Joint inversion of surface wave dispersion and receiver functions: a Bayesian Monte-Carlo approach

Weisen Shen, Michael H. Ritzwoller, Vera Schulte-Pelkum and Fan-Chi Lin

1 Center for Imaging the Earth’s Interior, Department of Physics, University of Colorado at Boulder, Boulder, CO 80309, USA. E-mail: Weisen.Shen@colorado.edu
2 Cooperative Institute for Research in Environmental Sciences and Department of Geological Sciences, University of Colorado at Boulder, Boulder, CO 80309, USA
3 Seismological Laboratory, Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA

Accepted 2012 October 29. Received 2012 October 16; in original form 2012 May 01

SUMMARY
A non-linear Bayesian Monte-Carlo method is presented to estimate a Vsv model beneath stations by jointly interpreting Rayleigh wave dispersion and receiver functions and associated uncertainties. The method is designed for automated application to large arrays of broad-band seismometers. As a testbed for the method, 185 stations from the USArray Transportable Array are used in the Intermountain West, a region that is geologically diverse and structurally complex. Ambient noise and earthquake tomography are updated by applying eikonal and Helmholtz tomography, respectively, to construct Rayleigh wave dispersion maps from 8 to 80 s across the study region with attendant uncertainty estimates. A method referred to as ‘harmonic stripping method’ is described and applied as a basis for quality control and to generate backazimuth independent receiver functions for a horizontally layered, isotropic effective medium with uncertainty estimates for each station. A smooth parametrization between (as well as above and below) discontinuities at the base of the sediments and crust suffices to fit most features of both data types jointly across most of the study region. The effect of introducing receiver functions to surface wave dispersion data is quantified through improvements in the posterior marginal distribution of model variables. Assimilation of receiver functions quantitatively improves the accuracy of estimates of Moho depth, improves the determination of the Vsv contrast across Moho, and improves uppermost mantle structure because of the ability to relax a priori constraints. The method presented here is robust and can be applied systematically to construct a 3-D model of the crust and uppermost mantle across the large networks of seismometers that are developing globally, but also provides a framework for further refinements in the method.

Key words: Inverse theory; Surface waves and free oscillations; Seismic tomography; Crustal structure; North America.

1 INTRODUCTION
The construction of crustal and uppermost mantle velocity models over extended regions is critical to an understanding of continental tectonics and the thermal and compositional structure of the lithosphere as well as to provide the structural framework for an assessment of natural hazards. Surface waves provide spatially continuous information that is useful in developing 3-D shear wave velocity (V_s) models at regional (e.g. Levshin & Ritzwoller 1995; Ritzwoller et al. 1998; Levin et al. 2002; Ritzwoller et al. 2003; Levshin et al. 2005; Yang & Forsyth 2006; Shapiro et al. 2008; Yang et al. 2011), continental (e.g. Ritzwoller & Levshin 1998; Ritzwoller et al. 2001; Villasenor et al. 2001; Bensen et al. 2009) and global (e.g. Shapiro & Ritzwoller 2002; Ritzwoller et al. 2004) scales. Receiver functions (RFs), in contrast, provide the spatially discrete local response of seismic waves to discontinuities beneath receiver locations (Vinnik 1977; Langston 1979). RFs also have been applied to construct local 1-D (Ammon & Zandt 1993) and 3-D models (Vinnik et al. 2004, 2006). The use of each type of data alone presents significant non-uniqueness problems (e.g. Ammon et al. 1990; Shapiro & Ritzwoller 2002): surface wave data do not image discontinuities whereas RFs do not strongly constrain absolute shear velocities between discontinuities (although they do constrain the contrast in S-wave velocity across the discontinuity). As a consequence, combining the two complementary data types was a natural direction for efforts to determine structure near to Earth’s surface and was introduced more than a decade ago (e.g. Last et al. 1997; Ozalaybey et al. 1997), with numerous realizations of the idea.
having subsequently been developed. Linearized joint inversions have been particularly popular (e.g. Du & Foulger 1999; Julia et al. 2000, 2003; Endrun et al. 2004; Horspool et al. 2006; Tkalčič et al. 2006; Gok et al. 2007; Pasyanos et al. 2007; Yoo et al. 2007; Tokam et al. 2010; Salah et al. 2011; Bailey et al. 2012).

Non-linear inversions based on model-space sampling methods such as Markov Chain Monte Carlo perhaps augmented by Bayesian approaches to construct ensembles of models that are used to quantify the degree of belief about earth structures (Mosegaard & Tarantola 1995; Sambridge 2001; Mosegaard & Sambridge 2002; Sambridge & Mosegaard 2002) also have been applied in both RF inversions (e.g. Shibutani et al. 1996; Zhao et al. 1996; Levin & Park 1997; Sambridge 1999a; Clitheroe et al. 2000; Piana Agostinetti et al. 2002; Bannister et al. 2003; Frederiksen et al. 2003; Nicholson et al. 2005; Vinnik et al. 2004, 2006; Lucente et al. 2005; Hetenyi & Bus 2007; Piana Agostinetti & Chiarabba 2008; Piana Agostinetti & Malinverno 2010) and surface wave inversions (e.g. Shapiro & Ritzwoller 2002; Yoshizawa & Kennett 2002; Socco & Boiero 2008; Maraschin & Futi 2010; Molnar et al. 2010; Khan et al. 2011) with varying degrees of theoretical sophistication, generality, spatial extent and data quantity. Joint inversions of RFs and surface wave dispersion or other geophysical data within the context of a non-linear, model-space sampling scheme are more rare and recent (An & Assumpção 2004; Chang et al. 2008; Piana & Wiens 2004; Vinnik et al. 2006; Liu et al. 2010; Moorokamp et al. 2010; Tokam et al. 2010; Basuyau & Tiberi 2011; Bodin et al. 2012).

Despite these advances, the joint interpretation of surface wave dispersion and RFs is still faced with at least four significant challenges. First, as traditionally applied, surface waves and RFs have yielded information at different lateral length scales. Secondly, also traditionally, surface wave dispersion information has derived from teleseismic earthquakes and has been produced mostly at long periods (>20 s), which are not ideally sensitive to the Earth’s crust. Thirdly, joint quantitative interpretation requires meaningful uncertainties for both data types. Fourthly, the uncertainties in the resulting model must be quantified. Inextricably related to this is the need to find a model parametrization that possesses all and only the detail of structure necessary to fit both types of data. The purpose of this paper is to develop an approach to joint interpretation of surface wave dispersion and RFs that addresses these challenges in a practical, computationally efficient way that allows for automated application to large arrays of seismometers.

The first challenge has been ameliorated by the recent deployment of large seismic arrays such as the Earthscope/USArray Transportable Array (TA), CEArray in China (Zheng et al. 2010), the Virtual European Broadband Seismic Network and PASSCAL and Flexible Array experiments in which relatively close station spacing (50–100 km) allows much better resolution from surface waves. Recent developments in surface wave tomography have been stimulated to exploit these arrays. Examples of array-based surface wave tomography methods include the two-plane wave and related methods (Pollitz 2008; Yang et al. 2008a,b; Pollitz & Snoke 2010), eikonal tomography (Lin et al. 2009) and Helmholtz tomography (Lin & Ritzwoller 2011). These new array methods applied to data from the TA (e.g. Lin et al. 2011; Ritzwoller et al. 2011) and CEArray (Zhou et al. 2012) generate surface wave dispersion maps of unprecedented resolution (50–75 km) across large regions. In addition, tomographic methods based on ambient noise have been developed to augment information from teleseismic earthquakes. Ambient noise tomography (ANT) produces relatively short period (8–40 s) surface wave information that agrees with information from teleseismic earthquakes in the period band of overlap (25–40 s; e.g. Yang & Ritzwoller 2008; Ritzwoller et al. 2011). Dispersion information from ambient noise, therefore, constrains the crust in a way that is directly complementary to RF analyses. The second challenge facing the joint interpretation of surface wave and RF information, therefore, is addressed by the development of ANT. Indeed, inversions for 3-D models of the crust and uppermost mantle based on surface wave dispersion information from ANT (perhaps with earthquake tomography) have been increasingly common in the past few years (e.g. Yang et al. 2008a,b; Bensen et al. 2009; Moschetti et al. 2010a,b; Lin et al. 2011; Zheng et al. 2011; Zhou et al. 2012).

The third challenge, determining uncertainty estimates for both types of data, has been partially solved recently for surface waves by the eikonal and Helmholtz tomography methods, which produce reliable error estimates for both ambient noise and earthquake tomography. We discuss in the Appendix a method to construct RFs referred to as harmonic stripping based on azimuthal harmonic analysis (e.g. Girardin & Farra 1998; Bianchi et al. 2010) that we use both to construct azimuthally independent RFs and estimate their uncertainty. Bodin et al. (2012) insightfully and in considerable detail discuss the importance of reliable uncertainty information, with particular emphasis on the need to estimate full data covariance matrices in the joint inversion. Finally, the fourth challenge, estimating model uncertainties, is addressed by the Bayesian Monte Carlo method that we develop and describe here. We show along the way that a quite simple model parametrization based on smooth B-splines in the crust and mantle can fit both data types at most locations across the western United States.

Therefore, we present here a non-linear Bayesian Monte-Carlo algorithm to estimate a V , model beneath stations by jointly interpreting surface wave dispersion and RFs and associated uncertainties. The method is designed specifically to be applied automatically to large numbers of seismic stations. As a testbed for the method, we use 185 stations from the TA in the Intermountain West (Fig. 1). We first update the ambient noise and earthquake tomography to construct Rayleigh wave dispersion maps from 8 to 80 s across the study region. Eikonal and Helmholtz tomography are applied to

Figure 1. Stations of the Earthscope USArray Transportable Array (TA) used in this study are shown with blue triangles. The main geological provinces are outlined with red contours and titled with red abbreviations [the Basin and Range (BR) province, the Colorado Plateau (CP), the Rocky Mountains (RM) and the Great Plains (GP)]. Other regional geological features and basins are also identified with abbreviations [Snake River Plain (SNP), Green River Basin in Southern Wyoming (GRB), Uinta Basin in Northwestern Utah (UB), Piceance Basin in northwestern Colorado (PB) and Denver Basin in Colorado (DB)]. Four stations that are used to demonstrate the methods are identified with red triangles. The black line indicates transect A–A’ ranging from the BR to the CP.
estimate uncertainties in ambient noise and earthquake dispersion information and the harmonic stripping method is used to generate backazimuth independent RFs with uncertainty estimates for each station. After applying the joint inversion (and for comparison the inversion based on surface wave dispersion alone) at each station, the final 3-D model of the crust and uppermost mantle is produced by interpolating between stations using simple-kriging at each model depth. Prior and posterior marginal distributions are presented to visualize changes in uncertainties for each data set. The study region is geologically diverse and structurally complex and provides an excellent proving ground for the methodology.

2 OVERVIEW OF THE METHOD

Monte Carlo and related parameter search algorithms (e.g. Sambridge 1999a) are designed to map data misfit across a broad range of model space. Within a Bayesian framework, these models are interpreted by computing the a posteriori (or posterior) probability distribution, which is the probability distribution of the model parameters given the observed data. Bayes’ theorem allows the probability distribution of the model by the likelihood functional

\[ L(m; D) \]  

the vector \( D \) and the likelihoods \( L(m_i) \) and \( L(m_j) \) are computed. The Metropolis law defines the probability of acceptance for model \( m_i \),

\[ P_{\text{accept}} = \begin{cases} 1 & \text{if } L(m_i) \geq L(m_j) \\ \frac{L(m_j)}{L(m_i)} & \text{if } L(m_i) < L(m_j) \end{cases} \]  

That is, model \( m_i \) is accepted if its likelihood is greater than that of model \( m_j \) (i.e. its misfit is lower). Even if its likelihood is smaller than that of model \( m_j \) it may still be accepted, however. For example, if \( P_{\text{accept}} = 0.4 \), we use a uniformly distributed random deviate generated between 0 and 1. If the value is less than 0.4 we accept the model. If the new model \( m_i \) is accepted, we define a new perturbation based on this model and search on. If the new model is not accepted, we make a record of it and its associated misfit information and start a new perturbation based on the previously accepted model \( m_j \) to choose the next step. The evolution of the algorithm in terms of improving misfit for surface wave data alone is shown in Fig. 8 and for the joint inversion in Fig. 11.

An additional model acceptance criterion then is applied to define the posterior distribution. In this we diverge from strict Bayesian methodology. If the data covariance matrix were known accurately, the choice of the acceptance criterion would be straightforward. Misfits to surface wave and RF data would be commensurable and the acceptance criterion would involve only choosing a probability threshold in the posterior distribution. As described later, because of difficulty in estimating the off-diagonal elements of the data covariance matrices, we have assumed that the surface wave dispersion data and the RFs have error processes that are independently but not identically distributed so that each covariance matrix is diagonal with elements \( \sigma_i^2 \) and \( s_j^2 \), respectively, for phase velocity at period \( i \) and RF at time \( j \). This assumption has the practical effect that the misfits between the two data types are incommensurable. We are, therefore, forced to introduce a misfit scaling parameter \( \kappa \) to normalize the misfits between the two data types, as described in Section 6.1. The resulting joint misfit function, therefore, has the following form:

\[ S_{\text{joint}}(m) = S_{SW} + S_{RF} = \sum_{i=1}^{N} \left[ \frac{[g_i(m) - D_{i,\text{obs}}]^2}{\sigma_i^2} \right] + \sum_{j=1}^{M} \left[ \frac{[R_j(m) - A(t_j)]^2}{s_j^2} \right], \]  

where \( g_i(m) \) is the phase velocity predicted for model \( m \) at period \( i \) on a discrete grid of \( N \) periods, \( D_{i,\text{obs}} \) is the observed phase velocity, the function \( A(t_j) \) is the observed RF at time \( t_j \) on a discrete grid of \( M \) times and \( R_j(m) \) is the predicted RF for model \( m \) at time \( t_j \). Somewhat different model acceptance criteria are applied in the inversion with surface wave data alone in Section 5 and in the joint inversion in Section 6. Example posterior marginal distributions are shown in Figs 9 and 12.

The methods that we present here differ from those of Bodin et al. (2011) in two principal ways. First, we choose smooth velocity profiles between specified boundaries at variable depths and fix the parameterization across the region of application. The nominal “transdimensionality” of the method of Bodin et al. refers to the variable parameterization of their method. We do not use an adaptive parameterization because, for the most part, we find that the
parametrization we choose is sufficiently flexible to fit the data. Secondly, as discussed above, we do not estimate the off-diagonal elements of the covariance matrices. Inspired by the hierarchical Bayesian formalism of Bodin et al. we attempted to estimate full covariance matrices for RFs, but found first that the estimated covariance matrices were singular and did not yield to matrix regularization methods designed to approximate the inverse matrices. Secondly, the covariance matrices that we derived were appropriate for a raw RF not the estimated azimuthally independent RF $A_{0}(t)$ that we use in the inversion. This is a subtle point that requires further analysis. For both reasons, we moved forward with diagonal covariance matrices for both dispersion data and RFs. Although these differences with the more general formulation of Bodin et al. (2011), which attempted to estimate the full covariance matrices for RFs, were motivated in part by our inability to estimate satisfactory covariance matrices, they do accelerate the inversion, which facilitates the application of the method to a large array like the USArray. In addition, as we show, the method works well to improve 3-D models relative to those produced from surface wave dispersion data alone.

3 DATA PROCESSING

3.1 Rayleigh wave data processing

3.1.1 Ambient noise tomography (ANT)

The method of ANT is now well established, including cross-correlation of long pre-processed time-series to generate empirical Green’s functions, measuring both Rayleigh and Love wave phase and group velocity curves and producing dispersion maps at particular periods (Shapiro et al. 2005; Yao et al. 2006; Moschetti et al. 2007; Yang et al. 2007). The data processing procedures that we adopt follow those of Bensen et al. (2007) and Lin et al. (2008), which we only briefly summarize here. First, raw vertical component seismograms recorded from 2005 to 2010 are downloaded from the IRIS DMC for the USArray TA stations in the western United States and are cut into 1-d time-series. The time-series length of each station is usually about 2 yr, but time-series lengths are variable and we do not apply a minimum time-series length restriction. Secondly, earthquake signals and other types of interference are removed by time domain normalization, in which the reciprocal of the mean of the absolute value of the waveform in a moving 80 s time window is used to weight the data point at the centre of the window. The weights are determined in the “earthquake band” between periods of 15 and 50 s period, but then are applied to the unfiltered data. Thirdly, the data are bandpass filtered between 5 and 100 s period and cross-correlation is performed between all station pairs. Then, Rayleigh wave group and phase velocity dispersion measurements are obtained from the symmetric component of each interstation cross-correlation by performing automatic frequency-time analysis (FTAN; Bensen et al. 2007). In this study, only Rayleigh wave phase velocity measurements are used. The automated FTAN dispersion measurements are winnowed by applying two criteria to select reliable measurements for surface wave tomography: (1) The interstation distance must be greater than three wavelengths at each period to ensure the far-field approximation and sufficient separation from pre-cursory noise and (2) signal-to-noise ratio (SNR) must be greater than 10 at each period for the measurement at that period to be accepted. Once all measurements are obtained, eikonal tomography (Lin et al. 2009) is then applied to produce phase velocity maps from 8 to 40 s period. Eikonal tomography takes account of ray bending (off-great circle propagation) but does not model finite frequency effects (Lin & Ritzwoller 2011). One of the signature features of eikonal tomography is that it produces meaningful uncertainty estimates at each geographical location. This is done by accumulating phase velocity estimates at a given location from many central stations, fitting and then removing a truncated Fourier Series over azimuth to estimate azimuthal anisotropy, and then measuring the standard deviation of the mean of the residual. Figs 2(a)–(c) presents phase velocity maps from ANT at periods of 8, 20 and 36 s. By 40 s period, the SNR of ambient noise has decreased enough that high resolution phase velocity maps are no longer generated. Earthquake data are introduced to produce the longer period dispersion maps.

3.1.2 Earthquake tomography (ET)

The data processing procedure for teleseismic earthquake data is discussed by Lin & Ritzwoller (2011) and is only briefly summarized here. First, teleseismic records at the USArray TA stations in the western United States following earthquakes with surface wave magnitudes $M_{s} \geq 5.0$ are downloaded from the IRIS DMC and cut according to the arrival window of surface waves before instrument responses are removed. Secondly, FTAN is applied to the cut seismograms and Rayleigh wave front group and phase traveltimes and amplitudes are measured as a function of period. Phase ambiguity is resolved using a network-based approach in which traveltimes are compared with measurements at nearby stations using the method of Lin & Ritzwoller (2011). Then Helmholtz tomography is performed to produce Rayleigh wave phase velocity maps from 32 to 80 s period. Uncertainties are estimated similar to eikonal tomography applied to ambient noise data, but by accumulating phase velocities over many different earthquakes. In Helmholtz tomography, finite frequency effects are corrected through a term involving the local Laplacian of the amplitude field. Figs 2(d) and (e) shows the phase velocity maps from ET at periods of 36 and 70 s, respectively. A comparison of the 36 s phase velocity maps obtained from ANT and ET is shown in Fig. 2(f). The average difference is 5 m/s with slightly faster phase velocities emerging from ET.

3.1.3 Construction of phase velocity curves at station locations

After Rayleigh wave phase velocity maps are generated on a $0.2^\circ \times 0.2^\circ$ grid from both ANT and ET, phase velocity curves are interpolated to each station location. At short periods (8–36 s), phase velocity measurements and associated uncertainties are taken exclusively from ANT maps. At long periods (40–80 s) results are taken solely from ET maps. Between periods of 36 and 40 s, we weight the two velocity measurements by their local uncertainties and a smooth curve is produced. Fig. 3 shows the dispersion curves and uncertainty estimates for TA stations R11A, T18A, Q22A and O25A whose locations are shown on Fig. 1. In the overlapping period band, the phase velocity measurements from ANT and ET are overplotted, demonstrating the coherence between ANT measurements (red error-bars) and ET measurements (blue error-bars).

3.2 RF data processing

In the Appendix, we describe a method to estimate the azimuthally independent RF, $A_{0}(t)$, and associated uncertainty, $s(t)$, for each
Figure 2. (a)–(c) Phase velocity maps at periods of 8, 20 and 36 s from ambient noise eikonal tomography (ANT). (d) and (e) Phase velocity maps at 36 and 70 s from earthquake Helmholtz tomography (ET). (f) Difference between the ANT and EQ results at 36 s period.

We refer to the method as ‘harmonic stripping’ because a truncated harmonic function is fit to the azimuthally dependent raw RFs for a given station at each time $t$,

$$H(\theta, t) = A_0(t) + A_1(t) \sin[\theta + \theta_1(t)] + A_2(t) \sin[2\theta + \theta_2(t)],$$

where the $A_i$ ($i = 0, 1, 2$) are the amplitudes of the three harmonic components and the $\theta_i$ are initial phases for the azimuthally dependent components. This procedure approximately strips the azimuthal dependence from the RFs and produces the azimuthally independent RF, $A_0$. As shown in the Appendix, dipping crustal interface and azimuthal anisotropy dominantly produce signals with periodicitites of $2\pi$ and $\pi$, respectively, which are largely removed via harmonic stripping so that the azimuthally independent component of the fit is a reliable estimate of the RF for a horizontally layered isotropic effective medium. Without complete and uniform azimuthal coverage, simply taking the average of the RFs over
azimuth as is commonly done does not guarantee an unbiased estimate. The data processing and quality control (QC) procedures for the RFs are also discussed in the Appendix.

In this study we use Ps RFs (teleseismic P-to-S converted waves) exclusively, but Sp RFs could also be used in the context of the methods we develop, at least in principle (Vinnik et al. 2004). We select all teleseismic events in the NEIC PDE catalogue with a body wave magnitude $m_b \geq 5.1$ within the epicentral distance range of $30^\circ$–$99^\circ$ from each station. All sets of 3-component seismograms are rotated to form the radial and transverse components and are then bandpass filtered between 0.03 and 4 Hz using a Butterworth filter (2 poles, 2 passes). The harmonic stripping method and QC procedures have been applied to the 185 TA stations in the study region (Fig. 1). An example is shown in Fig. 4 for the TA station R11A. The QC’ed RFs presented over azimuth are shown in Fig. 4(a) and the harmonic function $H(\theta, t)$ fit to these RFs is shown in Fig. 4(b) for comparison. As described in the Appendix, the residual between these functions is used to define the uncertainty, $s(t)$, of the azimuthally independent RF and is presented in Fig. 4(c). The individual harmonic components are shown in Figs 4(d)–(f). The functions $A_r(t)$ and $A_t(t)$ quantify the azimuthally dependent signals that may be produced by tilts on internal interfaces and anisotropy, respectively. Some of these arrivals are observed to undergo a moveout at different azimuths. These functions could be applied to fine-tune the model presented here, but are not currently used for this purpose.

Several examples of the final azimuthally independent RFs $A_r(t)$ and associated uncertainties are shown in Fig. 5. The RFs display significant station-to-station variations. Those observed at station R11A in the Basin and Range province and station Q22A in the Rocky Mountains have strong P-to-S converted positive amplitude phases between 3 and 7 s delay time, suggesting a sharp rather than gradient transition from crustal to mantle velocities. The RF for TA station T18A in the Colorado Plateau, in contrast, shows no Moho P-to-S conversion, which means that a gradient Moho is probable at this point. The RF for station O25A in the Great Plains has a strong negative arrival at $\sim 4$ s caused by sedimentary layer reverberations at this location which probably interferes with observation of the Moho conversion.

### 4 Model Parametrization and Determining the Prior Distribution

#### 4.1 Model parametrization

Surface wave phase velocity curves and azimuthally averaged RFs are sensitive to the local average 1-D $V_s$ structure beneath each location. For this reason, the inversion for a 3-D model reduces to a set of 1-D $V_s$ inverse problems. We invert only for shear wave speed in the 200 km beneath the surface (reliable only to about 150 km) because the longest period of surface waves that we use here is 80 s. Below 200 km, the model is assumed to be constant. Because we use only Rayleigh waves, which are pre-dominantly sensitive to $V_s$, we assume an isotropic $V_{sv}$ model where $V_R = V_s = V_{sv}$. We set the $V_R/V_s$ ratio to 2.0 in the sedimentary layer and 1.75 in the crystalline crust and mantle. The $V_R/V_s$ ratio is important in RF analysis and is discussed further in Section 8.1. For density we use the scaling relation (for 10 km) advocated by Christensen & Mooney (1995) in the crust and by Karato (1993) in the mantle. We apply a physical dispersion correction (Kanamori & Anderson 1977) using the Q model from PREM (Dziewonski & Anderson 1981), and the resulting model is reduced to 1 s period.

In a traditional surface wave inversion a 1-D model may be parametrized with smooth functions in the crust and mantle (Melnik & Romanowicz 2000), as a stack of layers (Yang & Forsyth 2006), or a combination of both (e.g. Shapiro & Ritzwoller 2002; Yang et al. 2008b; Moschetti et al. 2010). RF inversions are, however, typically parametrized with a set of fine layers (Cassidy 1992; Julia et al. 2000). We parametrize our model similar to some surface wave inversions in which B-splines represent structure in both the crystalline crust and the uppermost mantle as follows. (1) There is one sedimentary layer with a linear gradient velocity. Three parameters are used to describe this layer: layer thickness and $V_s$ at the top and bottom of the layer. (2) There is one crystalline crustal layer. Five parameters are used to describe this layer: layer thickness (km) and four B-spline coefficients for $V_s$. (3) There is one uppermost mantle layer to a depth of 200 km. Five parameters are used to describe this layer: five B-spline coefficients for $V_s$. The thickness of this layer is controlled by the thicknesses of the top two layers.

This model-parametrization contains 13 free parameters (Fig. 6). We explicitly seek vertically smooth models that fit the data. Here, smooth means that the model has no more vertical structure than required to fit the data within a specified tolerance and also that the model is continuous within some depth ranges. In this model parametrization, two discontinuities are introduced because the predominant signals in RFs are from the Moho and, in the presence of sediments, from the sediment-basement contact. We demonstrate that a smooth parametrization can explain both the surface wave and RF data within data uncertainties across almost the entire study region. In other areas where a discrete higher-velocity lower crustal layer is observed (e.g. immediately north of our study area; Gorman et al. 2002), an additional crustal interface may be needed. The
Joint Bayesian Monte-Carlo inversion

Figure 4. (a) The quality controlled receiver functions are plotted along backazimuth for station R11A. (b) The estimated receiver functions, $H(\theta, t)$, from harmonic stripping. (c) The difference between (a) and (b), from which uncertainties in the azimuthally independent receiver function are determined. (d)–(f) The three estimated components from harmonic stripping.

smoothness of the model is imposed by the parametrization so that ad hoc damping is not needed during the inversion.

With this model parametrization, we construct the model space in which the prior and posterior distributions are determined. The construction of the model space is based on perturbations to a reference model $m_0$, where sedimentary structure is taken from Mooney & Kaban (2010), crustal thickness from Bassin et al. (2000), and shear wave speeds in the crust and mantle from Shapiro & Ritzwoller (2002), as summarized in Table 1. At each location models are sampled around the reference in the range prescribed in the table. For example, the crystalline crustal thickness variation is $\pm 25$ per cent of the input value. If the reference thickness were 40 km, values would be considered from 30 to 50 km.

4.2 Determining the prior distribution

The prior distribution reflects the state of knowledge before data are introduced. Prior information that we apply in choosing models includes the following seven prior assumptions. (1) Models exist in the model space $M$, which is defined as the reference model subject to allowed perturbations (Table 1). (2) The model is continuous between Vsv discontinuities at the base of the sediments and Moho and is continuous in the mantle. (3) Velocity in the sedimentary layer increases with depth. (4) Velocity in the crystalline crust increases with depth (monotonicity constraint). (5) Velocity contrasts across the sedimentary basement and across the Moho discontinuity are positive. (6) $V_i < 4.9$ km s$^{-1}$ throughout the model. (7) When surface wave data are used alone we apply the assumption of a positive velocity gradient in the uppermost mantle. Up to constraint (7), the prior distributions for inversions using only surface wave data or surface wave data jointly with RFs are identical. We discuss in Section 7.3 how constraint (7) may introduce bias in the estimated model, and how it can be eliminated when RFs are introduced. Together, these assumptions reduce model complexity and also ameliorate some of the velocity-depth trade-offs that occur in the surface wave inversion.

To determine the prior distribution, we perform a random walk in model space with the following steps. First, we initiate the model search at a random point in model space $M$. If this random point does not obey the assumptions listed above, it is rejected and another random point is chosen. Secondly, we introduce a random jump in model space. We simultaneously perturb all 13 parameters by selecting a new value randomly governed by a Gaussian probability distribution near the old value subject again to the previous assumptions. These Gaussian widths have been chosen empirically to ensure the efficiency of the random walk in sampling the prior and posterior distributions. If the model produced is not contained in model space $M$ (i.e. does not satisfy the seven constraints in the previous paragraph and does not fall within the ranges presented in Table 1), then the model is rejected and the process re-initiates.

With a sufficient number of sample points ($>100,000$), the prior distribution of each parameter can be viewed as histograms of marginal distributions of model characteristics at different depths such as those shown in Fig. 7 at the location of the TA station R11A. For this figure we generated Vsv models from such parameter distributions by converting the B-spline coefficients to Vsv values as a function of depth. Due to the lack of prior information imposed
Figure 5. The estimated azimuthally independent receiver function ($A_0(\tau)$, red curve) and uncertainty ($\sigma(\tau)$, grey corridor) for four stations whose locations are identified in Fig. 1. The uncertainty is the rms of the residual remaining after the harmonic fitting (eq. (A2)) to the azimuthally dependent receiver functions, reduced by a factor of two between 3 and 8 s.

Figure 6. Model parametrization illustrating the 13 model parameters used in the Monte Carlo sampling of model space.

5 MONTE CARLO INVERSION OF SURFACE WAVE DISPERSION DATA

In this section, we discuss the Bayesian Monte Carlo inversion based on surface wave data alone for later comparison with the joint inversion with RFs, which is presented in Section 6. As above, we use the USArray TA station R11A in the Basin and Range province for examples and the input data for the inversion are shown in Fig. 3(a). At each location, Rayleigh wave phase velocities exist on a fixed discrete grid from 8 to 80 s period. Prior constraint (7) is not applied in this section.

5.1 The likelihood functional

For a problem in which the measurements are observed Rayleigh wave phase speeds at a particular location, which have independent but not identically distributed Gaussian observational uncertainties $\sigma_i$ (period index $i$), the likelihood functional is computed from eq. (2) where the misfit function comes from the first term on the right hand side of eq. (5). The assumption of the frequency independence of error processes in surface wave dispersion is made here and in
other studies (Liu et al. 2010; Bodin et al. 2012), because the covariance over frequency is not well understood. The correlation of the dispersion velocities over frequency deserves further investigation.

5.2 Determining the posterior distribution

The information about model space extracted by introducing the data is described by the posterior distribution. As defined by eq. (1), the posterior distribution is the product of the prior distribution and the evaluation of fit to the observed data, which is summarized by the likelihood functional \( L \). During model space sampling, when a new model \( m_i \) is generated by perturbing a given model \( m_j \) under the prior assumptions, the likelihood functional of this model is calculated through forward computation using the Thomson–Haskell method (computed using the code of Herrmann, http://www.eas.slu.edu/eqc/eqccps.html) with an earth-flattening transformation. Both of the likelihoods \( L(m_i) \) and \( L(m_j) \) are computed as discussed in Section 5.1, where the Metropolis law, eq. (4), defines the probability of acceptance for model \( m_i \), as described in Section 2.

Fig. 8 shows an example of how the likelihood functional and misfit function of the models accepted during sampling of the posterior distribution evolve for TA station R11A. Accepted models are identified with circles, rejected models are not plotted. The trend shows that after about 100 samples, the likelihood functional will fluctuate near unity during the remaining iterations. We choose to stop iterating after 3000 sampling steps. To ensure that the starting point does not affect the sampling, we initiate the Monte Carlo sampling at ten different random models and the resulting statistics reflect these ten independent samplings. Fig. 8 also shows how the root mean square (RMS) misfit function, \( \chi = \sqrt{\frac{1}{S} \sum_i S(m_i)} \), evolves. When \( \chi < 1 \), the fit lies within the estimated uncertainties, on average.

As discussed in Section 2, an additional model acceptance criterion is introduced as the basis for accepting models to form the posterior distribution. The minimum misfit \( \chi_{\text{min}} \) is defined as the minimum value of \( \chi \) found for all models visited during model space sampling. Once \( \chi_{\text{min}} \) is found, we define the threshold of acceptance \( \chi_{\text{crit}} \) for the surface wave inversion as follows:

\[
\chi_{\text{crit}} = \begin{cases} 
2 \chi_{\text{min}} & \text{if } \chi_{\text{min}} \geq 0.5 \\
\chi_{\text{min}} + 0.5 & \text{if } \chi_{\text{min}} < 0.5
\end{cases}
\]  

Thus, to define the ensemble of accepted models \( \text{a posteriori} \), we accept any model \( m \) that is visited during the inversion as long as \( \chi(m) \leq \chi_{\text{crit}} \). The choice of the critical threshold level below which models are accepted to form members of the posterior distribution is admittedly ad-hoc and deviates from strict Bayesian practice.
Figure 9. (a)–(f) The same as Fig. 7 for TA station R11A, but here the posterior distributions after Monte Carlo sampling using surface wave data alone (models fit the surface wave data) are plotted. The prior distributions from Fig. 7 are plotted as write histograms in the background. Surface wave data reduce the spread of structural velocities between boundaries, but have less effect on the depth to boundaries or velocity jumps across boundaries due to trade-offs between the parameters near the boundaries.

Eq. (7) succeeds to produce posterior distributions that capture our degree of belief in the final models, however.

From the set of accepted models in the posterior distribution, we compute the distribution of each parameter and the marginal distribution of Vsv at each depth. Fig. 9 presents example histograms of marginal distributions for several structural variables derived using surface wave data alone to compare with the prior marginal distributions shown in Fig. 7. Typically, the distributions of velocities between boundaries are narrower. For example, Vsv at 10 km depth in the crust is narrowed; its standard deviation decreases by about a factor of six from ∼0.3 to ∼0.05 km s⁻¹. However, posterior distributions of discontinuity depths (Moho, sediments) or velocity jumps across the discontinuities do not decrease as appreciably. For example, crustal thickness and Vsv contrast across Moho are not changed strongly. These histograms demonstrate the sensitivity of Rayleigh wave dispersion to Vp structure. (1) Sedimentary thickness does not change because of the lack of very short period (<8 s) surface wave dispersion measurements. (2) Vsv at 10 km depth shows a much narrower distribution compared with the prior distribution because ambient noise provides information about this depth. (3) The mean of the crustal thickness distribution is about 33 km with a 1σ width of ∼3.6 km, showing that surface wave dispersion at this location possesses only weak sensitivity to the depth of the Moho discontinuity. The distribution is bimodal and not strongly peaked. (4) The marginal distributions of Vsv in the lower crust and the Vsv contrast across the Moho are narrowed somewhat but retain considerable uncertainties (0.15–0.2 km s⁻¹) and the lower crustal distribution is also bimodal. (5) Vsv in the uppermost mantle (120 km) is narrowed appreciably because the earthquake data are sensitive to this depth, but it is not constrained as well as at 10 km in the crust. This demonstrates that the vertical resolution of surface waves to local structure degrades with depth. Therefore, we only report structure to 150 km. These findings make intuitive sense and are quantified with the posterior distributions.

An alternative view of the results is provided by computing the mean, the median, the 1σ uncertainty and the full range of Vsv at each depth after all accepted models are identified. Fig. 10 shows the resulting extent of accepted models for station R11A. The grey corridors in Figs 10(a) and (b) outline the extent of all accepted models that fit the data, whereas the red lines present the 1σ width around the mean model. Note that without assimilating RFs, the crustal thickness (Moho depth referenced to the surface in Fig. 10b) is not well resolved and the predicted RFs (red waveforms in Fig. 10c) do not fit the observed azimuthally independent RF well, on average. There is also a strong trade-off between the lower crustal velocity and the uppermost mantle velocity. We show in Section 6 that these problems are ameliorated with the addition of RFs in the inversion.

6 MONTE CARLO INVERSION OF SURFACE WAVE DISPERSION AND RF DATA

When RFs are assimilated into the Monte Carlo algorithm, the prior distribution of models largely remains the same, although as we discuss in Section 7.3 we release the constraint on the positivity of the uppermost mantle velocity gradient. Here, we introduce the joint likelihood functional and then describe the posterior marginal distribution at TA station R11A and then for three other TA stations.

6.1 The joint likelihood functional

The joint likelihood functional is defined by eqs (2) and (5), but because we do not know the diagonal elements of the data covariance matrices we need to estimate the relative scaling of the misfits for surface waves and RFs that is quantified through parameter κ. We consider values of κ ranging from 1 to 40, and choose 2.5 as a value that strikes the appropriate balance between the two data sets. By choosing this value both data sets are approximately weighted equally in the inversion and the resulting misfits of the two data sets are comparable. The joint misfit function of eq. (5), therefore, becomes,

\[ S_{\text{joint}}(m) = S_{\text{SW}} + \frac{1}{2.5} S_{\text{RF}} = N \chi^2_{\text{SW}} + \frac{M}{2.5} \chi^2_{\text{RF}} \]

\[ = \sum_{i=1}^{N} \frac{[g_i(m) - D_{i\text{obs}}]^2}{\sigma_i^2} + \frac{1}{2.5} \sum_{j=1}^{M} \left[ R_j(m) - A_0(t_j) \right]^2 s_j^2, \]

(8)

where we have implicitly defined the \( \chi^2 \) misfit for surface waves and RFs. RMS misfit is the square root of \( \chi^2 \).

6.2 Determining the posterior distribution in the joint inversion

Fig. 11 shows how the joint likelihood functional for station R11A evolves as model space is sampled. The convergence duration for this sampling is about 500 samples, which requires about five times more models than considered when surface wave data are used alone (Fig. 8). As with the surface wave inversion, we repeat the procedure starting from ten different random points to find the range of models that fit the data acceptably. The forward calculation of RFs is based on the code developed by Shibutani et al. (1996), which has also been used by Sambridge (1999a). We introduce...
Joint Bayesian Monte-Carlo inversion

Figure 10. (a) and (b) The model ensemble at the location of TA station R11A in the Basin and Range province resulting from the Monte Carlo inversion of surface wave data alone is shown in full width (black lines with grey fill) and $1\sigma$ width (red corridor). Average Moho depth is identified as a dashed line at $\sim 32$ km. (c) The observed receiver function (white line) is plotted with predicted receiver functions (maroon lines) computed from all accepted models (eq. (7)), showing that receiver functions are not well fit, on average, by models constrained by surface wave data alone. (d) The observed Rayleigh wave phase velocity dispersion data (black error bars) are plotted with predicted surface wave phase velocity curves computed from all accepted models (eq. (7), grey lines). The red curve is the predicted phase velocity curve from the best fitting model.

For the joint inversion a somewhat different criterion than for the surface wave inversion as the basis for accepting models to form the posterior distribution. For each location, minimum RMS misfits for surface wave dispersion $\chi^{SW}_{\text{min}}$ and RFs $\chi^{RF}_{\text{min}}$ are identified, although the models that minimize misfits for these two types of data are generally not the same. The joint normalized relative RMS function $\chi^{\text{joint}}$ is defined as follows:

$$\chi^{\text{joint}} = \frac{1}{2} \left( \frac{\chi^{SW}}{\chi^{SW}_{\text{min}}} + \frac{\chi^{RF}}{\chi^{RF}_{\text{min}}} \right).$$

This function is a combination of relative RMS misfits for the two data sets. The criterion for model acceptance is $\chi^{\text{joint}} < \chi_{\text{crit}}$, where $\chi_{\text{crit}}$ is defined as,

$$\chi_{\text{crit}} = \chi^{\text{joint}}_{\text{min}} + 0.5.$$  

For example, in the particular inversion for TA station R11A, the minimum misfit for the surface wave data is $\chi^{SW}_{\text{min}} = 0.51$, the minimum misfit for the RF data is $\chi^{RF}_{\text{min}} = 0.53$ and the minimum misfit to both data in the joint inversion is $\chi^{\text{joint}}_{\text{min}} = 1.15$. Thus, the best fitting model that emerges from the joint inversion does not fit either data type optimally, but fits them both acceptably. In total, about 1000 models are found below the critical threshold $\chi_{\text{crit}} = 1.65$.

Example posterior marginal distributions are plotted and compared with prior (marginal) distributions in Fig. 12. All the marginal distributions change shape relative to the prior distributions, especially for Vsv at different depths and for crustal thickness. Compared with the posterior distributions using surface wave data alone (Fig. 9), the distributions of velocities in the crust and upper-mantle change only subtly (Figs 12b and f). The other distributions change profoundly. For example, crustal thickness at station R11A sharpens (from $\sigma = 3.6$ km to $\sigma = 1.3$ km). Concerning the velocity difference across the Moho, the mean increases and the standard deviation approximately halves. Such significant changes are expected because RFs are most sensitive to velocity contrasts across layer boundaries beneath the station; therefore, the position of the Moho is better determined as are the values of model variables that trade-off with Moho depth in the surface wave inversion.
6.3 Examples of model ensembles

The model ensemble as well as the fit to the data at TA station R11A are shown in Fig. 13. The 1σ width is less than half of the full width of the model ensemble, because the Vsv distribution at any depth for the joint inversion is approximately Gaussian except for the sedimentary layer (see Fig. 12). Fig. 13(c) presents the azimuthally independent RF where the corridor outlined by black lines indicates the uncertainty range. The predicted RFs from the ensemble of accepted models are shown with grey lines and the model that fits the RF best is shown with the red line. The most prominent signal is the peak at about 4 s, which is fit quite well. However, the small trough near 2 s is not fit because doing so would require introducing another crustal discontinuity, which would violate monotonicity. The fact that the algorithm does not automatically adapt the parametrization to accommodate other discontinuities is discussed further in Section 8.2. Fig. 13(d) identifies the model that fits the surface wave data best, but the model that fits the RF best misfits the surface wave data somewhat between about 40 and 60 s period. The model that minimizes the joint misfit strikes a balance between these models.

Fig. 14 presents another example of the joint inversion for station T18A in the Colorado Plateau. In the observed RF, no clear peak is found from 3 to 8 s where the Moho P-to-S converted phase would be seen in “normal” RFs. After inversion, the RF data is well fit without a P-to-S converted arrival based on the model ensemble shown in Fig. 14(b). The Moho discontinuity is not well defined in the ensemble of models and Vsv changes smoothly from the lower crust to the uppermost mantle.

Fig. 15 summarizes the joint inversion result for station Q22A in the Colorado Rocky Mountains. In the observed RF, a peak at about 6 s suggests that the Moho is deeper than at station R11A in the Basin and Range province. As a result, the crustal thickness distribution is centred at about 49 km depth with an uncertainty of 2.3 km. The relatively small uncertainty of crustal thickness generates a sharp transition between the lower crust and uppermost mantle, particularly compared with station T18A in the Colorado Plateau.

Fig. 16 shows the joint inversion result for station O25A in a sedimentary basin, the Denver Basin region of the Great Plains. The peak at 1 s on the RF and the trough at 4 s indicate a thick sedimentary layer at shallow depths. After inversion, a thick sedimentary layer (~4 km) with a strong vertical velocity gradient is found and the crustal thickness is estimated to be 51.1 ± 3.9 km. Signatures in RFs that are common for sedimentary basins are a broadened direct P-pulse or an apparent lack of a direct P-arrival at zero delay time followed by a large amplitude arrival in the first second or so—the latter feature being due to bending to vertical incidence and a strong conversion to shear energy at the sediment-basement contact. The high-amplitude apparent mid-crustal negative arrival is modelled here as a reverberation within the sedimentary basin with two shear and one compressional legs within the sediment layer. The constraint of monotonically increasing velocities within the crust aids suppression of sediment multiples in favour of imaging true deeper crustal structure. In a few other locations in which there are true mid-crustal low-velocity layers, such as the Rocky Mountain location shown in Fig. 15, this may suppress a true feature, however. Extreme examples may be magma chambers (e.g. Wilson et al. 2003). Vsv in the sediments at O25A increases from less than 1 km s⁻¹ to more than 3 km s⁻¹ at depth of ~3 km, representing the compaction of sediments in this layer. Vsv in the upper crust is ~3.4 km s⁻¹ and reaches higher than 4 km s⁻¹ in the lower crust, perhaps indicating a mafic lower underplated crust (‘7.x layer’) that has been proposed for parts of the High Plains (e.g. Gorman et al. 2002). A fast lithosphere (>4.5 km s⁻¹) is observed in the upper mantle.
Figure 13. (a) and (b): The same as Fig. 10, but for results from the joint inversion of Rayleigh wave phase velocities and receiver functions at the same TA station (R11A). (c) The synthetic receiver functions from the accepted model ensemble are plotted with grey lines, with the best fitting receiver function shown as the red curve. The parallel black lines are the estimated uncertainty of the receiver function. There is a clear P-to-S conversion near 4 s period, necessitating a large velocity jump at a shallow Moho. (d) The predicted surface wave dispersion curves from all accepted models are plotted with grey lines with the best fitting curve identified in red.

7 SYSTEMATIC APPLICATION OF THE JOINT INVERSION ACROSS THE INTERMOUNTAIN WEST

In previous sections, the joint Bayesian Monte Carlo inversion of surface wave dispersion and RFs was shown to estimate reasonable 1-D models with attendant uncertainty information for stations situated in a wide variety of structural environments. We now report on the application of the joint inversion method to the 185 TA stations across the intermountain west (Fig. 1).

7.1 Construction of a 3-D Vsv model for the Intermountain West

We apply the Monte Carlo inversion to the 185 stations shown in Fig. 1. For each joint inversion, we identify the model from the resulting model ensemble that fits the surface wave data (SW) or RFs best as the SW or RF best fitting model, respectively. The misfits to SW data by the SW best fitting models for the stations tested are plotted with the white histogram in Fig. 17(a), whereas the misfits to the RFs by the RF best fitting models are plotted in Fig. 17(b).

For each station, the model with minimum joint misfit is identified, and is called the joint best fitting model. Misfits to SW and RF data from the joint best fitting models are shown in Fig. 17 with red histograms. At most stations, the joint best fitting models fit the SW or RF data only slightly worse than the model that fits each single data type best. This indicates that there is some tension in fitting the two data types. Secondly, the joint best fitting models have a misfit < 1 for both data sets for almost all stations, which indicates that the simple model parametrization that we use in the joint inversion can reproduce the RFs with misfits below the RF uncertainties for most of the stations tested. Thirdly, larger RF misfits appear for stations near the corner of Utah/Colorado/Wyoming and are sparsely distributed through the Basin and Range. The RFs at those stations have sharp backazimuthal variations that vitiate the harmonic striping method’s attempt to estimate an azimuthally independent RF. Cases with sharp backazimuthal signals in the RFs are discussed further in the Appendix, but in these cases the uncertainty in the RF increases substantially and the inversion reverts for the most part to fitting surface wave data. Fewer than three stations show a large misfit (> 0.9) to surface wave data and these stations are sparsely distributed across the map. Finally, the comparable minimum and
joint misfits for the two data sets also indicate that neither data set is overweighted during the inversion.

After inversion is performed at all stations, we obtain Vs\textsubscript{s} model ensembles beneath the 185 stations with means and uncertainties at all depths. The distribution of stations forms an irregular grid. To produce a smooth Vs\textsubscript{s} model on a regular grid, simple kriging (Schultz et al. 1999) interpolation is applied to smooth the Vs\textsubscript{s} values at each depth based on the mean value at that depth and the estimated uncertainties. For a given depth, at each grid node we search for stations within a 1-degree radius. We weight the average Vs\textsubscript{s} of the model ensemble for at given station using a weighting function defined as follows:

\[
W_i = \begin{cases} 
\frac{1}{(1 + d_i)\sigma_i} & \text{if } d_i \leq 1^\circ \\
0 & \text{otherwise} 
\end{cases},
\]

(11)

where \(d_i\) and \(\sigma_i\) denote the distance to station \(i\) from the grid node and the uncertainty of the model at the specified depth for that station, respectively. By performing this simple-kriging interpolation, map views of the 3-D model on a 0.25° × 0.25° spatial grid are produced.

7.2 Characteristics of the 3-D model

In the upper crust (10 km depth, Fig. 18a), high velocities are observed beneath the Colorado Plateau and Great Plains, whereas the Rocky Mountains show low Vs\textsubscript{s}. Near the northern boundary of the Colorado Plateau low velocities are also observed, presumably caused by very thick sediments in the Green River Basin, Uinta Basin, etc. At these locations, the maximum allowed sediment thickness may be less than the true thickness, leading to smearing of low velocities into the upper portion of the crystalline crustal layer. In contrast, the Denver Basin is much thinner than 10 km and the structure there is well constrained by RF data. Uncertainties at 10 km depth are approximately homogeneous across the region, averaging about 27 m s\textsuperscript{–1}, which is about 0.8 per cent.

In the lower crust (averaged from 4 km above the Moho to the Moho, Fig. 18c), the most prominent feature is the slow anomaly (<3.6 km s\textsuperscript{–1}) encompassing the eastern Basin and Range province near the northwestern Colorado Plateau. Slow lower crust is found across the entire Basin and Range and also near the western edge of the Rocky Mountain province in southwestern Colorado. There is coincidence between the ~3.85 km s\textsuperscript{–1} Vs\textsubscript{s} contour and the eastern Rocky Mountain topographic high in Colorado, implying a strong relationship between topography and crustal structure.
Figure 15. The same as Fig. 13, but for the result at TA station Q22A located in the Colorado Rocky Mountains. A clear P-to-S conversion is seen near 6 s, requiring a large velocity jump at a deeper Moho than in Fig. 13.

A fast lower crustal anomaly extends through the Colorado Great Plains and southern Wyoming and also penetrates into the Colorado Plateau. Uncertainties are larger and more variable in the lower crust (Fig. 18d) than the upper crust, ranging from about 40 m/s in parts of the Colorado Plateau to more than 160 m s$^{-1}$ in the Rocky Mountains of southern Colorado. Higher uncertainties have two causes. First, they appear where there is a large jump in velocities across the Moho, due to a trade-off between Moho depth and uppermost mantle structure. The trade-off has been ameliorated but not entirely eliminated through the addition of RFs in the inversion. Secondly, larger uncertainties also occur where RFs have a larger uncertainty in the amplitude of P-to-S conversions.

At 60 km depth (Fig. 18e), which is in the uppermost mantle across the region, a strong contrast is observed between the eastern and western sides of the study region. A strong anomaly as low as 4.0–4.1 km s$^{-1}$ is seen at the eastern edge of the Basin and Range, which may be related to the Cenozoic magmatism in this region (Roy et al. 2009). Much higher velocities are observed beneath the Great Plains, reaching up to 4.8 km s$^{-1}$. Uncertainties are largely homogeneous across the study region, averaging about 50 m s$^{-1}$ (Fig. 18f). At a depth of 120 km (Fig. 18g), Vsv beneath the Great Plains is very high, but the strongest high velocity anomaly is detected beneath the Wyoming Craton in southern Wyoming. This high velocity feature continues beneath the Colorado Plateau, suggesting a strong, thick lithosphere beneath the northern Colorado Plateau except for its northwestern periphery. The Basin and Range and Rocky Mountains show relatively homogeneous low Vsv compared with other regions. Uncertainties (Fig. 18h) are fairly homogeneous and average about 65 m s$^{-1}$ ($<1.5$ per cent) across the entire study region. Uncertainties are larger deeper in the mantle (120 km versus 60 km) because the surface wave dispersion information is less sensitive to deeper structure.

7.3 Changes in the 3-D model compared with the surface wave inversion

There are several significant advantages to adding RFs to surface wave dispersion data in the Monte Carlo inversion. We focus on three topics: (1) determining crustal thickness and uncertainties, (2) determining the velocity jump across the Moho and (3) constraining uppermost mantle structure.

By utilizing RFs in the inversion, there is a natural increase in the accuracy of estimates of Moho depth or crustal thickness. This is apparent on comparison between the crustal thickness distribution for station R11A from the inversion with surface wave data alone and that from the joint inversion of both data sets (Figs 9c and 12c). This observation holds for all stations with clear P-to-S converted
Figure 16. The same as Fig. 13, but for the result at TA station O25A located in the Great Plains. Strong sedimentary reverberations dominate the receiver function adding uncertainty in the location of and velocity jump at the Moho.

Figure 17. (a) Misfit histogram showing the misfits to surface wave data over the spatial grid covering the area of study. The white histogram is for the model that best fits the surface wave data at each point and the red histogram is for the model that jointly best fits both surface wave and receiver function data. At most gridpoints, the jointly best fitting models fit the surface wave data only slightly worse than the model that fits the surface wave data best. (b) Same as (a), but for receiver functions. The white histogram is for the model that best fits the receiver function data at each location and the red histogram is for the model that jointly best fits both data sets. At most gridpoints, the jointly best fitting models fit the receiver function data only slightly worse than the model that fits the receiver function data best.
Figure 18. Map views of Vsv at different depths (left) with uncertainties (right). (a) and (b) Vsv at 10 km depth. (c) and (d) Vsv in the lower crust, 4 km above Moho. (e) and (f) Vsv at 60 km depth. (g) and (h) Vsv at 120 km depth.

Signals. Figs 19(a) and (b) presents crustal thickness and associated uncertainties from inverting surface wave dispersion data alone. The uncertainty level scales with crustal thickness and is $\sim 5$ km on average with smaller values in the Basin and Range and larger values in the Great Plains. In contrast, crustal thickness and uncertainty determined from the joint inversion appears in Figs 19(c) and (d). The uncertainty level from the joint inversion decreases in the Basin and Range, the central Rocky Mountains and parts of the Great Plains, where P-to-S converted phases are well observed. In the northern part of the Colorado Plateau and southern Wyoming...
craton where thick sediments exist, crustal thickness uncertainties are not reduced by adding RFs or may even increase where the P-to-S signals in the RFs are muted by sedimentary reverberations. In addition to improving the determination of Moho depth, RFs also help to determine whether a sharp Moho discontinuity exists or not. An example is shown in the inversion at station T18A in the Colorado Plateau (Fig. 14), where the raw RF has no dominant arrival from 3 to 7 s where a P-to-S conversion should appear. The joint Monte Carlo inversion thus produces a model that has a gradient in Vsv at the depth where Moho is expected. The resulting crustal thickness distribution has larger uncertainty than when a clear P-to-S phase is observed, but none of the models in the distribution shows a sharp Moho.

Secondly, introducing RFs improves the determination of the Vsv contrast across the Moho, which is related to the amplitude of the P-to-S phase in the RFs. For instance, the RF at station R11A (Fig. 13) shows a strong Moho conversion whereas the converted phase in the RF at T18A is weak, resulting in a Vsv contrast at Moho that is stronger at R11A than at T18A (or perhaps a gradational crust–mantle transition at station T18A). A map of the Vsv contrast and uncertainty across the region is shown in Figs 19(e) and (f). The features shown are coherent with geological province. A high Vsv contrast across Moho is observed beneath the Basin and Range as well as parts of the Rocky Mountains and the Great Plains east of the Denver Basin. In contrast, beneath the Colorado Plateau, especially under the northwestern Colorado Plateau, the Vsv contrast is very low. This feature is consistent with observations made by earlier studies (Sheehan et al. 1997; Gilbert & Sheehan 2004; Levander et al. 2011; Bailey et al. 2012). Beneath the southern Wyoming Craton, we observe a small Moho Vsv contrast as well as high uncertainties in crustal thickness and thick sediments. This is because sedimentary reverberations dominate the RF where a Moho peak would be expected and the Vsv constrast across Moho is difficult to resolve. Uncertainties in the velocity contrast across Moho range between about 80 to 160 m s⁻¹, being smallest in the Colorado Plateau where the Vsv contrast is low in all accepted models.
8 DISCUSSION

8.1 Sensitivity to the \( V_p/V_s \) ratio

Rayleigh wave phase velocity dispersion curves are only weakly sensitive to \( V_p \) compared to \( V_s \); hence, the \( V_p/V_s \) ratio is usually a fixed parameter in the inversion of surface wave data. However, the \( V_p/V_s \) ratio is important in RF analysis because it is needed to map the P-to-S conversion time delay to depth. The \( V_p/V_s \) ratio can be determined when Moho reverberations such as the PpPs or PsPs phases are used as discussed by Zhu & Kanamori (2000). However, these phases arrive later than the 10 s time window used here in the Monte Carlo inversion. We omit these phases from consideration because at many locations in the western US Moho reverberations cannot be isolated cleanly due to lateral heterogeneity (e.g. the large scatter in the results of Lowry & Perez-Gussinye 2011, Fig. 2a) and in some places a gradational crust–mantle transition. Rayleigh wave ellipticity and local amplification may also potentially be used to constrain the crustal \( V_p/V_s \) ratio (Lin et al. 2012a,b). However, this is beyond the scope of this study.

Because we ignore crustal reverberation phases, we are unable to determine the three relevant parameters (\( V_p \) in the crust, crustal thickness and \( V_p/V_s \) ratio) simultaneously. To demonstrate this result, we add an extra degree of freedom to the model space, the \( V_p/V_s \) ratio in the crystalline crust, which we allow to vary between 1.65 and 1.85 during the inversion. First, the posterior distribution of the \( V_p/V_s \) ratio as well as its prior distribution are plotted in Fig. 21(a). The posterior distribution of the \( V_p/V_s \) ratio does not centre at any particular value. Secondly, once the \( V_p/V_s \) ratio is introduced as a variable in the inversion, the posterior distribution of the crustal thickness broadens (Fig. 21b) about 20 per cent compared with the same posterior distribution from the inversion with fixed \( V_p/V_s \) ratio (Fig. 12c) from 1.4 to ∼1.7 km. Finally, we observe a strong trade-off between crustal thickness and the \( V_p/V_s \) ratio (Fig. 21c), which indicates that given a higher \( V_p/V_s \) ratio the estimated crustal thickness will be lower and vice versa.

The fact that we are unable to determine a preferred \( V_p/V_s \) ratio from our data at station R11A also holds for other stations. Figs 21(d)–(f) shows the posterior distributions of \( V_p/V_s \) ratio at stations T18A in CP, Q22A in RM and O25A in GP, respectively. Overall, we poorly constrain the \( V_p/V_s \) ratio from the data we are using. The direct effect is that when we present the 3-D model in Section 7.2, the uncertainty of crustal thickness is underestimated by up to ∼20 per cent due to the fact that the \( V_p/V_s \) ratio is set to 1.75. To improve the determination of the \( V_p/V_s \) ratio, other information such as reverberation phases arriving after 10 s in the RFs and Rayleigh
wave ellipticity and local amplification (Lin et al. 2012a,b) would have to be included in the Monte Carlo sampling.

8.2 Limitations of the current method and potential refinements

The procedure that we have developed is intended to provide a practical method to invert surface wave dispersion and RFs jointly over large areas. As it currently exists, the method can be applied fruitfully across large arrays that have been developed around the globe. This includes all of the USArray in the United States, as well as the Chinese Earthquake Array, the Virtual European Seismic Network, various PASSCAL experiments around the world and F-net in Japan. However, the method also serves as a framework for future enhancements and improvements. In this regard we highlight four known limitations with the method as it is currently effected and four that may call for modification. A fifth limitation is discussed in Section 8.1 regarding a variable that have to be included in the Monte Carlo sampling.

First, the parametrization that we have defined is applied rigidly in the inversion. The algorithm has not been designed to sense misfit to the data and to adapt the parametrization accordingly, although other joint inversion algorithms include this feature (e.g. Bodin et al. 2011). For instance, in the inversion at station R11A (Fig. 13), a mid-crustal discontinuity was not introduced automatically to fit the negative arrival at ∼1 s on the RF. An adaptive parametrization would help the algorithm fit aspects of the data that are currently being ignored, but introducing more structure would increase uncertainties and may lead to overinterpretation of the data.

Secondly, the traveltime variation of P-to-S conversions due to dipping interfaces (notably on the Moho) is not fit in our inversion and is removed in the harmonic stripping algorithm to estimate the azimuthally independent RF that we use in the joint inversion, \( A_0(t) \). Not modelling the delay time variation explicitly means that the amplitude of P-to-S conversions will be underestimated if the variation is significant (i.e. \( \geq 0.5 \) s). The delay time variation does appear in the \( A_1(t) \) and \( A_2(t) \) components of the RF that are estimated in the harmonic stripping algorithm but which are not used in the inversion. Optimally, the algorithm would employ information that exists in these two components to correct for the underestimation of the amplitude of the phase conversions from dipping interfaces. In principle, this information as well as transverse component RF amplitudes and delay time variations could also be used to estimate the dip on the interfaces, but this information would probably have to be interpreted independently from surface wave data.

Thirdly, we do not use the RF produced from a multistation imaging technique [e.g. the Common Conversion Point (CCP) stacked RF] but rather a single-station RF which is an average of the 3-D variation near each single station. In the future, this procedure could be replaced with the RF stacked at the CCP rather than at the station. This procedure is inappropriate for the TA because the CCPs beneath nearby stations do not overlap, as Fig. 22(a) illustrates. The procedure would be appropriate for a denser array such as PASSAL or EarthScope Flexible Array experiments. Fig. 22(b) also presents a comparison between RFs computed using the TA and the CD-ROM PASSCAL experiment. The blue RFs are from three TA stations along the green transect in Fig. 22(a), and the red RFs are binned CD-ROM+TA RFs produced according to the Moho Conversion Points (MCP), where radial RFs that pierce the same bin at Moho depth are averaged laterally. The similarity between MCP RFs and the harmonic stripping \( A_0(t) \) RFs demonstrate that the single station azimuthally independent RFs that we construct are consistent with the multistation RFs. A smoothed image of the multistation RFs is presented in Fig. 22(c), demonstrating the resolution that RFs from a tighter network geometry provide. Higher resolution features such as the mid-crustal signal at ∼2.5 s beneath the Cheyenne Belt (CB) are not captured by the sparse TA array. However, using higher resolution RFs would necessitate accommodating the lower resolution of surface wave dispersion maps so as not to alias sharp features from the RFs into larger scale features derived in the joint inversion. It remains unclear what maximum resolution may be possible for surface waves from a tighter array spacing than the TA, for example from the Flexible Array component of USArray. It is likely, however, that surface wave resolution will always lag resolution from RFs. Although the installation of the TA has ameliorated the differential

![Figure 21](image_url)
Joint Bayesian Monte-Carlo inversion

9 CONCLUSIONS

We present a new method for joint inversion of surface wave dispersion data and RFs based on a Bayesian Monte Carlo scheme. When applied to RFs and surface wave data from ambient noise and earthquakes that are now emerging from extended broad-band seismic arrays, the method produces a 3-D model of the crust and uppermost mantle to a depth of about 150 km with associated uncertainties. The method is designed to be used in an automated fashion across a large number of stations and has been applied to data from 185 USAArray TA stations in a geologically diverse part of the Intermountain West. The effect of the introduction of RFs to surface wave dispersion data is visualized through improvements in the posterior marginal distribution of model variables. By comparing the statistics of the posterior distributions, we find that adding RF data quantitatively improves the accuracy of estimates of Moho depth, improves the determination of the Vsv contrast across Moho and improves uppermost mantle structure. Knowledge of uppermost mantle structure is improved because the assimilation of RF data makes it possible to relax ad-hoc structural constraints that are commonly invoked in inversions based on surface wave data alone.

Although the inversion method we describe can be applied robustly across large regions, there remain aspects of the method where refinements may prove beneficial. Three are particularly noteworthy: (1) development of an adaptive parametrization, particularly in the crust where low velocity layers may exist, (2) further investigation of the estimation of the full data covariance matrix and its inverse for both RFs (covariance over time) as was done by Bodin et al. (2011) and surface wave dispersion (covariance over frequency), (3) extension of the RFs past 10 s to recover reverberations that may help to constrain the $V_p/V_s$ ratio. These and other potential refinements to the method may reduce small biases that derive from current assumptions, but the current method produces results that are preferable to the use of surface wave dispersion or receiver function data alone.

ACKNOWLEDGMENTS

The authors gratefully acknowledge insightful reviews from Thomas Bodin, Malcolm Sambridge, an anonymous reviewer and the Associate Editor, Gabi Laske, that helped to improve this paper. They are also grateful to Craig Jones for insights into receiver function analyses and to Anne Sheehan for discussions concerning the history of joint inversions with receiver functions and surface wave dispersion. The facilities of the IRIS Data Management System, and specifically the IRIS Data Management Center, were used to access the waveform and metadata required in this study. The IRIS DMS is funded through the National Science Foundation and specifically the GEO Directorate through the Instrumentation and Facilities Program of the National Science Foundation under Cooperative Agreement EAR-0552316. This research was supported by NSF grants EAR-0711526, EAR-0844097, EAR-0750035 and EAR-1053291 at the University of Colorado at Boulder. F.-C. Lin is supported by the Director’s Post-Doctoral Fellowship of the Seismological Laboratory at the California Institute of Technology.

REFERENCES


### APPENDIX: CONSTRUCTING AZIMUTHALLY INDEPENDENT RECEIVER FUNCTIONS BY ‘HARMONIC STRIPPING’

#### A1 Receiver function data collection

We apply the time-domain iterative deconvolution method (Ligorria & Ammon 1999) to a time window between 20 s before and 30 s after the predicted P-wave arrival to calculate radial component receiver functions and filtered to produce a pulse width of approximately 1 s. P-waves at different distances have different incidence angles that affect both the timing and the amplitude of the P-to-S converted phase (Levin & Park 1997; Bostock 1998; Jones & Phinney 1998). We make corrections to the receiver functions in both time and amplitude by normalizing to a reference slowness of 0.06 deg s−1. The correction is based on a two-layer model with a crustal layer (Vs = 3.7 km s−1) of 40 km thickness and a half-space mantle layer (Vs = 4.3 km s−1). Because this correction is designed to normalize the P-to-S conversion at the Moho discontinuity and reverberated phases have a different time and amplitude dependence on incidence angle, we discard the portion of the receiver function beyond 10 s after the direct P-signal. Any shallow sedimentary multiples that are present in the receiver functions are still retained in this time-series, but they are stacked down because of their different moveout behaviour. After application of this correction, remaining variations in receiver functions over azimuth are mainly due to systematic effects such as dipping interfaces or anisotropy beneath the receiver, structural heterogeneity, scattering (Abers 1998) and random noise in the seismograms. An example of quality controlled receiver functions observed at station R11A in the Basin and Range region is shown in Fig. A1(a), where the receiver functions are arrayed according to event backazimuth.

#### A2 Harmonic stripping

Harmonic analysis of receiver functions is designed to analyse the azimuthal dependence of receiver function arrivals such that azimuthally smooth structural affects are detected and removed, notably potential dipping interfaces (e.g. Savage 1998; Bianchi et al.)
Figure A1. (a) Raw receiver function (RF) waveforms for station R11A in the Basin and Range are plotted as a function of backazimuth, where 6 s time is identified with the dashed line. (b) Red dots are RF amplitude measured at 6 s time with azimuthally binned amplitudes plotted with black error-bars. The harmonic fitting result is shown with the blue curve and the amplitudes of the estimated coefficients $A_0$, $A_1$ and $A_2$ (eq. (A1)) are shown. (c) The azimuthally independent $A_0(t)$ component of the RF that results from harmonic stripping is plotted (red curve) with the uncertainty range (grey corridor), which is the rms of the residual remaining after the harmonic fitting (eq. (A2)).

2010) and azimuthal anisotropy (e.g. Girardin & Farra 1998). This is possible because both dipping interfaces and anisotropy generate sinusoidal variations as a function of backazimuth with different azimuthal periodicity (Jones & Phinney 1998). Decomposition into azimuthal components isolates contributions from horizontal symmetry axis anisotropy (two amplitude peaks and troughs over the complete backazimuthal range) and a much stronger plunging symmetry axis anisotropy signal or dipping isotropic interfaces (one peak and trough over backazimuth). The azimuthally independent component of the resulting harmonic function represents an isotropic, horizontally layered average of structure that is relatively unbiased by uneven and incomplete backazimuthal sampling, and is useful for joint inversion with surface waves:

$$H(\theta, t) = A_0(t) + A_1(t) \sin[\theta + \theta_1(t)] + A_2(t) \sin[2\theta + \theta_2(t)],$$  \hspace{1cm} (A1)

where the $A_i$ ($i = 0,1,2$) are amplitudes of the three harmonic components, and the $\theta_i$ are initial phases for the azimuthally dependent components. An example of the amplitude versus backazimuth for the receiver functions observed at TA station R11A at 6 s is shown in Fig. A1(b).

Under realistic circumstances where dipping interfaces and anisotropy beneath the receiver frequently exist, $A_1$ and $A_2$ will not be zero whereas $A_0$ will be the azimuthally independent receiver function (Fig. A1c). Here, $A_0(t)$ is treated as the receiver function that is sensitive to the average horizontally layered isotropic structure beneath the receiver and is, therefore, comparable to surface wave dispersion data. We refer to the method by which the azimuthally independent receiver function, $A_0(t)$, is estimated by fitting and removing the harmonic function $H(\theta, t)$ from the raw receiver functions as ‘harmonic stripping’.

To estimate the uncertainty $s(t)$ in $A_0(t)$ we use the RMS difference over azimuth between the observed receiver functions and the harmonic function defined by eq. (A1),

$$s(t) = \{N^{-1}\sum_{i=1}^{N} [R_i(\theta, t) - H(\theta, t)]^2 \}^{1/2}. \hspace{1cm} (A2)$$

Here, $R_i(\theta, t)$ represents an observed quality-controlled (see below) receiver function at discrete azimuth $\theta_i$ for event $i$ and $N$ is the number of such receiver functions. This residual is a measure of the difference between the red dots and the fit line in Fig. A1(b), averaged over azimuth. Fig. 4(c) shows the difference between the observed receiver functions and the harmonic function for station R11A. The uncertainty estimated in this way is presented in Fig. A1(c) as a one standard deviation corridor about $A_0(t)$. Examples of receiver functions and uncertainties are presented for other stations in Fig. 5.

The RMS of the residuals over time for the receiver functions in Fig. A1(c) is $\sim 0.035$, which is about twice the estimated noise in receiver functions ($\sim 0.015$) found by Bodin et al. (2012), and is also about twice the RMS of the misfit between the observed receiver function and the receiver function from the best fitting model. We believe, therefore, that this definition of the uncertainty in receiver functions overestimates the uncertainty of $A_0(t)$ that is produced from harmonic stripping. To compensate for this overestimation of the uncertainties, we make two corrections. First, we reduce uncertainties in the azimuthally independent receiver function between 3 and 8 s by a factor of two to emphasize the fit to the $P$-to-$S$
converted signal in the receiver functions. Subsequent plots of receiver function uncertainties possess this reduction (Figs 5, 13 and 16). Secondly, we introduce a factor \( \kappa \) in the Monte Carlo inversion that acts to normalize the misfit to the surface wave and receiver function data as seen in eq. (5). The determination of the appropriate value for \( \kappa \) is discussed in Section 6.1.

### A3 Quality control (QC)

The observed azimuthally dependent receiver functions, \( R_i(\theta, t) \), that go into the harmonic stripping procedure must first be quality controlled. We build on the quality control applied to the Earthscope Automated Receiver Study (EARS) database and seek only receiver functions that are consistent with the automatically QC’ed receiver functions from EARS used in their stacking procedure that are available through IRIS (Crotwell & Owens 2005). The EARS data selection criterion is quite conservative (variance reduction in the iterative deconvolution of at least 80 per cent) so that significant azimuthal gaps are left in the EARS receiver function station sets (e.g. Fig. A2b). We seek to expand the receiver function database relative to the EARS database, but retain the quality of the constitutive receiver functions in the EARS database. To control the quality of the raw receiver functions (e.g. Fig. A2a for station R11A) we, first, perform harmonic stripping using the EARS data alone and get a preliminary estimate of the harmonic function \( \hat{H}(\theta, t) \). Then we consider each of the raw receiver functions in the database, filtered identically with those in the EARS database (the time windows used in the deconvolution differ somewhat between the EARS and our database, but this has little effect on the resulting receiver functions). Those that are similar to the preliminary estimate \( \hat{H}(\theta, t) \) from EARS are retained, whereas those that are dissimilar are rejected. There are numerous ways to define similarity that would yield similar results. We define the difference functional between each raw receiver function at azimuth \( \theta_i \) and the preliminary estimate of the harmonic fit from EARS as follows:

\[
D(\theta_i) = \sqrt{\frac{1}{M-1} \sum_{j=1}^{M} \left[ R_i(\theta_i, t_j) - \hat{H}(\theta_i, t_j) \right]^2},
\]

where \( M \) is the number of discrete time points in the receiver functions. We find that if we retain receiver functions where \( D(\theta_i) < 0.05 \), then we retain many more receiver functions than in the EARS database but at the same time do not greatly increase the noise level of the database. Fig. A2(c) presents the results of this QC procedure for TA station R11A. The resulting database has been expanded by about a factor of two relative to EARS, but most of the noisy raw receiver functions have been discarded. We refer to this as the QC’ed receiver function database and it is used to estimate the final harmonic fit and azimuthally independent receiver function \( A_0(t) \) with associated uncertainties \( s(t) \).
A4 Synthetic test of harmonic stripping

The goal of harmonic stripping is to remove the azimuthal periodicity of receiver function waveforms due to dipping interfaces, azimuthal anisotropy, or other structures beneath the receiver. Synthetic tests are presented here to illuminate the capabilities of harmonic stripping to remove structural effects and recover the azimuthally independent receiver function for a horizontally layered, isotropic effective medium.

First, we test the affects of structures that produce azimuthally smooth signals in the receiver functions; namely (1) a dipping Moho interface and (2) crustal azimuthal anisotropy, and show that the effects of these types of structures are dealt with effectively by harmonic stripping.

1) We define a horizontally layered, isotropic model $M_0$ without a dipping Moho or anisotropy in Table A1. The model, $M_d$, is the same as $M_0$ but with Moho tilted by 10° with the up-dip azimuth of 165° relative to north. Synthetic seismograms for every 10° in azimuth are computed (Frederiksen & Bostock 2000), noise is added to the waveform before the time-domain deconvolution is performed to generate Ps receiver functions, $R(\theta, t)$, which are plotted in Fig. A3(a). The dipping Moho creates a sinusoidal variation to the $P$–$S$ conversions at ~5.5 s in both amplitude and arrival time.

Harmonic stripping is applied to the synthetic receiver functions and the resulting estimates $A_0, A_1, A_2$ are shown in Figs A3(d)–(f). As described by Jones & Phinney (1998), the dipping interface generates a $2\pi$ periodicity with the strongest amplitude at the updip azimuth and the arrival time will be delayed most at this azimuth. This phenomenon is observed in the harmonic stripping results. $A_1$ shows the strongest perturbation at 5–6 s where the $P$–$S$ conversion occurs and $A_2$ is subtle in this time range. Fig. A3(b) and (c) show the harmonic fit, $H(\theta, t)$, and the difference $R$ and $H$, respectively. The random noise added to the seismograms is not mapped into the $A_1$, or $A_2$ estimates, but resides exclusively in the residuals, which we take as the uncertainty of the final azimuthally independent receiver function. For comparison, the receiver functions computed from the horizontally layered model $M_0$ (blue waveform) and the estimate of $A_0$ (red dashed line) are shown in Fig. A3(a). The arrival times of the $P$–$S$ conversion from both are similar, but the $A_0$ component recovered from the dipping model (red dashed line) displays a somewhat smaller amplitude because the harmonic stripping method does not fully reproduce the sinusoidal variation in arrival time of the input model (blue line). This inaccuracy, however, is encompassed by the enlargement of the uncertainties near the $P$–$S$ conversion time as discussed further in Section 8.2.

2) We also test an azimuthally anisotropic model, $M_{anis}$, in which a 4 per cent azimuthally anisotropic layer between 15 and 45 km depth range with a fast direction at 75° to north is added to $M_0$. The resulting synthetic receiver functions contaminated by random noise and the results of harmonic stripping are shown in Fig. A4. Azimuthal anisotropy introduces a $\pi$ periodicity into the arrival

### Table A1. Horizontally layered isotropic model $M_0$.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Vs (m s$^{-1}$)</th>
<th>Vp (m s$^{-1}$)</th>
<th>Density (kg m$^{-3}$)</th>
<th>Thickness (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3500</td>
<td>6125</td>
<td>2746</td>
<td>45</td>
</tr>
<tr>
<td>2</td>
<td>4300</td>
<td>7525</td>
<td>3350</td>
<td>~</td>
</tr>
</tbody>
</table>

---

Figure A3. Synthetic experiment performed with Moho dipping 10° with an up-dip azimuth of 165° clockwise from north. (a) The synthetic receiver functions $R(\theta, t)$ including additive noise plotted at the appropriate backazimuth. (b) The estimated receiver functions from harmonic stripping $H(\theta, t)$. (c) The difference between (a) and (b). (d and e) The estimated $A_0, A_1, A_2$ components from harmonic stripping, respectively.
Figure A4. The same as Fig. A3, but for synthetic receiver functions computed from a model with 4 per cent azimuthal anisotropy in the mid-to-lower crust with a fast-axis direction at 75° clockwise from North.

Figure A5. The same as Fig. A3, but for synthetic receiver functions from a model with a 5 km step of crustal thickness such that Moho is at 40 km depth east of the station and 45 km depth west of the station.
Figure A6. The same as Fig. A3, but for synthetic receiver functions from a model with a 2 km jump of sedimentary layer thickness such that there are no sediments east of the station and 2 km of sediments west of the station.

Figure A7. (a) The azimuthally independent $A_0$ receiver function estimated from harmonic stripping computed from a model with the dipping Moho is shown with the red dashed line. For comparison, the receiver function from the azimuthally averaged model (horizontal Moho) is shown with the blue line. Uncertainties $s(t)$ are outlined by red lines with a light-red fill. (b) The same as (a), but for the $A_0$ receiver function estimated from synthetic receiver functions computed using a model with 4% anisotropy in the mid-to-lower crust. (c) The same as (a), but for the $A_0$ receiver function estimated from a model with a step in Moho depth of 5 km at the station. (d) The same as (a), but for a model with a 2 km step in sediment thickness at the station.

From these tests for structures that vary smoothly with azimuth, a dipping Moho or azimuthal anisotropy in the crust, we conclude that the harmonic stripping technique effectively retrieves the azimuthally independent component which is an accurate estimate of times of the $P$–$S$ conversions (Fig. A4a), which appears in the estimated $A_2$ component (Fig. A4f). The estimated $A_0$ component is shown in Fig. A7(b) to agree with the receiver function computed from model $M_0$. 
the receiver function for an effective horizontally layered, isotropic medium.

Secondly, we test another type of model that possesses azimuthally discontinuous features: (3) a jump in Moho with azimuth and (4) a jump in sediment thickness with azimuth. We show that for azimuthally discontinuous structures, harmonic stripping does not reliably estimate the horizontally layered, isotropic effective medium. The method identifies this problem, however, by enlarging the uncertainties in the time band in which the estimated receiver function diverges from the receiver function of the effective medium.

(3) On the west side of the receiver (backazimuth angle between 180° and 360°) the model is the isotropic horizontally layered model $M_0$ with a Moho depth of 45 km, but on the east side Moho is 5 km shallower. Synthetic receiver functions from this model with additive noise are plotted in Fig. A5(a), and the harmonic stripping results are shown in Figs A5(d) and (f). The variations of the receiver functions with backazimuth are not sinusoidal; thus harmonic stripping cannot fit them fully. As a result, the estimated $A_0$ component deviates from the receiver function of the average model (a model with a Moho depth of 42.5 km), which is shown as the blue waveform in Fig. A7(c). However, uncertainties in the $A_0$ component grow in the time window where $A_0$ differs from the receiver function of the average model. Thus, in this time band the receiver function will be down-weighted relative to surface wave data, as desired.

(4) Finally, we test of the model with a step in sediment thickness (Fig. A6). The model has a 2 km sedimentary layer west of the receiver and no sediments on its east side. Both models have the same Moho depth. The average model, therefore, has 1 km of sediments. The synthetic receiver functions in Fig. A6(a) display a dichotomous pattern such that the Moho $P$–$S$ conversion arrival times jumps near 180° azimuth. The estimated $A_0$ component of the receiver function (Fig. A7d) shows an erroneous double Moho conversion peak as well as relatively large uncertainties that reduce the weight of the receiver function relative to surface wave data in the inversion.