REVERSE-TIME MIGRATION-BASED REFLECTION TOMOGRAPHY USING TELESEISMIC FREE SURFACE MULTIPLES

S. BURDICK[1], M. V. DE HOOP[2], S. WANG[2,3], R. D. VAN DER HILST[1]
[1] DEPARTMENT OF EARTH, ATMOSPHERIC, AND PLANETARY SCIENCES, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MA, USA
[2] DEPARTMENT OF MATHEMATICS, PURDUE UNIVERSITY, WEST LAFAYETTE, IN, USA
[3] CURRENTLY AT CONOCOPHILLIPS, HOUSTON, TX, USA

ABSTRACT. Converted and multiply reflected phases from teleseismic events are routinely used to create structural images of the crust-mantle boundary (Moho) and the elasticity contrasts within the crust and upper mantle. The accuracy of these images is to a large extent determined by the background velocity model used to propagate these phases to depth. In order to improve estimates of 3-D velocity variations and, hence, improve imaging, we develop a method of reverse-time migration-based reflection tomography for use with wavefields from teleseismic earthquakes recorded at broad-band seismograph arrays. Reflection tomography makes use of data redundancy – that is, the ability to generate numerous structural images of the subsurface with different parts of the wavefield. In exploration seismology (where it is known as migration velocity analysis) reflection tomography typically involves the generation of an extended image (e.g. offset- or angle-gathers), and the fitness of the background model is evaluated through the application of image-domain annihilators. In regional-scale passive source seismology, however, annihilation-based methods are inadequate because the sparse and irregular distribution of teleseismic sources is not likely to produce illumination over a sufficient range of angles. To overcome this problem we turn towards a source-indexed moveout scheme. Instead of extended image annihilation, we determine the success of the tomographic velocity model by cross correlating images produced with multiply-scattered waves from different teleseismic sources. The optimal velocity model is the one that minimizes correlation power between windowed images away from zero depth shift. We base our inversion scheme on the seismic adjoint method and a conjugate gradient solver. For each image pair, the update direction is determined by correlations between downgoing wavefields with upgoing adjoint wavefields for both images. The sensitivity kernels used in this method is similar to those found in other forms of adjoint tomography, but their shapes are controlled by the spatial distribution of the error function. We present the method and a proof-of-concept with 2-D synthetic data.

1. INTRODUCTION

Since the pioneering papers by Langston (1979) and Vinnik et al. (1979), receiver function analysis of teleseismic free-surface multiples and phase conversions has become the staple of regional scale crust and upper mantle studies with data from seismograph arrays. With this method, structural images are created using data where the direct teleseismic *P* and *S* arrivals have been separated from phases that have converted or multiply scattered at discontinuities in velocity. To provide an image of the discontinuities at depth, these so-called receiver functions used to be subjected to simple stacking at common conversion points (CCP stacks), but lately more sophisticated imaging methods like the generalized Radon transform (Bostock et al., 2001) have been put to use. Several recent studies focused with substantial success on wave-equation migration of converted phases and multiples (Shragge et al., 2006; Chen et al., 2009; Shang et al., 2012).

In most studies to date, the accuracy of the images is controlled by the velocity variations in and simplifying assumptions about the background medium. For instance, popular methods for receiver function analysis of crustal structure (Rondenay, 2009) or upper mantle transition zone studies with SS precursors (Deuss, 2009) assume that the medium is 1D and involve stacking across interfaces that are locally planar. In order to use the wealth of broadband data recorded at increasingly dense arrays and image more complex heterogeneity we must bypass these obstructions and use explicitly 2- or 3-D methods that are able to handle irregular discontinuities. Reflection tomography is one such method that can be used to estimate both the 3-D background velocities in addition to the locations of discontinuities.

Reflection tomography based on reverse time migration (RTM) optimizes a smooth velocity model using the redundancy in reflection data. This redundancy allows for the creation of extended images – that is, image gathers formed using different angular or subsurface offset components of the data – via RTM in a 3-D model. The fitness of the model can then be judged through the application of so-called annihilators. In effect, these annihilators provide a measure of image consistency (in the angle domain) or image focusing (in the offset domain). In the absence of caustics this is equivalent to differential semblance (Symes & Carazzone, 1991). Explicit constructions and implementations of

annihilators for cases with caustics in the wavefield can be found in (Stolk & de Hoop, 2006; De Hoop et al., 2006), which use the principle of downward continuation (Claerbout, 1985). For more recent implementations of annihilators developed for hydro-carbon exploration and production we refer to Shen & Symes (2008) and Sava & Vasconcelos (2011). The method of teleseismic reflection tomography developed here adopts concepts from these active source applications. The regional teleseismic applications that we have in mind are lithospheric studies with dense linear and areal arrays or the upper mantle studies with USArray or similar such arrays in Europe and Asia.

The formation and annihilation of offset or angle gathers requires a continuous range of source incidences. Due to the nature of global seismicity, it is unlikely that this requirement can be met using teleseismic data, though it is possible that an interferometric approach could fill in the gaps in arrival angle with the scattered wavefield (Schuster, 2010). Furthermore, the low frequency and roughly planar geometry of the teleseismic arrival can cause crosstalk in the offset gathers which renders the degree of focusing an insensitive measure of error. Annihilation-based approaches are, therefore, inadequate for wave-equation tomography with wavefields produced by teleseismic earthquakes.

Xie & Yang (2008) present a method for using source redundancy based on measuring residual moveout, or the difference in depth between structure imaged with different sources. The measurement of residual moveout (RMO) relies on the determination of a correlation maximum between these single-source images. In active source experiments, where the source signature is similar for all events, the correlation maximum reliably yields the moveout. In the teleseismic case, however, differences in the source signatures between different events can map into unwarranted moveout in spite of efforts to deconvolve them from the data. Additionally, the RMO method is intrinsically asymptotic. In order to build an inversion scheme, we must know the direction that the images move due to the perturbation. This moveout direction is not generally normal to the reflector, and must be solved for via a continuation ray tracing system (Duchkov & de Hoop, 2009).

For application to teleseismic wavefields we propose an alternate approach using a misfit criterion in the image domain. The wave equation reflection tomography presented here is based on a RTM inverse scattering transform (Op 't Root et al., 2012) and aims to use the similarity of images produced from different sources as the criteria to find a smooth velocity model that best explains the data. To avoid problems arising from unknown finite-bandwidth source functions, we move towards a correlation power norm in the image domain; here, we adapt the cross correlation power measure introduced by Van Leeuwen & Mulder (2008, 2010). When used in the image domain, this power norm approach penalizes correlation power between images away from zero depth shift. The resulting error function is robust and has a broad basin of attraction. We note that the cross correlation power has some similarities with the criterion used in wave-equation shear-wave splitting tomography (Long et al., 2008).

This paper consists of two main parts. In Section 2 we describe the geometry and mathematical analysis of the problem. Here we develop an inverse scattering theory based on the scalar Helmholtz equation and an image correlation power error function. We also develop an adjoint method for determining the gradient of the error function that is to be optimized, which we present in an Appendix. In Sections 3 and 4, which contain synthetic "experiments," we investigate the error function and present inversion results from 2-D synthetic subduction model.

2. RTM-BASED REFLECTION TOMOGRAPHY: GEOMETRY AND THEORY

2.1. **Teleseismic free-surface reflections.** We apply RTM-based reflection tomography to teleseismic *P*-wave free-surface multiple reflections. The geometry of this problem is shown in Figure 1. We consider incident *P*-waves from earthquakes recorded at seismograph arrays at epicentral distances between 30° and 90°. The surface-reflected waves then propagate downwards, scatter at major discontinuities like the Moho, subducted oceanic crust, or the mantle transition zone, and the back-scattered waves (or teleseismic multiples) are recorded at the same array. We use the incident wavefield as a "source" and the multiple wavefield as "data" to perform RTM and estimate the non-smooth contrasts in elastic properties beneath the array. Since the incident field used for reflection tomography can be estimated directly from the recorded wavefield, we do not need to know the precise location or focal mechanism of the earthquakes.

In the teleseismic setting we incorporate the free surface boundary condition in the fundamental solution to the wave equation. That is, the solution to the source-side problem includes both the initial teleseismic leg of propagation and the downward propagation between the free surface and the target discontinuities. Here, however, we suppress the propagation of the incident field along the teleseismic leg for simplicity of computation and to limit the sensitivity of the method outside of the area of interest. Instead, we intercept the source wavefield as a single-layer potential field at the surface estimated from the data, and we only consider its propagation following the free-surface reflection.

For the purposes of establishing a bridge between previous work in the active source setting and future application to teleseismic data, we restrict the analysis in this paper to *P*-polarized waves and, hence, use the scalar wave equation.

While *P* to *S* converted phases dominate the teleseismic scattered field, recent studies (Chen et al., 2009; Pearce et al., 2012) have shown that it is possible to create images using *P* backscattered phases such as Pp_mp. Future development of this method will consider the elastic case so that converted phases can be included.

Teleseismic sources tend to have complicated source-time functions. Owing to rupture processes and reverberation from near-source structure, the length of these source-time functions can exceed the difference in arrival time between the direct phase and multiples from structures in the lithosphere and upper mantle. This fact makes it difficult to decouple the direct *P* from *P* backscattered phases, particularly for flat discontinuities. Common preprocessing approaches (as in Rondenay 2009) address this by separating the phases using cross correlation based alignment and principle component analysis and removing the source time function by deconvolution. Images created with such preprocessed data will inevitably still have some coupling between imaged structures and source-time functions. Our search for a robust reflection tomography error measure is in part motivated by this limitation.

2.2. **Inverse scattering.** We introduce coordinates $(\mathbf{x}, z)$ such that $\mathbf{x} = (x_1, \ldots, x_{n-1})$ and $x_n = z$, with n the dimension of the problem (that is, n= 2 or 3) and z the depth coordinate. We assume that the data are recorded at $z = 0$. For the purposes of the analysis and proof of concept below, we develop our inverse scattering theory with a point source in mind. Absent polarity effects due to the double couple source – a reasonable assumption for measurements at a regional-scale array at teleseismic distances – we can treat an earthquake source (with its source-time function deconvolved) as a Dirac pulse at $(\mathbf{x}_\mathrm{s}, z_\mathrm{s})$. With G the Green's function, the fundamental solution of the wave equation is then given by

$$\begin{aligned} [c(\mathbf{x},z)^{-2}\partial_t^2 - \Delta]G(\mathbf{x},z,t) &= \delta(\mathbf{x}-\mathbf{x}_\mathrm{s}, z-z_\mathrm{s})\delta(t), \\ \text{subject to initial values } G(\mathbf{x},z,0) = 0 \text{ and } \quad \partial_t G(\mathbf{x},z,0) &= 0. \end{aligned} \tag{1}$$

Here, c is a smooth wavespeed, which we expand into basis functions ψ_k (wavelets or splines, for example):

$$c(\mathbf{x},z) = \sum_{k=1}^{N_k} \gamma_k \psi_k(\mathbf{x},z), \tag{2}$$

with the weights γ_k to be determined by inversion. As mentioned above, we will suppress the teleseismic leg of propagation and represent G as single-layer potential at the free surface.

Following the Born approximation we represent the medium as a combination of smooth wavespeed variations c and a non-smooth perturbation δc. That is, the scattering problem is obtained by linearization of (1) with velocity $(1 + r(\mathbf{x}, z))c(\mathbf{x}, z)$, where $r(\mathbf{x}, z) = \delta c(\mathbf{x}, z)/c(\mathbf{x}, z)$ is the (frequency independent) reflectivity function. The non-smooth character of $r(\mathbf{x}, z)$ gives rise to the scattered or reflected wave exploited here. We multiply (1) with $c(\mathbf{x}, z)^2$ and find

$$\begin{aligned} [\partial_t^2 - c(\mathbf{x},z)^2\Delta]u(\mathbf{x},z,t) &= 2r(\mathbf{x},z)\,\partial_t^2 G(\mathbf{x},z,t), \\ \text{subject to } u(\mathbf{x},z,0) = 0 \text{ and } \quad \partial_t u(\mathbf{x},z,0) &= 0. \end{aligned} \tag{3}$$

The scattered wave field $u(\mathbf{x}, z, t)$, caused by interactions of the surface reflected waves with subsurface contrasts $r(\mathbf{x}, z)$, is defined as the solution of this problem and will be the synthetic analog the observed wavefield. The mapping of $r(\mathbf{x}, t)$ to u can be represented by a forward (modeling) operator, say $\mathcal{F}$. Since we develop the theory here with a point source in mind, we make the common RTM assumption (Stolk, 2000) that no caustics or multi-pathing occur between the source and the scatter points. For the application of source-indexed reflection tomography this is of particular importance since artefacts created due to caustics will not average out through stacking. Note that in practice, by intercepting the source field at the surface with three component receivers, we can determine directional information about the wavefield. This effectively allows us to determine which path is taken between the surface and the scatterer, and we then would not need to make the no-caustic assumption.

We then define Σ as the set of receiver locations at $z = 0$, so that R_Σ is an operator that restricts the wavefield to the part that is actually recorded at the receiver array. The modeled data is then $d(\mathbf{x}, t) = R_\Sigma u(\mathbf{x}, z, t)$.

We now introduce the reverse-time propagated field, u_r, as the anticausal solution to

$$[c(\mathbf{x},z)^{-2}\partial_t^2 - \Delta]u_\mathrm{r}(\mathbf{x},z,t) = \delta(z)(Nd)(\mathbf{x},t), \tag{4}$$

with the operator N acting on data d the composition of preprocessing operators (which will be described below) and the up/down decomposition operator (De Hoop, 1996)

$$-2\mathrm{i}D_t\, c_0(\mathbf{x})^{-1}\sqrt{1-c_0(\mathbf{x})^2 D_t^{-2} D_x^2} \tag{5}$$

that acts on the wavefield at the free surface. Here, $D_t = \mathrm{i}^{-1}\partial_t$, $D_x = \mathrm{i}^{-1}\partial_x$, and $c_0(\mathbf{x}) = c(\mathbf{x}, 0)$ is the wavespeed at the surface (which is assumed to be fixed).

We proceed from here in the frequency domain, but where our method is concerned there are no inherent advantages or disadvantages to propagation and inverse scattering using time versus frequency domain formulation. We then obtain the image, $\mathcal{I}$, from the action of the inverse scattering transform, say $\mathcal{H}$, on the data:

$$\begin{aligned}\mathcal{I} = (\mathcal{H}d)(\mathbf{x}, z) &= \frac{1}{2\pi}\int_B \frac{\chi(\omega)}{\mathrm{i}\omega}\left(\frac{1}{\widehat{G}(\mathbf{x}, z, \omega)}\,\widehat{u}_{\mathrm{r}}(\mathbf{x}, z, \omega) - \frac{c(\mathbf{x}, z)^2}{\omega^2}\,\nabla\frac{1}{\widehat{G}(\mathbf{x}, z, \omega)}\cdot\nabla\widehat{u}_{\mathrm{r}}(\mathbf{x}, z, \omega)\right)\mathrm{d}\omega \\ &\approx \frac{1}{2\pi}\int_B \frac{\chi(\omega)}{\mathrm{i}\omega|\widehat{G}(\mathbf{x}, z, \omega_0)|^2}\frac{|\omega_0|^{n-3}}{|\omega|^{n-3}}\left(\overline{\widehat{G}(\mathbf{x}, z, \omega)}\,\widehat{u}_{\mathrm{r}}(\mathbf{x}, z, \omega) - \frac{c(\mathbf{x}, z)^2}{\omega^2}\,\nabla\overline{\widehat{G}(\mathbf{x}, z, \omega)}\cdot\nabla\widehat{u}_{\mathrm{r}}(\mathbf{x}, z, \omega)\right)\mathrm{d}\omega,\end{aligned} \tag{6}$$

where $\widehat{\ }$ denotes the Fourier transform with respect to time, $\overline{\ }$ denotes the complex conjugate, $\chi(\omega)$ is a smooth function that ensures that zero frequency is left out of the integration, and B signifies the available bandwidth. The approximation to the inverse scattering transform (right hand side of (6)) is introduced here for ease in calculating the adjoint state below. In effect, we choose the wavefield $\widehat{G}(\mathbf{x}, z, \omega_0)$ for one frequency ω_0 in the upper part of our frequency band to provide the source illumination factor, $|\widehat{G}(\mathbf{x}, z, \omega_0)|^{-2}$.

Op 't Root et al. (2012) show that, for a given source, the inverse scattering transform $\mathcal{H}$ is asymptotically the inverse of the modeling operator $\mathcal{F}$. We note that $\mathcal{H}$ and $\mathcal{F}$ both depend on the given data set (that is, the configuration of the array). The gradient terms in Equation 6 are important for the suppression of low-frequency artifacts (Wang et al., 2010).

2.3. **Correlation power functional.** As mentioned above, the sparse and irregular distribution of teleseismic sources requires pairwise comparison between images formed with different sources. The construction of single-source images is, therefore, critical for the development of an error function using teleseismic data.

We indicate the dependencies on the source as follows. For a source i located at $(\mathbf{x}_{s;i}, z_{s;i})$, and which generates data $d_i(\mathbf{x}, t)$, with $i = 1, \ldots, N_s$, we write G_i and $u_{\mathrm{r};i}$ for G and u_{r}, respectively. For source i the image $\mathcal{I}_i$ – the reconstruction of the non-smooth medium perturbations $r(\mathbf{x}, z)$ – is

$$\begin{aligned}\mathcal{I}_i(\mathbf{x}, z) = \frac{1}{2\pi}\int_B \frac{\chi(\omega)}{\mathrm{i}\omega|\widehat{G}_i(\mathbf{x}, z, \omega_0)|^2}\frac{|\omega_0|^{n-3}}{|\omega|^{n-3}} \\ \left(\overline{\widehat{G}_i(\mathbf{x}, z, \omega)}\,\widehat{u}_{\mathrm{r;i}}(\mathbf{x}, z, \omega) - \frac{c(\mathbf{x}, z)^2}{\omega^2}\,\nabla\overline{\widehat{G}_i(\mathbf{x}, z, \omega)}\cdot\nabla\widehat{u}_{\mathrm{r;i}}(\mathbf{x}, z, \omega)\right)\mathrm{d}\omega, \quad i = 1, \ldots, N_s.\end{aligned} \tag{7}$$

We note that the set of receiver locations used can also depend on the source location, that is, on i.

In the case of idealized illumination, the location of the images $\mathcal{I}_i(\mathbf{x}, z)$ should be equal for different sources i . That is, the imaging is redundant in $(\mathbf{x}_s, z_s)$. We exploit this redundancy to estimate $c(\mathbf{x}, z)$. To this end, we identify $N_{\mathcal{I}}$ target reflection points in particular depth intervals. We let $\varphi_\alpha(\mathbf{x}, z)$ be the spatial tapers which will effectively act as correlation windows that localize the relevant parts of the image:

$$(\varphi_\alpha\mathcal{I}_i)(\mathbf{x}, z) = \varphi_\alpha(\mathbf{x}, z)\mathcal{I}_i(\mathbf{x}, z).$$

We then form cross correlation between the images for different sources i and j,

$$\mathcal{C}_{ij}^{\alpha}(\mathbf{x}, \Delta z) = \int_{\mathbb{R}_{\geq 0}} (\varphi_\alpha\mathcal{I}_i)(\mathbf{x}, z)\,(T_{\Delta z}(\varphi_\alpha\mathcal{I}_j))(\mathbf{x}, z)\,\mathrm{d}z, \quad i \neq j,\ i, j = 1, \ldots, N_s,\ \alpha = 1, \ldots, N_{\mathcal{I}}, \tag{8}$$

where

$$(T_{\Delta z}(\varphi_\alpha\mathcal{I}_j))(\mathbf{x}, z) = (\varphi_\alpha\mathcal{I}_j)(\mathbf{x}, z + \Delta z)$$

represents a translation in depth.

We introduce the weighted correlation power $\mathcal{J}$ – as a function of wavespeed or, according to (2), as a function of γ – as the energy functional that is to be optimized:

$$\mathcal{J}[\gamma] = \sum_{\alpha=1}^{N_{\mathcal{I}}}\sum_{i\neq j, i, j=1}^{N_s}\frac{1}{S_{ij}^{\alpha}}\int_X\int_E W_B(\Delta z)\left|\mathcal{C}_{ij}^{\alpha}(\mathbf{x}, \Delta z)\right|^2\mathrm{d}\Delta z\,\mathrm{d}\mathbf{x}. \tag{9}$$

E is a closed interval containing 0 over which we vary Δz and W_B is a Gaussian weighting function which penalizes correlation energy away from zero depth shift:

$$W_B(\Delta z) = -\exp\left[-\frac{\Delta z^2}{\sigma_B^2}\right] \tag{10}$$

This weighting function is dependent on the frequency content of the data. In order to best capture shifts in depth, we choose σ_B such that W_B has a width that is of the same order as the first peak of the frequency-dependent correlation. Furthermore,

$$S_{ij}^{\alpha} = \int_X \int_E \left|\mathcal{C}_{ij}^{\alpha}(\mathbf{x}, \Delta z)\right|^2 \mathrm{d}\Delta z \, \mathrm{d}\mathbf{x} \tag{11}$$

represents a normalization which ensures that the functional is not minimized simply by correlating low-amplitude background and shifting the relevant reflectivity in the images outside the correlation windows.

For the error measure described here we develop a model gradient using the augmented Lagrangian method and optimize the velocity iteratively by a conjugate gradient algorithm. The details of the gradient construction can be found in Appendix A.

3. Inversion example – subduction model

3.1. **Generation of synthetic dataset.** For a first test of our inversion method we consider a target structure (figure 2(a)) based loosely on the subduction model of Audet et al. (2007). The model consists of a slow continental crust, a subducted slab with a relatively slow oceanic crust and faster lithosphere, and a mantle wedge between the slab and the continental crust. Since modeling the initial teleseismic leg of the wave propagation is difficult to achieve without resorting either to high frequency approximations or a great computational expense, we limit our computational domain to the area beneath the array. The model domain is spanned by a 300 km offset at the surface and a depth range of 100 km with grid spacing of 250 m.

We illuminate the model structure with planar waves arriving with different angles of incidence from the bottom of the domain. We use the 2-D Helmholtz solver (Wang et al., 2010) to generate a dataset up to 2 Hz that includes 29 different teleseismic events with incoming incidence angles between -30° and 30°. The Helmholtz solver accounts for multiple forward and backscattering. The direct incoming wave is altered by velocity heterogeneity as it travels through the model. It is then reflected back downwards at the top of the domain by a free surface boundary condition, and the back-scattered surface-reflections are recorded at the surface. We store data at every computational node at the free surface within the bounds of the array.

Following the generation of synthetic data, several preprocessing steps are needed before the data can be used for reflection tomography. We first time-shift the traces according to their predicted arrival times based on the 1D model so that they are nearly aligned according to the direct *P*. Next, we use multichannel cross correlation on a window around the first arrival to align the wavefield and then perform a principal component analysis. The incoming wavefront is only affected weakly by the model heterogeneity before it reaches the array. Empirically, then, we find that retaining the first two eigenvectors as the incident "source-side" wavefield yields the best results. The rest of the eigenvectors are used as the scattered multiple field. In the presence of horizontal reflectors in the medium, the estimate of the incident field will contain arrivals of the multiples. Both wavefields are then shifted back to their absolute times. The results of this process are displayed in Figure 3.

Since the incident field and the multiply scattered field are both recorded at the free surface, they are each the composition of an upgoing incoming field and downgoing reflected field. We wish to utilize the downgoing component of the incident field and the upgoing component of the multiples for our imaging. It is therefore necessary to perform a directional decomposition to separate these components. This is accomplished through application of the decompostion operator described in Equation 5. We follow Van Stralen et al. (1998) and chose a discretization of this operator in the frequency domain based on a symbol approximation in order to ensure proper amplitudes for our propagations.

We note that for the application of this method to 3-component teleseismic data from regional scale arrays, other preprocessing steps are needed as outlined in Section 4.4.

3.2. **Imaging with synthetic data.** As the first step in the inversion procedure, a reverse time migration is performed on the data from each separate teleseismic source. Using the inverse scattering transform (equation 7) we are able to recover an accurate estimate of the non-smooth part of the velocity model (Figure 2(b)) up to illumination effects. Figure 4(a) shows the image resulting from the traditional stacking approach for the true smooth model (Figure 2(c)). Figure 4(b) shows the imaging result when using a 1D reference model. Due to the overestimation of the model

velocity near the top of the region, the upper reflectors (continental Moho, top of slab and slab Moho) come in below the correct location while the bottom reflector (base of slab) registers as too shallow due to the underestimation of velocities in the slab. Note that without *a priori* knowledge of the reflector locations it would be difficult to distinguish which of the two images is correct based on the stacked images alone.

The accuracy of the imaging for a given bandwidth of data depends on the spacing of the stations and the ability to interpolate the wavefield between them. For a station spacing of greater than 2.5 km, images formed with data with 2 Hz peak frequency exhibit aliasing. At shallow depths, where the Fresnel zones do not overlap, the sparsity of stations introduces artifacts which makes it both difficult to interpret the image and to use the result to perform tomography. For the frequency content prevalent in teleseismic data, however, a station spacing or interpolation of 5 km is sufficient for forming an accurate image. In practice, this interpolation can be effectively done via a curvelet decomposition (Häuser & Ma, 2013). For the imaging results presented here, the wavefield at the surface is defined on all grid nodes within the array area.

In order to apply the tomographic error function to the image, we eschew stacking it over the whole dataset and instead compare between images formed by different events. Figure 5 show single source images formed in the true model for a variety of incidence angles. Events with high angles of incidence (Figure 5(a) and (c)) have difficulty illuminating the discontinuity between the oceanic lithosphere and the mantle due to the geometry of the incoming wave and baseline length of the array.

3.3. **Error function.** It can be expected that, given the correct background model, images from different events will form images focused at the same locations, up to illumination effects. Figure 6 demonstrates the effect of the background model on the source-indexed image gathers. The image gathers display the residual moveout between different single-source images by setting them side by side. The shape of these gathers gives information on the success of the imaging. If the model is correct, as in Figure 6(b), the image gathers should be essentially flat. If the model is incorrect, moveout appears in the image gathers – a slow model leads to images from high incidence angle events forming at shallower depths than low incidence angle events. The reality of teleseismic data dictates that these images will, however, have differing shape due to varied frequency content, stretching effects, and coupling between imaged structure and imperfectly deconvolved source-time functions. For this reason, we look for an image-based error function sensitive only to the kinematics of the image.

Van Leeuwen & Mulder (2010) introduced the correlation power norm as a robust way to characterize the differences in depth moveout in between images. This is illustrated by Figure 6(a). In practice, we pairwise cross correlate partial images generated by distinct sources. The images essentially vary continuously with source position along a curve (in 2.5D, or surface in 3-D) in any reasonable velocity model. Thus, even if the velocity model used for imaging is far from the true model, we need not ensure that the images generated by different sources are sufficiently close to one another.

The efficacy of the correlation power measure depends on a variety of parameters. Foremost among these is weighting function, $W_B(\Delta z)$. The purpose of the weighting function is to penalize correlation energy far away from zero shift, but at the same time we wish to avoid penalizing the image for the shape of its reflectors. We therefore look for a weighting function with a width of the same order as the first correlation peak which dependends on the bandwidth of the data.

Figure 8 illustrates the error function using the correlation power in depth. This correlation is multiplied by the weight function, and the sum of the weighted correlation is added to the error function. In order to ensure that we are not mapping noise into the error function we must limit the image correlation to areas where a strong, consistent image exists for the events considered. The selection of correlation windows is based on the amplitude, coherence, and linearity of the images. Typically, we take into account image points where the sum of the correlation power between two images exceeds a threshold value relative to the average strength of image autocorrelation. This set of image points can be updated during the iterative optimization if necessary. Figure 7 shows the location of image points and the correlation windows around them that were used in our inversion.

The use of a correlation error function in depth only may bias the inversion towards reflectors with shallow dips. In principle, the correlation should ideally take place in the direction that the image moves under a perturbation of the velocity model, as determined by the solution of the velocity continuation ray tracing system (Duchkov & de Hoop, 2009). While it is possible to adjust the correlations in this way, if the target reflectors are horizontal or near horizontal and if the moveout of the images is almost completely in the vertical direction, it is sufficient to consider only vertical shifts in the correlations and apply this measure iteratively. For the case of non-planar scattering points or steeply-dipping reflectors it is also possible to design a correlation function in two dimensions, which will allow for

a measurement of image moveout in a general direction. The weighting function in equation 10 then takes the form:

$$W_B(\Delta x, \Delta z) = -\exp\left[-\frac{\Delta x^2}{\sigma_{B,x}^2}\right] - \exp\left[-\frac{\Delta z^2}{\sigma_{B,z}^2}\right] \tag{12}$$

where weighting parameters $\sigma_{B,x}$ and $\sigma_{B,z}$ can be fixed based on frequency content and reflector dip (and need not be the same).

To demonstrate the effectiveness of the error function we test its ability to distinguish the correct model from an incorrect model, which, however, retains the general 1-D structure of the model. We vary the strength a perturbation δc given by

$$\delta c[\alpha] = c_{true}(x, z) - \alpha c_{true}(X - x, z) \tag{13}$$

by a parameter α from -1 to 1. In effect, we start with an initial guess that is the mirror image of the actual subduction model since $c_{true} + \delta c[-1] = c_{true}(X - x, z)$. Figure 9 shows the endmember models and their resulting error measures. Even for such a severe model perturbation, the correlation power error function proves to be convex and robust.

3.4. **Adjoint calculation.** We construct the model gradient using the seismic adjoint method. For each image pair correlation, there are four contributions to the gradient – one from the source-side leg of propagation (Equation 25) and one from the receiver-side (Equation 26) for both images involved. The construction of these gradients involves a solution to the Helmholtz equation for each of the four contributions with an adjoint source that is distributed over the entire correlation window. Correlation of these adjoint fields with each of their respective forward fields then determines the gradient. For a single correlation window, these four contributions to the gradient take the form of deformed "bananas", where the kernel extends in the direction of propagation, widening at the surface since all receivers are included in the calculation for both source- and receiver-side gradients.

Previous methods use such adjoint calculations to determine a sensitivity kernel that allows the projection of an error measure, e.g. traveltime delay (Dahlen et al., 2000; De Hoop & Van der Hilst, 2005), residual moveout (Xie & Yang, 2008) or extended image annihilation (De Hoop et al., 2006), back onto a model space. For the reflection tomography considered here, the shape of the sensitivity kernel is inextricably linked to the error measurement itself. The spatial extent and characteristics of the "sources" in equations 25 and 26 are determined by the correlations and choices of weighting functions.

To confirm the accuracy of our inverse theory, we compare the kernels determined by an adjoint state calculation with ones computed via a finite difference approach. Beginning with a smooth 1D model, c_0, we solve the forward problem once with a perturbation of $\delta\gamma_k = 50$ m/s to each basis function $\psi_k(\mathbf{x}, z)$ and find the perturbed error function. The finite difference gradient,

$$\frac{\partial \mathcal{J}}{\partial \gamma_k} = \delta\gamma_k^{-1}\left(\mathcal{J}[c_0] - \mathcal{J}[c_0 + \delta\gamma_k\psi_k(\mathbf{x}, z)]\right), \tag{14}$$

should then be approximately equal to the one computed by the adjoint method. The kernels for one correlation pair are shown in Figure 10 with the adjoint method gradient projected onto the basis functions for comparison. Above the correlation window, the kernels are quite similar, but the finite difference version does not include a sensitivity below and directly beside the correlation window due to a lack of backscattering from the smooth perturbations. These kernels demonstrate the sensitivity of the error function to scattering in the transmission regime between the discontinuities and the surface. It is this sensitivity that renders the method apt for optimizing the smooth, long wavelength velocity model.

3.5. **Inversion results.** We ran a series of test inversions to examine the resolving power of the method. For the first series of tests, we consider updates only the central 150 km of the model where the structure is illuminated by all sources. The initial model (Figure 11(a)) has 1D structure in the target region and a smooth version of the true subduction zone model in Figure 2(a) elsewhere. The perturbation between the true and initial models is shown in Figure 11(b). To cover the target region we choose as our basis functions quadratic B-splines with 6 km spacing, with 23 bases in offset and 14 in depth.

Figure 12 shows several iterations from the inversion with the most complete set of data parameters. This inversion was performed using data up to 2 Hz in frequency and for the full set of events with incident angles ranging from -30° to 30°. The error function correlation windows used, shown in Figure 7, were picked based on coherent single-source image formation in the starting model and are not updated during the process. The error function parameters $\sigma_{B,z}$

were of the order of 0.5-1.0 km. The maximum Δz for the correlations was typically ± 10 km but varied according to the width of the image and nearness of other reflectors.

We perform the inversion via non-linear conjugate gradient. After 5 iterations (Figure 12(a)), the general outline of the test perturbation comes into picture. The boundary between the crust and mantle wedge on the right side of the model begins to be delineated and the subducted oceanic crust can be seen clearly. In general, the amplitudes are underestimated and the perturbation below the second reflector has yet to fill in. After 16 iterations the continental crust and wedge outlines begin to become well defined. A fast "hole" remains in the center of the crust due to the complexity of the image below. The connection between the tip of the mantle wedge and the subducted lithospheric mantle is a consistent feature in all high-frequency inversions. After 23 iterations the hole in the crust is ameliorated, and the boundaries of the model sections are sharpened. The spurious connection remains even if the inversion is near its final result. By this point, the amplitudes of the inferred perturbations are already quite accurate. Further iterations result in minor improvements, notably the fine structure of the upper subducted crust and its connection to the lower part of the slab, as seen in Figure 12(d). The final result has limited accuracy above the bottommost reflector where only one set of crossing sensitivities are available.

The sensitivity of the error function to model perturbations depends on the model used for propagation and thus changes with each iteration. Figure 14 shows the evolution of the sensitivity kernel for one correlation pair between the first and final iterations.

4. Discussion

4.1. **Effects of limited incidence angles.** Studies using teleseismic free surface multiples typically include data from earthquake sources between 30° and 90° epicentral distance from the array. Arrivals from these events have incident angles between 15° and 30°, which leaves a considerable gap in angular coverage compared to the above best-case scenario. Although the absence of the near vertical incidence events from the data leads to a decrease in model resolution, the advantage of a source-indexed approach is the ability to handle sparse coverage. The quality of the inversion will degrade near the base of the model and around the edges of the array due to inadequate illumination by multiple reflected events. However, the center of the model remains well resolved due to the existence of multiple crossing paths.

Figure 15 shows the results of the inversion with a realistic distribution of teleseismic sources, including one vertically incident phase, such as *PKIKP*. Though the number of image correlations done for the limited dataset is halved, the misfit for this case is not much worse than for the model derived with the full sweep of incident angles. The spurious connection between the mantle wedge and slab mantle is stronger, and the inversion fails to resolve the bottom of the slab at the right side of the target, but the result is not otherwise degraded.

4.2. **Effects of limited array aperture.** The capability to produce tomography updates of the model at certain depths derives from the ability to form images using a variety of incidence angles. For tomography with reflections in the upper mantle we require an array baseline long enough to ensure both the direct wave and the multiples will be recorded from these reflectors. To illustrate this point we perform an inversion with an array limited to only the central 130 km of the model space. The image created with this truncated array is displayed in Figure 16(a). The color of the image points represents the log of the total number of image correlations that were performed at each image point. The lack of angular illumination reduces the number of correlations that can be done, particularly at the deeper reflectors, beneath the mantle wedge and towards the edges of the image, as shown by the colored points at the centers of the windows. The resulting model reconstruction 16(b) remains accurate at shallower depths and where the distribution of correlations is similar to the previous inversions, but is degraded in the lower half and at the margins. The subducted crust, visible in the other inversions, is not recovered here.

4.3. **Effects of source spectrum.** Due to the nature of the teleseismic problem, data must be gathered from events with a variety of source-time functions and frequency content. Even with sophisticated techniques for deconvolving the source signature from the scattered wavefield, images from different events will be formed using different frequency peaks and noise content. This can be mediated, in part, by stacking images from common source regions prior to applying the image correlation. To simulate the difficulties involved with teleseismic data we create an alternate series of images with different frequency content.

Figure 17(a) shows an image created using data only up to 1Hz, a reasonable bandwidth for teleseismic data. A new set of error parameters were determined for use with this image. The correlation windows are widened and $\sigma_{B,z}$ and Δz both scale roughly linearly with frequency. The number of correlations performed is similar to the higher frequency case except in regions where the new, "thicker" images of reflectors are more likely to overlap, as in the tip

of the mantle wedge. The resulting model (b) has some advantages over the high frequency inversions. The connection between the upper and lower sections of the subducted crust is more continuous and the shape of the subducted mantle is better recovered. The hole-like anomalies in the center of the crust are filled in earlier in the iterative process. Areas with spottier correlation coverage are somewhat worse, as in the mantle wedge, and the fine structure of the subducted crust on the left side of the model is not recovered.

As a final test of the method, we generate a series of single source images with different frequency characteristics for every source. The amplitude spectra used in the test fall between the endmembers from the previous inversions, with peak frequencies ranging between 0.7 and 1.2 Hz, as shown in figure 18 (a). The resulting source-time functions vary by a factor of 2 in length. The bandwidth is simply indexed to the source number, with the lowest frequency event coming from -30° and the highest frequency event coming from 30°.

Inversions were performed for these images using the sets of error parameters both from the 2 Hz tests and 1 Hz tests, and the results are displayed in Figure 18(b) and (c), respectively. 18(b) shows the result from (c) iterated further with the 2 Hz error parameters. All model results share characteristics with the other inversions using the same error functions. While none perform quite as well as the uniform frequency cases, both manage to reconstruct the major elements of the test model. The inversion using high frequency error parameters recovers the sharply defined boundaries between the model elements, but suffers from a greater prevalence of holes within them. The inversion done with the low frequency parameters does well at recovering the smooth variations of the model, but it retains the anomaly between the wedge and subducted mantle common to inversions with high frequency data. While this points to a tradeoff in the selection of the error function parameters, the fact that both sets reconstructed the models with low misfit demonstrates the robustness of the method. Figure 18(d) shows the model from (c) updated for another 10 iterations with the high frequency set of parameters. In this paper we have focused on using uniform sets of error functions, but in practice these results could be further improved by selecting parameters based on individual source pairs.

4.4. **Preprocessing for application to regional-scale array data.** For the application of the method presented here to broadband 3-component array data, additional preprocessing is required. Before we can separate the incident and multiple wavefields, we must separate the data by polarization. This can be done by rotating the 3-component data into *P*, *SV*, and *SH* directions through the application of a free-surface transfer operator (Rondenay, 2009). Such analysis would allow for the application of the method to either *P* to *P* or *SV* to *SV* free surface multiples while still using the scalar wave equation. Additionally, the source-time function of teleseismic events must be estimated from the data and carefully deconvolved from the incident and multiple wavefields. The source-time function can most accurately be represented by the (time-windowed) first principal component. Deconvolution can then be performed in using a number of approaches, including the frequency domain method with optimized water levels (Chen et al., 2010).

4.5. **Comparison with Full Waveform Inversion.** Although they have characteristics in common, the reflection tomography method presented here is fundamentally different from full waveform inversion (FWI). FWI seeks to find a model of elastic properties that best explains the phase and amplitude of seismic observations (Pratt, 1999; Virieux & Operto, 2009). Indeed, the inverse scattering computation that comprises the first step of our inversion method is identical to the initial gradient in a waveform inversion (Wang et al., 2010), but the two approaches diverge thereafter. While an FWI approach would use the results of inverse scattering to model data to compare with observations, reflection tomography uses an image-domain scheme to measure the fitness of the background velocity. The gradient in FWI is thus sensitive to backscattering generated by sharp contrasts in elastic properties while the reflection tomography gradient is sensitive forward scattering between the discontinuities and the surface that are due to smooth variations in the background model.

Each method has advantages and disadvantages with regards to teleseismic data. As currently implemented, reflection tomography can only deal with singly backscattered free surface multiples, while a full waveform approach (Pageot et al., 2013) implicitly incorporates forward scattering along the initial teleseismic leg, in addition to higher order multiples. One of the most persistent difficulties encountered in FWI is cycle skipping, that is, the initial estimate of the velocity structure must model data within a quarter wavelength of observations or the optimization problem will not be convex. Reflection tomography in principle avoids this issue by considering image correlation over an adequate range of depth shifts, which allows for convergence starting from relatively less suitable starting models. For this reason, reflection tomography and FWI are often used in tandem in exploration seismics, with the reflection tomography inversion recovering a model adequate to avoid cycle skipping. In the application to the teleseismic problem, the cycle skipping issue may be ameliorated due the long wavelength content of the data compared to lithospheric scales, but the problem likely persists on scales that include the mantle transition zone.

Due to the difficult properties of teleseismic, reflection tomography may be more apt as a starting point for model building than FWI. The long baselines of teleseismic arrays can lead to variable noise and amplitude properties from receiver to receiver, making the amplitude estimation of the discontinuities (and thus the modeled data) less reliable. For this reason, the reflection tomography seeks only to match the phase information of the images. Additionally, both methods rely on estimation of the source time function. Since some coupling inevitably remains between the source time function and near-receiver structure, the power norm based tomographic error function seeks to reduce its impact. A simple *L2* minimization of data residuals for FWI would be strongly influenced by this coupling, although Pageot et al. (2013) suggests alternative metrics that would be more robust.

4.6. **Treatment of *P* to *S* conversions.** We have considered only *P* to *P* free surface multiples in order to link our work with acoustic studies performed in exploration seismology. In fact, imaging with *P* to *S* (and *S* to *P*) converted phases play a much larger role in teleseismic studies. Furthermore, mixed-phase free surface multiples like Pp_ms and Ps_ms can provide strong additions to the image. The extension of our methodology to the elastic case is therefore a first-order priority for future work.

This can be accomplished in one of two ways. The more straightforward route is to separate the *P* and *S* wavefields and propagate both using a scalar wave equation and the *P* and *S* velocity models can be updated independently of each other or with constraints. *P* to *S* multiple scattering can then be treated in much the same way as *P* multiples, while upgoing parts of both of the incident *P* and *Ps* arrivals can be backpropagated to produce an image from the converted phase.

Another way would be to backpropagate both polarizations at once using an elastic solver Wang et al. (2011). *Ps* converted phases can be treated using the methodology of Shang et al. (2012), wherein the (source-time function deconvolved) three component data is propagated in reverse time. The resulting wavefield is then subjected to a polarization decomposition, and an image is obtained by applying the elastic inverse scattering transform (Brytik et al., 2012) to the separated *P* and *S* wavefields.

4.7. **Extension to 3-D.** In 2-D, the applicability of this method is limited to arrays with earthquakes at a variety of epicentral distances aligned with the strike of the array. Events at too great an angle to the strike will necessitate an onerous 2.5D assumption. An relatively high density linear array like the LA RISTRA experiment in New Mexico (Gao et al., 2004), which has major earthquakes generated at subduction zones at the teleseismic distances along its great circle arc, is ideal for our application, but for many linear arrays sufficient coverage will be unavailable.

Fortunately, this limitation will be ameliorated by the extension to 3-D. Tomography using areal array data will be able to consider events from all azimuths, and no matter how sparse that set is, it is more likely to generate paths crossing at a variety of different directions beneath the array. Figure 19 illustrates the geometric requirements of the 3-D problem. The tomographic update, as is true for body wave tomography, does not rely on a full set of incoming incidences and backazimuths, but rather a sufficiently variable set of crossing paths.

The extension to 3-D will be aided by the advantages in illumination the teleseismic geometry has over exploration geometry. The locally planar geometry of the "source" we use ensures each source more or less illuminates the subsurface evenly across the entire array. This is only limited by the incidence angle the lower the incidence angle, the more the illumination tapers off at high offsets at depth. The result for 3-D applications is that there will be a hemisphere-like region where the subsurface is illuminated by multiple events sufficiently for tomography, similar to the 2-D case shown in Figure 16(b).

5. CONCLUSIONS

We have developed a method for RTM-based wave equation reflection tomography using teleseismic free surface multiples that is able to handle several issues with teleseismic data. In response to the sparsity of source locations due to the limits of global seismicity we have based our error function not on waveform fits but on the similarity of images formed from single sources or sets of sources in the same location. In order to deal with irregular source-time functions and frequency content, as well as limitations of preprocessing in removing these effects, we compare different single source images by means of a weighted correlation of image windows which penalizes correlation power away from zero depth shift.

The robustness of the error measure in the face of large deviations from the correct model gives this method an advantage over full waveform inversion in regions where an accurate starting model cannot be defined. Synthetic tests show that we can begin with a 1-D model and still recover the correct heterogeneity in the parts of the model where multiple events are reflected. The smooth model resulting from this method can be used in tandem with the model of singular velocity variations given by inverse scattering to achieve a complete characterization of the heterogeneity.

REFERENCES

Audet, P., Bostock, M. G., & Mercier, J.-P., 2007. Teleseismic waveform modelling with a one-way wave equation, *Geophysical Journal International*, **171**(3), 1212–1225.

Bostock, M. G., Rondenay, S., & Shragge, J., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves 1. Theory for oblique incidence, *Journal of Geophysical Research-Solid Earth*, **106**(B12), 30771–30782.

Brytik, V., de Hoop, M. V., & van der Hilst, R. D., 2012. Elastic-wave inverse scattering based on reverse time migration with active and passive source reflection data, in *Inverse Problems and Applications: Inside Out II*, vol. 60, pp. 411–453, Cambridge Univ. Press, Cambridge, U. K.

Chen, C. W., Miller, D. E., Djikpesse, H. A., Haldorsen, J. B. U., & Rondenay, S., 2010. Array-conditioned deconvolution of multiple-component teleseismic recordings, *Geophysical Journal International*, **182**(2), 967–976.

Chen, L., Cheng, C., & Wei, Z., 2009. Seismic evidence for significant lateral variations in lithospheric thickness beneath the central and western North China Craton, *Earth and Planetary Science Letters*, **286**(1-2), 171–183.

Claerbout, J. F., 1985. *Imaging the Earth's Interior*, Blackwell Scientific Publications, Inc., Palo Alto, CA.

Dahlen, F. A., Hung, S. H., & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes - I. Theory, *Geophysical Journal International*, **141**(1), 157–174.

De Hoop, M. V., 1996. Generalization of the Bremmer coupling series, *Journal of Mathematical Physics*, **37**, 32463282.

De Hoop, M. V. & Van der Hilst, R. D., 2005. On sensitivity kernels for 'wave-equation' transmission tomography, *Geophysical Journal International*, **160**(2), 621–633.

De Hoop, M. V., van der Hilst, R. D., & Shen, P., 2006. Wave-equation reflection tomography: annihilators and sensitivity kernels, *Geophysical Journal International*, **167**(3), 1332–1352.

Deuss, A., 2009. Global observations of mantle discontinuities using SS and PP precursors, *Surveys in Geophysics*, **30**(4-5), 301–326.

Duchkov, A. A. & de Hoop, M. V., 2009. Velocity continuation in the downward continuation approach to seismic imaging, *Geophysical Journal International*, **176**(3), 909–924.

Gao, W., Grand, S. P., Baldridge, W. S., Wilson, D., West, M., Ni, J. F., & Aster, R., 2004. Upper mantle convection beneath the central Rio Grande rift imaged by P and S wave tomography, *Journal of Geophysical Research-Solid Earth*, **109**(B3).

Häuser, S. & Ma, J., 2013. Seismic data reconstruction via shearlet-regularized directional inpainting, *Preprint*.

Keller, J. B. & Givoli, D., 1989. Exact non-reflecting boundary-conditions, *Journal of Computational Physics*, **82**(1), 172–192.

Langston, C. A., 1979. Structure under Mout Rainier, Washington, inferred from teleseismic body waves, *Journal of Geophysics Research*, **84**, 4749–4762.

Long, M. D., de Hoop, M. V., & van der Hilst, R. D., 2008. Wave-equation shear wave splitting tomography, *Geophysical Journal International*, **172**(1), 311–330.

Op 't Root, T. J. P. M., Stolk, C. C., & de Hoop, M. V., 2012. Linearized inverse scattering based on seismic reverse time migration, *Journal De Mathematiques Pures Et Appliquees*, **98**(2), 211–238.

Pageot, D., Operto, S., Vallee, M., Brossier, R., & Virieux, J., 2013. A parametric analysis of two-dimensional elastic full waveform inversion of teleseismic data for lithospheric imaging, *Geophysical Journal International*, **193**(3), 1479–1505.

Pearce, F. D., Rondenay, S., Sachpazi, M., Charalampakis, M., & Royden, L. H., 2012. Seismic investigation of the transition from continental to oceanic subduction along the western Hellenic Subduction Zone, *Journal of Geophysical Research-Solid Earth*, **117**.

Plessix, R. E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophysical Journal International*, **167**(2), 495–503.

Pratt, R. G., 1999. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model, *Geophysics*, **64**(3), 888–901.

Rondenay, S., 2009. Upper mantle imaging with array recordings of converted and scattered teleseismic waves, *Surveys in Geophysics*, **30**(4-5), 377–405.

Sava, P. & Vasconcelos, I., 2011. Extended imaging conditions for wave-equation migration, *Geophysical Prospecting*, **59**(1), 35–55.

Schuster, G. T., 2010. *Seismic interferometry*, Cambridge University Press.

Shang, X., de Hoop, M. V., & van der Hilst, R. D., 2012. Beyond receiver functions: Passive source reverse time migration and inverse scattering of converted waves, *Geophysical Research Letters*, **39**.

Shen, P. & Symes, W. W., 2008. Automatic velocity analysis via shot profile migration, *Geophysics*, **73**(5), VE49–VE59.

Shragge, J., Artman, B., & Wilson, C., 2006. Teleseismic shot-profile migration, *Geophysics*, **71**(4), SI221–SI229.

Stolk, C. C., 2000. Microlocal analysis of a seismic linearized inverse problem, *Wave Motion*, **32**(3), 267–290.

Stolk, C. C. & de Hoop, M. V., 2006. Seismic inverse scattering in the downward continuation approach, *Wave Motion*, **43**(7), 579–598.

Symes, W. & Carazzone, J., 1991. Velocity inversion by differential semblance optimization, *Geophysics*, **56**, 654–663.

Van Leeuwen, T. & Mulder, W. A., 2008. Velocity analysis based on data correlation, *Geophysical Prospecting*, **56**(6), 791–803.

Van Leeuwen, T. & Mulder, W. A., 2010. A correlation-based misfit criterion for wave-equation traveltime tomography, *Geophysical Journal International*, **182**(3), 1383–1394.

Van Stralen, M. J. N., de Hoop, M. V., & Block, H., 1998. Generalized Bremmer series with rational approximation for the scattering of waves in inhomogeneous media, *Journal of Acoustic Society of America*, **104**(4), 1943–1963.

Vinnik, L. P., Mikhailova, N. G., & Avetisian, R. A., 1979. Analysis of observations of the mantle PS waves, *Doklady Akademii Nauk Sssr*, **248**(3), 573–576.

Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.

Wang, S., de Hoop, M. V., & Xia, J., 2010. Acoustic inverse scattering via Helmholtz operator factorization and optimization, *Journal of Computational Physics*, **229**(22), 8445–8462.

Wang, S., de Hoop, M. V., & Xia, J., 2011. On 3D modeling of seismic wave propagation via a structured parallel multifrontal direct Helmholtz solver, *Geophysical Prospecting*, **59**(5), 857–873.

Xie, X.-B. & Yang, H., 2008. The finite-frequency sensitivity kernel for migration residual moveout and its applications in migration velocity analysis, *Geophysics*, **73**(6), S241–S249.

APPENDIX A. THE GRADIENT

To efficiently optimize the smooth velocity model, we develop an adjoint method for determining the gradient of the error function with respect to the model parameters. We first extend the computational domain Ω to $\mathbb{R}^n$ and describe the gradient in terms of solutions to the Helmholtz equation. To begin with, the fields $\widehat{G}_i(\mathbf{x}, z, \omega)$ and $\widehat{u}_{\mathrm{r};i}(\mathbf{x}, z, \omega)$ satisfy the inhomogeneous Helmholtz equations supplemented with the Sommerfeld radiation condition,

$$\begin{aligned} [\Delta + \omega^2 c(\mathbf{x}, z)^{-2}]\widehat{G}_i(\mathbf{x}, z, \omega) &= -\delta(\mathbf{x} - \mathbf{x}_{s;i}, z - z_{s;i}), \\ \lim_{r\to\infty} r^{(n-1)/2}\left(\frac{\partial}{\partial r} + \mathrm{i}\frac{\omega}{c_\infty}\right)\widehat{G}_i &= 0, \end{aligned} \tag{15}$$

and

$$\begin{aligned} [\Delta + \omega^2 c(\mathbf{x}, z)^{-2}]\widehat{u}_{\mathrm{r};i}(\mathbf{x}, z, \omega) &= -\delta(z)\widehat{N_\Sigma d_i}(\mathbf{x}, \omega), \\ \lim_{r\to\infty} r^{(n-1)/2}\left(\frac{\partial}{\partial r} + \mathrm{i}\frac{\omega}{c_\infty}\right)\widehat{u}_{\mathrm{r};i} &= 0, \end{aligned} \tag{16}$$

respectively. By introducing an exterior Dirichlet-to-Neumann map on $\partial\Omega$, we can restrict our analysis to Ω (Keller & Givoli, 1989).

We follow the techniques from the theory of optimization with equality constraints (Plessix, 2006), and introduce the Lagrangian, $\mathcal{L}[c]$, associated with our minimization problem. The quantities, λ, below are Lagrange multipliers.

Following the building up of the cross correlation power, we introduce

(17)
$$
\begin{aligned}
\mathcal{L} = &\sum_{\alpha=1}^{N_{\mathcal{I}}} \sum_{i\neq j,i,j=1}^{N_s} \frac{1}{S_{ij}^{\alpha}} \int_X \int_E W_B(\Delta z) \left|\mathcal{C}_{ij}^{\alpha}(\mathbf{x},\Delta z)\right|^2 \mathrm{d}\Delta z\, \mathrm{d}\mathbf{x} \\
&+ \sum_{\alpha=1}^{N_{\mathcal{I}}} \sum_{i\neq j,i,j=1}^{N_s} \int_X \int_E \lambda_{\mathcal{C};ij,\alpha}(\mathbf{x},\Delta z) \left\{ \mathcal{C}_{ij}^{\alpha}(\mathbf{x},\Delta z) - \int_{\mathbb{R}_{\geq 0}} (\varphi_\alpha \mathcal{I}_i)(\mathbf{x},z)\, (T_{\Delta z}(\varphi_\alpha \mathcal{I}_j))(\mathbf{x},z)\, \mathrm{d}z \right\} \mathrm{d}\Delta z\, \mathrm{d}\mathbf{x} \\
&+ \sum_{i=1}^{N_s} \int_\Omega \lambda_{\mathcal{I};i}(\mathbf{x},z) \left\{ \mathcal{I}_i(\mathbf{x},z) - \frac{1}{2\pi} \int_B \frac{\chi(\omega)}{\mathrm{i}\omega} \right. \\
&\qquad \left. \left(\frac{1}{\widehat{G}_i(\mathbf{x},z,\omega)}\, \widehat{u}_{\mathrm{r};i}(\mathbf{x},z,\omega) - \frac{c(\mathbf{x},z)^2}{\omega^2} \nabla \frac{1}{\widehat{G}_i(\mathbf{x},z,\omega)} \cdot \nabla \widehat{u}_{\mathrm{r};i}(\mathbf{x},z,\omega) \right) \mathrm{d}\omega\ \Psi_{\Omega_0}(\mathbf{x},z) \right\} \mathrm{d}\mathbf{x}\mathrm{d}z \\
&+ \sum_{i=1}^{N_s} \int_B \int_\Omega \overline{\lambda_{G;i}(\mathbf{x},z,\omega)} \left\{ [\Delta + \omega^2 c(\mathbf{x},z)^{-2}] \widehat{G}_i(\mathbf{x},z,\omega) + \delta(\mathbf{x}-\mathbf{x}_{s;i}, z-z_{s;i}) \right\} \mathrm{d}\mathbf{x}\mathrm{d}z\, \mathrm{d}\omega \\
&+ \sum_{i=1}^{N_s} \int_B \int_\Omega \overline{\lambda_{\mathrm{r};i}(\mathbf{x},z,\omega)} \left\{ [\Delta + \omega^2 c(\mathbf{x},z)^{-2}] \widehat{u}_{\mathrm{r};i}(\mathbf{x},z,\omega) + \delta(z) \widehat{N_\Sigma d_i}(\mathbf{x},\omega) \right\} \mathrm{d}\mathbf{x}\mathrm{d}z\, \mathrm{d}\omega \\
&\qquad + \int_\Omega \lambda_c(\mathbf{x},z) \left\{ c(\mathbf{x},z) - \sum_{k=1}^{N_c} \gamma_k \psi_k(\mathbf{x},z) \right\} \mathrm{d}\mathbf{x}\mathrm{d}z.
\end{aligned}
$$

To the inverse scattering term in the Lagrangian we have added a cutoff term $\Psi_{\Omega_0}(\mathbf{x},z)$ with support contained in $\Omega_0 \subset \Omega$. We now derive the adjoint state equations. The derivatives follow to be:

(18)
$$
\frac{\partial \mathcal{L}}{\partial \mathcal{C}_{ij}^{\alpha}} = 2\frac{1}{S_{ij}^{\alpha}} \left[W_B(\Delta z) - \sum_{\alpha=1}^{N_{\mathcal{I}}} \sum_{i\neq j,i,j=1}^{N_s} \frac{1}{S_{ij}^{\alpha}} \int_X \int_E W_B(\Delta z) \left|\mathcal{C}_{ij}^{\alpha}(\mathbf{x},\Delta z)\right|^2 \mathrm{d}\Delta z\, \mathrm{d}\mathbf{x} \right] \mathcal{C}_{ij}^{\alpha}(\mathbf{x},\Delta z) + \lambda_{\mathcal{C};ij,\alpha}(\mathbf{x},\Delta z),
$$

for the state variable $\mathcal{C}_{ij}^{\alpha}$, and

(19)
$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathcal{I}_i} = -\sum_{\alpha=1}^{N_{\mathcal{I}}} \sum_{j=1}^{N_s} \int_E \left\{ \lambda_{\mathcal{C};ij,\alpha}(\mathbf{x},\Delta z) \varphi_\alpha(\mathbf{x},z)\, (T_{\Delta z}(\varphi_\alpha \mathcal{I}_j))(\mathbf{x},z) \right. \\
\left. + \lambda_{\mathcal{C};ji,\alpha}(\mathbf{x},\Delta z) (T^*_{\Delta z}(\varphi_\alpha \mathcal{I}_j))(\mathbf{x},z)\, \varphi_\alpha(\mathbf{x},z) \right\} \mathrm{d}\Delta z + \lambda_{\mathcal{I};i}(\mathbf{x},z),
\end{aligned}
$$

for the state variable $\mathcal{I}_i$. Here, $T^*_{\Delta z}$ is obtained through a change of variable of integration. We use that the supports of φ_α do not contain a neighborhood of the acquisition surface. For obtaining the adjoint state equation for the state variable $\widehat{G}_i$, we first rewrite the terms containing the gradient of $\widehat{G}_i$:

(20)
$$
\begin{aligned}
&\int_\Omega \lambda_{\mathcal{I};i}(\mathbf{x},z)\ \Psi_{\Omega_0}(\mathbf{x},z) \frac{1}{2\pi} \frac{\chi(\omega)}{\mathrm{i}\omega}\ \frac{c(\mathbf{x},z)^2}{\omega^2} \nabla \frac{1}{\widehat{G}_i(\mathbf{x},z,\omega)} \cdot \nabla \widehat{u}_{\mathrm{r};i}(\mathbf{x},z,\omega)\, \mathrm{d}\mathbf{x}\mathrm{d}z \\
&\qquad = \int_{\partial\Omega} \lambda_{\mathcal{I};i}(\mathbf{x},z)\ \Psi_{\Omega_0}(\mathbf{x},z) \frac{1}{2\pi} \frac{\chi(\omega)}{\mathrm{i}\omega}\ \frac{c(\mathbf{x},z)^2}{\omega^2} \frac{1}{\widehat{G}_i(\mathbf{x},z,\omega)}\ \frac{\partial \widehat{u}_{\mathrm{r};i}}{\partial \nu}(\mathbf{x},z,\omega)\, \mathrm{d}A(\mathbf{x},z) \\
&\qquad\quad - \int_\Omega \frac{1}{\widehat{G}_i(\mathbf{x},z,\omega)}\ \nabla \cdot \left[\lambda_{\mathcal{I};i}(\mathbf{x},z)\ \Psi_{\Omega_0}(\mathbf{x},z) \frac{1}{2\pi} \frac{\chi(\omega)}{\mathrm{i}\omega}\ \frac{c(\mathbf{x},z)^2}{\omega^2} \nabla \widehat{u}_{\mathrm{r};i}(\mathbf{x},z,\omega) \right] \mathrm{d}\mathbf{x}\mathrm{d}z,
\end{aligned}
$$

and

$$\int_\Omega \overline{\lambda_{G;i}(\mathbf{x},z,\omega)}\,[\Delta+\omega^2 c(\mathbf{x},z)^{-2}]\widehat{G}_i(\mathbf{x},z,\omega)\,\mathrm{d}\mathbf{x}\mathrm{d}z = \int_{\partial\Omega}\left[\,\overline{\lambda_{G;i}(\mathbf{x},z,\omega)}\,\frac{\partial\widehat{G}_i}{\partial\nu}(\mathbf{x},z,\omega)-\overline{\frac{\partial\lambda_{G;i}}{\partial\nu}(\mathbf{x},z,\omega)}\,\widehat{G}_i(\mathbf{x},z,\omega)\,\right]\mathrm{d}A(\mathbf{x},z) + \int_\Omega \overline{([\Delta+\omega^2c(\mathbf{x},z)^{-2}]\lambda_{G;i})(\mathbf{x},z,\omega)}\,\widehat{G}_i(\mathbf{x},z,\omega)\,\mathrm{d}\mathbf{x}\mathrm{d}z. \tag{21}$$

Using the exterior Dirichlet-to-Neumann map, we find that the boundary integral on the right-hand side of this equation does not play a role. In view of the cut-off function, Ψ_{Ω_0}, the boundary integral in (20) vanishes. Hence,

$$\frac{\partial\mathcal{L}}{\partial\widehat{G}_i}=\lambda_{\mathcal{I};i}(\mathbf{x},z)\frac{1}{2\pi}\frac{\chi(\omega)}{\mathrm{i}\omega\widehat{G}_i(\mathbf{x},z,\omega)^2}\,\Big(\,\Psi_{\Omega_0}(\mathbf{x},z)\,\widehat{u}_{\mathrm{r};i}(\mathbf{x},z,\omega) + \nabla\cdot\left[\,\lambda_{\mathcal{I};i}(\mathbf{x},z)\;\Psi_{\Omega_0}(\mathbf{x},z)\,\frac{c(\mathbf{x},z)^2}{\omega^2}\,\nabla\widehat{u}_{\mathrm{r};i}(\mathbf{x},z,\omega)\,\right]\Big)+[\Delta+\omega^2c(\mathbf{x},z)^{-2}]\overline{\lambda_{G;i}(\mathbf{x},z,\omega)}. \tag{22}$$

In a similar fashion, we first rewrite the terms containing the gradient of $\widehat{u}_{\mathrm{r};i}$,

$$\begin{aligned}\int_\Omega \lambda_{\mathcal{I};i}(\mathbf{x},z)\;\Psi_{\Omega_0}(\mathbf{x},z)\,\frac{1}{2\pi}\frac{\chi(\omega)}{\mathrm{i}\omega}\,\frac{c(\mathbf{x},z)^2}{\omega^2}\,\nabla\frac{1}{\widehat{G}_i(\mathbf{x},z,\omega)}\cdot\nabla\widehat{u}_{\mathrm{r};i}(\mathbf{x},z,\omega)\,\mathrm{d}\mathbf{x}\mathrm{d}z&\\ =-\int_{\partial\Omega}\lambda_{\mathcal{I};i}(\mathbf{x},z)\;\Psi_{\Omega_0}(\mathbf{x},z)\,\frac{1}{2\pi}\frac{\chi(\omega)}{\mathrm{i}\omega\widehat{G}_i(\mathbf{x},z,\omega)^2}\,\frac{c(\mathbf{x},z)^2}{\omega^2}\,\frac{\partial\widehat{G}_i}{\partial\nu}(\mathbf{x},z,\omega)\,\widehat{u}_{\mathrm{r};i}(\mathbf{x},z,\omega)\,\mathrm{d}A(\mathbf{x},z)&\\ -\int_\Omega\nabla\cdot\left[\,\lambda_{\mathcal{I};i}(\mathbf{x},z)\;\Psi_{\Omega_0}(\mathbf{x},z)\,\frac{1}{2\pi}\frac{\chi(\omega)}{\mathrm{i}\omega}\,\frac{c(\mathbf{x},z)^2}{\omega^2}\,\nabla\frac{1}{\widehat{G}_i(\mathbf{x},z,\omega)}\,\right]\,\widehat{u}_{\mathrm{r};i}(\mathbf{x},z,\omega)\,\mathrm{d}\mathbf{x}\mathrm{d}z,&\end{aligned} \tag{23}$$

and then obtain

$$\frac{\partial\mathcal{L}}{\partial\widehat{u}_{\mathrm{r};i}}=-\lambda_{\mathcal{I};i}(\mathbf{x},z)\frac{1}{2\pi}\frac{\chi(\omega)}{\mathrm{i}\omega}\,\Big(\,\Psi_{\Omega_0}(\mathbf{x},z)\,\frac{1}{\widehat{G}_i(\mathbf{x},z,\omega)} + \nabla\cdot\left[\,\lambda_{\mathcal{I};i}(\mathbf{x},z)\;\Psi_{\Omega_0}(\mathbf{x},z)\,\frac{c(\mathbf{x},z)^2}{\omega^2}\,\nabla\frac{1}{\widehat{G}_i(\mathbf{x},z,\omega)}\,\right]\Big)+[\Delta+\omega^2c(\mathbf{x},z)^{-2}]\overline{\lambda_{\mathrm{r};i}(\mathbf{x},z,\omega)}. \tag{24}$$

The computation of the adjoint sources in equations (22) and (24) become somewhat complicated by the presence of the divergence terms. For practical purposes, we can instead follow the approximation introduced in Equation (6). In this approximation, we need not take the derivative of terms containing $\widehat{G}_i(\mathbf{x},z,\omega_0)$. This substitution modifies the first right hand side terms of (22) and (24). Furthermore, finite difference tests (Figure 10) show that the divergence terms are negligible, and we therefore omit the second right hand side terms from the calculations. The source side adjoint field in equation (22) then becomes:

$$\frac{\partial\mathcal{L}}{\partial\widehat{G}_i}=\lambda_{\mathcal{I};i}(\mathbf{x},z)\frac{1}{2\pi}\frac{\chi(\omega)}{\mathrm{i}\omega|\widehat{G}_i(\mathbf{x},z,\omega_0)|^2}\,\frac{|\omega_0|^{n-3}}{|\omega|^{n-3}}\,\Psi_{\Omega_0}(\mathbf{x},z)\,\widehat{u}_{\mathrm{r};i}(\mathbf{x},z,\omega)+[\Delta+\omega^2c(\mathbf{x},z)^{-2}]\overline{\lambda_{G;i}(\mathbf{x},z,\omega)} \tag{25}$$

and the receiver side adjoint field in (24) becomes:

$$\frac{\partial\mathcal{L}}{\partial\widehat{u}_{\mathrm{r};i}}=\lambda_{\mathcal{I};i}(\mathbf{x},z)\frac{1}{2\pi}\frac{\chi(\omega)}{\mathrm{i}\omega|\widehat{G}_i(\mathbf{x},z,\omega_0)|^2}\,\frac{|\omega_0|^{n-3}}{|\omega|^{n-3}}\,\Psi_{\Omega_0}(\mathbf{x},z)\,\overline{\widehat{G}_i(\mathbf{x},z,\omega)}+[\Delta+\omega^2c(\mathbf{x},z)^{-2}]\overline{\lambda_{\mathrm{r};i}(\mathbf{x},z,\omega)} \tag{26}$$

Finally, we have

$$\frac{\partial\mathcal{L}}{\partial c}=\sum_{i=1}^{N_s}\int_B\left(-\frac{2\omega^2}{c^3}\right)\left\{\,\overline{\lambda_{G,i}(\mathbf{x},z,\omega)}\,\widehat{G}_i(\mathbf{x},z,\omega)+\overline{\lambda_{\mathrm{r},i}(\mathbf{x},z,\omega)}\,\widehat{u}_{\mathrm{r};i}(\mathbf{x},z,\omega)\,\right\}\mathrm{d}\omega+\lambda_c(\mathbf{x},z), \tag{27}$$

and

$$\frac{\partial\mathcal{L}}{\partial\gamma_k}=\frac{\partial\mathcal{J}}{\partial\gamma_k}-\int_\Omega\lambda_c(\mathbf{x},z)\psi_k(\mathbf{x},z)\,\mathrm{d}\mathbf{x}\mathrm{d}z. \tag{28}$$

At a stationary point, all of these derivatives go to zero, and we can obtain the model update direction, $\frac{\partial\mathcal{J}}{\partial\gamma_k}$, by eliminating the multipliers through successive substitution. In other words, we solve (A.4) for $\lambda_{\mathcal{C};ij,\alpha}$ and substitute

it into (A.5), which is then solved for $\lambda_{\mathcal{I};i}$. We then need to solve two Helmholtz equations, (25) and (26), with $\overline{\lambda_{G;i}}$ corresponding to the adjoint "source field" and $\overline{\lambda_{r;i}}$ corresponding to the adjoint "receiver field." Equation (27) has the form of a traditional cross correlation imaging condition between these adjoint fields and their corresponding forward fields. Finally, equation (28) just entails a projection onto the relevant basis functions. Thus, from (18), (19), (25), (26), (27) and (28) we obtain a system of equations which we solve from top to bottom; in this process we encounter the fields which make up the images of $r(z, \mathbf{x})$ as well as the images themselves.

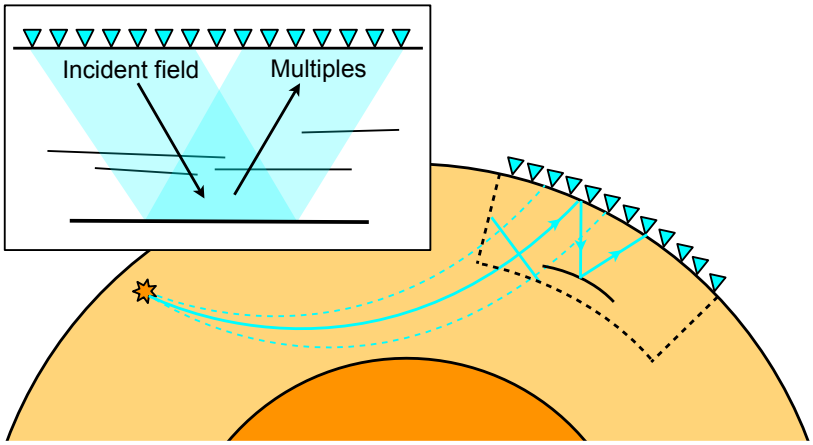
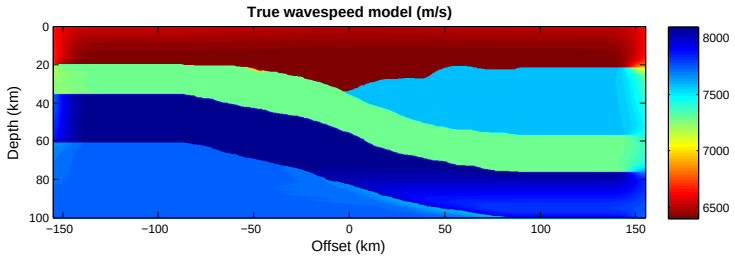
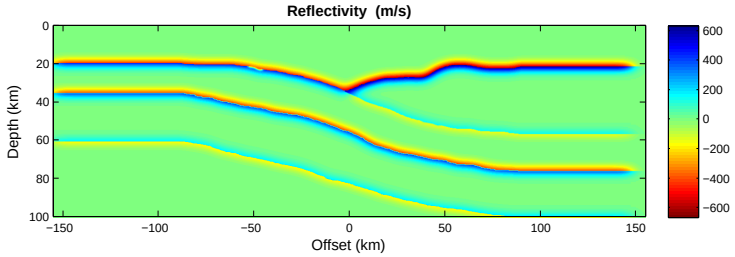


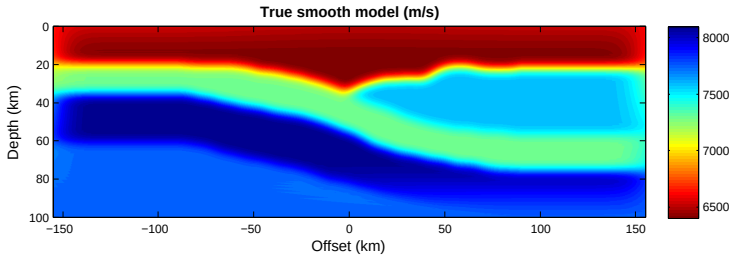
FIGURE 1. Geometry of the teleseismic problem. The direct P waves from distant earthquakes travel through the mantle and are recorded at a seismic array. The waves are reflected at the free surface, are scattered at discontinuities, and are recorded again by the same array. The teleseismic RTM approach uses the direct incident field as a source and the scattered field as data. For the synthetic teleseismic data presented in this paper, we omit the initial leg of travel and give the teleseismic source as a localized plane wave within the computational domain, represented here by the dashed black box.



(a)

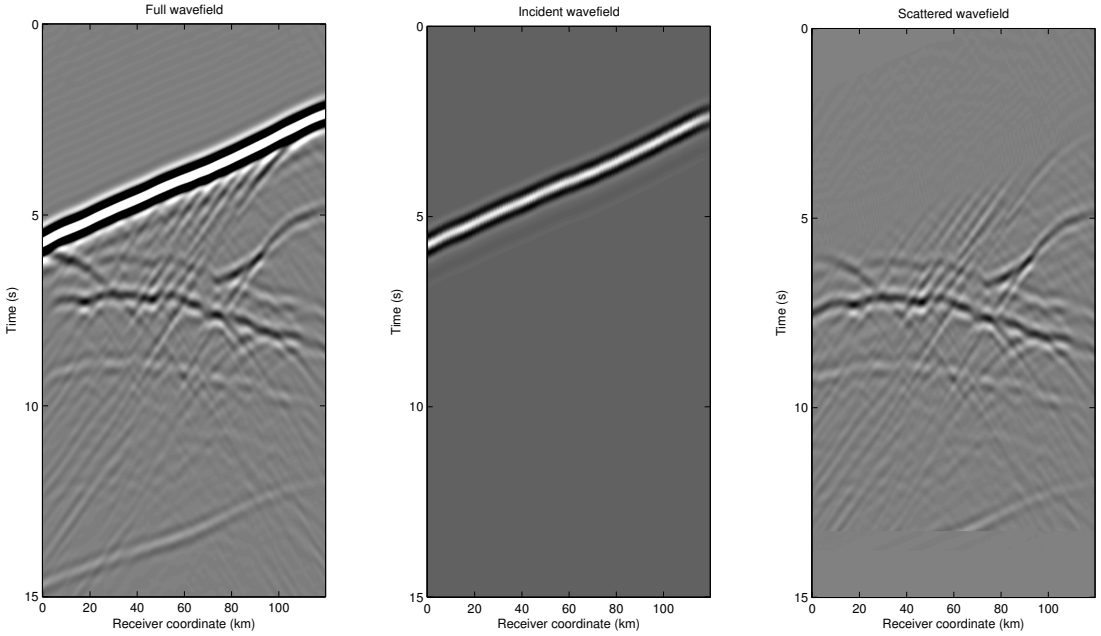


(b)



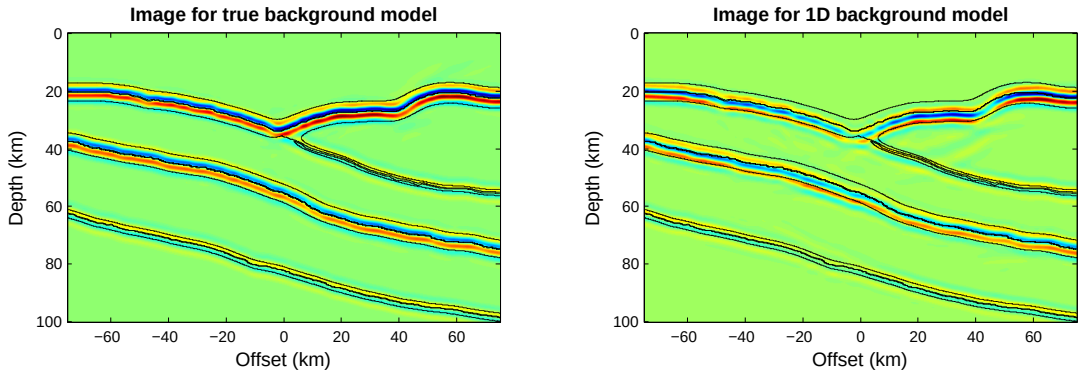
(c)

FIGURE 2. (a) Subduction zone model used to create synthetic teleseismic data. (b) Singular part of the subduction zone model. Values in percent perturbation from smooth model. (c) Smooth wavespeed perturbation to the smoothed 1D model.model used for wave propagation in imaging.



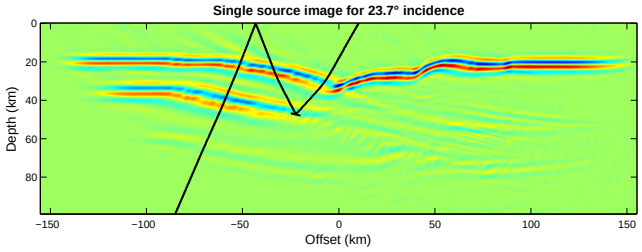
(a) (b) (c)

FIGURE 3. Synthetic data (a) generated by a Helmholtz solver in the velocity model displayed in 2(a) is split into the direct arrival, (b) and the free surface multiples, (c). The model is illuminated using a plane wave with horizontal slowness 5×10^{-5} s/m impinging on the bottom of the domain. The wave travels upwards, is reflected by the free surface boundary condition on at the top of the domain and is multiply scattered. The direct and multiple fields are separated by aligning the traces by cross correlation and eigenvalue analysis. Since the source-time function is known in this synthetic case, we do not deconvolve for the source-time function.

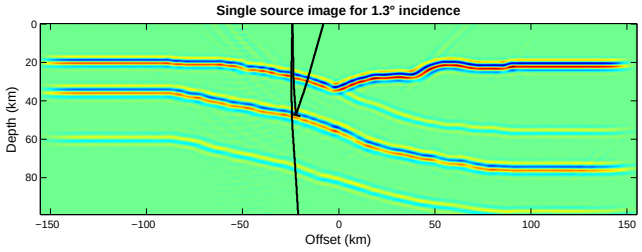


(a) (b)

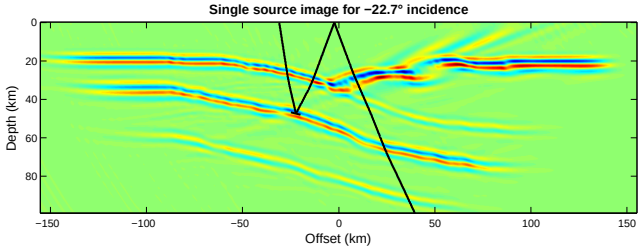
FIGURE 4. (a) The central 150 km of stacked image formed in the true model. Black outlines represent contours of reflectivity in the subduction model as shown in Figure 2(b). The image fits well within the contours in this case. (b) Stacked image for 1D model. The top reflectors are shifted downwards from their true location due to an overestimated velocity.



(a)

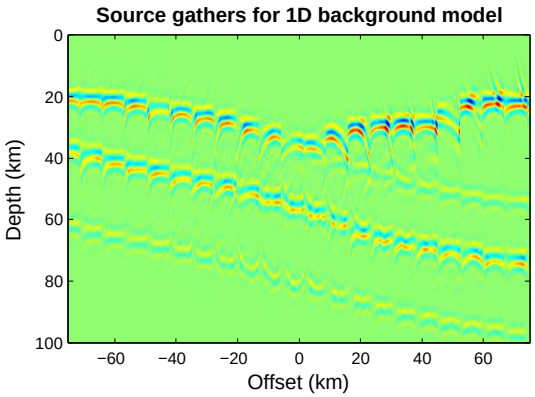


(b)

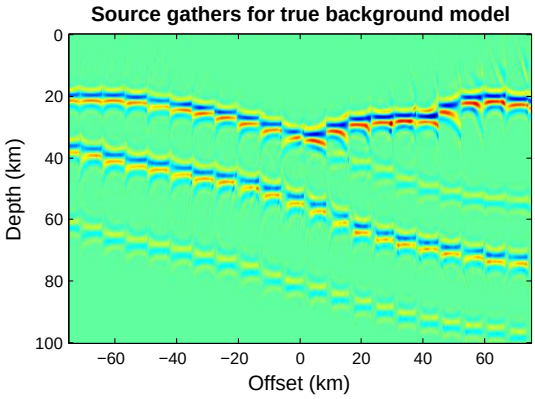


(c)

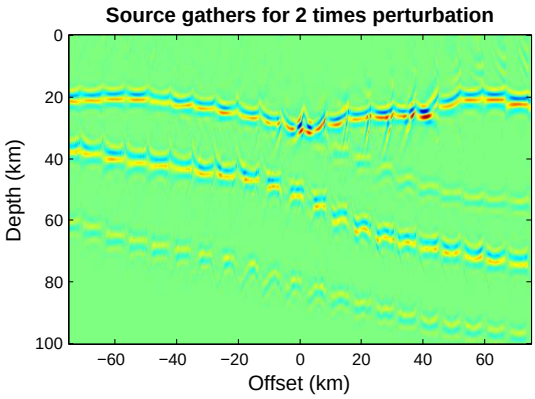
FIGURE 5. Single source images formed using inverse scattering in true model for teleseismic sources arriving from (a) 23° from the left, (b) near vertical incidence, and (c) 23° from the right. The black lines show the ray geometry for one reflection point.



(a)



(b)



(c)

FIGURE 6. Source gathers from the central 150 km section of the image. For a series of horizontal offsets spaced 7.5 km apart, the single-source images from all events are set next to each other to visualize difference in depth moveout in the images. (In the inversion process, measurements are made at all illuminated points along the discontinuities.) (a) Gathers formed using the 1D model. Due to the overestimated velocity in the crust, the images of the uppermost reflector from lower incidence events form deeper. (b) Source gathers for correct smooth model show no residual depth moveout between images. (c) Source gathers for model with double the smooth perturbation from 1D model. The underestimated crustal velocity leads to a convex gathers for the uppermost reflector.

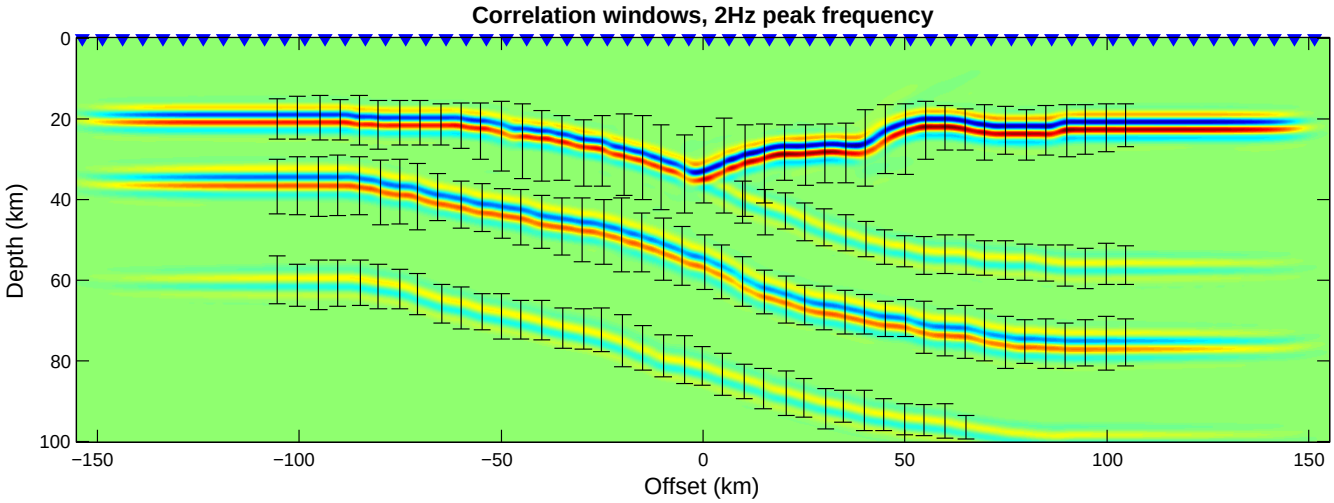


FIGURE 7. Map of correlation windows. Black bars correspond to representative window locations. Correlations are taken at every offset where coherent structure is imaged. Triangles indicate the extent of the receiver array.

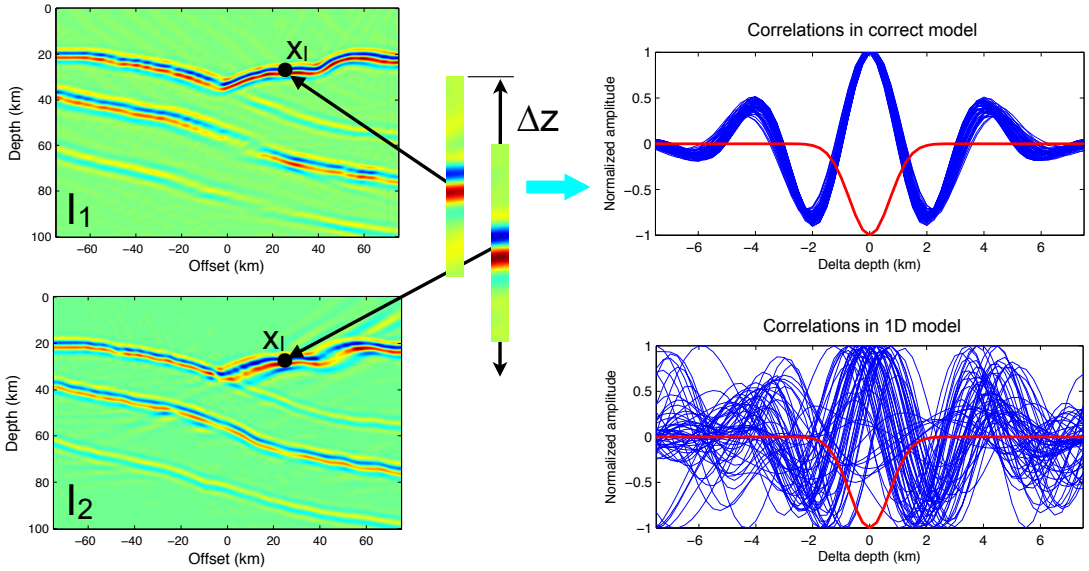


FIGURE 8. An explanation of the correlation power error metric. At each relevant image point, pairs images are correlated on a depth window over Δz. This correlation is multiplied by a weighting function $W_B(\Delta z)$ (shown as a red line) which is chosen based on the bandwidth of the data used to form the image. If the smooth velocity used to propagate the wavefields is correct, the two images should best align at zero depth shift and the sum of the weighted correlation should be at its minimum value. If the model is incorrect along the paths of the waves, the correlations will typically not be centered at zero shift, as shown in the bottom plot. If the image point is poorly illuminated and two images do not meet some threshold value, they can give false information about the model, and the correlation must be discarded

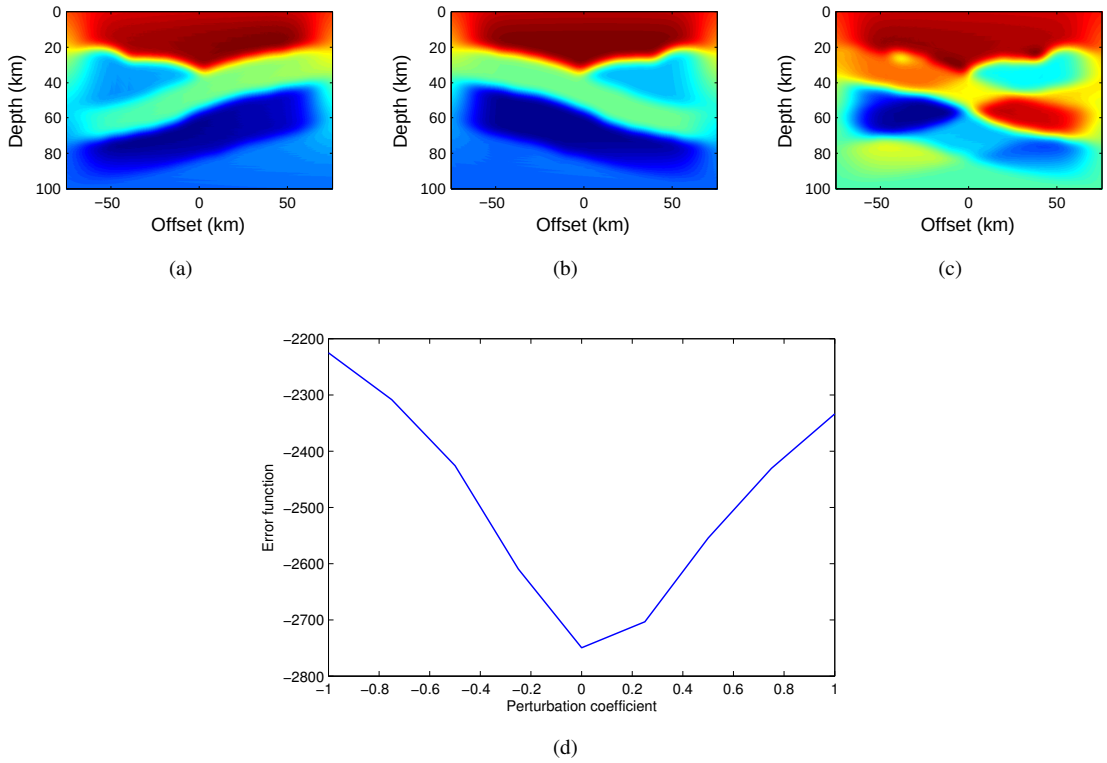
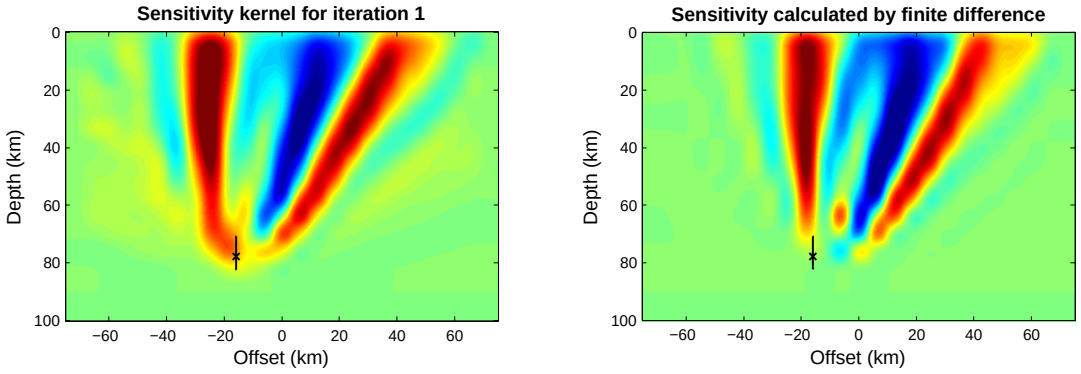
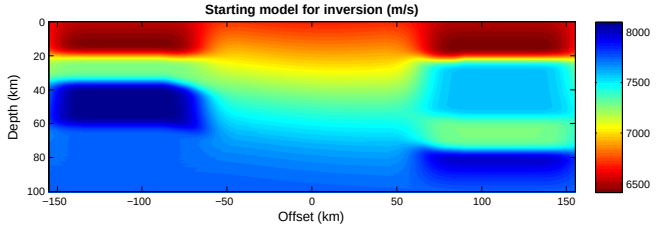


FIGURE 9. We test the error function by varying a perturbation to the true model. The perturbation is the difference between true model and the true model flipped from left to right. (a) True model minus perturbation. (b) True model. (c) True model plus perturbation. (d) Correlation power norm error function with weighting function displayed in figure 8. The error function is minimized by the correct subduction model and has a broad basin of attraction.

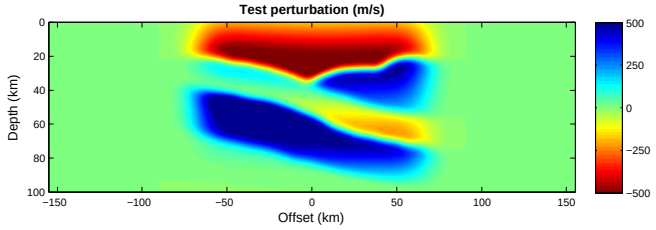


(a) (b)

FIGURE 10. Sensitivity kernels for one correlation window and a single pair of events with incidence angles of 0° and 23° from the right, as shown in Figure 5. Black **x**'s represent location of image point centered on bottommost reflector and line shows extent of correlation window. The red indicates areas where a positive perturbation to the velocity results in an increase in the error function while the blue indicates that the perturbation will reduce the error function. (a) Kernel computed with adjoint method and projected on to spline basis. (b) Kernel computed via finite differences for same correlation window. For each basis function, the smooth starting model in figure 11 is perturbed by the function with an amplitude of 50 m/s.



(a)



(b)

FIGURE 11. (a) The initial model for test inversions has 1D structure in the 150 km target region at the center of the model and a smooth version of the true subduction model elsewhere. Tomographic updates are performed in the target region. (b) The perturbation between the true smooth model and the initial model that will be recovered by the inversion.

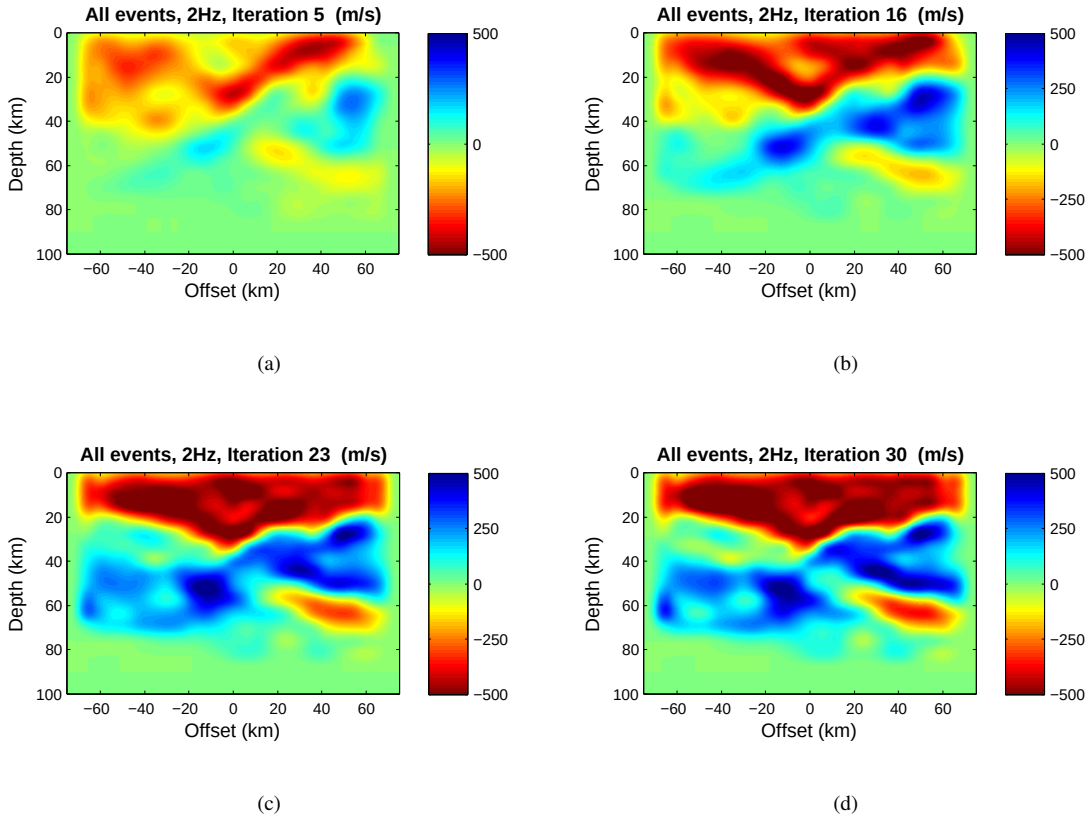


FIGURE 12. Inversions results for set of parameters shown in Figure 7 using full range of incidence angles and data bandwidth up to 2 Hz. All results are plotted only for the spatial extent of test perturbation for clarity.

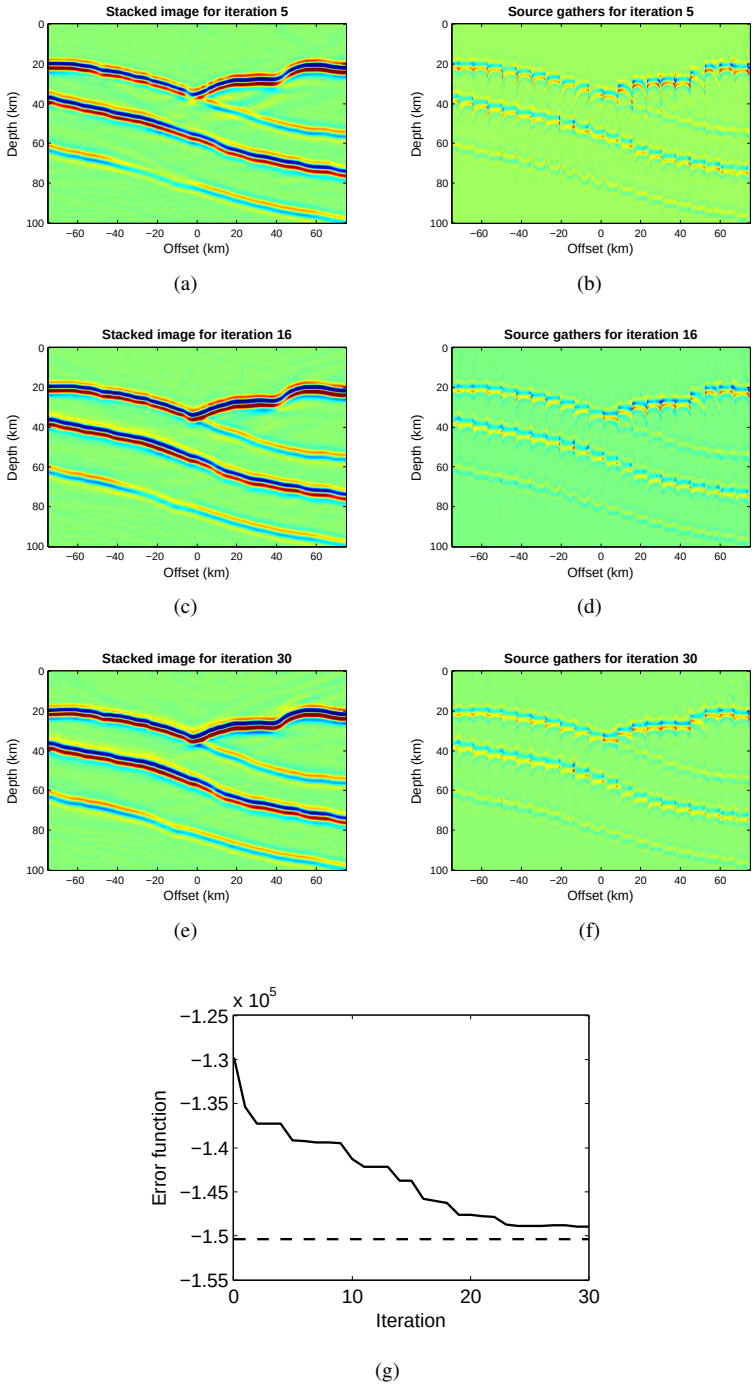


FIGURE 13. (a),(c),(e) Stacked images for model iterations corresponding to figure 12. (b),(d),(f) Image gathers for same model iterations. For gathers formed in the starting model, refer to figure 6(a). (g) Error function reduction for all iterations. Dashed line represents the error measure for the true model.

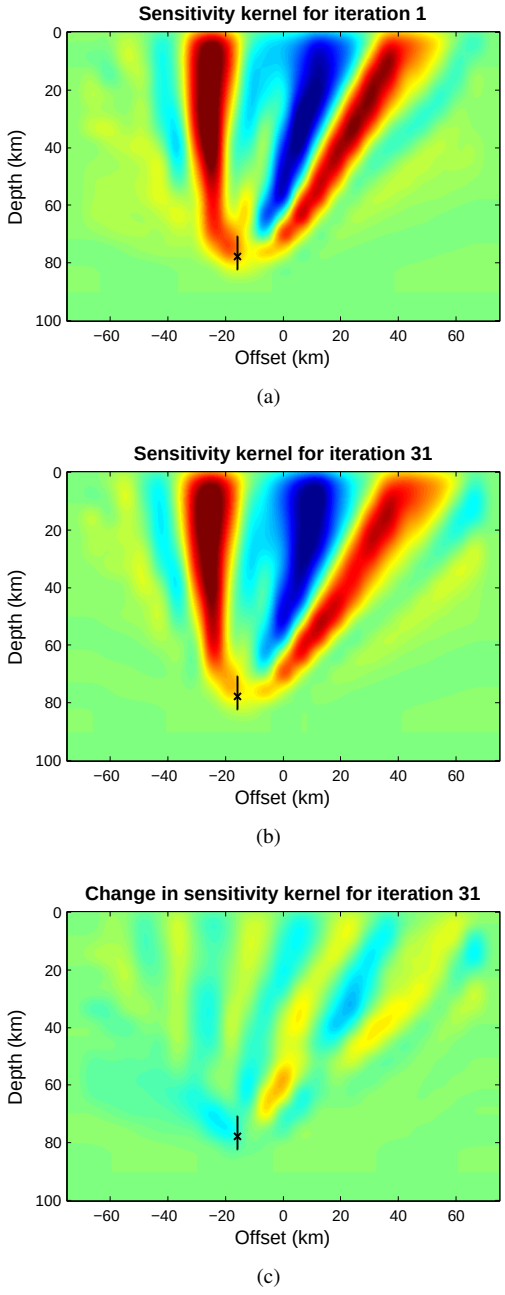


FIGURE 14. Evolution of the sensitivity kernel for correlation window from Figure 10 after (a) the first iteration and (b) the final iteration. (c) The difference between the first and final iterations.

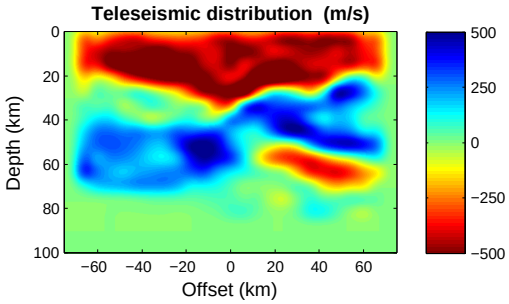
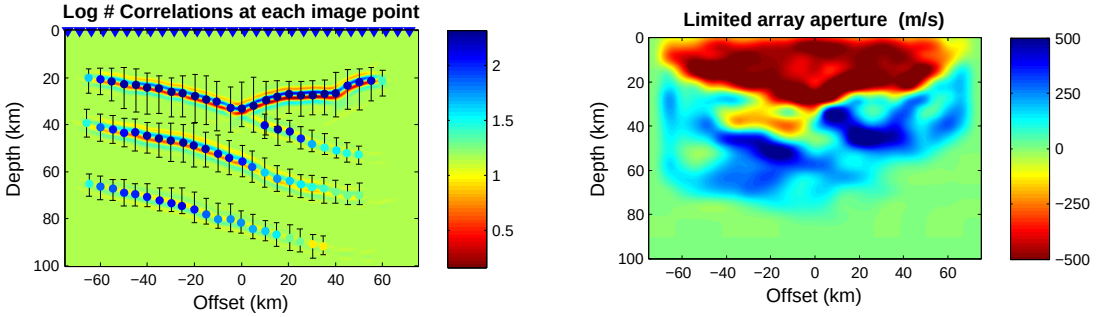
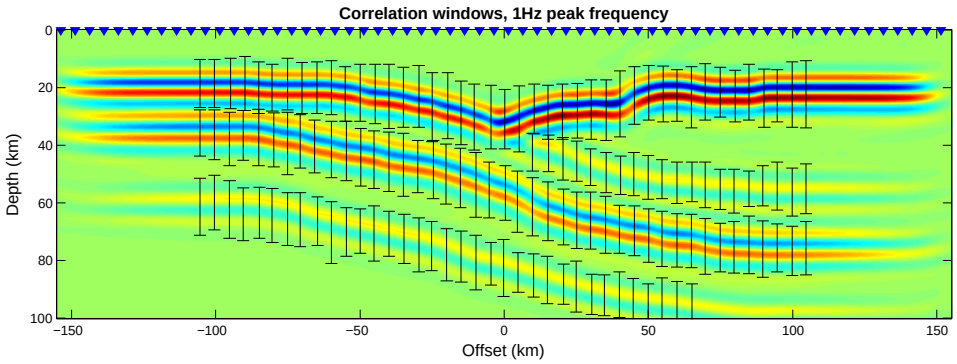


FIGURE 15. Inversion result for more realistic teleseismic distribution of incidence angles. Arrivals range from $15 - 30^\circ$ from normal, in addition to one vertically incident *PKIKP*.

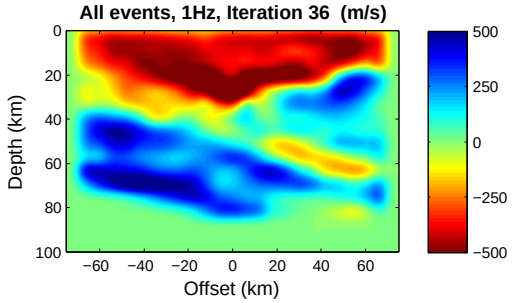


(a) (b)

FIGURE 16. Effects of a limited array aperture on the inversion. (a) Map of correlation windows projected over image formed with limited aperture array. Inverted triangles show the extent of the recordings. The number of error function measurements is severely decreased at the bottom reflector and towards the edges. (b) Inversion result.



(a)



(b)

FIGURE 17. (a) Map of correlation windows for lower frequency image. Black bars correspond to window extent. The limited bandwidth of the image reduces the number of acceptable windows relative to the 2 Hz image in areas where reflectors are close together. (b) Inversion result with data up to 1 Hz after 36 conjugate gradient iterations.

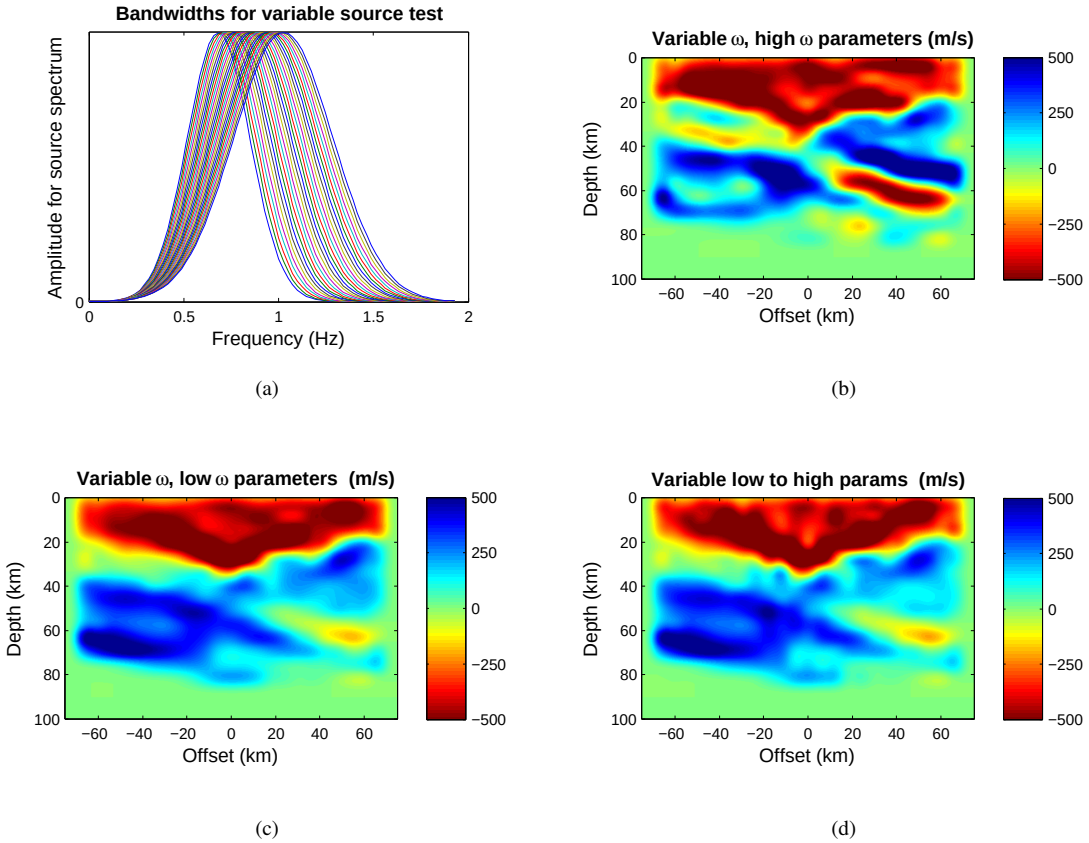


FIGURE 18. Inversion results with variable frequency information. (a) Source spectra for the 29 incident wavefields vary with peak frequencies between 0.7 and 1.2 Hz. Inversions were performed with the error function parameters displayed in (b) Figure 7 and (c) Figure 17. (d) Beginning from the result with low frequency parameters in (c), additional iterations were performed with high frequency parameters as in (b).

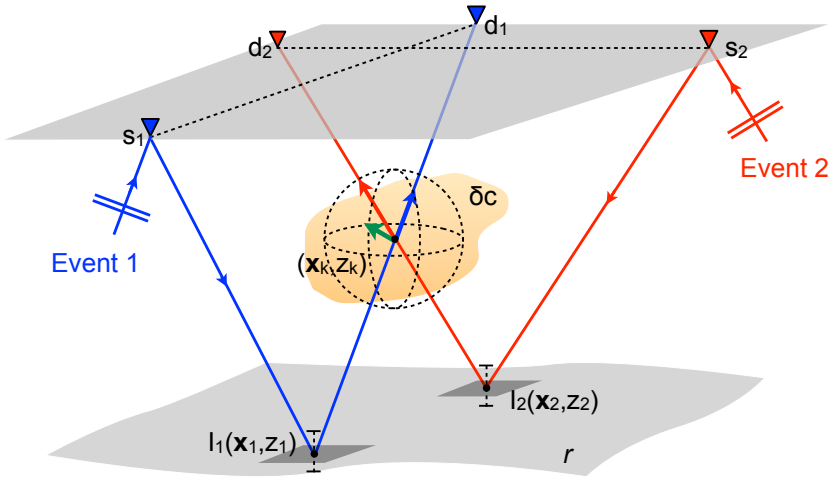


FIGURE 19. Geometry of 3-D reflection tomography problem. A teleseismic wave from event 1 (blue) is recorded at s_1, reflects off of the free surface, scatters off discontinuity *r* at $(\mathbf{x}_1, z_1)$ and is recorded at d_1. A wave from event 2 (red) arrives from a different backazimuth, reflects downward at s_2, is scattered at $(\mathbf{x}_2, z_2)$ and is recorded again at d_2. The two "rays," in addition to a third "ray" shown in green, intersect at a point $(\mathbf{x}_k, z_k)$ within a smooth velocity heterogeneity δc. In order to constrain the heterogeneity at $(\mathbf{x}_k, z_k)$, we require wavepaths through the point in three sufficiently different directions.