, *J. geophys. Res.,* **104,** 7229–7242.

Shearer, P.M., 1990. Seismic imaging of upper mantle structure and new evidence for a 520 km discontinuity, *Nature,* **344,** 121–126.

Shearer, P.M., 1991. Constraints on upper mantle discontinuities from observations of long-period reflected and converted phases, *J. geophys. Res.,* **96,** 18 147–18 182.

Shearer, P.M., 1996. Transition zone velocity gradients and the 520-km discontinuity, *J. geophys. Res.,* **101,** 3053–3066.

Shearer, P.M., 2000. Upper mantle seismic discontinuities, in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale,* Vol. 117 of Geophysical Monograph, American Geophysical Union.

Simute, S., Steptoe, H., Cobden, L., Gokhberg, A. & Fichtner, A., 2016. Full-waveform inversion of the Japanese islands region, *J. geophys. Res.,* **121**(5), 3722–3741.

Tackley, P.J., Stevenson, D., Glatzmaier, G. & Schubert, G., 1994. Effects of multiple phase transitions in a three-dimensional spherical model of convection in earth's mantle, *J. geophys. Res.,* **99,** 15 877–15 901.

Talagrand, O. & Courtier, P., 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: theory, *Q. J. R. Meteorol. Soc.,* **113,** 1311–1328.

Tape, C., Liu, Q. & Tromp, J., 2007. Finite-frequency tomography using adjoint methods - methodology and examples using membrane surface waves, *Geophys. J. Int.,* **168**(3), 1105–1129.

Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2010. Seismic tomography of the southern California crust based on spectral-element and adjoint methods, *Geophys. J. Int.,* **180,** 433–462.

Tauzin, B., Debayle, E. & Wittlinger, G., 2008. The mantle transition zone as seen by global *Pds* phases: no clear evidence for a thin transition zone beneath hotspots, *J. geophys. Res.,* **113,** doi:10.1029/2007JB005364.

Tauzin, B., van der Hilst, R., Wittlinger, G. & Ricard, Y., 2013. Multiple transition zone seismic discontinuities and low velocity layers below western United States, *J. geophys. Res.,* **118,** 2307–2322.

Trampert, J. & Snieder, R., 1996. Model estimations by truncated expansions: possible artifacts in seismic tomography, *Science,* **271,** 1257–1260.

Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160**(1), 195–216.

Valentine, A. & Trampert, J., 2015. The impact of approximations and arbitrary choices on geophysical images, *Geophys. J. Int.,* **204,** 59–73.

van der Lee, S., Paulssen, H. & Nolet, G., 1994. Variability of *P660s* phases as a consequence of topography of the 660 km discontinuity, *Phys. Earth planet. Inter.,* **86,** 147–164.

Vinnik, L.P., 1977. Detection of waves converted from P to SV in the mantle, *Phys. Earth planet. Inter.,* **15,** 39–45.

Xu, F., Vidale, J., Earle, P. & Benz, H., 1998. Mantle discontinuities under Southern Africa from precursors to P'P'˳df, *Geophys. Res. Lett.,* **25,** 571–574.

Zhao, L. & Chevrot, S., 2003. SS-wave sensitivity to upper mantle structure: implications for the mapping of transition zone discontinuity topographies, *Geophys. Res. Lett.,* **30**(11), 1590.
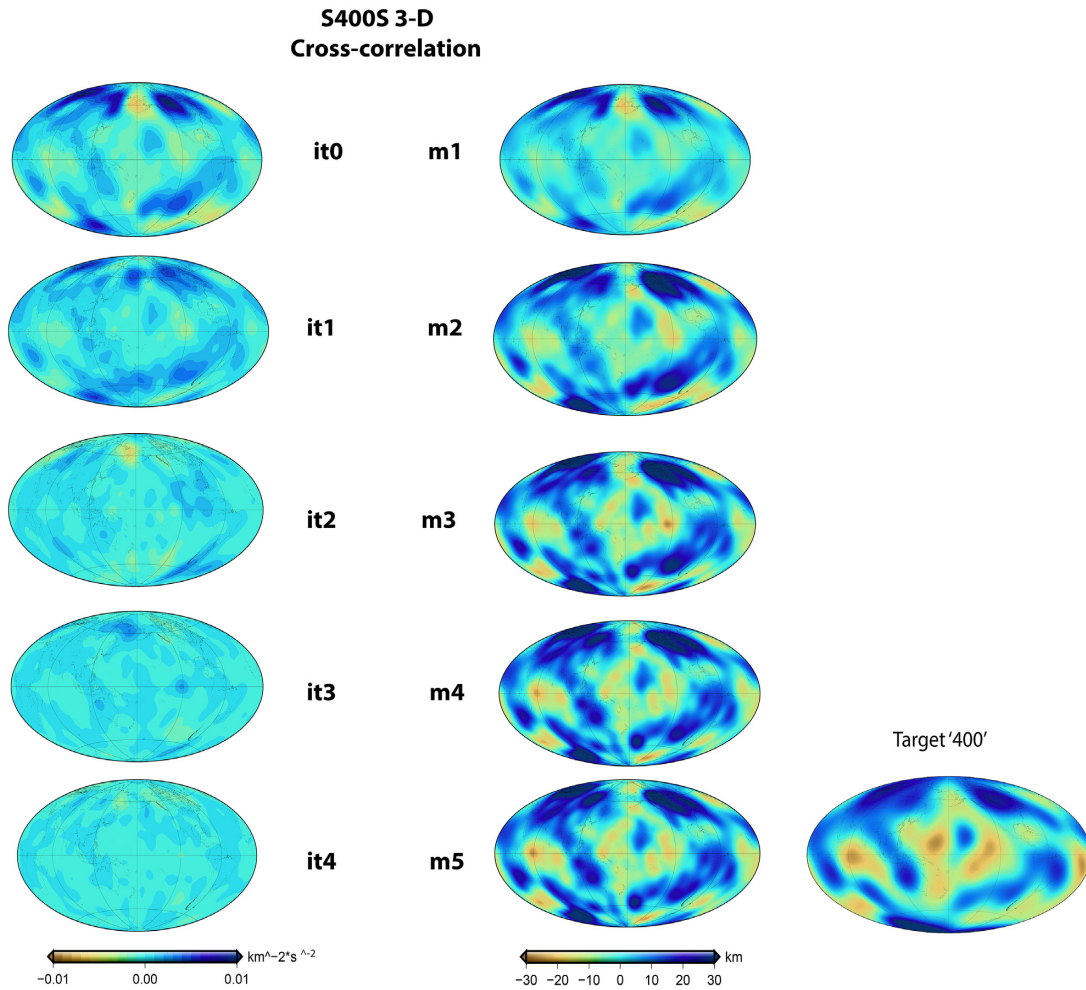
APPENDIX

A1 Successive model updates
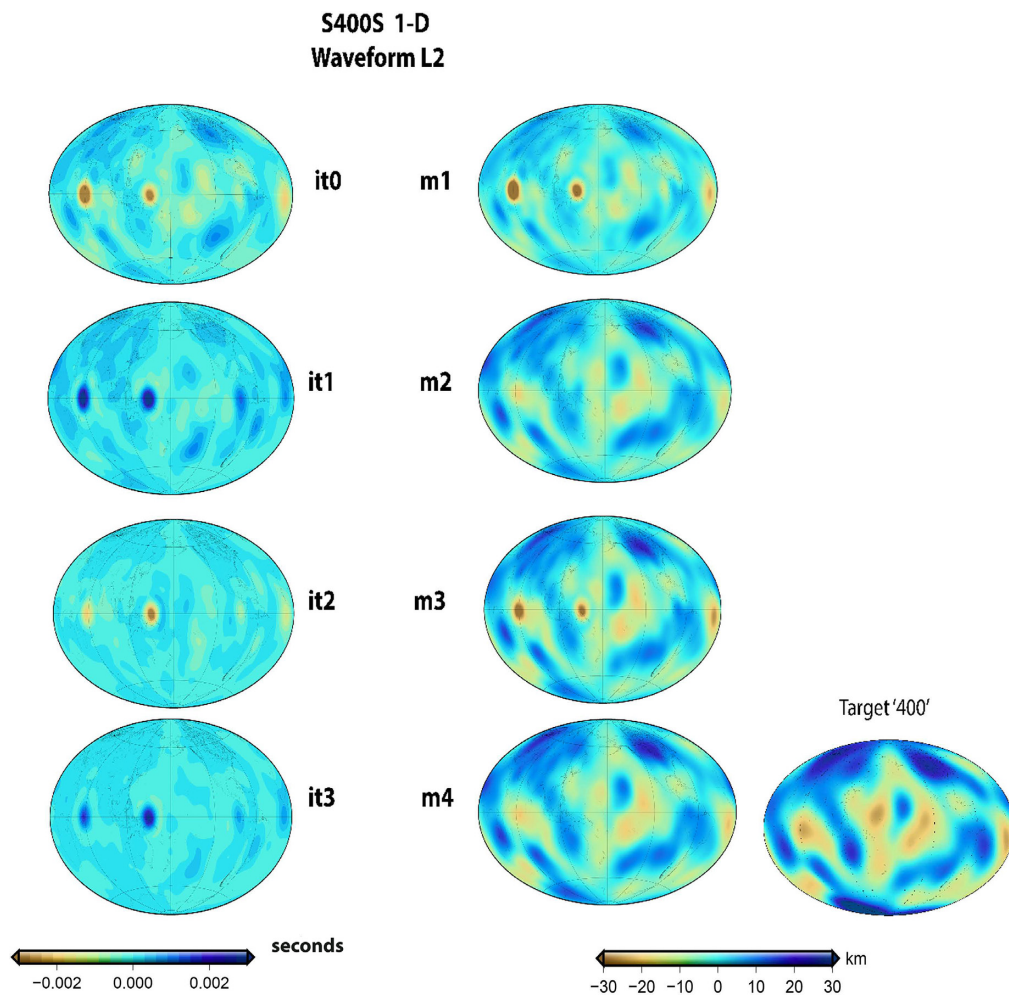
Figure A1. Successive smoothed boundary misfit kernels (left-hand column) per iteration for a waveform misfit in a 1-D elastic background model. The misfit is measured in the *S400S* time windows. The units of the boundary misfit kernels are seconds. In the right-hand column, the five successive model updates for the 400 km discontinuity are shown. The target model is shown to the right of the last update.
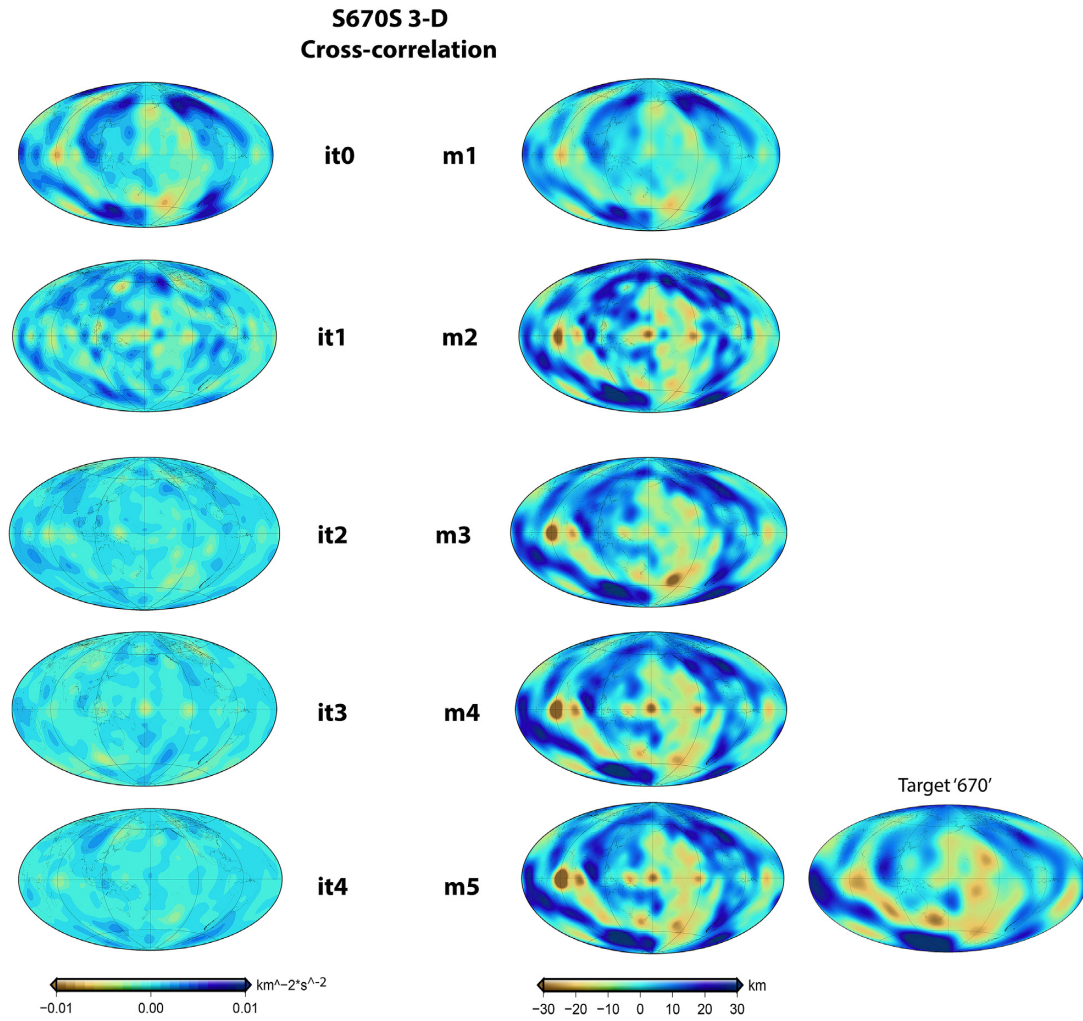
S670S 1-D
Waveform L2
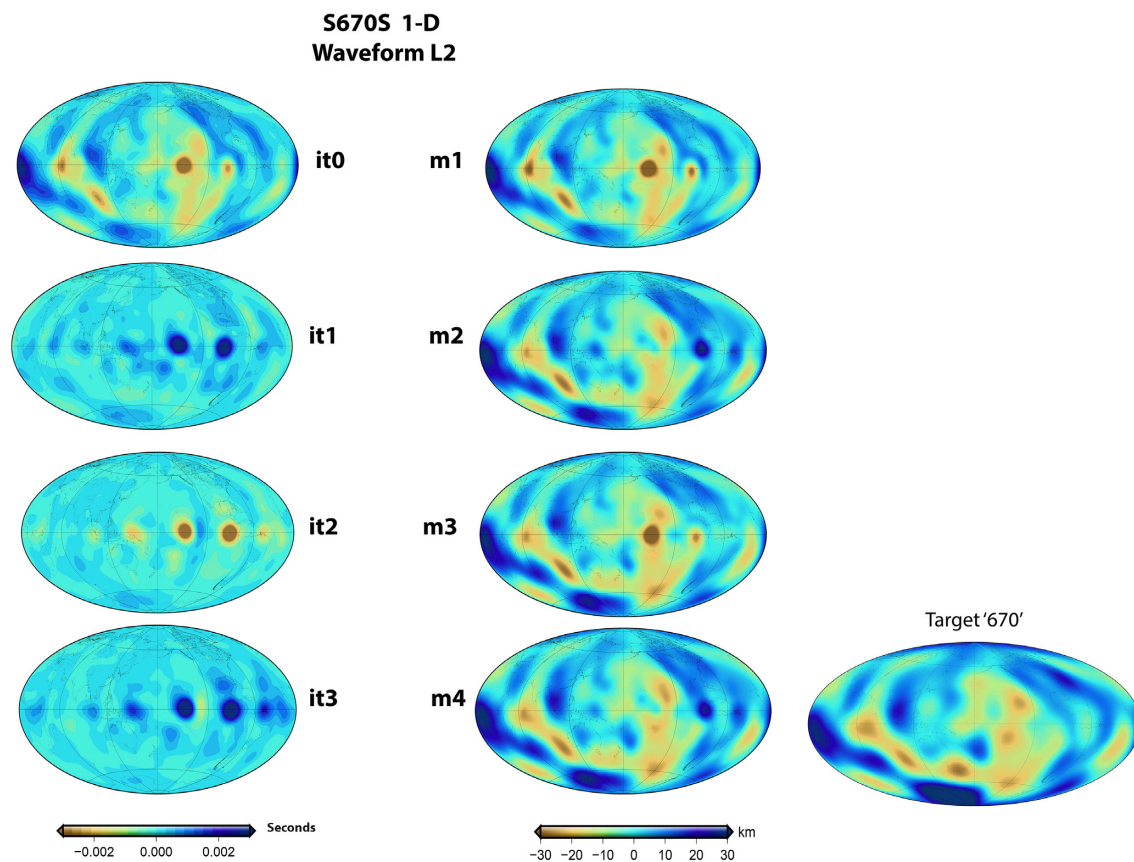
Figure A2. The same as Fig. A1, but for the 670 km discontinuity.
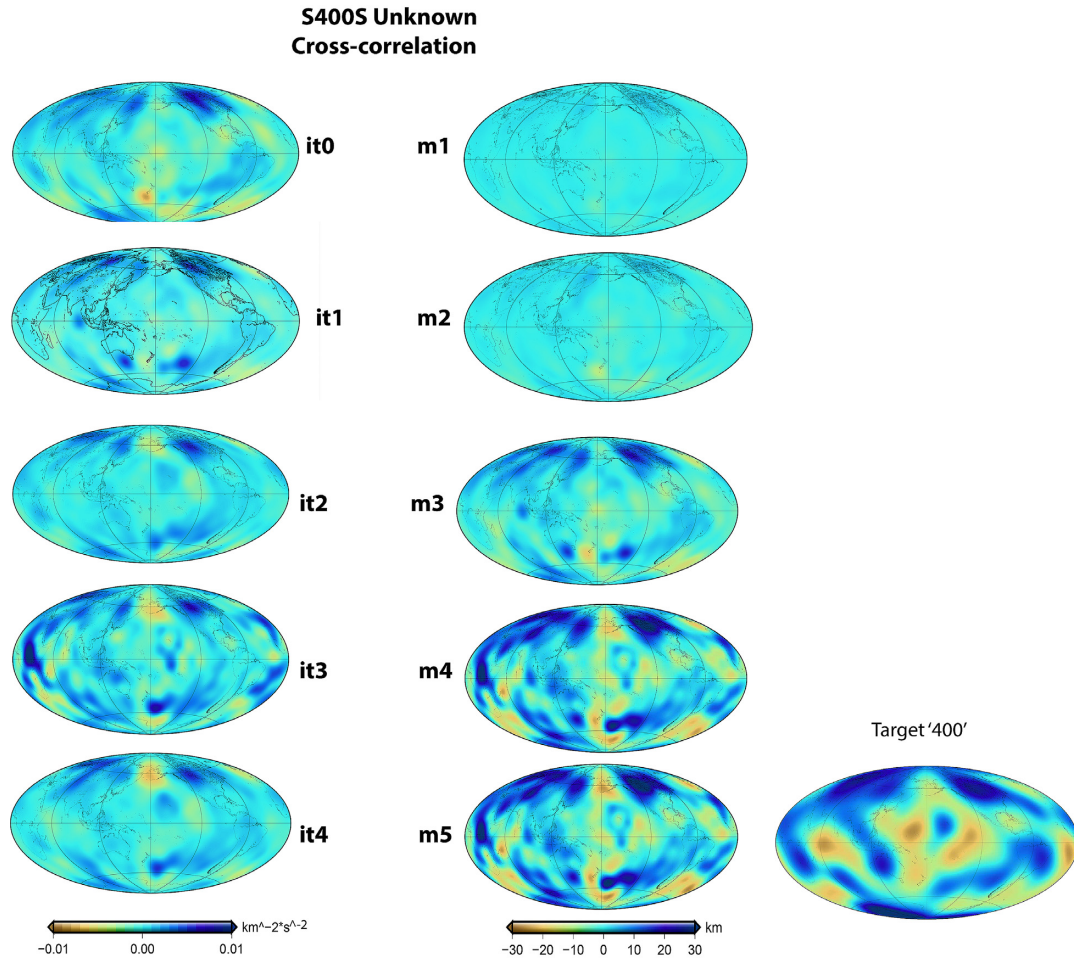
**S400S 3-D
Cross-correlation**

Figure A3. Successive smoothed boundary misfit kernels (left-hand column) per iteration for a cross-correlation misfit in a 3-D elastic background model. The misfit is measured in the *S400S* time windows. The units of the boundary misfit kernels are $s^{-2} \cdot km^{-2}$. In the right-hand column, the five successive model updates for the 400 km discontinuity are shown. The target model is shown to the right of the last update.

**S670S 3-D
Cross-correlation**

Figure A4. The same as for Fig. A3, but for the 670 km discontinuity.

S400S Unknown
Cross-correlation

Figure A5. Successive smoothed boundary misfit kernels (left-hand column) per iteration for a cross-correlation misfit in a fixed 1-D elastic background model. The data however are calculated in a 3-D elastic background model. This test simulates an unknown background model. The misfit is measured in the *S400S* time windows. The units of the boundary misfit kernels are $s^{-2} \cdot km^{-2}$. In the right-hand column, the five successive model updates for the 400 km discontinuity are shown. The target model is shown to the right of the last update.
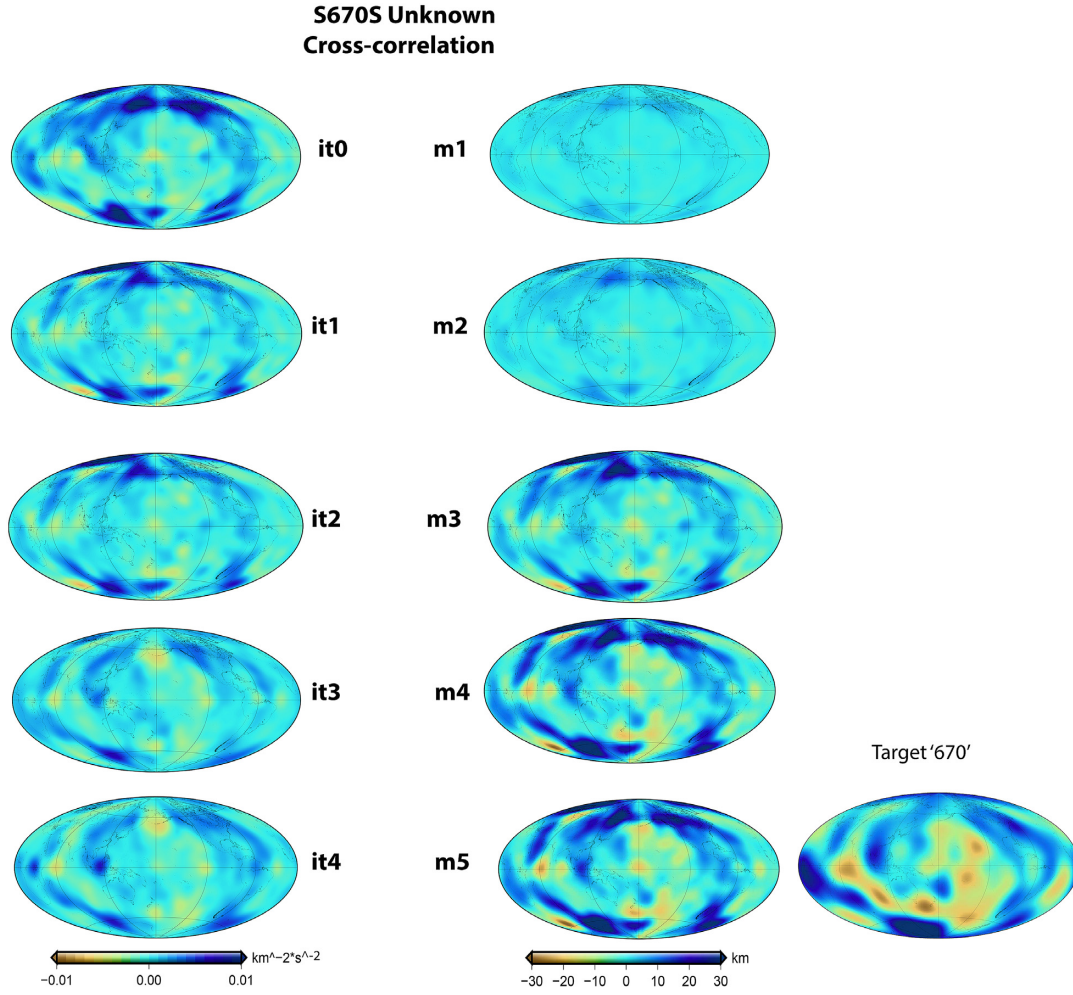
Figure A6. The same as for Fig. A5, but for the 670 km discontinuity.

A2 Topography updates in SPECFEM3D_GLOBE

SPECFEM3D_GLOBE provides relative radius boundary kernels expressed as:

$$d \ln h_k = \ln h_{k+1} - \ln h_k = \ln \frac{h_{k+1}}{h_k}, \tag{A1}$$

h stands for the radius of the model discontinuity, k is the iteration number. The update then becomes:

$$\ln \frac{h_{k+1}}{h_k} = \alpha p_k, \tag{A2}$$

where α is the step length and is a real positive constant and $p_k - \delta \chi_k$ is the smoothed descent direction. SPECFEM3D_GLOBE work with absolute depth perturbations dz from the reference model, rather than radius perturbations, that is d$h = -$dz.

$$h_{k+1} = h_{\text{PREM}} - \text{d}z_{k+1} \tag{A3}$$

and

$$h_k = h_{\text{PREM}} - \text{d}z_k \tag{A4}$$

Eq. (A4) then becomes:

$$r_{\text{PREM}} - \text{d}z_{k+1} = (h_{\text{PREM}} - \text{d}z_k) \exp(\alpha p_k) \tag{A5}$$

or

$$\text{d}z_{k+1} = h_{\text{PREM}} - (h_{\text{PREM}} - \text{d}z_k) \exp(\alpha p_k). \tag{A6}$$

The reference depth for the perturbations is h_{PREM}, that is 400 and 670 km. The resulting models consist of spherical harmonic coefficients up to degree 20 that represent absolute depth perturbations with respect to the reference depth in PREM.