



**HAL**  
open science

# Accelerating full waveform inversion via source stacking and cross-correlations

Barbara Romanowicz, Li-Wei Chen, Scott W. French

► **To cite this version:**

Barbara Romanowicz, Li-Wei Chen, Scott W. French. Accelerating full waveform inversion via source stacking and cross-correlations. *Geophysical Journal International*, 2020, 220, pp.308-322. 10.1093/gji/ggz437 . insu-03584816

**HAL Id: insu-03584816**

**<https://insu.hal.science/insu-03584816>**

Submitted on 24 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Accelerating full waveform inversion via source stacking and cross-correlations

Barbara Romanowicz,<sup>1,2,3</sup> Li-Wei Chen<sup>1</sup> and Scott W. French<sup>1,4</sup>

<sup>1</sup>University of California, Department of Earth and Planetary Science, Berkeley, CA, USA

<sup>2</sup>Collège de France, Paris, France

<sup>3</sup>Institut de Physique du Globe de Paris, France. E-mail: [barbara.romanowicz@gmail.com](mailto:barbara.romanowicz@gmail.com)

<sup>4</sup>Now at Google Inc., Mountainview, CA, USA

Accepted 2019 September 16. Received 2019 August 9; in original form 2019 February 18

## SUMMARY

Accurate synthetic seismic wavefields can now be computed in 3-D earth models using the spectral element method (SEM), which helps improve resolution in full waveform global tomography. However, computational costs are still a challenge. These costs can be reduced by implementing a source stacking method, in which multiple earthquake sources are simultaneously triggered in only one teleseismic SEM simulation. One drawback of this approach is the perceived loss of resolution at depth, in particular because high-amplitude fundamental mode surface waves dominate the summed waveforms, without the possibility of windowing and weighting as in conventional waveform tomography.

This can be addressed by redefining the cost-function and computing the cross-correlation wavefield between pairs of stations before each inversion iteration. While the Green's function between the two stations is not reconstructed as well as in the case of ambient noise tomography, where sources are distributed more uniformly around the globe, this is not a drawback, since the same processing is applied to the 3-D synthetics and to the data, and the source parameters are known to a good approximation. By doing so, we can separate time windows with large energy arrivals corresponding to fundamental mode surface waves. This opens the possibility of designing a weighting scheme to bring out the contribution of overtones and body waves. It also makes it possible to balance the contributions of frequently sampled paths versus rarely sampled ones, as in more conventional tomography.

Here we present the results of proof of concept testing of such an approach for a synthetic 3-component long period waveform data set (periods longer than 60 s), computed for 273 globally distributed events in a simple toy 3-D radially anisotropic upper mantle model which contains shear wave anomalies at different scales. We compare the results of inversion of 10 000 s long stacked time-series, starting from a 1-D model, using source stacked waveforms and station-pair cross-correlations of these stacked waveforms in the definition of the cost function. We compute the gradient and the Hessian using normal mode perturbation theory, which avoids the problem of cross-talk encountered when forming the gradient using an adjoint approach. We perform inversions with and without realistic noise added and show that the model can be recovered equally well using one or the other cost function.

The proposed approach is computationally very efficient. While application to more realistic synthetic data sets is beyond the scope of this paper, as well as to real data, since that requires additional steps to account for such issues as missing data, we illustrate how this methodology can help inform first order questions such as model resolution in the presence of noise, and trade-offs between different physical parameters (anisotropy, attenuation, crustal structure, etc.) that would be computationally very costly to address adequately, when using conventional full waveform tomography based on single-event wavefield computations.

**Key words:** Seismic Tomography; Computational Seismology; Waveform Inversion.

## 1 INTRODUCTION

Global seismic tomography traditionally relies on constraints from traveltimes of a limited number of seismic ‘phases’, which allow very fast forward computations, even when finite frequency kernels are considered. The seismic phases considered are those that are easy to isolate on seismograms, such as fundamental mode surface waves or teleseismic  $P$  and  $S$  waves, and in some studies, their surface and core reflections, leaving a large part of the information contained in seismic records unexploited.

On the other hand, full waveform inversion should ultimately make it possible to exploit every wiggle in a seismogram. It was first introduced by Tarantola (1984) in the world of exploration geophysics and by Woodhouse & Dziewonski (1984) for the study of global earth structure. In this paper, we focus on global structure applications using natural earthquake sources.

Full waveform inversion requires the computation of synthetic seismograms in the current 3-D earth model, at each iteration of the model. Woodhouse & Dziewonski (1984) derived this in the context of the simplest asymptotic approximation to first order normal mode perturbation theory, that was shown to be equivalent to the ‘path-average’ approximation (PAVA) used in surface wave dispersion studies (Mochizuki 1986; Park 1987; Romanowicz 1987). This approximation assumes that the sensitivity of the seismic waves is limited to the 1-D average structure between the source and the receiver. A similar approach, also based on the PAVA, but differing in the way data are processed, was introduced by Nolet (1990), the so-called ‘partitioned waveform inversion’. These approaches have been widely used, continuing up to this day, sometimes in conjunction with body wave traveltimes, resulting in several generations of global shear wave velocity models of the whole mantle (e.g. Su & Dziewonski 1997; Kustowski *et al.* 2008; Schaeffer & Lebedev 2013; Moulik & Ekström 2014; Durand *et al.* 2017) or the upper mantle (e.g. Debayle & Ricard 2012; Schaeffer & Lebedev 2014).

The PAVA approach is computationally fast, but it is not strictly appropriate for modelling body waves because it does not account for the concentration of their sensitivity along the infinite frequency, infinitesimal curved ray path. Li & Tanimoto (1993) showed how the sensitivity of body waves to structure may be accounted for more accurately by including coupling across mode branches (not included in PAVA), while Li & Romanowicz (1995) combined this with the PAVA to develop the NACT (non-linear asymptotic coupling theory). This allowed these authors to construct whole mantle global shear velocity models, based entirely on long period waveforms, including surface waves, overtones as well as body waves down to 30 s period (e.g. Li & Romanowicz 1996; Mégnin & Romanowicz 2000; Panning & Romanowicz 2006).

Nevertheless, first order normal mode perturbation theory has its limitations, in that it is based on the assumption that heterogeneities are smooth (i.e. long wavelength) and relatively weak. There are also effects of wave propagation outside of the great circle path, due to smaller scale heterogeneity, that causes amplitude variations (e.g. Ruan & Zhou 2012) not accounted for within either PAVA or NACT theory. A review of the relative strengths and weaknesses of these theories can be found in Romanowicz *et al.* (2008).

As higher resolution is sought to resolve finer scale structure across the mantle, two different approaches have been proposed. One approach posits that massive amounts of data, as are now available from global and regional scale high quality broadband networks installed around the world over the last 30 yrs, can compensate for errors in the wave propagation theory used, and there is certainly still much to be done with the efficient approximate computations

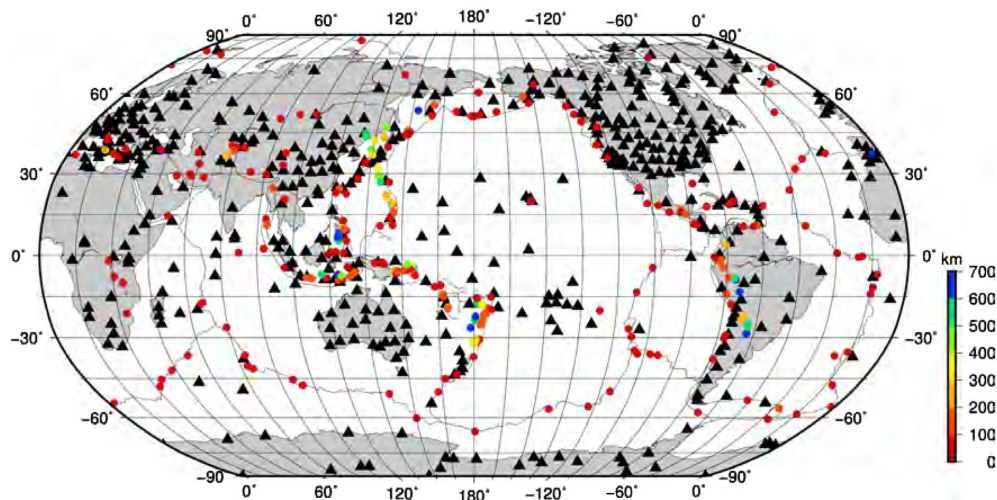
of the wavefield using ray theory or PAVA. The other approach considers that, in order to fully exploit the information contained in the scattered wavefield as well as such effects as wave front healing, and thus attain higher resolution, it is necessary to compute the wavefield more accurately. This has become possible recently, with the introduction to global seismology of the spectral element method (SEM, e.g. Komatitsch & Vilotte 1998; Komatitsch & Tromp 1999, 2002), an efficient numerical algorithm that allows the computation of accurate regional and teleseismic wavefields in 3-D earth models containing realistic wavelengths and strengths of heterogeneity. While efficient compared to other numerical methods, the computational cost is much higher than ray theory or PAVA/NACT, in particular as it increases roughly as the cube of the highest frequency considered.

An additional computational challenge comes with the use of an adjoint methodology for the inverse part of the modelling, which requires additional SEM wavefield computations to obtain the ‘exact’ gradient (Tromp *et al.* 2005), while convergence in the absence of a good approximation to the Hessian remains slow. Therefore, modelling based on a completely numerical approach to full waveform tomography has been so far applied mostly at the regional or continental scale (e.g. Fichtner & Trampert 2011; Rickers *et al.* 2013; Zhu *et al.* 2015, 2017), with application to global whole mantle shear velocity models emerging only recently (Bozdag *et al.* 2016).

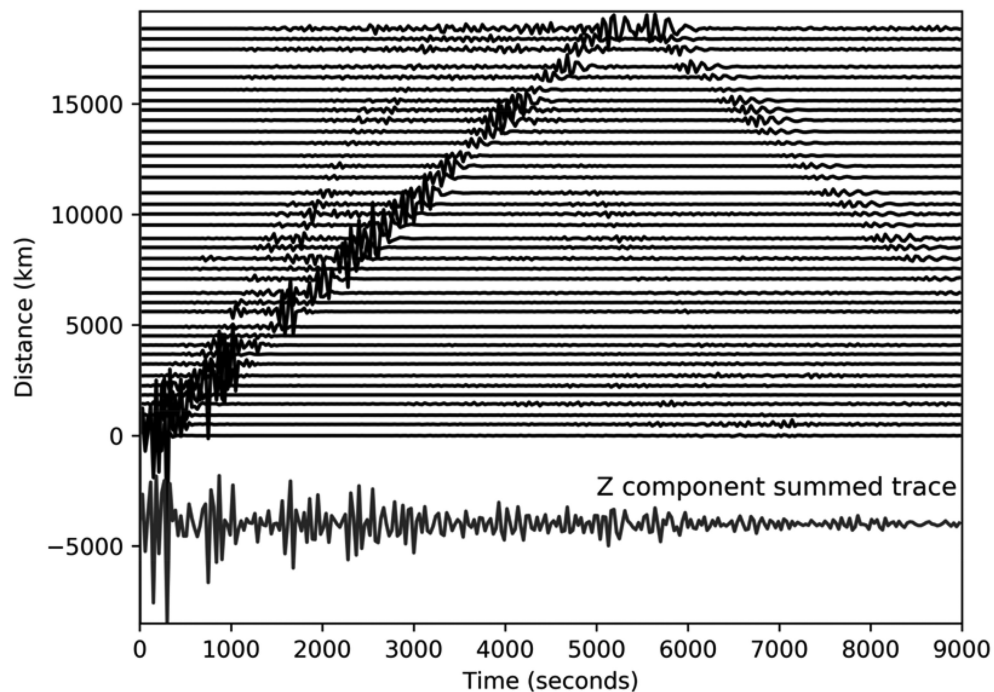
The computations can be significantly reduced, if only the forward wavefield is computed using the SEM, while Frechet kernels are computed using the approximate NACT, which also allows the computation of an approximation to the full Hessian. Three generations of global models have so far been developed based on this computationally more efficient approach in which SEM is used for the forward computation and NACT for the calculation of the gradient and Hessian (Lekic & Romanowicz 2011; French *et al.* 2013; French & Romanowicz 2014), revealing intriguing low velocity fingers in the oceanic upper mantle, and broad plumes beneath major hotspot volcanoes in the lower mantle (French & Romanowicz 2015).

Various techniques have been or are being developed to render full waveform tomography based on SEM (and other numerical methods) more efficient, such as assuming cylindrical symmetry, as in the code AxiSEM (Nissen-Meyer *et al.* 2014), more recently extended to allow slowly varying structure outside of the great circle plane, or hybrid methods, that couple efficient approximations outside of a target region with exact computations within that region (e.g. Kawai *et al.* 2014; Monteiller *et al.* 2015) and most recently, ‘box tomography’, which takes advantage of the reciprocity theorem to compute the 3-D teleseismic wavefield only once, and confine iterations to a target region (Masson *et al.* 2014; Masson & Romanowicz 2017; Clouzet *et al.* 2018; Masson & Romanowicz 2018). Similar approaches have recently been proposed in Wang *et al.* (2016) and Beller *et al.* (2018). These promising directions make it possible to increase the maximum frequency to which the wavefield is computed, and therefore achieve higher resolution in target areas of the mantle, in the near future.

One of the challenges of numerical wavefield computation is that the wavefield needs to be computed for a large number of events at each iteration. French & Romanowicz (2014) used 273 globally distributed earthquakes of magnitude  $M_w$  6–7. Altogether, starting from a 1-D model (to avoid biases due to pre-assumed 3-D structure) and taking into account the successive construction of upper mantle models SEMum (Lekic & Romanowicz 2011) and SEMum2 (French *et al.* 2013), the building of whole mantle radially anisotropic shear velocity model SEMUCB-WM1 (French &



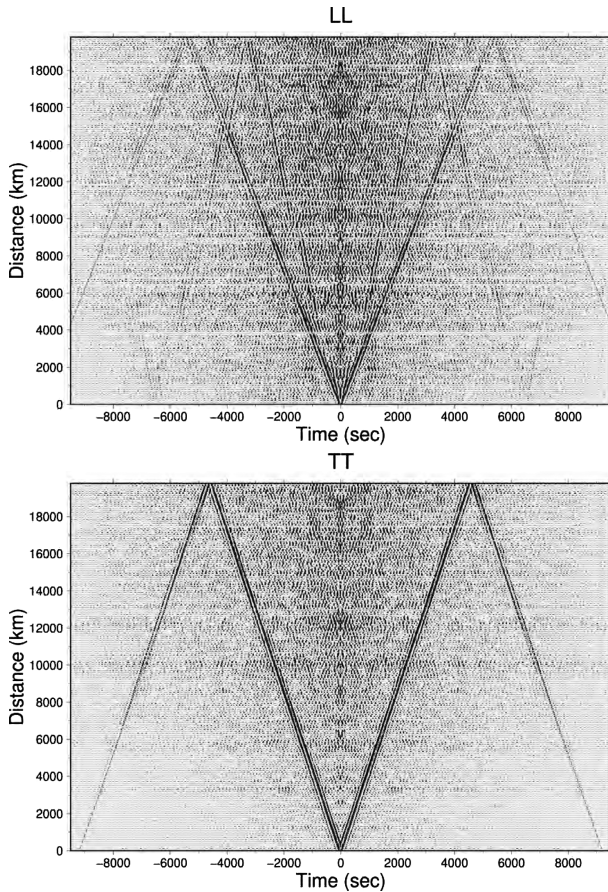
**Figure 1.** The distribution of events and stations used for the synthetic experiments presented in this paper. It consists of the 273 earthquake sources and 537 stations used for the development of global mantle model SEMUCB-WM1. The colours indicate the depths of sources.



**Figure 2.** An example of vertical component 10 000-s long traces calculated at station TATO in the 1-D average model derived from SEMUCB-WM1 (French & Romanowicz 2014). 38 different events are shown here, aligned on origin time and ordered as a function of epicentral distance. The frequency band of the calculation is 400–60 s. The bottom trace is the sum of the 38 aligned traces.

Romanowicz 2014) took 13 iterations. This is significantly fewer iterations than necessary when using the adjoint method, owing to the faster quadratic convergence that makes use of the full Hessian, however, in practice, it still took almost a decade, including all intermediate steps, adjustments to the approach used and such, to develop the whole mantle model SEMUCB-WM1. Notably, it is very challenging to assess model uncertainty and resolution in a rigorous manner without redoing all the iteration steps on synthetic models, a daunting proposition. This is even more challenging for adjoint-based tomography (e.g. Fichtner & Trampert 2011).

Here, we present and illustrate an approach, based on source stacking, as first introduced by Capdeville *et al.* (2005), which can reduce computational time by at least an order of magnitude. In this approach, many different sources are first aligned at their origin time, normalized by their seismic moment. The resulting summed wavefield is computed only once per iteration and compared with the corresponding observed summed wavefield at each station. Capdeville *et al.* (2005) showed that this could work with a relatively small number of events and stations, for the retrieval of very long wavelength whole mantle structure. This approach however, was not pursued further, at least in global seismology, due to



**Figure 3.** Stacked L–L (top panel) and T–T (bottom panel) cross-correlations between station pairs for the entire synthetic data set. The strongest energy corresponds to group velocities of  $3.75 \text{ km s}^{-1}$  for L–L and  $4.35 \text{ km s}^{-1}$  for T–T, which are in the middle range of group velocities for fundamental mode Rayleigh and Love waves, respectively.

concerns that information was lost in the process of summation, and that the summed records are dominated by large amplitude fundamental mode surface waves, without the possibility of relative weighting of body waves, overtone and fundamental mode surface waves, nor weighting of redundant paths, as can be done when working with single records (e.g. Li & Romanowicz 1996; Maggi *et al.* 2009).

In exploration geophysics, such an approach was perfected and is still an active area of research. It was shown in particular that ‘source encoding’, that is the use of multiple sums over the same group of events, to which a random perturbation to the source phase is applied, can help improve the resolution of the inverted model (e.g. Krebs *et al.* 2009; Schuster *et al.* 2011)—particularly by reducing ‘cross-talk’, that is, the appearance of interference terms in the computation of the gradient by the adjoint method, which introduces noise in the solution. We note that this is not an issue when computing the gradient and Hessian using NACT, so no source encoding should be necessary.

As noted above, another key concern when dealing with summed traces is the risk of information loss. However, information needs not be lost when using source encoding, as illustrated in Zhang *et al.* (2018) where the authors use a source encoding method that allows them to decode single records from the sum, albeit at the particular frequency used for encoding. Recovering the complete

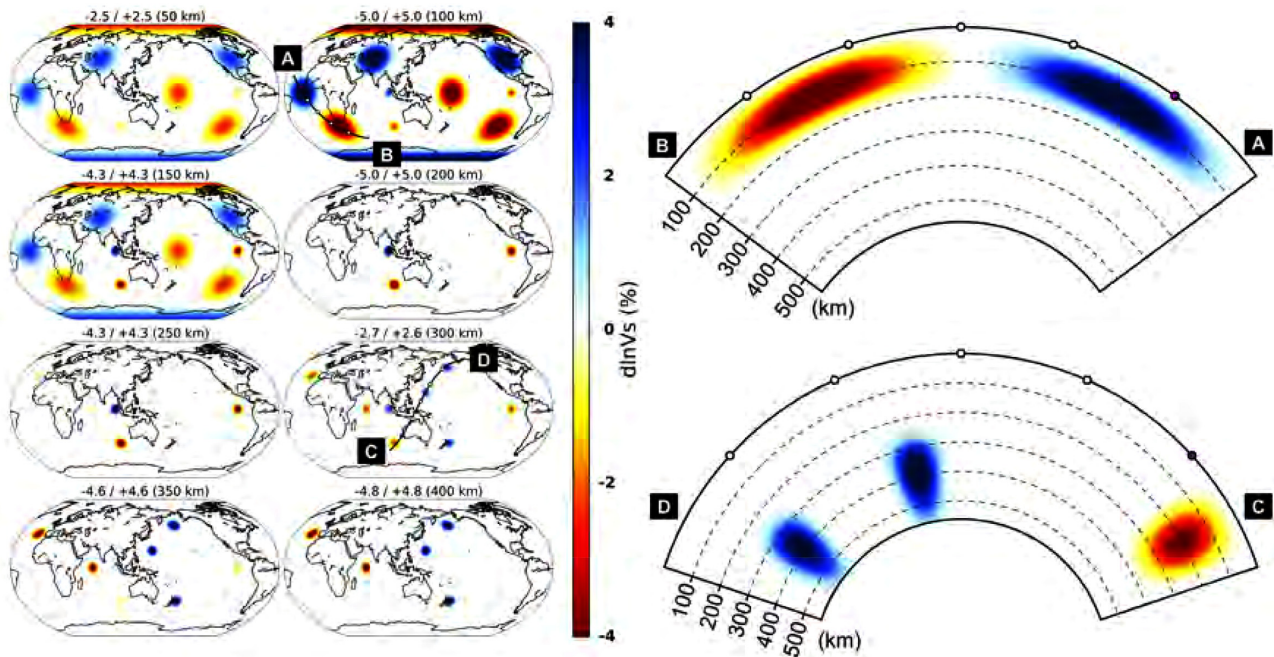
record requires multiple runs with different frequency encoding, which can be done, but at the expense of increasing computational time.

### 1.1 Proposed approach

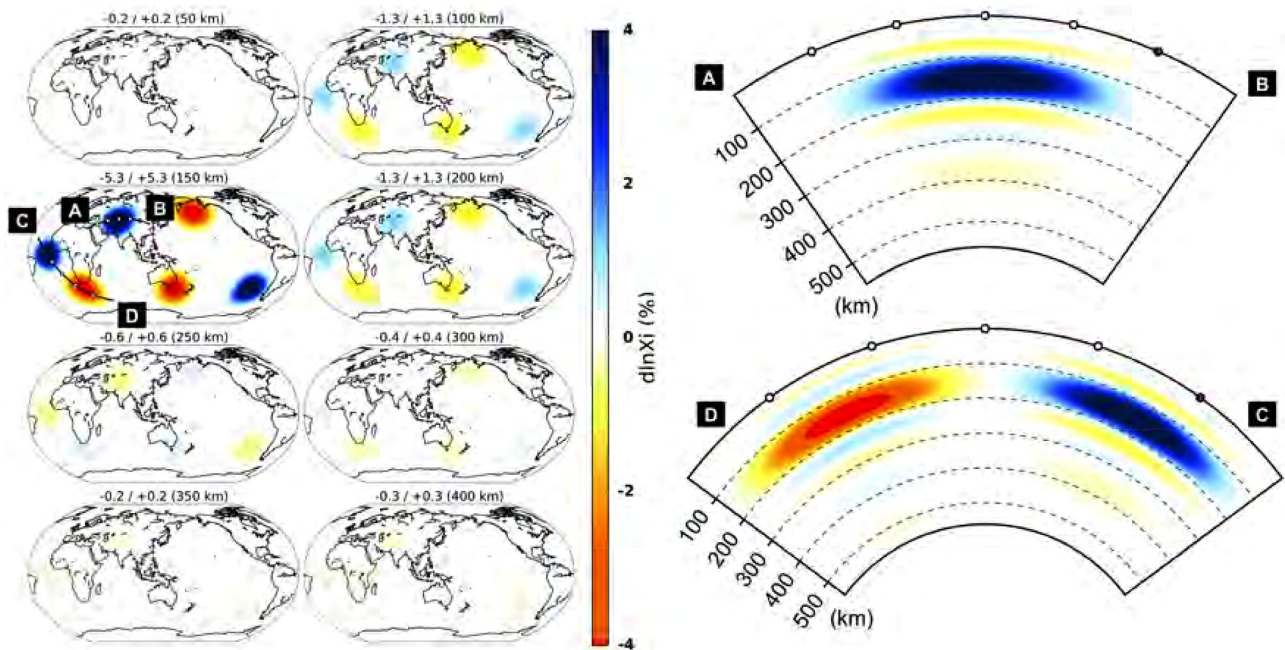
In this proof-of-concept paper, we further develop the idea of source stacking for global seismology, where we propose to account for the fact that surface waves cannot be down-weighted in the summed seismograms. For this, we borrow from the now widely popular field of ambient noise tomography (ANT, e.g. Shapiro & Campillo 2004). Indeed, it has been known since the work of Aki (1957), that the Green’s function between two stations can be retrieved from the cross-correlation of the noise wavefield at these two stations, provided the noise sources are uniformly distributed. The resulting methodology, which involves stacking of time-series over significant time periods, has been particularly successful for the retrieval of crustal structure, owing to the strong prevalent seismic noise in the microseismic band (3–25 s), made up primarily of short period surface waves generated in the oceans. Attempts at retrieving deeper mantle structure, requiring the use of weaker, longer period noise, have shown potential at retrieving the very longest wavelengths of upper mantle structure (Nishida 2013; Haned *et al.* 2016). Body wave energy has been identified in these cross-correlation stacks (e.g. Gerstoft *et al.* 2008; Lin *et al.* 2013), however many issues remain before an approach based solely on ANT can be used confidently for global tomography (e.g. Tsai 2009; Pham *et al.* 2018).

The idea developed and illustrated in this paper, first presented in Romanowicz *et al.* (2013), consists in adding a second step after source stacking, namely computing the cross-correlation of the source-stacked wavefield between pairs of stations and using these cross-correlations as data in the definition of the cost function for the inversion. This step is computationally negligible. Because the distribution of earthquake sources is not uniform around the globe, the Green’s function between two given stations will not be retrieved perfectly. However, this is not an issue for our approach, since in this case, unlike in ANT, we know the location and source mechanism of the sources (at least to a good approximation), so that the synthetic sums and following cross-correlations can be formed in exactly the same way as the observed ones. An additional advantage of using earthquake sources is that the depth distribution of earthquakes is broader than that of the primary noise sources, which are confined to the ocean-seafloor interface, thus providing stronger excitation of overtones and body waves in a wide frequency band. Stated differently, recovery of the Green’s function is *not* the objective of the technique described here although motivated by ANT, the cross-correlation is simply used as a *derived data type* that has specific advantages over plain summed traces. Namely: (1) the dominant energy corresponding to the fundamental mode surface waves is more localized in time, so that appropriate windowing and weighting could be applied, in order to more effectively bring out the contribution of smaller amplitude overtones and body waves to the wavefield and (2) it is possible to introduce path weighting, to avoid the dominance of redundant paths in the inversion.

In what follows, we first illustrate how the additional cross-correlation step can help separate the large amplitude waveforms due to fundamental mode surface waves, thus mitigating the primary source of information loss associated with inversion of summed traces. We then present synthetic tests first for inversion using the



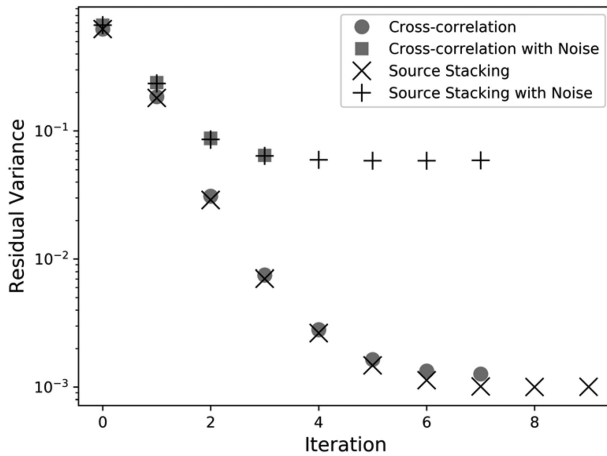
**Figure 4.** Isotropic  $V_s$  part of the synthetic ‘BLOB’ upper mantle model used as input model for the synthetic experiments discussed in this paper. Left-hand panels show the isotropic  $V_s$  model at several depths, presented in percent with respect to the global mean at that depth. The extrema in  $V_s$  are shown above each map. Right-hand panels present depth cross-sections along two profiles  $A - B$  and  $C - D$  as indicated on the 100 and 300 km maps, respectively.



**Figure 5.** Anisotropic parameter  $\xi$  variations in the ‘BLOB’ model, plotted in map view at several depth (left-hand panel), in percent relative to the global mean (isotropic) at that depth. Extrema of each map are shown above the corresponding map. Right-hand panels present depth cross-sections along two profiles  $A - B$  and  $C - D$  as indicated on the 150 km map.

stacked records themselves, and then for inversion using cross-correlations of stacked records. The computation of partial derivatives of the cross-correlations is outlined in the Appendix. These kernels have been computed using NACT, however, extending them to an adjoint or any other inversion approach is straightforward. The

frequency band considered in this paper (400–60 s) is the same one used to develop the upper mantle models SEMum (Lekic & Romanowicz 2011) and SEMum2 (French *et al.* 2013). Consequently, our synthetic tests are designed to evaluate the retrieval of long-wavelength upper mantle structure only.



**Figure 6.** The evolution of variance reduction as a function of iteration for each of the experiments performed. Note that for iteration 1 with noise, the residual variances for source-stacking inversion and when the additional cross-correlation step is included, plot on top of each other.

## 2 SEPARATION OF DISTINCT WAVE TRAINS BY CROSS-CORRELATION

Fig. 1 shows the distribution of events and stations used for the synthetic experiments presented in this paper. It consists of the 273 earthquake sources and 537 stations used for the development of SEMUCB-WM1. However, we removed 22 stations that were located at a distance smaller than 150 km from one or more events, in order to avoid very large amplitudes at the beginning of the time-series. Here we assume no records are missing. This assumption will be discussed in the discussion section.

Fig. 2 shows an example of vertical component 10 000 s long traces calculated at station TATO in the 1-D average model of SEMUCB-WM1, for 38 different events, aligned on origin time and ordered as a function of epicentral distance. The resulting summed trace is shown at the bottom. This figure illustrates how long period surface waves dominate the individual records and therefore the sum, hiding the lower amplitude overtone energy, and making it difficult to apply the same kind of weighting scheme as is applied in conventional waveform tomography, where larger amplitude phases are weighted down compared to weaker ones, in order to bring out the information contained in the latter (e.g. Appendix in Li & Romanowicz 1996). As we will show below, this is not a major drawback for the simple tests at long periods presented here, but could become an issue when higher resolution lower mantle structure is targeted, which requires the inclusion of shorter period, relatively low amplitude, body waves.

In Fig. 3, we show L–L and T–T station pair cross-correlations for the entire data set, where L is the radial and T the transverse horizontal component, defined with respect to the great circle path between the two stations. The Z–Z cross-correlations are shown in Fig. S1. The cross-correlations have been ordered as a function of distance between stations, and stacked in bins of 100 km in distance. Fig. 3 shows the emergence of fundamental mode Rayleigh wave energy on the L component and Love wave energy on the T component, both for the minor and major arcs. The strong energy in front of these surface waves contains information on earth’s deeper structure. In particular, on the L–L component, reconstructed overtone energy is also distinguishable.

We stress that the main goal of computing the cross-correlations is not to extract the Green’s function, for which the distribution of

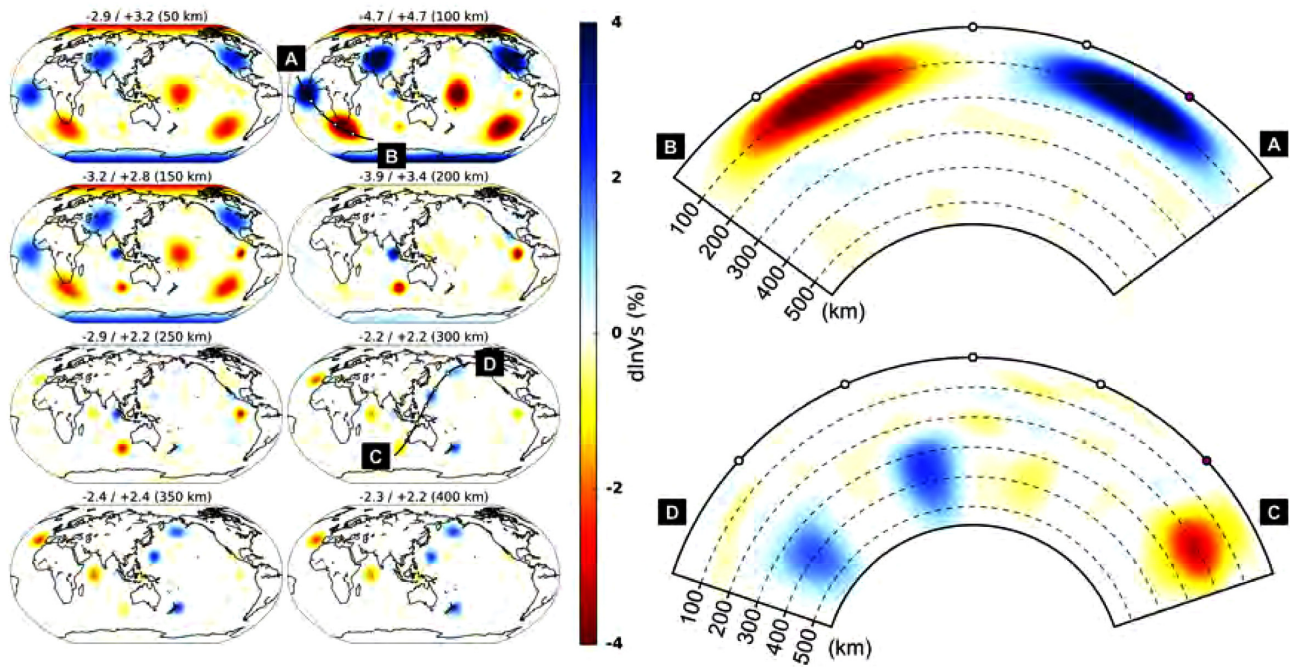
stacked events is clearly insufficient, but to enable the capability of windowing out the strong energy corresponding to fundamental mode surface waves, giving better access to the information contained in overtones (and ultimately, as we extend the approach to shorter periods, body waves). Because we are performing the same operations on the synthetics and the actual data, how well we form the Green’s function is not an issue.

We note that there are two ways of bringing out the pseudo-Green’s function by cross-correlation. The first one is presented in Fig. 3, stacking over distance bins. The other one is obtained by stacking individual station-pair cross-correlations obtained in several realizations of summed randomly 1-bit source-encoded synthetics. For example, Fig. S2 shows the results of an experiment in which the Z–Z cross-correlations have been obtained from stacking 10 realizations of random 1-bit source-encoded synthetics (e.g. Krebs *et al.* 2009). Aligned by interstation distance, the energy corresponding to the fundamental mode Rayleigh wave is strongly localized. In this way of stacking, the computational cost is larger, since we need to compute a number of realizations of the stacked seismograms. Because the stacking is performed over a given interstation path, it may result in better recovery of structure than when cross-correlations are stacked over specific distance bins, where station pairs may correspond to different locations on the globe. However, since the same operations are applied to the data and the synthetics, the inversion should recover structure in the correct geographical location also in the latter case. This will be evaluated more precisely in forthcoming work.

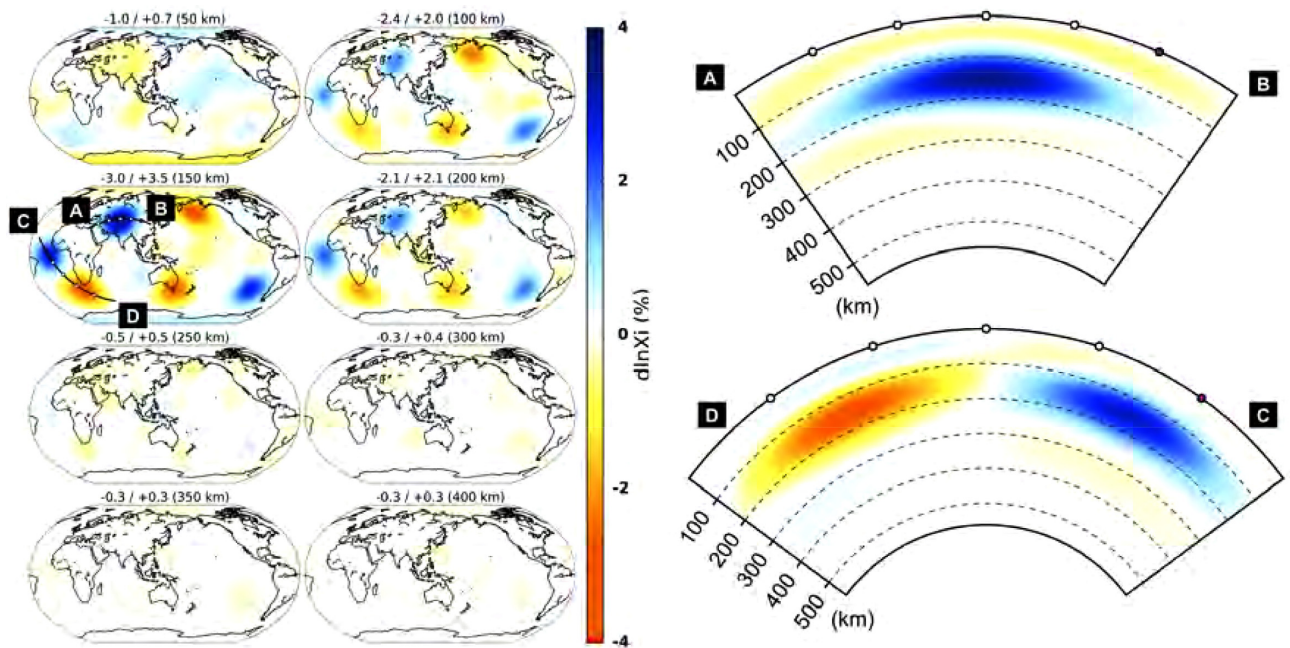
## 3 CONSTRUCTION OF THE SYNTHETIC DATA SETS AND EXPERIMENTS

In order to develop the methodology and for the initial tests on a synthetic data set, we constructed a simple global shear velocity model, thereafter called ‘BLOB’, which contains both large and small anomalies in shear wave velocity ( $V_s$ ) distributed around the globe, and centred at different depths in the upper mantle and transition zone (Fig. 4). It also includes heterogeneities in the anisotropic parameter  $\xi = (V_{sh}/V_{sv})^2$ , with a different distribution and size than for the isotropic anomalies, and centred at a depth of 150 km (Fig. 5). In the synthetic tests presented here, we did not consider lateral variations in crustal structure. Three component summed synthetic seismograms of 10 000 s duration were computed in these models for the 273 events used in the development of SEMUCB-WM1 (French & Romanowicz 2014), and the 515 stations considered here. The sources were aligned at their origin time, normalized by the seismic moment of each source, with source parameters obtained from the CMT catalogue (Dziewonski *et al.* 1981; Ekström *et al.* 2012). Both minor and major arc Love and Rayleigh waves are included in the 10 000 s computation, although we note that the corresponding cross-correlations are complete only up to  $\pm 5000$  s. Synthetics were computed using the CSEM (Capdeville *et al.* 2003), a version of a spectral element code which couples the SEM computation with 1-D normal mode computation in the core, originally developed for computational efficiency. In what follows, this synthetic data set will be referred to as ‘data’, while the predictions from successive iterations of modelling will be called ‘synthetics’. These synthetics are also computed using CSEM, at each iteration.

The starting model for the inversion is the 1-D average model of model BLOB, which is the 1-D radially anisotropic part of model SEMUCB-WM1, a model which, contrary to PREM (Dziewonski



**Figure 7.** The recovered model in isotropic  $V_s$  after six iterations of direct inversion of three component summed seismograms, without noise. Plots are presented as in Fig. 4.



**Figure 8.** Same as Fig. 7 for the parameter  $\xi$ .

& Anderson 1981) does not contain a discontinuity at 220 km depth. Spatial correlation lengths of 2000 km for  $V_s$  and 6000 km for  $\xi$  are used for the computation of the model covariance matrix in the inversion.

In a first experiment, the summed time domain seismograms were inverted directly, applying as many iterations as was necessary until convergence. In a second experiment, at each iteration of the inversion, we first applied cross-correlation to all available pairs

of stations, and inverted the cross-correlograms instead. In both experiments, we used NACT to compute the gradient and Hessian. We stress that in NACT the partial derivatives and Hessian depend non-linearly on the 3-D model, and thus are recomputed at every iteration, based on the evolving 1-D model and 3-D perturbations around it. The derivation of misfit function gradient and Jacobian, in the case of cross-correlations, is given in the Appendix, and is independent of the methodology used for the computation of



gradient. In Section 4.1 we compare the results of inversion using stacked sources and that obtained after adding the cross-correlation step, without noise. In Section 4.2, we perform the same experiments with real noise added.

Fig. 6 shows the evolution of variance reduction in the data as a function of iteration for each of the experiments performed. Note that we performed eight iterations of the inversion of summed seismograms without noise, but will be showing the resulting models after six iterations, when the variance reduction stabilized. We performed fewer iterations for other experiments.

## 4 INVERSION RESULTS

In what follows, we present synthetic tests of inversion using two different definitions of the cost function: one based directly on the source-stacked traces, and another based on the resulting station-pair cross-correlations. We stress that in this study, we implemented path-weighting, but we stopped short of applying windowing to the cross-correlation traces. Given the simplicity of the toy model considered and the very long period filter applied, we show that using cross-correlations without windowing recovers the input model at least as well as using straight stacked records. The additional advantage of windowing/path weighting would emerge for more complex models and/or when including shorter periods. This will be considered in the near future.

### 4.1 Experiments on synthetic data without noise

Figs 7 and 8 show the results of inverting three component summed seismograms directly, for isotropic  $V_s$  and  $\xi$ , respectively, after six iterations. The residual variance does not evolve significantly after the 6th iteration (see Fig. 6). We find that the larger, shallow anomalies emerge first, constrained by the dominant fundamental mode energy, while the deeper, narrower anomalies come into focus in the later iterations. We find that all anomalies are well recovered, with little distortion in shape, although with some loss in amplitude, as commonly observed in tomographic inversions, and due to the introduction of damping. As expected, and not unlike what is observed in conventional tomography, the anisotropic part is less well resolved, with larger amplitude loss and some vertical smearing.

Figs 9 and 10 show the corresponding results for the inversion of cross-correlations, with the variance reduction evolution shown in Fig. 6. The variance reduction tracks that of the summed seismogram inversion. It is slightly higher, due likely to the slight difference in the damping scheme used. There is no significant difference for the variance reduction in the results obtained by both methods.

### 4.2 Experiments on synthetic data with noise

#### 4.2.1 Effect of noise on the signal to noise ratio of summed synthetics and cross-correlations

An important issue to consider is the effect of noise in the data. This is particularly important given the fact that much of the low frequency background noise is due to surface waves generated by ocean-solid earth interactions, the so-called ‘seismic hum’ (e.g. Suda *et al.* 1998; Nishida 2013), the energy of which is equivalent to about one  $M$  5.8 earthquake per day. Since sources of seismic hum are not included in the computation of the synthetics, it is important to evaluate whether they might significantly contribute and bias the inversion. It is *a priori* unlikely, because months of data are usually

stacked for the Green’s function to emerge in cross-correlation ANT studies. Here, we are stacking a relatively small number of event records corresponding to different time periods, and we expect the hum sources will not stack coherently.

To assess the potential effect of background noise, and particularly the vertical component noise due to the earth’s ‘hum’, we first estimated the signal to noise level in the cross-correlations. For this, we considered three stations in Eurasia for which a large number of real records were available, stations ECH, HIA and MDJ. For these stations, we collected 10 000 s of noise waveforms on the vertical component before the origin time of each of the 179 events jointly recorded by both stations and compared them to 10 000 s of real data starting at the origin time of each event. The distribution of these events and location of the three stations is shown in Fig. S3. Fig. 11(right-hand panels) also shows a comparison of the station pair cross-correlations of summed noise (red traces) and summed signals (black traces) for these data. Fig. 12 shows the cross-correlation spectra, filtered with corner frequencies of 20 and 800 s. The corresponding spectra for the summed traces are shown in Fig. S4. It is clear that in the period band considered and beyond (down to at least 20 s and up to  $\sim$ 400 s), the signal-to-noise ratio is large. We thus see that the background noise does not significantly affect the recorded signals, at least in the frequency band considered here and up to frequencies that one would consider for whole mantle tomography including body-waveforms.

#### 4.2.2 Inversion with realistic noise added to synthetic data

We also performed another experiment in which we inverted a synthetic data set constructed by adding realistic noise to each record. Here, we could not add the real noise preceding each record as in the previous experiment, because in the real data set considered for the development of SEMUCB-WM1, not all stations recorded all events. Instead, we downloaded the probability density functions (PDFs) provided on the IRIS website (<http://service.iris.edu/mustang/noise-pdf/1/>), separately for vertical and horizontal components, for the entire period of operation of each station. For the 178 stations for which this is not available, we conservatively considered the average between the NLNM (new low noise model) and NHLM (new high noise model, Peterson 1993) for both vertical and horizontal components. We then added a random phase to the amplitude spectrum thus obtained at each station, converted back to the time domain, and applied the filter used in all our experiments (60-80-250-400s). We then added this noise time-series  $N$  times to the synthetic summed data, where  $N$  is the number of events stacked ( $N = 273$ ).

Figs 13 and 14 show the results of inversion of stacked seismograms after six iterations with this noise added. We see that the anomalies are still recovered although their shape is a little distorted, and a stronger background level of spurious structure is introduced, as reflected in the much poorer variance reduction. Also, as expected, the effect of noise is stronger on the recovery of the anisotropic parameter  $\xi$ . The evolution of variance reduction as a function of iteration number is shown in Fig. 6. The variance reduction, as expected, is smaller than in the case without noise and stops evolving after the 4th iteration. We also show the corresponding variance reduction for the cross-correlations after three iterations, which tracks that of the stacked-event inversion, as do the modelling results, as shown in Figs S5–S10. We thus did not pursue these inversions further, to save computational time for more important tasks. The recovery of the model is thus only slightly worse

in the case with noise than in the case without noise, whether or not the cross-correlation based misfit function is considered. Moreover, we consider this as a worst case scenario experiment. First, the stations were not selected based on their performance in terms of background noise (which would be done in the case of real data), second, we are including a conservatively high noise level for 1/3 of the stations (for which no noise statistics are available).

#### 4.2.3 Evolution of model misfit as a function of iteration number

With access to more extensive synthetic tests than is possible in conventional full waveform tomography, we can follow the evolution of the model misfit as a function of iteration. We quantify the evolution model residual misfit for all the experiments described above, defined as:

$$res_i = 100 \times \left[ 1 - \frac{\|m_0 - m_i\|_2^2}{\|m_0\|_2^2} \right], \quad (1)$$

where  $m_0 = \text{dln}X_T$  represents the distribution of relative perturbations in the target model  $X_T$  with respect to the 1-D reference model  $X_0$ , with  $X = V_s$  or  $\xi$ . The corresponding distribution of relative perturbations in the model obtained at iteration  $i$  is  $m_i = \text{dln}X_i$ . A particularity of the target ‘BLOB’ model considered is that only a very small volume of the upper mantle contains non-zero 3-D perturbations. Fig. 15a shows that when the model misfit is computed over the entire global volume of the model down to 550 km depth (below which no 3-D structure exists in the target model), then, in the case of source stacking with noise, the residual misfit decreases until iteration 3, but stabilizes or starts slightly increasing at further iterations. The reason for this is that, while the model continues to improve in the regions where 3-D structure actually exists, low amplitude spurious structure is introduced elsewhere because of the presence of noise, in particular due to the importance of the norm damping term of the regularization. Because the 3-D structure occupies only a small portion of the model space, the noise in the model plays an increasingly important role when quantifying the misfit. Indeed, Fig. 15b shows the evolution of the residual model misfit when only those parts of the model space that correspond to 3-D anomalies of amplitude larger than 1 per cent are included in the calculation. In that case, the model misfit decreases continuously as a function of iteration. On the other hand, Fig. 15c shows that the increase in residual misfit after iteration 3 originates from the regions where weak or no 3-D structure exists in the target model. The same behaviour is found for  $\xi$  (Fig. S11).

The difference between the inverted  $V_s$  model and the target one is presented in Figs S12 and S13, for the cases of stacked-source inversion with and without adding cross-correlations, respectively, at iteration 6 (without noise added). We note a global distribution of the residual misfit with more or less uniform amplitude down to 300 km, while the amplitude recovery of the narrower anomalies at larger depth is poorer at least at this iteration, although slightly better in the case when cross-correlations are added (Fig. S13). Figs S14 and S15 show the same as Figs S12 and S13 but for  $\xi$ . We note the poorer recovery of the amplitudes of the target anomalies in  $\xi$ , and greater smearing with depth as already seen in Fig. 8, manifested here in the sidelobes on both sides of the target location of 3-D anomalies in  $\xi$ . The cross-correlation step does not seem to help, at least visually. Fig. S11bc indicates that the amplitude recovery of  $\xi$  in the regions where the target model has non-zero 3-D structure is slightly better with cross-correlation, but that more noise is introduced elsewhere. However, this is subtle and may be due to the difficulty of choosing equivalent damping schemes

for the two types of models (with or without cross-correlation). Finer adjustment of the damping scheme at each iteration as well as experimenting with other damping schemes could help inform tomographic experiments with real data on how best to balance recovery of amplitudes of true structure while minimizing model noise. Including shorter period data would also help to improve the recovery of anomalies of smaller lateral extent in the deeper parts of the model.

## 5 DISCUSSION

We have shown that source stacking, as well as source stacking followed by cross-correlations between pairs of stations provides a promising approach for significantly reducing the number of SEM wavefield computations, at least for resolving structures at the scale presented in this paper. While path-weighting is already included in our implementation, we expect that the additional step of windowing allowed by stacking cross-correlations will help achieve sufficient resolution when shorter period body wave constraints on deep mantle structure are included. We also posit that the uncertainties in source parameters will be reduced when using stacked cross-correlations, just as they are in ANT, and particularly so when using source encoding. These will be topics which we will address in a forthcoming manuscript.

A particularly useful application of this approach is for exploring modelling uncertainties such as due to background noise, as illustrated here, in ways that cannot be addressed using conventional waveform tomography based on the inversion of single records, since performing full synthetic tests is too costly computationally and generally only the last iteration can be assessed using resolution matrix computations and bootstrapping (e.g. French & Romanowicz 2014) or by sampling the Hessian (e.g. Fichtner & Trampert 2011). This also opens the way to exploring various trade-offs, such as, for example, between crustal structure and radial anisotropy, which are known to be important (e.g. Ferreira *et al.* 2010; Lekic *et al.* 2010), as well as uncertainties in source parameters and other trade-offs due to choices of parametrization. Also, the choice of damping schemes generally involves some level of arbitrariness. This could be informed by tests performed on synthetic data sets, and involving several iterations, as illustrated here. At the very least, insights gained in such tests can inform uncertainties associated with conventional waveform inversions. The ‘blob’ model considered here is very simple, and the frequency range rather low. In the case of conventional waveform tomography, one could envisage designing more realistic synthetic tests based on the methodology presented here.

A concern with the source stacking approach voiced in the applied geophysics community, and more generally in the community that relies on the computation of gradients using an adjoint formalism, is the introduction of cross-terms when computing the gradient based on the adjoint of the summed source seismograms. This is generally addressed by performing random source encoding which adds to the number of SEM runs but is still worthwhile computationally (Schuster *et al.* 2011). Recently, a method has been proposed whereby the encoding is such that it allows the decoding of single traces from the stack at given frequencies (Zhang *et al.* 2018), although computational trade-offs still need to be evaluated comprehensively.

We note that in our approach so far, there is no cross-talk that reduces the quality of the partial derivatives, since we are not using an adjoint formalism for the inverse part of the modelling. Rather,

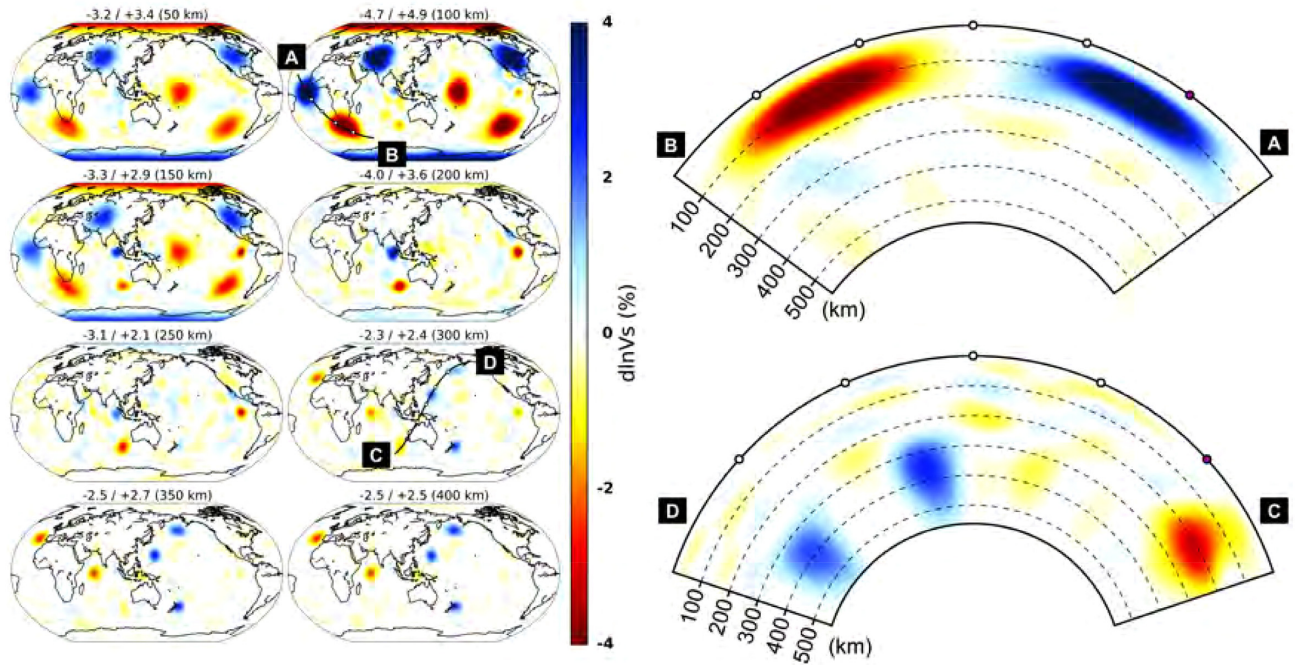


Figure 9. Isotropic  $V_s$  part of the model recovered after six iterations of inversion of station pair cross-correlations, without noise added.

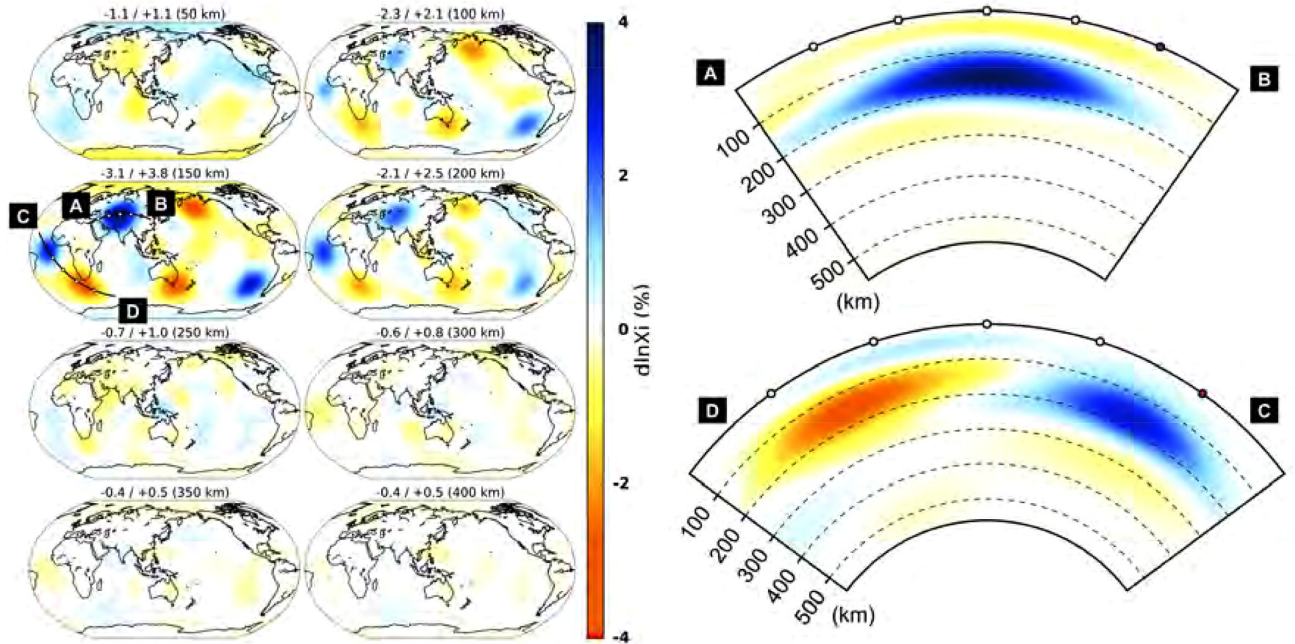
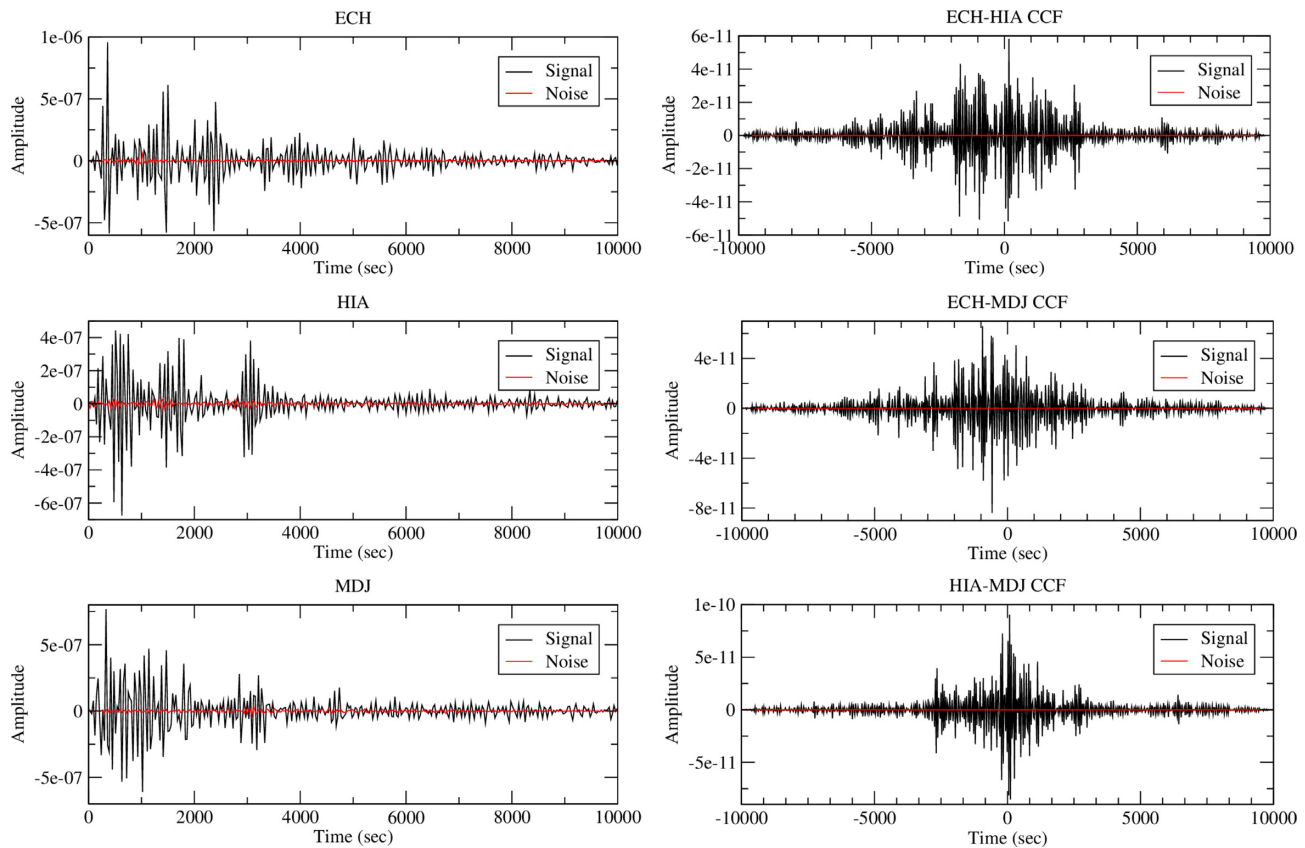


Figure 10. Same as Fig. 9 for the parameter  $\xi$ .

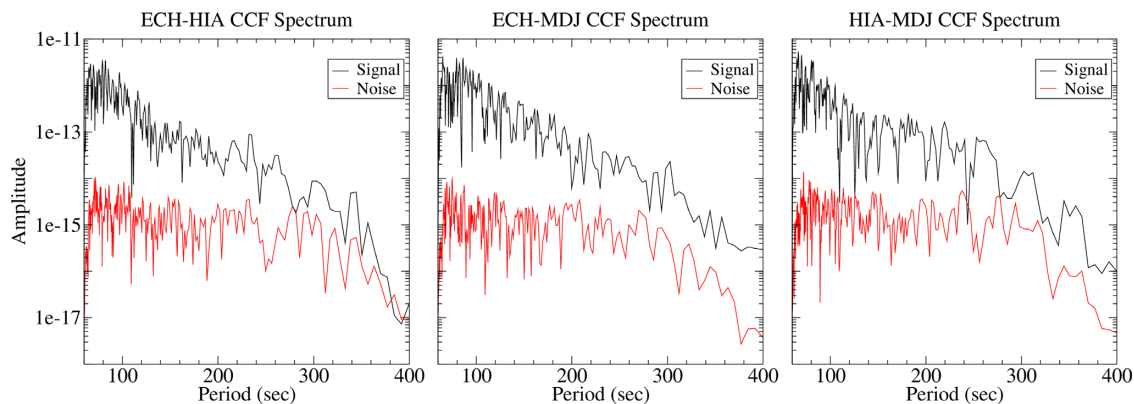
we compute the gradient and Hessian to quite a good approximation using normal mode perturbation theory: as the tests show, we can converge to a good approximation of the starting model with relatively few iterations starting from a 1-D model. The updates to the gradient and Hessian at each iteration can be rapidly computed.

In the case of real data, another obvious challenge is that not all stations, and not all three components of all stations are generally available. This is the case for the data set used for the construction of

SEMUCB-WM1, and not only because of limited operation times of stations, but also because of the careful data selection that was performed, to eliminate noisy data, and in particular nodal stations. However, the data set assembled for the construction of SEMUCB-WM1 spanned an interval of  $\sim 20$  yrs (from 1992 to 2012) during which broadband networks expanded significantly, so that records for older events were missing at stations installed more recently. When focusing on the last ten years, we can assemble event and



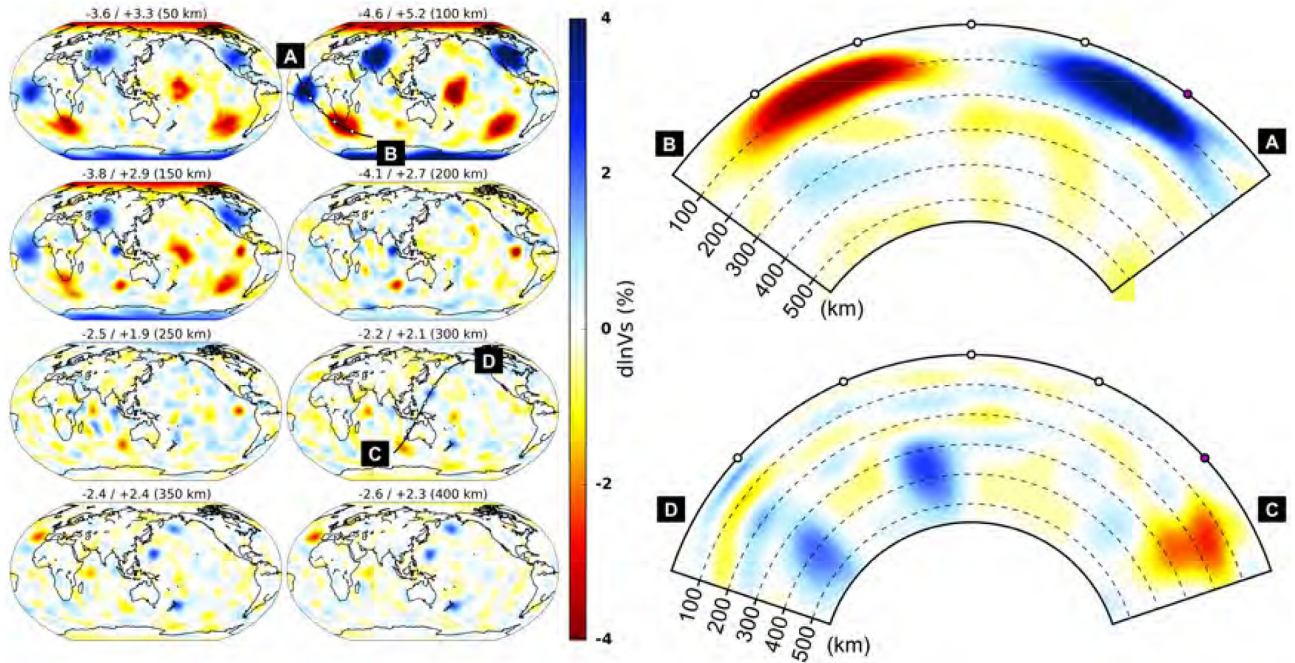
**Figure 11.** Left-hand panel: examples of comparisons of 10 000-s-long summed ‘noise’ and ‘signal’ records. Black traces show the summed signal. Red traces show summed noise. Right-hand panels: comparison of the station pair cross-correlations of summed noise (red traces) and signal (black traces) records for the three stations shown on the left-hand panels. The traces have been filtered with corner frequencies of 60 and 400 s. The distribution of the 179 events and location of the three stations for which records are shown is presented in Fig. S4.



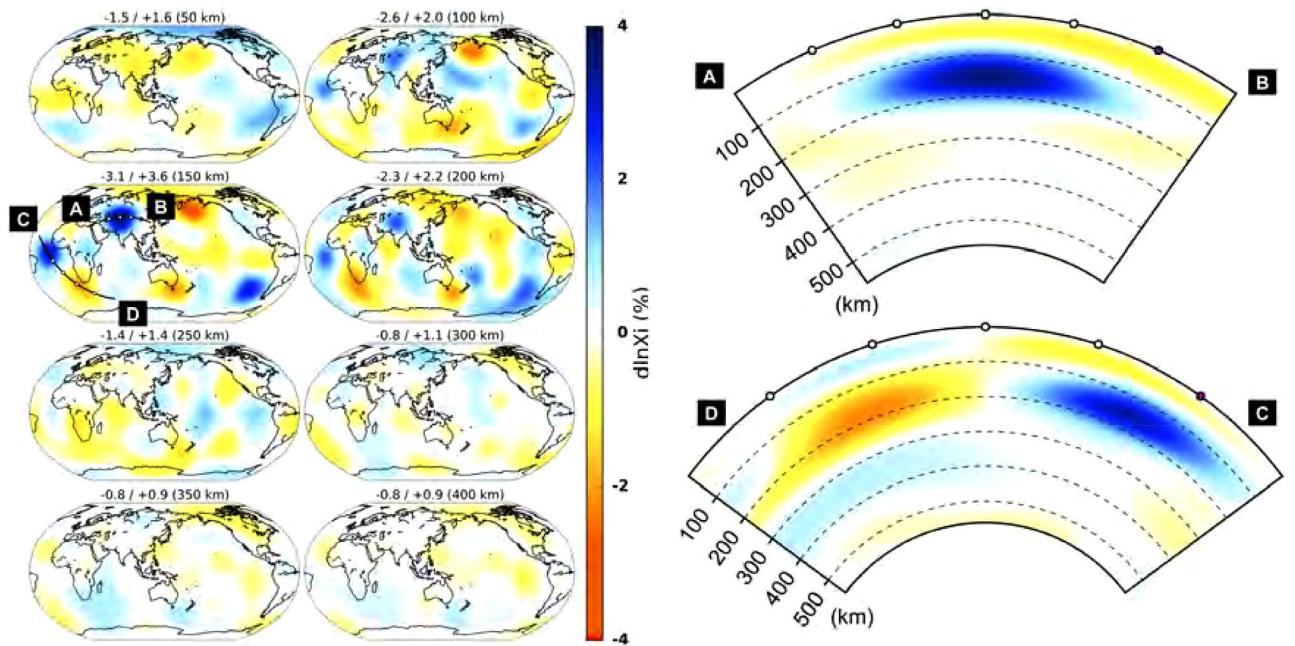
**Figure 12.** Comparison of the spectra of the cross-correlations traces shown in Fig. 11. The traces have been filtered with corner frequencies of 20 and 800 s. Note the logarithmic scale on the y-axis.

station groups in a way that can minimize missing records in any particular event/station group. Once a set of generally high quality stations is selected for performing a particular stack, we need not necessarily remove individual noisy records (due for example to a station being nodal for a particular event). Whether or not it is better to keep noisy records or replace them by synthetics computed in the current 3-D model can be assessed by synthetic tests we plan to

perform in the near future. Particular care will need to be taken to adjust the model parametrization in successive iterations (starting with a 1-D model), as well as the period range (progressing from longer periods to shorter periods), in order to avoid undesirable effects of cycle slipping, which will be difficult to detect explicitly in this approach.



**Figure 13.**  $V_s$  isotropic part of the model recovered after six iterations of inversion of three component summed seismograms, with noise added. The target model is presented as in Fig. 4.

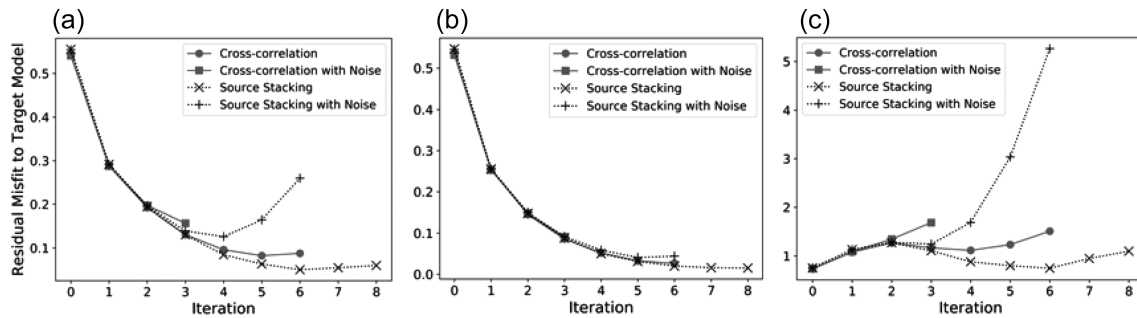


**Figure 14.** The  $\xi$  part of the model recovered from the same experiment as shown in Fig. 13.

## 6 CONCLUSIONS

Through a series of experiments based on synthetic waveform data with and without realistic noise added, we have shown that some of the drawbacks encountered when trying to invert stacked-source seismograms directly may be overcome by redefining the cost-function using station pair cross-correlations of the summed traces. Indeed, large amplitude fundamental surface waves that dominate

the latter can be, to a large extent, separated in time in the cross-correlations, allowing implementation of windowing and weighting to bring out overtone and body wave constraints to the inversion for deep mantle structure. Moreover, the use of cross-correlations allows straightforward implementation of path-weighting, an important step in balancing the available coverage in the inversion step, as well as wave-packet weighting (i.e. windowing), which is important to bring out the contributions of lower amplitude body



**Figure 15.** Residual Vs model misfit (normalized by variance in the target model), as a function of experiment and iteration when (a) the whole model volume down to 550 km depth is considered in the calculation of the misfit; (b) only the portion of the model space where  $d\ln V_s > 1$  per cent in the target model is considered in the calculation of the misfit (this corresponds to 7.9 per cent of the total volume); (c) only the portion of the model space where  $d\ln V_s < 1$  per cent is considered in the calculation of the misfit (this corresponds to 92.1 per cent of the total volume and is dominated by regions without 3-D structure in the blob model).

waves. While implementation of the wave-packet weighting and application to real data is in progress, the tools developed so far result in efficient computations, and can already help assess model resolution issues more thoroughly than is possible with conventional waveform tomography, at least, as shown in this paper, at long wavelengths. A particularly important application, which is the subject of a forthcoming manuscript, is to investigate trade-offs between imperfectly known crustal structure and deeper anisotropic structure.

## ACKNOWLEDGEMENTS

This work was supported by NSF grants EAR-1417229 and EAR-1460205. Computations were performed at the National Energy Research Scientific Computing Center (NERSC), supported by the U.S. Department of Energy Office of Science.

## REFERENCES

Aki, K., 1957. Space and time spectra of stationary stochastic waves, with special reference to microtremors, *Bull. Earthq. Res. Inst.*, **35**, 415–456.

Beller, S., Operto, S., Nolet, G., Paul, A. & Zhao, L., 2018. multiparameter teleseismic full-waveform inversion, *Geophys. J. Int.*, **212**, 1369–1388.

Bozdag, E. *et al.*, 2016. Global adjoint tomography: first-generation model, *Geophys. J. Int.*, 1739–1766.

Capdeville, Y., Chaljub, E., Vilotte, J.P. & Montagner, J.P., 2003. Coupling the spectral element method with a modal solution for elastic wave propagation in global earth models, *Geophys. J. Int.*, **152**, 34–67.

Capdeville, Y., Gung, Y. & Romanowicz, B., 2005. Towards global earth tomography using the spectral element method: a technique based on source stacking, *Geophys. J. Int.*, **162**, 541–554.

Clouzet, P., Masson, Y. & Romanowicz, B., 2018. Box tomography: first application to the imaging of upper-mantle shear velocity and radial anisotropy structure beneath the North American continent, *Geophys. J. Int.*, **213**, 1849–1875.

Debayle, E. & Ricard, Y., 2012. A global shear velocity model of the upper mantle from fundamental and higher Rayleigh mode measurements, *J. geophys. Res.*, **117**, 1–24.

Durand, S., Debayle, E., Ricard, Y., Zanolli, C. & Lambotte, S., 2017. Confirmation of a change in the global shear velocity pattern at around 1000 km depth, *Geophys. J. Int.*, **211**, 1628–1639.

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

Dziewonski, A.M., Chou, T. & Woodhouse, J.H., 1981. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. geophys. Res.*, **86**, 2825–2852.

Ekström, G., Nettles, M. & Dziewon, A.M., 2012. The global CMT project 2004–2010: centroid-moment tensors, *Phys. Earth planet. Inter.*, **201**, 1–9.

Ferreira, A.M.G., Woodhouse, J.H., Visser, K. & Trampert, J., 2010. On the robustness of global radially anisotropic surface wave tomography, *J. geophys. Res.*, **115**, 1–16.

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

French, S., Lekic, V. & Romanowicz, B., 2013. Waveform tomography reveals channeled flow at the base of the oceanic asthenosphere, *Science*, **342**, 227–230.

French, S.W. & Romanowicz, B., 2015. Broad plumes rooted at the base of the Earth's mantle beneath major hotspots, *Nature*, **525**, 95–99.

French, S.W. & Romanowicz, B.A., 2014. Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography, *Geophys. J. Int.*, **199**, 1303–1327.

Gerstoft, P., Shearer, P.M., Harmon, N. & Zhang, J., 2008. Global P, PP, and PKP wave microseisms observed from distant storms, *Geophys. Res. Lett.*, **35**(November), 4–9.

Haned, A., Stutzmann, E., Schimmel, M., Kiselev, S. & Davaille, A., 2016. Geophysical Journal International, *Geophys. J. Int.*, **204**, 1222–1236.

Kawai, K., Konishi, K., Geller, R.J. & Fuji, N., 2014. Methods for inversion of body-wave waveforms for localized three-dimensional seismic structure and an application to D" structure beneath Central America, *Geophys. J. Int.*, **197**, 495–524.

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

Komatitsch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation – I. Validation, *Geophys. J. Int.*, **150**, 390–412.

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.

Krebs, J.R., Anderson, J.E., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A. & Lacasse, M.-D., 2009. Fast full-wavefield seismic inversion using encoded sources, *Geophysics*, **74**(6), doi:10.1190/1.3230502.

Kustowski, B., Ekstro, G. & Dziewon, A.M., 2008. Anisotropic shear-wave velocity structure of the Earth's mantle: a global model, *J. geophys. Res.*, **113**, 1–23.

Lekic, V., Panning, M. & Romanowicz, B., 2010. A simple method for improving crustal corrections in waveform tomography, *Geophys. J. Int.*, **182**, 265–278.

Lekic, V. & Romanowicz, B., 2011. Inferring upper-mantle structure by full waveform tomography with the spectral element method, *Geophys. J. Int.*, **185**, 799–831.

Li, X.-D., Romanowicz, B., 1996. Global mantle shear velocity model developed using nonlinear asymptotic coupling theory, *J. geophys. Res.*, **101**(B10), 22245–22272.

- Lin, F.-c., Tsai, V.C., Schmandt, B., Duputel, Z. & Zhan, Z., 2013. Extracting seismic core phases with array interferometry, *Geophys. Res. Lett.*, **40**, 1049–1053.
- Li, X.-D. & Tanimoto, T., 1993. Waveforms of long-period body waves in a slightly aspherical Earth model, *Geophys. J. R. astr. Soc.*, **112**, 92–102.
- Li, X.D. & Romanowicz, B., 1995. Comparison of global waveform inversions with and without considering cross-branch modal coupling, *Geophys. J. R. astr. Soc.*, **121**, 695–709.
- Maggi, A., Tape, C., Chen, M., Chao, D. & Tromp, J., 2009. An automated time-window selection algorithm for seismic tomography, *Geophys. J. Int.*, **178**, 257–281.
- Masson, Y., Cupillard, P., Capdeville, Y. & Romanowicz, B., 2014. On the numerical implementation of time-reversal mirrors for tomographic imaging, *Geophys. J. R. astr. Soc.*, **196**, 1580–1599.
- Masson, Y. & Romanowicz, B., 2017. Fast computation of synthetic seismograms within a medium containing remote localized perturbations: a numerical solution to the scattering problem, *Geophys. J. Int.*, **208**, 674–692.
- Masson, Y. & Romanowicz, B., 2018. Box tomography: localised imaging of remote targets buried in an unknown medium, a step forward for understanding key structures in the deep Earth, *Geophys. J. Int.*, **211**, 141–163.
- Mochizuki, E., 1986. Free oscillations and surface waves of an aspherical earth, *Geophys. Res. Lett.*, **13**(13), 1478–1481.
- Monteiller, V., Komatitsch, D. & Wang, Y., 2015. Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the SEM – DSM hybrid method, *Geophys. J. Int.*, **202**, 811–827.
- Moulik, P. & Ekström, G., 2014. An anisotropic shear velocity model of the Earth’s mantle using normal modes, body waves, surface waves and long-period waveforms, *Geophys. J. Int.*, **199**, 1713–1738.
- Mégnin, C. & Romanowicz, B., 2000. The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms, *Geophys. J. Int.*, **143**, 709–728.
- Nishida, K., 2013. Earth’s background free oscillations, *Annu. Rev. Earth planet. Sci.*, **41**, 719–740.
- Nissen-Meyer, T., van Driel, M.V., Stähler, S.C., Hosseini, K., Hempel, S., Auer, L., Colombi, A. & Fournier, A., 2014. AxiSEM: broadband 3-D seismic wavefields in axisymmetric media, *Solid Earth*, **5**, 425–444.
- Nolet, G., 1990. Partitioned Waveform Inversion and Two-dimensional structure under the network of autonomously recording seismographs, *J. Geophys. Res.*, **95**, 8499–8512.
- Panning, M. & Romanowicz, B., 2006. A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophys. J. Int.*, **167**, 361–379.
- Park, J., 1987. Asymptotic coupled-mode expressions for multiplet amplitude anomalies and frequency shifts on an aspherical earth, *Geophys. J. R. astr. Soc.*, **90**, 129–169.
- Peterson, J., 1993. Observations and modelling of seismic background noise, US Geological Survey, open-file report 93-322.
- Pham, T.-S., Tkalcic, H., Sambridge, M. & Kennett, B.L.N., 2018. Earth’s correlation wavefield: late coda correlation, *Geophys. Res. Lett.*, **45**, 3035–3042.
- Rickers, F., Fichtner, A. & Trampert, J., 2013. The Iceland – Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: evidence from full-waveform inversion, *Earth planet. Sci. Lett.*, **367**, 39–51.
- Romanowicz, B.A., 1987. Multiplet-multiplet coupling due to lateral heterogeneity: asymptotic effects on the amplitude and, *Geophys. J. R. astr. Soc.*, **90**, 75–100.
- Romanowicz, B.A., French, S.W., Rickers, F. & Yuan, H., 2013. Source stacking for numerical wavefield computations: application to continental and global scale seismic mantle tomography, in *American Geophysical Union, Fall Meeting 2013, Abstract ID. S21E-05*.
- Romanowicz, B.A., Panning, M.P., Gung, Y. & Capdeville, Y., 2008. On the computation of long period seismograms in a 3-D earth using normal mode based approximations, *Geophys. J. Int.*, **175**, 520–536.
- Ruan, Y. & Zhou, Y., 2012. The effects of 3-D anelasticity (Q) structure on surface wave amplitudes, *Geophys. J. Int.*, **189**, 967–983.
- Schaeffer, A.J. & Lebedev, S., 2013. Global shear speed structure of the upper mantle and transition zone, *Geophys. J. Int.*, **194**, 417–449.
- Schaeffer, A.J. & Lebedev, S., 2014. Imaging the North American continent using waveform inversion of global and USArray data, *Earth planet. Sci. Lett.*, **402**, 26–41.
- Schuster, G.T., Wang, X., Huang, Y., Dai, W. & Boonyasiriwat, C., 2011. Theory of multisource crosstalk reduction by phase-encoded statics, *Geophys. J. Int.*, **184**, 1289–1303.
- Shapiro, N.M. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.*, **31**, 8–11.
- Suda, N., Nawa, K. & Fukao, Y., 1998. Earth’s background free oscillations, *Science*, **279**, 2089–2092.
- Su, W.-J. & Dziewonski, A.M., 1997. Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle, *Phys. Earth planet. Inter.*, **100**, 135–156.
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), 1259–1266.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**, 195–216.
- Tsai, V.C., 2009. On establishing the accuracy of noise tomography travel-time measurements in a realistic medium, *Geophys. J. Int.*, **178**, 1555–1564.
- Wang, Y. *et al.*, 2016. The deep roots of the western Pyrenees revealed by full waveform inversion of teleseismic P waves, *Geology*, **44**, 475–478.
- Woodhouse, J.H. & Dziewonski, A.M., 1984. Mapping the upper mantle’s three-dimensional modeling of earth structure by inversion of seismic waveforms, *J. geophys. Res.*, **89**, 5953–5986.
- Zhang, Q., Mao, W., Zhou, H., Zhang, H. & Chen, Y., 2018. Hybrid-domain simultaneous-source full waveform inversion without crosstalk noise, *Geophys. J. Int.*, **215**, 1659–1681.
- Zhu, H., Bozdag, E. & Tromp, J., 2015. Seismic structure of the European upper mantle based on adjoint tomography, *Geophys. J. Int.*, **201**, 18–52.
- Zhu, H., Komatitsch, D. & Tromp, J., 2017. Radial anisotropy of the North American upper mantle based on adjoint tomography with USArray, *Geophys. J. Int.*, **211**, 349–377.

## APPENDIX: MISFIT FUNCTION BASED ON CROSS-CORRELATIONS

### A1 Discrete misfit function

Let  $E$  be the set of seismic events to be summed. Further, let  $R$  be the common subset of receivers spanning event set  $E$ . For all receiver pairs  $(p, q) \in R \times R$ , we wish to minimize the discrete, weighted cross-correlogram difference misfit function

$$\Phi_E(\mathbf{m}) = \sum_{(p,q) \in R \times R} \Phi^{(p,q)}(\mathbf{m}), \quad (\text{A1})$$

where  $\mathbf{m}$  is the earth model and the inner, trace-pair misfit  $\Phi^{(p,q)}(\mathbf{m})$  is given by:

$$\Phi^{(p,q)}(\mathbf{m}) = \frac{1}{2} \sum_{i=-N+1}^{N-1} \mathbf{W}_{i,i}^{(p,q)} \left[ \Delta C_i^{(p,q)} \right]^2 \quad (\text{A2})$$

in which the difference of discrete cross-correlograms  $\Delta C_i^{(p,q)}$  (i.e. the residual) is given by

$$\Delta C_i^{(p,q)} = C_i(\mathbf{d}^{(p)}, \mathbf{d}^{(q)}) - C_i(\mathbf{G}^{(p)}(\mathbf{m}), \mathbf{G}^{(q)}(\mathbf{m})),$$

where  $\mathbf{d}$  and  $\mathbf{G}(\mathbf{m})$  are the summed data and synthetics for  $E$  (each a time-series of length  $N$ ),  $C_i(\cdot, \cdot)$  represents the discrete cross-correlation operator evaluated at time point  $t_i$ , and  $\mathbf{W}$  is a diagonal weighting matrix reflecting both path uniqueness [for station pair  $(p, q)$ ] and time-dependent windowing.

**A2 Minimization**

In order to solve this minimization problem in the context of the generalized least-squares formalism, we require expressions for both the gradient of eq. (A1) and the Jacobian of the discrete cross-correlogram, both with respect to the model  $\mathbf{m}$  (the former also implicitly depends upon the latter). After some manipulation of eq. (A2), we find that the misfit gradient for a single station pair is given by

$$\frac{\partial}{\partial \mathbf{m}_k} \Phi^{(p,q)}(\mathbf{m}) = - \sum_{i=-N+1}^{N-1} \mathbf{W}_{i,i}^{(p,q)} \Delta C_i^{(p,q)} \mathbf{G}_{i,k}^{(p,q)},$$

where the Jacobian of the discrete station-pair cross-correlogram is given by

$$\mathbf{G}_{i,k}^{(p,q)} = \frac{\partial}{\partial \mathbf{m}_k} \{C_i(\mathbf{G}^{(p)}(\mathbf{m}), \mathbf{G}^{(q)}(\mathbf{m}))\}.$$

For our purposes, it suffices to develop an expression for the latter.

**A3 Cross-correlogram Jacobian**

Defining the above discrete cross-correlation of two real time-series as

$$C_i(s^{(p)}, s^{(q)}) = \sum_j s_j^{(p)} s_{i+j}^{(q)}$$

it follows that

$$\begin{aligned} \mathbf{G}_{i,k}^{(p,q)} &= \frac{\partial}{\partial \mathbf{m}_k} \sum_j s_j^{(p)} s_{i+j}^{(q)} \\ &= \sum_j \frac{\partial s_j^{(p)}}{\partial \mathbf{m}_k} s_{i+j}^{(q)} + s_j^{(p)} \frac{\partial s_{i+j}^{(q)}}{\partial \mathbf{m}_k} \\ &= \sum_j \mathbf{G}_{j,k}^{(p)} s_{i+j}^{(q)} + s_j^{(p)} \mathbf{G}_{i+j,k}^{(q)}, \end{aligned}$$

where  $\mathbf{G}^{(p)}$  and  $\mathbf{G}^{(q)}$  are the traditional summed-wavefield Jacobians for event set  $E$ , evaluated at receivers  $p$  and  $q$ , respectively:

$$\mathbf{G}_{i,k}^{(p)} = \frac{\partial}{\partial \mathbf{m}_k} \mathbf{G}_i^{(p)}(\mathbf{m}) \text{ and } \mathbf{G}_{i,k}^{(q)} = \frac{\partial}{\partial \mathbf{m}_k} \mathbf{G}_i^{(q)}(\mathbf{m}).$$

Finally, it can be seen that

$$\mathbf{G}^{(p,q)} = \mathbf{S}_L^{(q)} \mathbf{G}^{(p)} + \mathbf{S}_R^{(p)} \mathbf{G}^{(q)},$$

where  $\mathbf{S}_L$  and  $\mathbf{S}_R$  are matrices containing left- and right-shifted synthetic time-series, of the form

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & s_0 \\ 0 & 0 & \cdots & s_0 & s_1 \\ 0 & s_0 & \cdots & s_{N-3} & s_{N-2} \\ s_0 & s_1 & \cdots & s_{N-2} & s_{N-1} \\ s_1 & s_2 & \cdots & s_{N-1} & 0 \\ s_{N-2} & s_{N-1} & \cdots & 0 & 0 \\ s_{N-1} & 0 & \cdots & 0 & 0 \end{pmatrix}$$

and

$$\begin{pmatrix} s_{N-1} & 0 & \cdots & 0 & 0 \\ s_{N-2} & s_{N-1} & \cdots & 0 & 0 \\ & & \ddots & & \\ s_1 & s_2 & \cdots & s_{N-1} & 0 \\ s_0 & s_1 & \cdots & s_{N-2} & s_{N-1} \\ 0 & s_0 & \cdots & s_{N-3} & s_{N-2} \\ & & \ddots & & \\ 0 & 0 & \cdots & s_0 & s_1 \\ 0 & 0 & \cdots & 0 & s_0 \end{pmatrix},$$

respectively. Thus,  $\mathbf{G}^{(p,q)}$  may be computed as a post-processing step, combining the individual-station Jacobians  $\mathbf{G}^{(p)}$  and  $\mathbf{G}^{(q)}$  with the corresponding synthetic traces through  $\mathbf{S}_R^{(p)}$  and  $\mathbf{S}_L^{(q)}$ .