Geophysical Journal International

Geophys. J. Int. (2017) **210,** 1517–1524
Advance Access publication 2017 June 7
GJI Seismology

doi: 10.1093/gji/ggx249

Full-waveform inversion of triplicated data using a normalized-correlation-coefficient-based misfit function

Kai Tao,[1,2] Stephen P. Grand[2] and Fenglin Niu[1,3]

[1] *State Key Laboratory of Petroleum Resource and Prospecting, and Unconventional Natural Gas Institute, China University of Petroleum at Beijing, Beijing* 102249, *China. E-mail: taotaokai@gmail.com*
[2] *Department of Geological Sciences, The University of Texas at Austin, Austin, TX* 78705, *USA*
[3] *Department of Earth Science, Rice University, Houston, TX* 77005, *USA*

Accepted 2017 June 6. Received 2017 June 4; in original form 2017 March 15

SUMMARY

In seismic full-waveform inversion (FWI), the choice of misfit function determines what information in data is used and ultimately affects the resolution of the inverted images of the Earth's structure. Misfit functions based on traveltime have been successfully applied in global and regional tomographic studies. However, wave propagation through the upper mantle results in multiple phases arriving at a given receiver in a narrow time interval resulting in complicated waveforms that evolve with distance. To extract waveform information as well as traveltime, we use a misfit function based on the normalized correlation coefficient (CC). This misfit function is able to capture the waveform complexities in both phase and relative amplitude within the measurement window. It is also insensitive to absolute amplitude differences between modeled and recorded data, which avoids problems due to uncertainties in source magnitude, radiation pattern, receiver site effects or even miscalibrated instruments. These features make the misfit function based on normalized CC a good candidate to achieve high-resolution images of complex geological structures when interfering phases coexist in the measurement window, such as triplication waveforms. From synthetic tests, we show the advantages of this misfit function over the cross-correlation traveltime misfit function. Preliminary inversion of data from an earthquake in Northeast China images a sharper and stronger amplitude slab stagnant in the middle of the transition zone than FWI of cross-correlation traveltime.

Key words: Waveform inversion; Computational seismology; Seismic tomography.

1 INTRODUCTION

With recent advances in computational seismology and large-scale computing facilities, seismic wave propagation modeling in realistic 3-D Earth structures can be easily accomplished [e.g. spectral element method (SEM), Komatitsch & Tromp 1999]. At the same time, continuing deployments of dense seismic networks across the globe have accumulated large amounts of high-quality seismic records. Recently, full-waveform tomography that combines the power and accuracy of 3-D waveform modeling with the abundance of high-quality broad-band seismic records has greatly improved the resolution of the seismic structure of Earth's interior (e.g. French & Romanowicz 2014; Chen *et al*. 2015; Zhu *et al*. 2015).

In full-waveform inversion (FWI), one finds a subsurface seismic model that minimizes the misfit between predicted data and observations. Misfit between data and observations, the misfit function, can be defined in different ways. The choice of misfit function plays a central role in how we model the data and has a direct control on model resolution (Luo & Schuster 1991; Rickers *et al*. 2012). A straightforward choice is to use the L2 norm of waveform difference between modeled and recorded seismograms. However, this approach has been shown to be highly non-linear in relating model parameters to data and inversions are prone to finding local minima in data fitting (Gauthier *et al*. 1986). Also, absolute amplitudes are often difficult to model due to uncertainty in factors other than the velocity model, such as source magnitude, radiation pattern, attenuation, receiver site effects and even miscalibrated instruments. To alleviate these problems, a misfit function based on cross-correlation traveltime has been proposed (Luo & Schuster 1991; Marquering *et al*. 1999). It has the advantage that it can handle large time-shifts and is less prone to the 'cycle skipping' problem. However, it is less sensitive to the details of the waveform shape other than an overall time-shift. For example, when multiple phases interfere in the measurement window as for the upper-mantle triplicated waves, the cross-correlation traveltime measurement is dominated by the large amplitude main phases and a single time-shift cannot represent observed waveform distortions. Missing the waveform details limits the resolving power of the data and hence the model resolution. Other types of misfit functions have been proposed, for example, time–frequency misfit (Fichtner *et al*. 2008), multitaper

frequency-dependent traveltime misfit (Zhou *et al*. 2004; Tape *et al.* 2010) and instantaneous phase measurement (Bozdağ *et al*. 2011). These methods separate phase and amplitude and extract frequency or time-varying phase information. In this case, more waveform details are included than in the single cross-correlation traveltime misfit, but the amplitude information is ignored. Although absolute amplitude is difficult to model, there are situations where the relative amplitude of different arrivals are less affected by the source or receiver site effects because the ray paths are much closer near the source and receiver (e.g. upper-mantle triplicated waveforms). By fitting this relative amplitude ratio, we gain extra constraints on the velocity model.

The misfit function based on the normalized correlation coefficient (CC) has been successfully applied in early studies using 1-D waveform fitting (Matzel & Grand 2004) and 3-D FWIs in exploration seismology (Routh *et al*. 2011; Choi & Alkhalifah 2012; Liu *et al*. 2017). This misfit function is sensitive to both phase and relative amplitude in the measurement window but insensitive to the absolute amplitude difference between the data and synthetics, and it is also easy to implement. Liu *et al*. (2017) made detailed comparison between the normalized CC and L2 misfit functions using synthetic data derived from the Marmousi model and concluded that CC-based FWI is a reliable and effective inversion method for exploration seismic data. Here, we show the feasibility of using this misfit function in 3-D FWI of triplicated waves and its potential for high model resolution of the mantle transition zone. In the following, we first formulate the misfit function and the corresponding adjoint source. Then, we show its advantages over the cross-correlation traveltime misfit from an inversion of synthetic data. In the last section, we present a simple real data example from Northeast China to illustrate the effectiveness of the CC-based misfit function.

2 METHOD

Inverting surface seismic displacements to estimate subsurface seismic properties is a difficult, highly non-linear inverse problem. Two general approaches to the problem are direct search algorithms (Sambridge 1999) and techniques based on the gradient of a misfit function (Tarantola 1984; Tromp *et al*. 2005; Fichtner *et al*. 2006). The direct search method samples the model parameter space (usually huge) in a stochastic way and tries to find acceptable models that produce misfits to the data below a certain threshold. The gradient method starts from an initial model and updates the model along a direction based on the model gradient of the misfit function. Since this approach involves an approximate linearization of a non-linear problem, model updates are made iteratively using techniques such as the non-linear conjugate gradient method (Hestenes & Stiefel 1952; Fletcher & Reeves 1964) or quasi-Newton methods, such as limited-memory BFGS (Liu & Nocedal 1989). In both methods, the definition of misfit function is important, as it defines what information we use and how we model the data. The starting model is vital to the success of the gradient method. If the starting model does not lie in the same 'basin of attraction' as the global minimum, the gradient method can only descend to the neighbouring local minimum (e.g. 'cycle-skipping' problem). The direct search method partly alleviates this problem due to its random nature in exploring the model space. However, in large 3-D waveform inversions, the direct search method is computationally prohibitive because of the large amount of forward modeling required. Thus, most current seismic 3-D FWIs are based on the gradient method.

The gradient of the misfit function numerically computes as the change in misfit function for a change in each model parameter. In 3-D waveform inversions, the number of model parameters tends to be very large (several million) and thus the direct calculation of the gradient is very expensive or even prohibitive. The adjoint method overcomes this problem (Tarantola 1984). The gradient is computed as the correlation between the forward wavefield $\mathbf{u}$ and an adjoint wavefield $\mathbf{v}$ (Fichtner *et al*. 2006)

$$\frac{\delta\chi}{\delta\mathbf{m}} = -\int_{t_0}^{t_1} (D_{\mathbf{m}}\mathbf{L}\mathbf{u})^{\dagger}\cdot\mathbf{v}dt, \tag{1}$$

where $\mathbf{m}$ denotes the model parameters, $\mathbf{L}$ is a linear differential operator representing the seismic wave equation $\mathbf{L}(\mathbf{m})\mathbf{u} = \mathbf{f}$, $\mathbf{u}(\mathbf{x}, t)$ is the modeled displacement wavefield excited by the force term $\mathbf{f}$, $D_{\mathbf{m}}\mathbf{L}$ is the derivative of $\mathbf{L}(\mathbf{m})$ with respect to model parameters, † denotes the adjoint of an operator and the time integral starts from the beginning of the earthquake (t_0) to the end of the recorded waveforms under consideration (t_1). The adjoint wavefield $\mathbf{v}$ is the solution of the adjoint wave equation $\mathbf{L}^{\dagger}(\mathbf{m})\mathbf{v} = \delta_{\mathbf{u}}\chi$, and the source term, $\delta_{\mathbf{u}}\chi$ (the variational derivative of misfit function with respect to the modeled waveform), is called the 'adjoint source'. Only two simulations are required then to give the misfit function gradient. In practice, an additional backward simulation is run simultaneously with the adjoint simulation to reconstruct the forward wavefield on the fly to avoid saving the complete wavefield on every gridpoint (Tromp *et al*. 2005).

The cross-correlation traveltime misfit is defined as

$$\chi_{\Delta T} = \frac{1}{2}\sum_{i=1}^{N}|\Delta T_i|^2 \tag{2}$$

where $\Delta T_i = T_i^{\mathrm{obs}} - T_i^{\mathrm{syn}}$ is the traveltime difference measured by cross-correlating between observed and synthetic seismograms in the ith time window for a total of N time windows. The corresponding adjoint source is (Tromp *et al*. 2005)

$$\frac{\delta\chi_{\Delta T}}{\delta\mathbf{u}}(\mathbf{x}, t) = \sum_{i=1}^{N}\Delta T_i\frac{\dot{\mathbf{u}}(\mathbf{x}_i, t)}{\int|\dot{\mathbf{u}}(\mathbf{x}_i, t)|^2dt}\delta(\mathbf{x}-\mathbf{x}_i), \tag{3}$$

where $\dot{\mathbf{u}}(\mathbf{x}_i, t)$ is the time derivative of the synthetic seismograms and $\mathbf{x}_i$ is the location of station on which the ith time window is measured. This adjoint source is completely determined by the traveltime differences and does not depend directly on the waveform differences. The L2 norm waveform misfit,

$$\chi_{L_2} = \frac{1}{2}\sum_{i=1}^{N}\int_{t_0}^{t_1}|\mathbf{d}(\mathbf{x}_i, t) - \mathbf{u}(\mathbf{x}_i, t)|^2dt, \tag{4}$$

which measures the L2 norm of the waveform differences between observed ($\mathbf{d}$) and synthetic ($\mathbf{u}$) seismograms, has an adjoint source equal to the waveform difference (Tromp *et al*. 2005)

$$\frac{\delta\chi_{L_2}}{\delta\mathbf{u}}(\mathbf{x}, t) = -\sum_{i=1}^{N}[\mathbf{d}(\mathbf{x}_i, t) - \mathbf{u}(\mathbf{x}_i, t)]\delta(\mathbf{x}-\mathbf{x}_i). \tag{5}$$

As discussed above, the absolute amplitude is often hard to model and the L2 norm misfit has been found to be difficult to apply. Here, we propose to use a misfit function based on normalized CC, which includes both phase and relative amplitude.

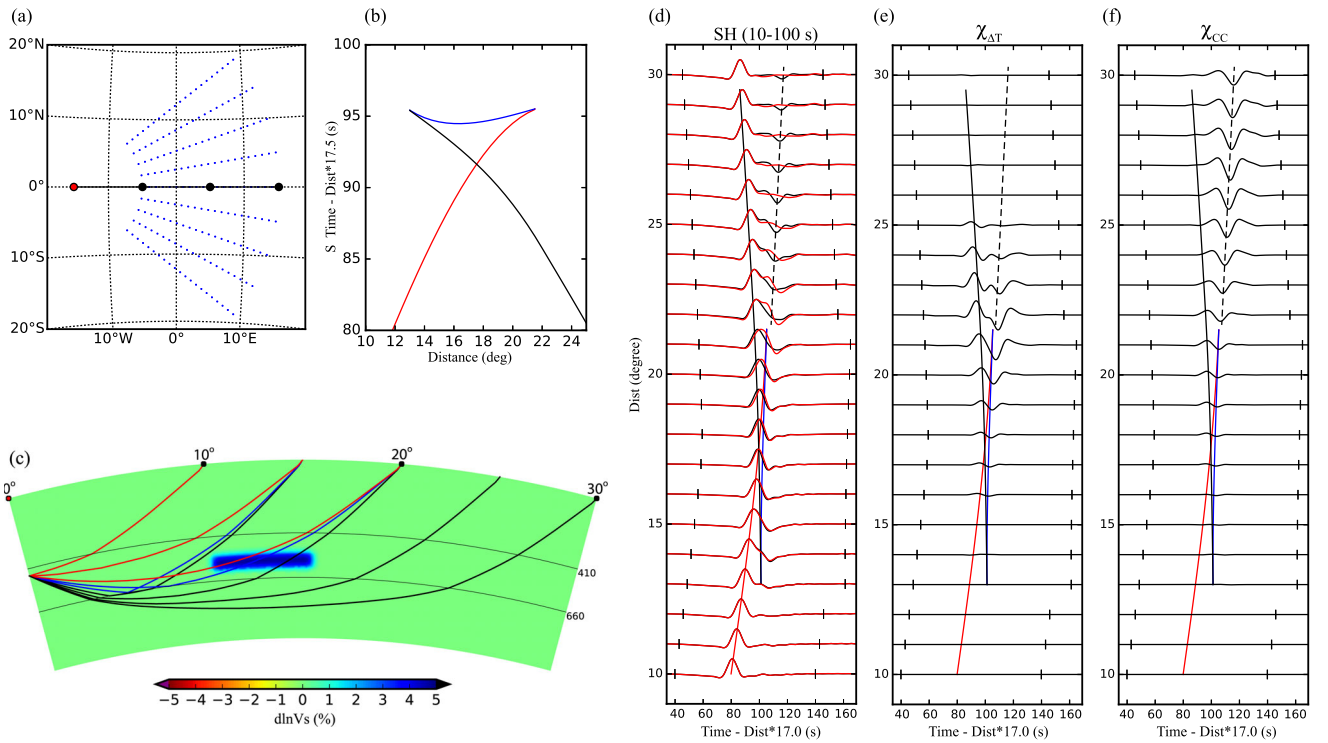

Figure 1. Synthetic model setup. (a) Map shows station distribution (blue dots) and location of the vertical cross-section (black line with dots). A 450 km deep earthquake is located at the left end of the cross-section (red dot). (b) Traveltime curve of the triplicated *S* waves predicted by IASP91 model (Kennett & Engdahl 1991). The red, blue and black colours denote ray paths that turn above, reflect off and turn below the 660 km discontinuity, respectively. (c) Cross-section of the synthetic model. The high V_s anomaly centres at 560 km depth with a width of 550 km in E–W direction and extends 1500 km perpendicular to the cross-section in both the north and south directions. The triplicated *S*-wave ray paths predicted by IASP91 are overlaid and colour coded based on the bottoming depths. (d) Synthetic *SH* seismograms calculated with the 3-D synthetic model (black) and the 1-D starting model (red, IASP91). Stations are located along the cross-section shown in (a). The amplitudes of all seismograms are normalized in individual time windows. The *S*-traveltime curve is overlaid on the synthetics. The dashed line denotes the extended second arrivals due to the high-velocity anomaly. The short vertical bars on the seismograms indicate the time window used in the inversion. (e) and (f) show the corresponding adjoint sources (multiplied by −1) calculated from the traveltime and CC misfit functions, respectively. The amplitudes are normalized by a constant factor for all traces in each case. Note that the adjoint sources for traveltime misfit function do not honour the waveform distortion marked by the dashed line in (d) at distances larger than ∼23°, as compared to the CC-based misfit function (f).

The normalized CC, which measures the similarity between the recorded $\mathbf{d}(\mathbf{x}, t)$ and modeled $\mathbf{u}(\mathbf{x}, t)$ seismograms, is computed as

$$\mathrm{CC} = \frac{\int \mathbf{d}(\mathbf{x}, t) \cdot \mathbf{u}(\mathbf{x}, t)\, dt}{\sqrt{\int |\mathbf{d}(\mathbf{x}, t)|^2 dt \cdot \int |\mathbf{u}(\mathbf{x}, t)|^2 dt}}. \tag{6}$$

We define a misfit function based on the normalized CC as

$$\chi_{\mathrm{CC}} = \sum_{i=1}^{N} (1 - \mathrm{CC}_i), \tag{7}$$

where i denotes the number of time windows used in the inversion. The corresponding adjoint source is

$$\frac{\delta \chi_{\mathrm{CC}}}{\delta \mathbf{u}}(\mathbf{x}, t) = -\sum_{i=1}^{N} W_i{}^{-1} \left(\mathbf{d}(\mathbf{x}_i, t) - A_i \mathbf{u}(\mathbf{x}_i, t)\right) \delta(\mathbf{x} - \mathbf{x}_i), \tag{8}$$

where $W_i = \sqrt{\int |\mathbf{d}(\mathbf{x}_i, t)|^2 dt \cdot \int |\mathbf{u}(\mathbf{x}_i, t)|^2 dt}$ is a normalization factor and $A_i = \int \mathbf{d}(\mathbf{x}_i, t) \cdot \mathbf{u}(\mathbf{x}_i, t) dt / \int |\mathbf{u}(\mathbf{x}_i, t)|^2 dt$ the amplitude ratio factor. The derivation of eq. (8) is given in the Supporting Information. This adjoint source is different from the waveform difference, the adjoint source corresponding to the L2 waveform misfit, by the addition of the magnitude normalization and amplitude ratio terms. The magnitude normalization term leads to equal weight for each time window, regardless of their amplitudes. The amplitude ratio term also removes the absolute amplitude ratio between data and synthetics from the misfit function. This feature is desirable because the absolute amplitude can be hard to model as it can be affected by many factors difficult to incorporate in simulations, as discussed above. However, the relative amplitudes within a time window still contain important information that constrains the velocity structure and is less affected by uncertainties in the source parameters or receiver site effects. For example, when multipath arrivals (like triplicated waves) interfere with each other, since their ray paths are close near the source and receiver, the relative amplitudes are mostly controlled by the regions where the multipath rays separate the most between the source and receiver. For triplicated waves caused by mantle transition zone discontinuities, which are mostly recorded at regional distances (10°–30°), it has long been appreciated that both the relative timing and the amplitude ratios between phases can be used to constrain upper-mantle structure (Tajima & Grand 1998; Wang *et al*. 2014). Thus by including relative amplitudes, more waveform details are modeled, which can further improve the resolution of subsurface structure. We also note that in some FWI studies in exploration seismology (e.g. Warner *et al*. 2013) the amplitude of the modeled data on each trace is normalized in a sliding time window, such that the resultant residual data are minimized. This effectively results in similar adjoint sources to the one calculated from eq. (8), except for the pre-factor that depends on the choice in weighting and amplitude normalization.

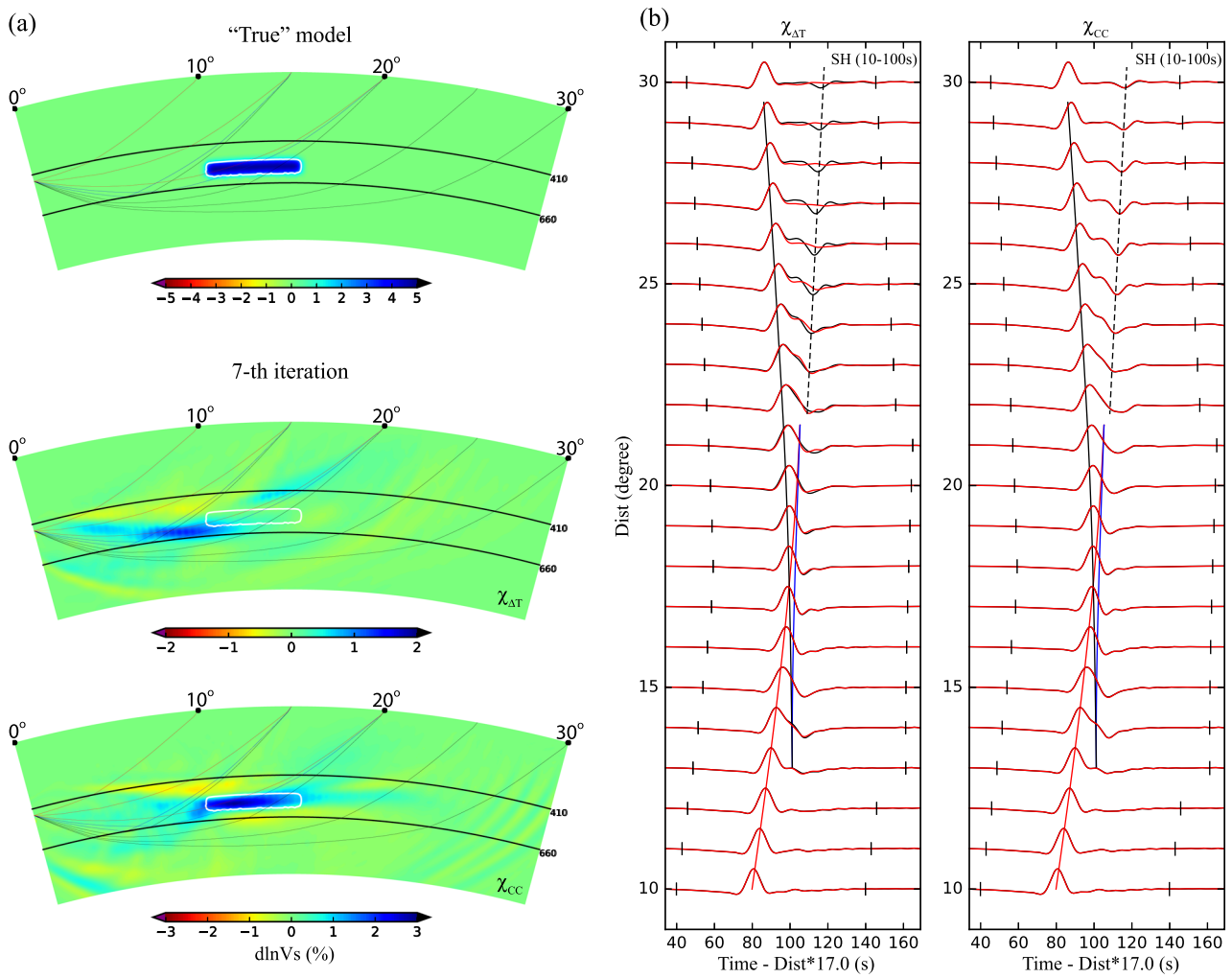

Figure 2. Comparison of inversion results. (a) Cross-sections of the synthetic (top) and inverted V_s perturbations after seven iterations from traveltime (middle) and CC (bottom) based misfit functions. The 1 per cent contour of the synthetic model (white) and the triplicated *S* ray paths predicted by IASP91 are overlaid. (b) *SH* seismograms predicted by the 'true' model (black) and the inverted models (red) from the traveltime (left) and CC (right) based misfit functions. The *S* traveltime curves are overlaid and the dashed lines denote the secondary refracted arrivals.

3 SYNTHETIC TEST

In this section, we run an inversion on synthetic data to compare the effectiveness of traveltime versus cross-correlation misfit functions described above. Synthetics were computed using the SEM (Komatitsch & Tromp 1999). In the SEM simulation, one full mantle wedge (−20–20°E, −20–20°N) is divided into 12 × 12 = 144 slices. Each slice uses one CPU with a total of 144 CPUs. The calculations were accurate down to periods about 10 s. For illustration purposes, we just use one source located at 0°N/−15° E and 450 km depth and we put stations every one degree between 10° and 30° distance from the source. The stations lie along great circles emanating from the epicentre in 10° azimuth increments (Figs 1a and c). The test model consists of the IASP91 model (Kennett & Engdahl 1991) with a superimposed high-velocity anomaly in the transition zone. The anomaly is centred at 560 km depth with a thickness of 80 km. It has a width of 550 km in the E–W direction, and extends 1500 km in the N–S direction (Fig. 1c).

Only the tangential component of the *S* waves (*SH*) are included in the inversion for illustration. All seismograms are bandpass filtered between 10 and 100 s using a second-order Butterworth filter. The pre-conditioned conjugate gradient algorithm (Fletcher & Reeves 1964) is used to iteratively update the starting model towards a model with a smaller misfit, using the information provided by the gradient of the misfit function (sensitivity kernel). The sensitivity kernel is calculated in each iteration via the adjoint method. The kernel amplitudes around the source and receivers are often extremely high. A proper pre-conditioning of the kernel balances the spatial sensitivity and speeds up the convergence rate. In this example, we downweight the regions around the source by applying a mask of ∼500 km in radius, which gradually decreases to zero when approaching the source point. We also taper the shallow part of the kernel beginning at 300 km depth with unit weight towards the surface with zero weight, which reduces the large amplitudes at the receiver side. The model update direction is computed by the non-linear conjugate gradient method from the pre-conditioned kernel and the model update in the previous iteration. We also need to determine the step length by which to update the model along the update direction. Usually the step length is found by the line search method, which requires several forward simulations with different step lengths. Here, we calculate the partial waveform derivatives by finite difference along the update direction and find the optimum step length by fitting the linearized approximated seismograms to the observed data using a grid search. In this way, only one forward

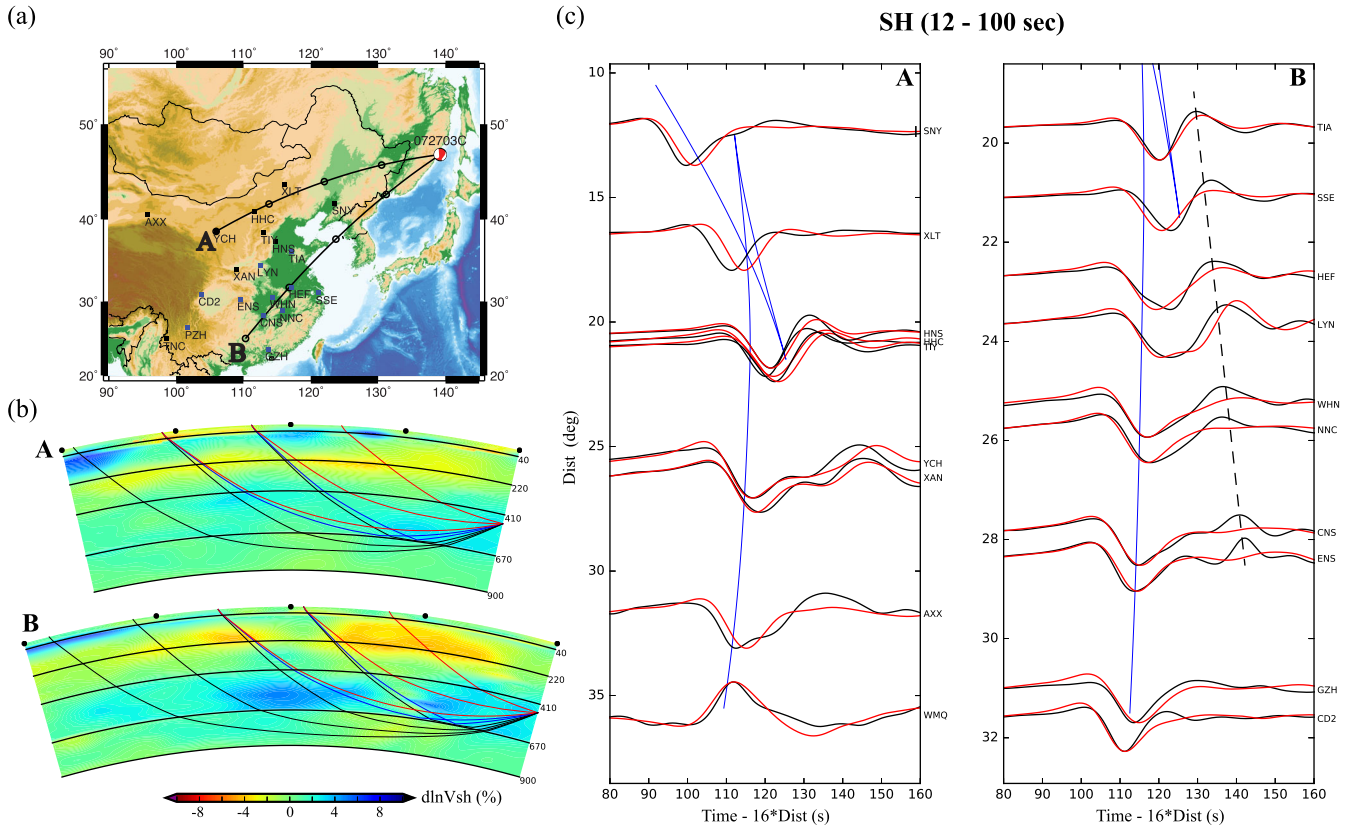

Figure 3. (a) Map shows the locations of event and stations used in the inversion. (b) Two cross-sections (black lines labeled as A and B) show the initial V_{sh} model perturbations relative to the 1-D STW105 model (Kustowski *et al*. 2008), together with the triplicated *S*-wave ray paths. (c) Synthetic *SH* waveforms (red) from the initial model are compared with observations (black). The blue lines show the IASP91 predicted traveltime curves. The dashed lines mark arrivals corresponding to an extension of a refracted arrival from just above the 660 km discontinuity not seen in the initial synthetics.

simulation is needed, and the grid search requires much less computation than the forward simulation. Since the model perturbation is usually small at each iteration step (less than 5 per cent), the linearized approximation is a good approximation to get a useful step length.

The inversion is started from an initial 1-D model (IASP91). Clear secondary arrivals are seen in the IASP91 synthetics (Fig. 1d) as well as waveforms of the 3-D perturbed synthetic model (Fig. 1c), due to the 660 km discontinuity. However, the 3-D synthetics show clear changes in the amplitude and timing of the second arrival. Because the traveltime misfit is dominated by the large amplitude first arrival, the corresponding adjoint sources do not contain waveform distortions associated with the secondary arrivals at distances larger than ∼23° (Fig. 1e), in contrast to the adjoint sources calculated with the CC misfit function (Fig. 1f). We observed that the inverted model and the predicted waveforms do not change much after five iterations and thus stopped at 7. The misfit function $\chi_{\rm CC}$ performs better than $\chi_{\Delta T}$ in both the model recovery and the waveform fit (Fig. 2). The inversion based on traveltime misfit places the high-velocity anomaly mostly near the bottoming depth of the ray paths that turn above the 660 km discontinuity. This misplaced fast anomaly leads to a good traveltime fit as seen on the left-hand panel of Fig. 2(b), but cannot fit the small diffracted energy from the true high-velocity anomaly in the middle of the transition zone (dashed line in Fig. 2b). In contrast, the inversion using the CC-based misfit function correctly places the high-velocity anomaly at the true location but with distortions and smaller amplitude compared with the true model (Fig. 2b, right-hand panel). The waveform fit is almost perfect, which is due to the fact that the CC-based misfit function is sensitive to the waveform similarity. In this one source inversion experiment, the sampling geometry is limited, for example, there are no crossing rays. The purpose is to show that with the same data coverage and computational cost the CC-based misfit function is able to achieve better model resolution than the traveltime misfit function.

4 REAL DATA EXAMPLE

To test the CC-based misfit function on real data, we perform a waveform inversion experiment using data from a single deep earthquake used by Wang *et al*. (2014) to investigate *SH* structure beneath NE China (Fig. 3). Chen *et al*. (2015) have developed a 3-D model for *P* and *S* velocity structure beneath East Asia (EARA2014) using adjoint tomography with a frequency-dependent traveltime misfit function. We use their model as the starting model for our test inversion. Their inversion used periods to 12 s (average grid spacing ∼7.7 km) and we use the same period range although we re-meshed the starting model to a smaller grid spacing (∼4.3 km) such that we can invert to 10 s period in the future. We use the relocated source parameters from Wang *et al*. (2014). The source time function is approximated by a Gaussian function which mimics a triangle with a half-duration of 5.5 s as determined from the Harvard CMT solution (Komatitsch & Tromp 2002). The two panels in Fig. 3(c) show initial waveforms predicted from EARA2014 overlain on the observed data. In this inversion, we use a time window cut around the *S*-wave

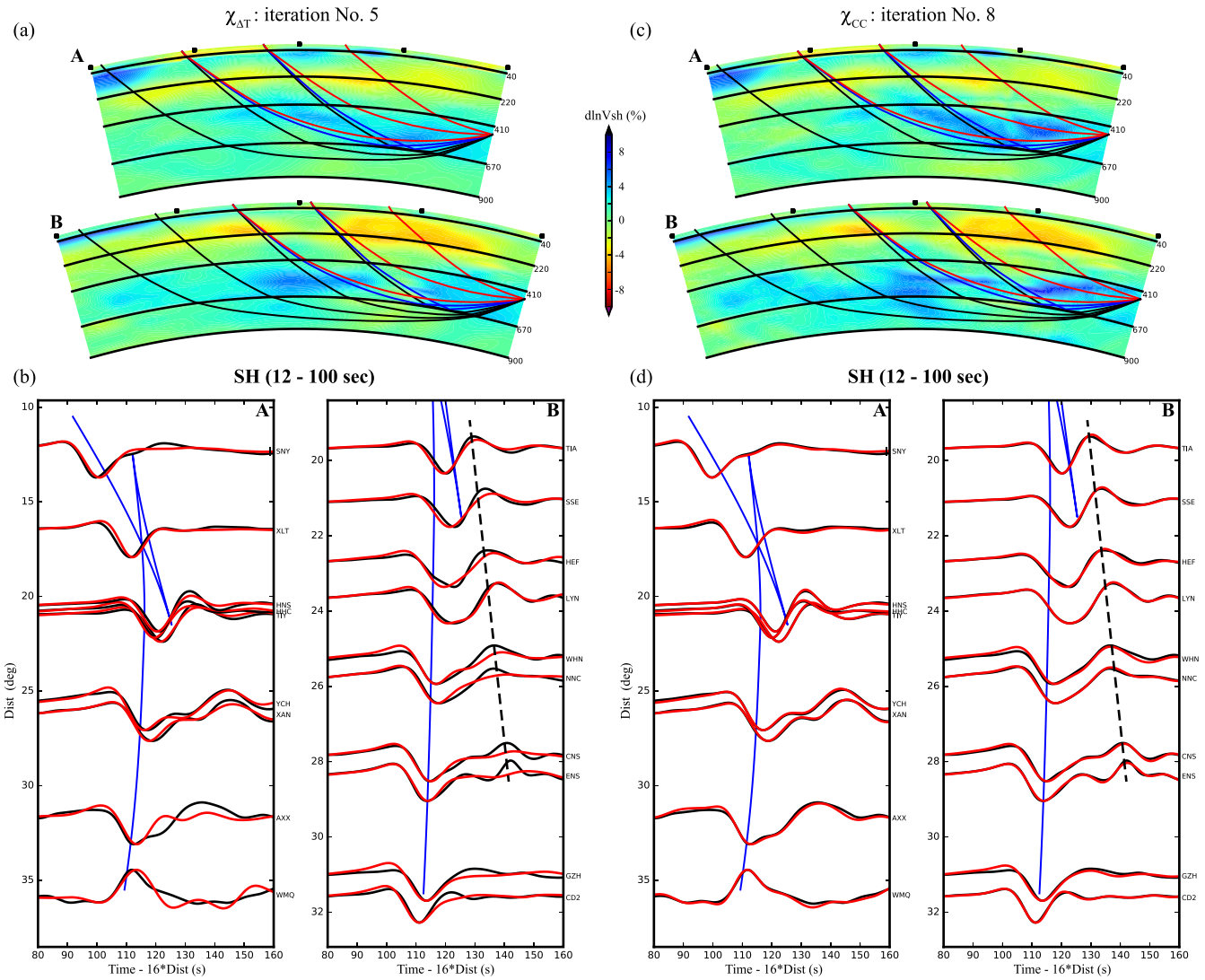

Figure 4. Comparison of inversion results between traveltime and CC-based misfit functions. The final V_{sh} model after five iterations with the traveltime-based misfit function and the corresponding *SH* waveforms are shown in (a) and (b), respectively. (c) and (d) show the results from inversions with the CC-based misfit function. Here, the final V_{sh} model was obtained after eight iterations.

arrival time and only invert the tangential component. The same inversion procedure is used as in the synthetic test. We find that the traveltime-based inversion becomes stable after four iterations, so we stop the inversion at the fifth iteration. However, the CC-based inversion continues to improve waveform fitting after four iterations, and we stop the inversion at the eighth iteration, in which the waveform residuals are comparable with the pre-event noise level. We compare the inversion results in Fig. 4. The traveltime-based inversion reduces the overall time residuals but does not fit the waveform details, most notably the weak second arrival (Fig. 4b). In contrast, the CC-based inversion achieves excellent fits in waveform, especially the relative amplitude and timing of the second arrival (Fig. 4d). This arrival is produced by refractions from the stagnant slab structures above 660 km depth, which extends the triplication branch to far greater distances than the initial model predicts, and places constraints on transition zone structure. Note how the sharper and higher amplitude fast anomalies are needed to fit the data (Fig. 4d). The CC inversion sharpens and increases amplitude in structures that can be seen in the starting model as rather diffuse anomalies (Fig. 3b), which remain more or less the same in the traveltime-based inversion (Fig. 4a). The inversion does not produce spurious short-wavelength structures that would indicate an instability in the approach.

5 DISCUSSION

In the above synthetic and real data experiments, we compare the performance of two misfit functions in waveform tomography. In both cases, the CC-based misfit function achieves better model resolution than the simple cross-correlation traveltime misfit function does. Because the CC-based misfit function exploits the waveform similarity within the measurement window, while the traveltime misfit only extracts one time-shift that may not represent the waveform complexities when secondary phases exist, like diffracted or triplicated waves. The recorded seismograms along profile B shown in Fig. 3(c) are good examples illustrating these waveform complexities. The second arrivals (marked by a dashed line), which are not fit by the initial model and the traveltime derived model (Fig. 4b),

constrain the slab sharpness and amplitude when the CC-based misfit function is used (Fig. 4c).

It has been suggested that a misfit function that mixes the phase and amplitude information could be highly non-linear with respect to model perturbations, and in this case a gradient method is more prone to finding the local minima if the starting model is moderately far from the actual model (Luo & Schuster 1991). In the inversions discussed above, the arrival times are fit reasonably well by the starting models and the traveltime misfit is less than the half-dominant period. The starting model is important for the success of the CC-based misfit function we use, and the method is unlikely to work well when large time-shifts (i.e. cycle skipping) exist between the starting model predictions and data (Gauthier *et al*. 1986). When the starting model does not produce a reasonably good fit in traveltime, one may adopt the multiscale approach that starts the inversion at longer periods and increase the frequency at later iterations, when the long-wavelength structure has been adjusted. Another choice is to stage the inversion using traveltime misfit first and later the waveform misfit when the traveltimes are reasonably fit. A hybrid misfit function that combines the traveltime and waveform misfit functions is also shown to be effective in the case of large starting model misfit (Zhou *et al*. 1995). The same idea can also be used to combine the traveltime and CC-based misfit functions.

6 CONCLUSIONS

We show the feasibility of the misfit function based on normalized CC in 3-D FWI with both synthetic and real data experiments. Compared with misfit based on cross-correlation traveltime, the CC-based misfit function is able to capture the waveform complexities in both phase and relative amplitudes and has higher model resolution. It is particularly suitable for inversions of upper-mantle structure. Data sampling the upper mantle are dominated by triplicated waveforms and by incorporating fitting of secondary arrivals in the misfit function, higher resolution imaging of structures in the mantle transition zone is possible.

ACKNOWLEDGEMENTS

The waveform data were obtained from China Earthquake Administration, and Incorporated Research Institutions for Seismology (IRIS). We used the parallel computing resources by Texas Advanced Computing Center (TACC) at UT Austin. We thank Dr Min Chen for providing the initial model and for helpful discussion. We thank Dr Hejun Zhu for help undertanding the SEM code. We also thank the editor and two reviewers for their constructive comments and suggestions, which significantly improved the quality of this paper. The work was supported by National Science Foundation (EAR-1547228, EAR-1547494 and CMMI-1028889), National Natural Science Foundation of China (41604044 and 41630209) and State Key Laboratory of Petroleum Resources and Prospecting (PRP/indep-03–1616).

REFERENCES

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

Chen, M., Niu, F., Liu, Q., Tromp, J. & Zheng, X., 2015. Multiparameter adjoint tomography of the crust and upper mantle beneath East Asia: 1. Model construction and comparisons, *J. geophys. Res.,* **120,** 1762–1786.

Choi, Y. & Alkhalifah, T., 2012. Application of multi-source waveform inversion to marine streamer data using the global correlation norm, *Geophys. Prospect.,* **60,** 748–758.

Fichtner, A., Bunge, H.P. & Igel, H., 2006. The adjoint method in seismology: II. Applications: traveltimes and sensitivity functionals, *Phys. Earth planet. Inter.,* **157,** 105–123.

Fichtner, A., Kennett, B.L., Igel, H. & Bunge, H.P., 2008. Theoretical background for continental-and global-scale full-waveform inversion in the time–frequency domain, *Geophys. J. Int.,* **175,** 665–685.

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

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.

Gauthier, O., Virieux, J. & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results, *Geophysics,* **51,** 1387–1403.

Hestenes, M.R. & Stiefel, E., 1952. Methods of conjugate gradients for solving linear systems, *J. Res. Natl. Bur. Stand.,* **49**(6), 409–436.

Kennett, B.L.N. & Engdahl, E.R., 1991. Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.,* **105,** 429–465.

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.,* **149**(2), 390–412.

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,** B06306, doi:10.1029/2007JB005169.

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

Liu, Y., Teng, J., Xu, T., Wang, Y., Liu, Q. & Badal, J., 2017. Robust time-domain full waveform inversion with normalized zero-lag cross-correlation objective function, *Geophys. J. Int.,* **209,** 106–122.

Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltime inversion, *Geophysics,* **56,** 645–653.

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

Matzel, E. & Grand, S.P., 2004. The anisotropic seismic structure of the East European platform, *J. geophys. Res.,* **109,** B01302, doi:10.1029/2001JB000623.

Rickers, F., Fichtner, A. & Trampert, J., 2012. Imaging mantle plumes with instantaneous phase measurements of diffracted waves, *Geophys. J. Int.* **190,** 650–664.

Routh, P.S. *et al.*, 2011. Full-wavefield inversion of marine streamer data with the encoded simultaneous source method, in *Proceedings of the 73rd EAGE Conference & Exhibition,* Expanded Abstracts, F032.

Sambridge, M., 1999. Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space, *Geophys. J. Int.,* **138,** 479–494.

Tajima, F. & Grand, S.P., 1998. Variation of transition zone high-velocity anomalies and depression of 660 km discontinuity associated with subduction zones from the southern Kuriles to Izu-Bonin and Ryukyu, *J. geophys. Res.,* **103,** 15 015–15 036.

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.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49,** 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.

Wang, T., Revenaugh, J. & Song, X., 2014. Two-dimensional/three-dimensional waveform modeling of subducting slab and transition zone beneath Northeast Asia, *J. geophys. Res.,* **119,** 4766–4786.

Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N. & Conroy, G., 2013. Anisotropic 3D full-waveform inversion, *Geophysics,* **78**(2), R59–R80.

Zhou, C., Cai, W., Luo, Y., Schuster, G.T. & Hassanzadeh, S., 1995. Acoustic wave-equation traveltime and waveform inversion of crosshole seismic data, *Geophysics,* **60,** 765–773.

Zhou, Y., Dahlen, F.A. & Nolet, G., 2004. Three-dimensional sensitivity kernels for surface wave observables, *Geophys. J. Int.,* **158,** 142–168.

Zhu, H., Bozdağ, E. & Tromp, J., 2015. Seismic structure of the European upper mantle based on adjoint tomography, *Geophys. J. Int.,* **201,** 18–52.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Derivation of adjoint source of CC-based misfit function (eq. 8).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.