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

Weisen Shen[1], Michael H. Ritzwoller[1], Vera Schulte-Pelkum[2] and Fan-Chi Lin[3]

1 – Center for Imaging the Earth's Interior, Department of Physics, University of Colorado at Boulder, Boulder, CO 80309, USA (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

Abstract

A non-linear Bayesian Monte-Carlo method is presented to estimate a Vs model beneath stations by jointly interpreting surface wave dispersion and receiver functions and associated uncertainties, which is designed for automated application to large arrays of broadband seismometers. As a proving ground 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 sec to 80 sec across the study region. A harmonic stripping method is applied as a basis for quality control and to generate back-azimuth independent receiver functions with uncertainty estimates for each station. A smooth parameterization 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.

Key words: Inverse theory, Seismic tomography, Surface waves and free oscillations, Composition of the continental crust

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 (Vs) models at regional (e.g., Levin *et al*., 2002; Levshin *et al*., 2005; Ritzwoller *et al*., 2005; Yang and Forsyth, 2006; Shapiro *et al*., 2008; Yang *et al*., 2011), continental (e.g., Ritzwoller *et al.*, 2001, Villasenor *et al*., 2001; Bensen *et al.*, 2009) and global (e.g., Shapiro and 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 (Langston, 1979). RFs also have been applied to construct local 1-D (Ammon and Zandt, 1993) and 3-D models (Vinnik *et al.*, 2004, Vinnik *et al.*, 2006). The use of each type of data alone presents significant non-uniqueness problems (e.g., Ammon et al., 1990; Shapiro *et al*., 2002): surface waves do not image discontinuities whereas RFs do not strongly constrain shear velocities between discontinuities. 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 since been developed. Linearized joints inversions have been particularly popular (e.g., Du and Foulger, 1999; Julia *et al.*, 2000, 2003; Endrun *et al.*, 2004, Horspool *et al.*, 2006, Gok *et al.*, 2007, Tkalcic *et al.*, 2006; 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 and Tarantola, 1995; Sambridge, 2001; Mosegaard and Sambridge, 2002; Sambridge and Mosegaard, 2002) also have been applied in both receiver function inversions (e.g., Shibutani *et al.*, 1996; Zhao *et al.*, 1996; Levin and Park, 1997; Sambridge 1999a; Clitheroe *et al*., 2000; Piana Agostinetti *et al.*, 2002; Bannister *et al*., 2003; Fredericksen *et al*., 2003; Nicholson *et al*., 2003; Vinnik *et al*., 2004, 2006; Lucente *et al*., 2005; Hetenyi and Bus, 2007; Piana Agostinetti and Chiarabba, 2008; Piana Agostinetti and Malinverno, 2010) and surface wave inversions (e.g., Shapiro *et al*., 2002; Socco

and Boiero, 2008; Maraschini and Foti, 2010; Molnar *et al*., 2010; Khan *et al*., 2011) with varying degrees of theoretical sophistication, generality, spatial extent, and data quantity. Joint inversions of receiver functions 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 and Assumpcao, 2004; Chang *et al*., 2004; Lawrence and Wiens, 2004; Vinnik *et al.*, 2006; Liu *et al.*, 2010; Moorkamp *et al*., 2010; Tokam *et al*., 2010; Basuyau and Tiberi, 2011; Bodin *et al*., 2011).

Despite these advances, the joint interpretation of surface wave dispersion and receiver functions is still faced with at least four significant challenges. First, as traditionally applied, the techniques have yielded information at different lateral resolutions. Second, also traditionally, surface wave dispersion information has derived from teleseismic earthquakes and has been produced mostly at long periods (> 20 sec), which are not ideally sensitive to the earth's crust. Third, joint quantitative interpretation requires meaningful uncertainties for both data types. Fourth, the uncertainties in the resulting model must be quantified. Inextricably related to this is the need to find a model parameterization 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 receiver functions that addresses these challenges.

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 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 and Snoke, 2010), eikonal tomography (Lin *et al.*, 2009) and Helmholtz tomography (Lin and 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 sec) surface wave information that agrees with information from teleseismic earthquakes in the period band of overlap (25-40 sec) (e.g., Yang and 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., Bensen *et al*., 2009; Lin *et al*., 2011; Moschetti *et al*., 2010a,b; Yang *et al*., 2008a,b; Yang *et al*., 2012; 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 here a method to construct receiver functions based on azimuthal harmonic analysis (Girardin and Farra,1998) that we use to estimate uncertainties for receiver functions. Bodin *et al*. (2011) discuss the importance of reliable uncertainty information in the joint inversion insightfully and at substantial length. Finally, the fourth challenge, estimating model uncertainties, is addressed by a Bayesian Monte Carlo method that we develop and describe here. We show along the way that a quite simple model parameterization based on smooth B-splines in the crust and mantle can fit both data types at most locations across the western US.

Therefore, we present here a non-linear Bayesian Monte-Carlo algorithm to estimate a Vs model beneath stations by jointly interpreting surface wave dispersion and receiver functions 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 sec to 80 sec across the study region. Eikonal and Helmholtz tomography are applied to estimate uncertainties in ambient noise and earthquake dispersion information and harmonic stripping method is used to generate back-azimuth 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 3D 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 new

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 posterior distribution $\sigma(m)$ for a model m to be computed from the prior information on model space (given by the prior probability density $\rho(m)$ for model m) and the observed data as represented by the likelihood functional $L(m)$ (the probability of observing the measured data given a particular model):

$$\sigma(m) \propto \rho(m)L(m) \qquad (1)$$

Geophysical applications of Bayesian inference have been presented by Tarantola and Valette (1982), Mosegaard and Tarantola (1995), and Sambridge (1999b). The key steps are the expression of prior information as a probability density function and the computation of the likelihood functional, which is determined from data misfit. From a set of models distributed like the posterior distribution, one can then determine certain properties of the ensemble (e.g., best fitting model, mean model, median model), the covariance between model parameters, or compute the marginal distribution of particular model parameters or quantities derived from the model parameters.

The procedure we have developed to jointly invert surface wave dispersion information and receiver functions is similar in motivation and even is some of the details to the methods described by Bodin *et al*. (2011). We present a general overview of our methodology here as a guidepost to the principal results in the paper. We will also point out some differences with the method of Bodin *et al*. (2011). The data used in the inversion are described in detail in section 2.

The model parameterization and the construction of the prior distribution are presented in section 3. The prior information includes ranges in which individual model parameters are allowed to vary as well as rules that govern the relative values of different model parameters. Examples of the prior distribution for several model parameters are shown in Figure 9.

The likelihood functional is simply related to the misfit function $S(m)$ as follows:

$$L(m) = \exp\left(-\frac{1}{2}S(m)\right) \tag{2}$$

where

$$S(m) = \left(g(m) - D^{obs}\right)^T C_e^{-1}\left(g(m) - D^{obs}\right) \tag{3}$$

and D^{obs} is a vector of measured data, g*(m)* is the vector of data predicted from model m, C_e is the data covariance matrix, and T represents transposition of a vector. For surface wave dispersion here the vector D^{obs} consists of Rayleigh wave phase velocities observed on a discrete set period grid and for receiver functions D^{obs} is the azimuthally independent receiver function over a discrete time grid.

The Markov Chain Monte Carlo sampling of model space to generate the posterior distribution evolves as follows. A model m_j is selected from the prior distribution. A second model m_i is drawn and the likelihoods $L(m_i)$ and $L(m_j)$ are computed. The Metropolis law defines the probability of acceptance for model m_i:

$$P_{accept} = \begin{cases} 1 & \text{if } L(m_i) \geq L(m_j) \\ L(m_i)/L(m_j) & \text{if } L(m_i) < L(m_j) \end{cases} \tag{4}$$

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 $P_{accept} = 0.4$, 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 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 Figure 10 and for the joint inversion in Figure 13.

An additional model acceptance criterion then is applied to define the posterior distribution. If the data covariance matrix were known accurately, the choice of the acceptance criterion would be straightforward. Misfits to surface wave and receiver function 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 receiver functions have error processes that are independently but not identically distributed so that each covariance matrix is diagonal with elements σ_i^2 and s_j^2, respectively, for phase velocity at period i and receiver function 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 κ to normalize the misfits between the two data types, as described in section 5.1. The resulting joint misfit function, therefore, has the following form:

$$S_{\text{joint}}(m) = S_{SW} + \frac{1}{\kappa} S_{RF} = \sum_{i=1}^{N} \frac{\left(g_i(m) - D_i^{obs}\right)^2}{\sigma_i^2} + \frac{1}{\kappa} \sum_{j=1}^{M} \frac{\left(R_j(m) - A_0(t_j)\right)^2}{s_j^2} \tag{5}$$

where $g_i(m)$ is the phase velocity predicted for model m at period i on a discrete grid of N periods, D^{obs} is the observed phase velocity, the function $R_j(m)$ is the observed receiver function and $A_0(t_j)$ is the predicted receiver function for model m at time t_j on a grid of M times. Somewhat different model acceptance criteria are applied in the inversion with surface wave data alone in section 4 and in the joint inversion in section 5. Example posterior marginal distributions are shown in Figures 11 and 14.

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 parameterization we choose is sufficiently flexible to fit the data. Second, as discussed above, we do not estimate the off-diagonal elements of the covariance matrix. Inspired by the hierarchical Bayesian formalism of Bodin et al., we did attempt to estimate full covariance matrices for RFs. We found, however, that the covariance matrices that we derived were appropriate for a raw receiver function not the estimated azimuthally independent receiver function $A_0(t)$ that we use in the inversion. This is a subtle point that requires further analysis, thus we forged forward with diagonal covariance matrices. Although these differences with the more general formulation of Bodin *et al*. (2011)

were motivated by the difficulty in effecting a data adaptive parameterization and 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 3D 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 ambient noise tomography 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, Yang *et al.*, 2007, Moschetti *et al.*, 2007, Yao *et al.*, 2006) . 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 TA stations in the western US and are cut into 1-day time series. The time series length of each station is usually about two years, but time series lengths are variable and we do not apply a minimum time series length restriction. Second, 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 sec time window is used to weight the data point at the center of the window. The weights are determined in the "earthquake band" between periods of 15 and 50 sec period, but then are applied to the unfiltered data. Third, the data are band-pass filtered between 5 and 100 sec 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 inter-station 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 inter-station distance must be greater than three wavelengths at each period to ensure the far-field approximation and sufficient separation from precursory 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 sec period. Eikonal tomography takes account of ray bending (off-great circle propagation) but does not model finite frequency effects (Lin and Ritzwoller, 2011) . One of the signature features of eikonal tomography is that it produces meaningful uncertainty estimates at each geographical location. Figure 2 (a-c) presents phase velocity maps from ambient noise tomography at periods of 8, 20 and 36 sec. By 40 sec period, the SNR of ambient noise has decreased enough that high resolution phase velocity maps are no longer generated. Earthquake data are brought in 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 and Ritzwoller, 2011) and are only briefly summarized here. First, teleseismic records at the TA stations in the western US following earthquakes with surface wave magnitudes Ms ≥ 5.0 are downloaded from the IRIS DMC and cut according to the arrival window of surface waves, before instrument responses are removed. Second, FTAN is applied to the cut seismograms and Rayleigh wavefront group and phase travel times and amplitudes are measured as a function of period. Phase ambiguity is resolved using a network based approach in which travel time measurements are compared with measurements at nearby stations using the method of Lin and Ritzwoller (2011). Then Helmholtz tomography is performed to produce Rayleigh wave phase velocity maps from 32 to 80 sec period. In Helmholtz tomography, finite frequency effects are corrected through a term involving the local Laplacian of the amplitude field. Figure 2d,e shows the phase velocity maps from ET at periods of 36 and 70 sec, respectively. A comparison of the 36 sec phase velocity maps obtained from ANT and ET is shown in Figure 2f. 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° x 0.2° grid from both ANT and ET, phase velocity curves are interpolated to each station location. At short periods (8-36 sec), phase velocity measurements and associated uncertainties are taken exclusively from ANT maps. At long periods (40-80 sec) results are taken solely from ET maps. Between periods of 36 sec and 40 sec, we weight the two velocity measurements by their local uncertainties and a smooth

curve is produced. Figure 3 shows four dispersion curves and uncertainty estimates for TA stations R11A, T18A, Q22A and O25A whose locations are shown on Figure 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 Receiver function data processing

3.2.1 Receiver function data collection

In this study we use P-wave receiver functions (teleseismic P-to-S converted waves) exclusively, but S-wave receiver functions could also be used in the context of the methods we develop, in principle (Vinnik *et al.*, 2004) . We select all teleseismic events in the NEIC PDE catalogue with a body wave magnitude of $m_b \geq 5.1$ within the epicentral distance range of 30°-99° from each station. All sets of 3-component seismograms are rotated to form the radial and transverse components and are then band-pass filtered between 0.03 and 4 Hz using a Butterworth filter (2 poles, 2 passes). We apply the time-domain iterative deconvolution method (Ligorria and Ammon, 1999) to a time window between 20 sec before and 30 sec after the predicted P-wave arrival to calculate radial component receiver functions filtered to result in a pulse width of approximately 1 sec.

Receiver functions from events at different distances to a receiver have a P-wave incidence angle dependence in both timing and amplitude for the P-to-S converted phase (Jones and Phinney, 1998; Levin and Park, 1997; Bostock, 1998). We make corrections to the RFs in both time and amplitude by normalizing to a reference slowness of 0.06 deg/sec. 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 sec after the direct P-signal. After application of the move-out/amplitude correction, effects from slowness variations on the P-to-S conversion have been minimized and the variations in receiver functions over azimuth are mainly due to systematic errors 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 Rocky Mountain region is shown in Figure 4a, where the RFs are arrayed according to event back-azimuth.

3.2.2 Harmonic stripping

Harmonic analysis of receiver functions is designed to analyze the azimuthal dependence of receiver function arrivals such that potential dipping interfaces (e.g. Savage, 1998; Bianchi et al., 2010) or anisotropy (e.g. Girardin and Farra, 1998) beneath the receiver can be detected, owing to the fact that both dipping interfaces and anisotropy will generate sinusoidal variations as a function of back-azimuth with different azimuthal periodicity (Jones and Phinney, 1998). Decomposition into azimuthal components isolates contributions from horizontal symmetry axis anisotropy (2 amplitude peaks and troughs over the complete backazimuthal range) and a much stronger plunging symmetry axis anisotropy signal or dipping isotropic interfaces (1 peak and trough over backazimuth); the azimuthally-independent component of the resulting harmonic function represents an isotropic, plane-layered average of structure that is relatively unbiased by uneven and incomplete backazimuthal sampling, and is useful for joint inversion with surface waves (Schulte-Pelkum,2012).

$$H(\theta,t_0) = A_0(t_0) + A_1(t_0)\sin(\theta + \theta_1(t_0)) + A_2(t_0)\sin(2\theta + \theta_2(t_0)) \qquad (6)$$

where the A_i (i=0,1,2) are amplitudes of the three harmonic components, and the θ_i are initial phases for azimuthally dependent components. An example of the amplitude vs. back-azimuth for the RFs observed at TA station R11A at 6 sec is shown in Figure 4b.

Under perfect circumstances (no dipping interface or anisotropy beneath the receiver, all events would have the same slowness for incoming waves), A_1 and A_2 would both be 0 and A_0 would be the amplitude of the receiver function. But 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 form an estimate of the component without variations along back-azimuth (Fig. 4c). Here, A_0 will be treated as the receiver function that is sensitive to the average planar isotropic structure beneath the receiver and is, therefore, comparable to surface wave dispersion data.

To estimate the uncertainty $s(t)$ in $A_0(t)$ we use the RMS difference over azimuth between the observed RFs and the harmonic function given by equation (6):

$$s(t) = \left[N^{-1} \sum_{i=1}^{N} (R_i(\theta_i, t) - H(\theta_i, t))^2 \right]^{1/2} \quad (7)$$

Here, $R_i(\theta_i, t)$ represents a quality-controlled (see below) receiver function at discrete azimuth θ_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 Figure 4b, averaged over azimuth. Figure 6c shows the difference between the observed RFs and the harmonic function for station R11A. The uncertainty estimated in this way is presented in Figure 4c as a one standard deviation corridor about $A_0(t)$. Examples of RFs and uncertainties will be presented for other stations later in the paper.

The RMS of the residuals over time for the RF in Figure 4c is ~ 0.035, which is about twice the estimated RF noise (~ 0.015) found by (Bodin *et al.*, 2012), and is also about twice the RMS of the misfit between the observed RF and the RF from the best fitting model. We believe, therefore, that this definition of the uncertainty in RFs is proper for the raw RF, but over-estimates the uncertainty of $A_0(t)$ which results from harmonic stripping. To compensate for this overestimation of the uncertainties, we do two things. First, we reduce uncertainties in the RF between 3 sec and 8 sec by a factor of two in order 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. 7, 15-18). Second, we introduce a factor (κ) in the Monte Carlo inversion that acts to normalize the misfit to the surface wave and RF data as seen in equation (5). The determination of the appropriate value for κ is discussed in section 6.1

3.2.3 Quality control (QC)

The RFs, $R_i(\theta_i, t)$, that go into the harmonic stripping procedure must first be quality controlled. To do this we build on the quality control applied to the EARS (Earthscope Automated Receiver Study) database and seek only RFs that are consistent with the automatically QC'ed RFs from EARS used in their stacking procedure that are available via IRIS (Crotwell and Owens, 2005). The EARS data selection criterion is quite conservative (variance reduction in the iterative deconvolution (Ligorria and Ammon, 1991) of at least 80%), so that significant azimuthal gaps are left in the EARS receiver function station sets (e.g., Fig. 5b). We seek to expand the RF database relative to the EARS database, but retain the quality of the constitutive RFs in the EARS database. The set of raw RFs at stations R11A is shown in Figure 5a to compare with the EARS RFs.

To control the quality of the raw RFs (Fig. 5a) 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 our database, calculated with the same Gaussian filter parameter (2.5) as 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 RFs). 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 RF at azimuth θ_i and the preliminary estimate of the harmonic fit from EARS as follows:

$$D(\theta_i) = \left\{ M^{-1} \sum_{j=1}^{M} \left[R_i(\theta_i, t_j) - \hat{H}(\theta_i, t_j) \right]^2 \right\}^{1/2} \qquad (8)$$

where M is the number of discrete time points in the RFs. We find that if we retain RFs where $D(\theta_i) < 0.05$, then we retain many more RFs than in the EARS database but at the same time do not greatly increase the noise level of the database. Figure 5c 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 RFs have been discarded. We refer to this as the QC'ed RF database and it is used to estimate the final harmonic fit and receiver function $A_0(t)$ with associated uncertainties $s(t)$.

The result of this RF analysis procedure for station R11A is summarized in Figure 6. The QC'ed RFs presented over azimth are shown in Figure 6a and the harmonic function $H(\theta,t)$ fit to these RFs is shown in Figure 6b for comparison. The residual between these functions, which is used to define the uncertainty in the final azimuthally independent RF $(A_0(t))$, is presented in Figure 6c. The individual harmonic components are shown in Figure 6d-f. The functions $A_1(t)$ and $A_2(t)$ quantify the azimuthally dependent signals that may be produced by tilts on internal interfaces and anisotropy. Some of these arrivals are observed to undergo move-out at different azimuths. These functions could be applied to fine-tune the model presented here, but are not currently used for this purpose.

The quality control procedure described here has been applied to the 185 TA stations in the study region (Fig. 1). Several examples of final receiver functions $A_0(t)$ and associated uncertainties are

shown in Figure 7. The receiver functions display significant 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 sec delay time, suggesting a sharp rather than gradient transition from crustal to mantle velocities. The receiver function for TA station T18A in the Colorado Plateau, in contrast, shows no Moho P-to-S conversion, which means there is a gradient Moho at this point. The receiver function for station O25A in the Great Plains has a strong negative arrival at ~ 4 sec caused by sedimentary layer reverberations at this location which likely interfere with the Moho conversion.

We note that a simple dipping interface does not only create a sinusoidal variation in amplitude (which can be removed by harmonic stripping), but also a sinusoidal variation in the arrival times of the conversion phases from the discontinuity. This sinusoidal variation in move-out is not accurately modeled by harmonic stripping. This means that the amplitude of the conversion phase will be underestimated, on average, in our analysis. This underestimation of the amplitude of converted phases will introduce a bias into the final models generated from the joint inversion of RFs and surface wave dispersion data, as discussed in section 8.2 below.

4. Model Parameterization and Determining the Prior Distribution

4.1 Model parameterization

Because surface wave phase velocity curves and azimuthally averaged receiver functions are sensitive to the local average 1-D Vs structure beneath each location, the inversion for a 3-D model reduces to a set of 1-D Vs inverse problems. Because the longest period of surface waves that we use here is 80 sec, we invert only for shear wave speed in the 200 km beneath the surface. Below 200 km, the model is assumed to be constant (but will only be reliable to about 150 km). Because we use only Rayleigh waves, which are predominantly sensitive to Vsv, we assume an isotropic Vsv model where Vs = Vsh = Vsv. We set the Vp/Vs ratio to 2.0 in the sedimentary layer and 1.75 in the crystalline crust and mantle. The Vp/Vs ratio is important in receiver function analysis and it is discussed further in section 8.1. For density we use the scaling relation advocated by Christensen and Mooney (1995): $\rho = 0.541 + 0.3601Vp$, where ρ is in g/cm^3 and Vp is in km/s. We apply a physical dispersion correction (Kanamori and Anderson, 1977) using the Q model from PREM (Dziewonski & Anderson, 1981), and the resulting model is reduced to 1 sec period.

In a traditional surface wave inversion a 1-D model may be parameterized with smooth functions in the crust and mantle (Megnin and Romanowicz, 2000), as a stack of layers (Yang and Forsyth, 2006), or a combination of both (e.g., Shapiro and Ritzwoller, 2002, Moschetti *et al.*, 2010, Yang *et al.*, 2008). Receiver function inversions are, however, typically parameterized with a set of fine layers (Cassidy, 1992, Julia *et al.*, 2000). We parameterize 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 Vsv 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 Vsv. (3) There is one upper-most mantle layer to a depth of 200 km. Five parameters are used to describe this layer: five B-spline coefficients for Vsv. The thickness of this layer is controlled by the thicknesses of the top two layers.

This model-parameterization contains 13 free parameters (Fig. 8). 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 parameterization, two discontinuities are introduced because that the predominant signal in receiver functions is from the Moho and in of the presence of sediments, the sediment-basement contact. We show here that a smooth parameterization can explain both the surface wave and receiver function 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 have to be introduced. The smoothness of the model is imposed by the parameterization so that ad hoc damping is not needed during the inversion.

With this model parameterization, 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 the reference model, where sedimentary structure is taken from Mooney and Kaban (2010), crustal thickness from Bassin *et al*. (2000), and shear wave speeds in the crust and mantle from Shapiro and Ritzwoller (2002), as stated 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 ±25% of the input value. If the reference thickness were 40 km, values would be

taken 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) Vs < 4.9 km/s 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 receiver functions are identical. We discuss in section 7.3 how this constraint may introduce bias into the estimated model, and how it can be eliminated when receiver functions are introduced. Together, these assumptions are made to reduce model complexity and also to remove 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. Second, 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. This Gaussian distribution is controlled by the width of the distribution (standard deviation), which is given in Table 2. 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 Figure 9 at the location of the TA station R11A. In 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 on the thickness of the sedimentary layer or the thickness of the crystalline crust, the marginal prior distributions of both appear as uniform distributions (Fig. 9a,c). However, the marginal distributions of Vs at different depths are not uniform because different parameters interact through the prior assumptions. The resulting marginal distributions for these variables are more similar to tilted Gaussian patterns because they are affected by prior restrictions to sampling such as the monotonicity constraint on crustal velocities.

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 receiver functions, 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 Figure 3a. At each location, Rayleigh wave phase velocities exist on a fixed discrete grid from 8 to 80 sec period.

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 σ_i, the likelihood functional is computed from equation (2) where the misfit function comes from the first term on the right hand side of equation (5). The assumption of random, independent noise in surface wave dispersion is made here and in other studies (Liu *et al.*, 2010, Bodin *et al.*, 2012), mainly because it is believed that the measurements in the dispersion curve collected from dispersion maps are independent. For simplicity we retain this assumption in this study, but 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 equation (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 the likelihoods $L(m_i)$ and $L(m_j)$ are computed as discussed in section 5.1, where the Metropolis law, equation (4), defines the probability of acceptance for model m_i, as described in section 2.

Figure 10 shows an example of how the likelihood functional and misfit function of the models accepted during sampling of the posterior distribution evolves 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. Figure 10 also shows how the RMS misfit function, $\chi = \sqrt{N^{-1}S(m)}$, evolves. When $\chi < 1$, the fit lies within the estimated uncertainties.

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 $\chi_{\min}$ is defined as the minimum value of χ found for all models visited during model space sampling. Once $\chi_{\min}$ is found, we define the threshold of acceptance χ_{crit} for the surface wave inversion as follows:

$$\chi_{crit} = \begin{cases} 2\chi_{\min} & \text{if } \chi_{\min} \geq 0.5 \\ \chi_{\min} + 0.5 & \text{if } \chi_{\min} < 0.5 \end{cases} \qquad (9)$$

Thus, to define the ensemble of accepted models a posteriori, we accept any model m that is visited during the inversion as long as $\chi(m) \leq \chi_{crit}$. The choice of the critical threshold level below which models are accepted to form members of the posterior distribution is admittedly ad-hoc. Equation (9) 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. Figure 11 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 Figure 9. Typically, the distributions of velocities between boundaries are narrowed. 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 km/sec to ~ 0.05 km/sec. However, posterior distributions of discontinuity depths (Moho, sediments) or velocity jumps across the discontinuities do not decrease appreciably. For example, crustal thickness and Vsv contrast across Moho are not changed strongly. These histograms demonstrate the sensitivity of Rayleigh wave dispersion to Vs structure. (1) Sedimentary thickness does not change because of the lack of very short period 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/sec) 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 Vsv at 10 km in the crust. This demonstrates that the vertical resolution of surface waves to local structure degrades with depth. 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 extent of Vsv at each depth after all accepted models are identified. Figure 12 shows the resulting extent of accepted models for station R11A. The gray corridor outlines the extent of all accepted models that fit the data, while the red lines present the 1σ width around the mean model. Note that without assimilating receiver functions, the crustal thickness (Moho depth referenced to the surface in Fig. 12b) is not well resolved and the predicted receiver functions (red waveforms in Fig. 12c) do not fit the observed $A_0(t)$ receiver function 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 receiver functions to the inversion.

6. Monte Carlo Inversion of Surface Wave Dispersion and Receiver Function Data

When receiver functions are assimilated into the Monte Carlo algorithm, the prior distribution of models does not change (although we discuss in section 7.3 releasing the constraint on the positivity of the uppermost mantle velocity gradient), so it is not introduced again. Here, we introduce the joint likelihood functional and then describe the posterior marginal distribution at TA station R11A and then for three other stations.

6.1 The joint likelihood functional

The joint likelihood functional is defined by equations (2) and (5), but we need to estimate the relative scaling of the misfits for surface waves and receiver functions 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 and the resulting misfits of the two data sets are comparable. The joint misfit function, therefore, becomes:

$$S_{\text{joint}}(m) = S_{SW} + \frac{1}{2.5} S_{RF} = N\chi^2_{SW} + \frac{M}{2.5}\chi^2_{RF} = \sum_{i=1}^{N} \frac{\left(g_i(m) - D_i^{obs}\right)^2}{\sigma_i^2} + \frac{1}{2.5}\sum_{j=1}^{M} \frac{\left(R_j(m) - A_0(t_j)\right)^2}{s_j^2} \quad (10)$$

where we have implicitly defined the χ^2 misfit for surface waves and receiver functions. RMS misfit is the square root of χ^2.

6.2 Determining the posterior distribution in the joint inversion

Figure 13 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 when surface wave data are used alone (Fig. 10). As with the surface wave inversion, we repeat the procedure starting from different random points to find the range of models that fit the data acceptably, but repeat this procedure ten times. The forward calculation of RFs is based on the widely used Thomson-Haskell matrix method. We use the code developed by Shibutani *et al*. (1996), which has also been used by Sambridge (1999a). We introduce for the joint inversion a somewhat different acceptance 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 $\chi^{SW}_{\min}$ and receiver functions $\chi^{RF}_{\min}$ are identified, although the models that minimize misfits for these two types of data are generally not the same ones. The joint normalized RMS function χ^{joint} is defined as follows:

$$\chi^{\text{joint}} = \frac{1}{2}\left(\frac{\chi^{SW}}{\chi^{SW}_{\min}} + \frac{\chi^{RF}}{\chi^{RF}_{\min}} \right) \tag{11}$$

This function is a combination of RMS misfits for the two data sets. The minimum value of the joint misfit function $\chi^{\text{joint}}_{\min}$ helps to diagnose when discrepancies between the data sets arise. The criterion for model acceptance is $\chi^{\text{joint}} < \chi_{crit}$, where χ_{crit} is defined as:

$$\chi_{crit} = \chi^{\text{joint}}_{\min} + 0.5 \tag{12}$$

For example, in the particular inversion for TA station R11A, the minimum misfit for the surface wave data is $\chi^{SW}_{\min} = 0.53$ while the minimum misfit to the surface wave data from the joint inversion is $\chi^{\text{joint}}_{\min}$ =1.15. In total, about 1000 models are found below the critical threshold $\chi_{crit} = 1.65$. The accepted models define the (marginal) distribution of model variables such as crustal thickness and velocity at a given depth.

Example posterior marginal distributions are plotted and compared with prior (marginal) distributions in Figure 14. All the marginal distributions change shapes 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. 11), the distributions of velocities in the crust and upper mantle change only subtly (Fig. 14b,f). The other distributions change profoundly. For example, crustal thickness at station R11A narrows (from σ = 4km to σ = 1.3km). The velocity difference across the Moho discontinuity sharpens with a mean shift changing from 0.5 km/sec to 0.8 km/sec. Such significant changes are expected because receiver functions 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 Figure 15. Note that the 1 σ width is about 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. 14). Figure 15c presents the $A_0(t)$ receiver function where the corridor outlined by black lines indicates the uncertainty range. The predicted receiver functions from the ensemble of accepted models are shown with gray lines and the model that fits the receiver function best is shown with the red line. The most prominent signal is the peak at about 4 sec, which is fit quite well. However, the small trough near 2 sec is not fit because doing so would require introducing another crustal discontinuity and violate monotonicity. The fact that the algorithm does not automatically adapt the parameterization to accommodate other discontinuities is discussed further in section 8.2. The misfit to the receiver function predicted by the model that fits the RF best is 0.53, which is similar to the optimal fit to the surface wave data. However, the surface wave dispersion predicted by the model that fits the receiver function optimally is not as good, with a misfit equal to about 1.15. Figure15d shows that the model that fits the receiver function best for this station does not ideally fit the surface wave data between about 40 and 60 sec period, a period range that is sensitive to the Moho depth and uppermost mantle structure. The model that minimizes the joint misfit strikes a balance between these models.

Figure 16 presents another example of the joint inversion for station T18A in the Colorado Plateau. In the observed receiver function, no clear peak is found from 3 sec to 8 sec where the Moho P-to-S conerted phase would be seen in "normal" receiver functions. After inversion, the receiver function data is well fit without a P-to-S converted arrival based on the model ensemble shown in Figure 16b. The Moho discontinuity is not well defined in the ensemble of models and Vsv changes smoothly from the lower crust to the uppermost mantle.

Figure 17 summarizes the joint inversion result for station Q22A in the Colorado Rocky Mountains. In the observed $A_0(t)$ receiver function, a peak at about 6 sec suggests that the Moho is deeper than at station R11A in the Basin and Range province. As a result, the crustal thickness distribution is centered at about 49 km depth with an uncertainty of 2.3 km. In the crust, Vsv is about 3.3 km/sec with an uncertainty of about 0.05 km/sec, and at about 100 km depth Vsv is about 4.4 km/sec. 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.

Figure 18 shows the joint inversion result for station O25A in a sedimentary basin, the Denver Basin region of the Great Plains. At long periods (>50sec) the surface wave dispersion demonstrates a fast lithospheric mantle Vsv, and the peak at 1 sec on the receiver function and the negative signal at 4 sec indicate a thick sedimentary layer at shallow depths. After inversion, a thick sedimentary layer (~4 km) with a strong velocity gradient is found and the crustal thickness is estimated to be 51.1 ± 3.9 km. Signatures in receiver functions 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 modeled here as a reverberation within the sedimentary basin with two shear and one compressional legs within the sediment layer (i.e., Yeck *et al*., 2012). The constraint of monotonically increasing velocities within the crust aids suppression of sediment multiples in favor 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 Figure 17, we may suppress a true feature, however. Extreme examples may be magma chambers (e.g., Wilson, 2003). Vsv in the sediments at O25A increases from 1 km/sec to ~ 2.6 km/sec at depth of ~ 4km, representing the compaction of sediments in this layer. Vsv in the upper crust is ~ 3.4km/sec and reaches a high of at 3.8-3.9 km/sec 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 (>4.6km.sec) is observed in the upper mantle.

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 receiver functions was shown to be effective at stations situated in a wide variety of structural environments. We now present the joint inversion method to the 185 TA stations in the study region in 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 Figure 1. For each joint inversion, we identify the model from the resulting model ensemble that fits the surface wave data (SW) or receiver functions (RF) 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 Figure 18a, while the misfits to the RFs by the RF best fitting models are plotted in Figure 18b. 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 best fitting models are shown in Figure 18 with red histograms. At most stations, the 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. Second, the joint best fitting models have a misfit < 1 for both data sets at most stations indicating that the simple model parameterization that we use in the joint inversion can reproduce the RFs with misfits below the raw RF noise for most of the stations tested. Third, 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 strong back-azimuthal variations, presumably indicating a significantly dipping Moho that renders the $A_0(t)$ receiver function a less representative 1-D average of structure beneath the station. The minimum misfits to SW data are about 0.5 for most of the stations (143 stations have a minimum misfit < 0.5). Fewer than three stations show a large misfit (>0.9) to surface wave data and those stations are sparsely distributed across the map. Finally, the relatively comparable minimum misfits for the two data sets also indicate that neither data set is over-weighted during the inversion.

After inversion is performed at all stations, we obtain Vsv 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 Vsv model on a regular grid, simple kriging (Schultz *et al.*, 1999) interpolation is applied to smooth the Vsv 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 1-degree radius. We weight the average Vsv of the model ensemble for at given station using a weighting function defined as

$$W_i = \begin{cases} \dfrac{1}{(1+d_i)\sigma_i} & \text{if } d_i \le 1^{\circ} \\ 0 & \text{otherwise} \end{cases} \qquad (13)$$

where d_i and σ_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, we produce map views of the 3-D model on a 0.5°x0.5° spatial grid.

7.2 Characteristics of the 3-D model

In the upper crust (10 km depth, Figure 20a), high velocities are observed beneath the Colorado Plateau and Great Plains, while the Rocky Mountains show a low Vsv. 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, which is about 0.8 %.

In the lower crust (averaged from 4 km above the Moho to the Moho, Fig. 20c), the most prominent feature is the slow anomaly (<3.6 km/sec) 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/sec Vsv contour and the eastern Rocky Mountain topographic high in Colorado, implying a strong relationship between topography and crustal anomalies. 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. 20d) than the upper crust, ranging from about 40 m/s in parts of the Colorado Plateau to more than 160 m/s 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-offs has been ameliorated but not entirely eliminated through the addition of receiver functions in the inversion. Second, larger uncertainties also occur where receiver functions have a larger uncertainty in the amplitude of P-to-S conversions.

At 60 km depth (Fig. 20e), 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 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. Uncertainties are largely homogeneous across the study region, averaging about 50 m/s. At a depth of 120 km (Fig. 20g), 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 are fairly homogeneous and average about 65 m/s (< 1.5%) 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 3D model compared with the surface wave inversion

There are several significant advantages to adding receiver functions 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 receiver functions in the inversion, we naturally increase 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 (Fig. 11c and Fig. 14c). This observation holds for all stations with clear P-to-S converted signals. Figure 21a,b presents crustal thickness and associated uncertainties from inverting surface wave dispersion data alone. The uncertainty level scales with crustal thickness and is ~ 5 km on average with smaller values in the Basin and Range and larger values in the east. In contrast, crustal thickness and uncertainty determined from the joint inversion appears in Figure 21c,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 receiver functions or may even increase where the P-to-S signals in the receiver functions are muted by sedimentary reverberations. In addition to improving the determination of Moho depth, receiver functions 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. 16), where the raw receiver function has no dominant arrival from 3 sec to 7 sec where a P-to-S conversion should appear. The joint Monte Carlo inversion thus results a model that has a gradient in Vsv at the depth where Moho is expected (e.g., Fig. 16). 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.

Second, introducing receiver functions improves the determination of the Vsv contrast across the Moho, which is related to the amplitude of P-to-S phase in the receiver functions. For instance, the $A_0(t)$ receiver function at station R11A (Fig. 15) shows a strong Moho conversion while 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. A map of the Vsv contrast and uncertainty across the region is shown in Figure 21e,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 and 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 of crustal thickness and thick sediments. This is because sedimentary reverberations are dominant in the receiver function signal where a Moho peak is usually expected and the Vsv constrast across Moho is difficult to resolve. Uncertainties in the velocity contrast across Moho range between about 80 m/s to 160 m/s, being smallest in the Colorado Plateau where the Vsv contrast is low in all accepted models.

The third advantage of the joint inversion method is a better determination of mantle structure below the Moho discontinuity. This is largely due to reduction of the trade-off between Moho depth and lower crustal velocity. Because of this trade-off, when surface wave dispersion is used alone to invert for a 3-D model, a prior constraint is often applied on the vertical velocity gradient in the uppermost mantle. For example, some studies have set the Vsv gradient in the uppermost mantle to be positive (Yang *et al.*, 2008, Moschetti *et al.*, 2010, Zheng *et al.*, 2011). Figure 22 shows the result of various upper mantle constraints along the transect identified in Figure 1. In the model constructed by using surface wave data alone with an imposed positive Vsv gradient in the uppermost mantle (Fig. 22a), a slow anomaly belt is found immediately

beneath the Moho. In contrast, if the constraint is changed in sign so that a negative Vsv gradient is imposed, a fast anomaly belt is observed (Fig. 22b). These two models show differences down to depths of more than 100 km and the placement of anomalies in depth is affected strongly. Lithospheric thickness and the depth to prominent asthenospheric anomalies are both changed by varying this constraint. However, both models fit the surface wave data equally well, which indicates that the surface wave data alone cannot distinguish between them. Figure 22c shows the model constructed from the joint inversion of surface wave dispersion and RF data with no constraint on the uppermost mantle velocity gradient. The result looks like a combination of Figure 22a and 22b. Beneath the Basin and Range where the Moho is prominent and the Vsv contrast across Moho is large, the model is more similar to Figure 22b constructed with a negative gradient constraint in the uppermost mantle. However, in places where a gradient Moho is expected, the model is more similar to models from the positive gradient constraint, Figure 22a. The assimilation of receiver functions in the inversion resolves the velocity-depth trade-off with data rather than by imposing ad hoc prior constraints.

8. Discussion

8.1 Sensitivity to the Vp/Vs ratio

Rayleigh wave phase velocity dispersion curves are only weakly senstive to Vp compared to Vsv; hence, the Vp/Vs ratio is usually a fixed parameter in the inversion of surface wave data. However, the Vp/Vs ratio is important in receiver function analysis because it is needed to map the P-to-S conversion time delay to depth. The Vp/Vs ratio can be determined when Moho reverberations such as the PpPs or PsPs phases are used as discussed by (Zhu and Kanamori, 2000). However, these phases arrive later than the 10 sec 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 because of lateral heterogeneity (e.g., the large scatter in results in Lowry and Perez-Gussinye, 2011, Fig. 2a). Rayleigh wave ellipticity and local amplification may also potentially be used to constrain the crustal Vp/Vs ratio (Lin et al. 2012a,b). However, this is is beyond of the scope of this study.

Because we ignore crustal reverberation phases, we are unable to determine the three relevant parameters (Vs in the crust, crustal thickness, and Vp/Vs ratio) simultaneously. To demonstrate this result, we add an extra degree of freedom to the model space, the Vp/Vs ratio in the

crystalline crust, which we allow to vary between 1.65 and 1.85 during the inversion. First, the posterior distribution of Vp/Vs ratio as well as its prior distribution is plotted in Figure 23a. The posterior distribution of the Vp/Vs ratio does not center at any particular value. Second, once the Vp/Vs ratio is introduced, the posterior distribution of the crustal thickness broadens (Fig. 23b) compared with the same posterior distribution from the inversion with fixed Vp/Vs ratio (Fig. 14c). The uncertainty of the crustal thickness increased ~ 30% from 1.4 km to ~ 1.8 km. Finally, we observe a strong trade-off between crustal thickness and the Vp/Vs ratio (Fig. 23c), which indicates that given a higher Vp/Vs ratio, estimated crustal thickness will be lower and vice versa.

The fact that we are unable to determine a preferred Vp/Vs ratio from our data also holds for other stations. Figure 23d-f shows the posterior distributions of Vp/Vs ratio at stations T18A in CP, Q22A in RM, and O25A in GP, respectively. We do not observe narrow Gaussian shaped distributions for any of them. Overall, we poorly constrain the Vp/Vs ratio from the data we are using. The direct effect is that, when we present the 3-D model in section 6.2, the uncertainty values of crustal thickness are under-estimated due to the fact that the Vp/Vs ratio is set to 1.75. To improve the determination of the Vp/Vs ratio, other information such as reverberation phases 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 general method to invert surface wave dispersion and receiver functions 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 US, 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 enhancement and improvements. In this regard we highlight four known limitations with the method as it is currently effected that may call for modification. A fifth limitation is discussed in section 8.1 regarding a variable Vp/Vs ratio.

First, the parameterization 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 parameterization accordingly, although other recent joint inversion algorithms include this feature (e.g., Bodin *et al.*, 2011). For

instance, in the inversion at R11A (Fig. 15), a mid-crustal discontinuity was not introduced automatically to fit the negative arrival at 2 sec on the receiver function, although it has a large uncertainty compared with the amplitude itself. An adaptive parameterization 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 over-interpretation of the data.

Second, the travel time 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 receiver function that we use in the joint inversion, $A_0(t)$. Not modeling the delay time variation explicitly means that the amplitude of P-to-S conversions will be underestimated if the variation is significant (i.e., ≥ 0.5 sec). The delay time variation does appear in the $A_1(t)$ and $A_2(t)$ components of the receiver function that are estimated in the harmonic stripping algorithm but which are not used in the inversion. Optimization of 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 receiver function 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 because the sensitivity of surface wave phase velocities to local slopes in structure is more complicated.

Third, we do not use the receiver function produced from a multi-station imaging technique (e.g., the Common Conversion Point (CCP) stacked receiver function) but rather a single-station RF which is an average of the 3-D variation beneath each single station. In the future, this procedure could be replaced with the receiver function stacked at the common convergence point (CCP) rather than at the station. This procedure is inappropriate for the TA because the CCPs beneath nearby stations do not overlap, as Figure 24a illustrates. The procedure would be appropriate for a denser array such as PASSAL or EarthScope Flexible Array experiments. Figure 24b 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 Figure 24a, and the red RFs are binned CD-ROM+TA RFs produced according to the Moho Conversion Points (MCP)), where radial receiver functions that pierce the same bin at Moho depth are averaged laterally. The similarity between MCP RFs and the harmonic stripping A_0 RFs demonstrate that

the single station A_0 RFs that we construct are consistent with the multi-station RFs. A smoothed image of the multi-station RFs is presented in Figure 24c, which demonstrating the resolution that RFs from a tighter network geometry provides. Higher resolution features such as the mid-crustal signal at ~ 2.5 sec 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 from 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 receiver functions. Although the installation of the TA has ameliorated the differential resolution problem in the US, the issue may ultimately need to be resolved by low-pass filtering the RFs.

Fourth, the full error covariance matrix for both surface wave dispersion data and receiver functions have not been utilized here. Rather, we have assumed that the matrices are diagonal (errors are independent) and balanced the two data sets by introducing a scaling parameter (κ) that effectively normalizes the misfit found for the two data sets. Estimating the covariance matrix for both data sets is not trivial, but it would improve the effectiveness of the algorithm.

9. Conclusions

We present a new method for joint inversion of surface wave dispersion data and receiver functions based on a Bayesian Monte Carlo scheme. When applied to receiver functions and surface wave data from ambient noise and earthquakes that are now emerging from extended broadband 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 USArray Transportable Array (TA) stations in a geologically diverse part of the Intermountain West. The effect of the introduction of receiver functions 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 receiver function 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 receiver function data makes it possible to relax 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 the method where refinements may prove beneficial. Three are particularly noteworthy: (1) development of an adaptive parameterization, particularly in the crust where low velocity layers may exist, (2) further investigation of the estimation of the full data covariance matrix for both receiver functions (correlation over time) as was done by Bodin *et al*. (2011) and surface wave dispersion (correlation over frequency), (3) extension of the receiver functions past 10 sec to recover reverberations that may help to constrain the Vp/Vs 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 are grateful to Craig Jones for insights into the receiver function analysis and 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. Lin is supported by the Director's Post-Doctoral Fellowship of the Seismological Laboratory at the California Institute of Technology.

Table 1. Model space and references.

Model Parameters	**Range**	**Reference**
Sediment thickness	0-2$\mathbf{m_0}$ (km)	Mooney and Kaban (2010)
Crystalline crustal thickness	$\mathbf{m_0}$±0.25$\mathbf{m_0}$ (km/s)	Bassin, C. *et al*. (2000)
Vsv, top of sedimentary layer	$\mathbf{m_0}$±0.2$\mathbf{m_0}$ (km/s)	Bassin, C. *et al*. (2000)
Vsv, bottome of sedimentary layer	$\mathbf{m_0}$±0.2$\mathbf{m_0}$ (km/s)	Bassin, C. *et al*. (2000)
B-spline coefficients, crust	$\mathbf{m_0}$±0.2$\mathbf{m_0}$ (km/s)	Shapiro and Ritzwoller (2002)
B-spline coefficients, mantle	$\mathbf{m_0}$±0.2$\mathbf{m_0}$ (km/s)	Shapiro and Ritzwoller (2002)

Table 2. Width of the Gaussian distribution in the model perturbation.

Parameters	Gaussian Width
Thickness, sedimentary layer	0.1 km
Thickness, crystalline crust layer	1 km
B-spline coefficients, crust	0.05 km/sec
B-spline coefficients, mantle	0.05 km/sec
Velocity at top/bottom of sedimentary layer	0.05 km/sec

References

Abers, G. A. (1998), Array measurements of phases used in receiver-function calculations: Importance of scattering, *Bulletin of the Seismological Society of America.*, *88*, 313-318

Ammon, C. J., G. E. Randall, and G. Zandt (1990), On the nonuniqueness of receiver function inversions, *J. Geophys. Res.*, *95*(B10), 15,303–15,318, doi:10.1029/JB095iB10p15303.

Ammon, C.J. & Zandt, G. (1993), Receiver structure beneath the southern mojave block, California, *Bulletin of the Seismological Society of America,* 83**,** 737-755.

An, M., and M. S. Assumpção (2004), Multi-objective inversion of surface waves and receiver functions by competent genetic algorithm applied to the crustal structure of the Paraná Basin, SE Brazil, *Geophys. Res. Lett.*, *31*, L05615, doi:10.1029/2003GL019179.

Bailey, I. W., M. S. Miller, K. Liu, and A. Levander (2012), V_S and density structure beneath the Colorado Plateau constrained by gravity anomalies and joint inversions of receiver function and phase velocity data, *J. Geophys. Res.*, *117*, B02313, doi:10.1029/2011JB008522.

Bannister, S., Yu, J., Leitner, B. & Kennett, B.L.N., (2003). Variations in crustal structure across the transition from West to East Antarctica, Southern Victoria Land, *Geophysical Journal International,* 155**,** 870-884.

Bassin, C., Laske, G. and Masters, G., The Current Limits of Resolution for Surface Wave Tomography in North America, (2000). EOS Trans AGU, 81, F897.

Basuyau, C. & Tiberi, C. (2011), Imaging lithospheric interfaces and 3D structures using receiver functions, gravity, and tomography in a common inversion scheme, *Computers & Geosciences,* 37**,** 1381-1390.

Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, M.P., Shapiro, N.M. & Yang, Y., (2007). Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophysical Journal International,* 169**,** 1239-1260.

Bensen, G.D., M.H. Ritzwoller, and Y. Yang (2009), A 3D shear velocity model of the crust and uppermost mantle beneath the United States from ambient seismic noise, *Geophysical*

Journal International, 177(3), 1177-1196.

Bianchi, I., J. Park, N. Piana Agostinetti, and V. Levin (2010), Mapping seismic anisotropy using harmonic decomposition of receiver functions: An application to Northern Apennines, Italy, *J. Geophys. Res.*, *115*, B12317, doi:10.1029/2009JB007061.

Bodin, T., M. Sambridge, H. Tkalčić, P. Arroucau, K. Gallagher, and N. Rawlinson (2012), Transdimensional inversion of receiver functions and surface wave dispersion, *J. Geophys. Res.*, *117*, B02301, doi:10.1029/2011JB008560.

Bostock, M. G. (1998), Mantle stratigraphy and evolution of the Slave province, *J. Geophys. Res.*, *103*(B9), 21,183–21,200, doi:10.1029/98JB01069.

Cassidy, J. F. (1992), Numerical experiments in broad-band receiver function-analysis, *Bulletin of the Seismological Society of America*, *82*(3), 1453-1474.

Chang, S.J., Baag, C.E. & Langston, C.A. (2004), Joint analysis of teleseismic receiver functions and surface wave dispersion using the genetic algorithm, *Bulletin of the Seismological Society of America,* 94**,** 691-704.

Christensen, N.I. & Mooney, W.D. (1995), Seismic velocity structure and composition of the continental crust: A global view*, J. Geophys. Res.,*100(B6): 9761–9788.

Clitheroe, G., O. Gudmundsson, and B. L. N. Kennett (2000), The crustal thickness of Australia, *J. Geophys. Res.*, *105*(B6), 13,697–13,713, doi:10.1029/1999JB900317.

Crotwell, H. P., and T. J. Owens (2005), Automated receiver function processing, *Seismological Research Letters*, *76*(6), 702-709.

Du, Z.J. & Foulger, G.R., (1999). The crustal structure beneath the northwest fjords, Iceland, from receiver functions and surface waves, *Geophysical Journal International,* 139**,** 419-432.

Dziewonski, A. and D. Anderson (1981), Preliminary reference Earth model, *Phys. Earth Planet. Int.,* 25(4): 297–356.

Endrun, B., Meier, T., Bischoff, M. & Harjes, H.P. (2004), Lithospheric structure in the area of Crete constrained by receiver functions and dispersion analysis of Rayleigh phase velocities, *Geophysical Journal International,* 158**,** 592-608.

Frederiksen, A.W., Folsom, H. & Zandt, G., (2003). Neighbourhood inversion of teleseismic Ps conversions for anisotropy and layer dip, *Geophysical Journal International,* 155**,** 200-212.

Gilbert, H. J., and A. F. Sheehan (2004), Images of crustal variations in the intermountain west, *J. Geophys. Res.*, *109*, B03306, doi:10.1029/2003JB002730.

Girardin, N. and V. Farra (1998), Azimuthal anisotropy in the upper mantle from observations of P-to-S converted phases: application to southeast Australia, *Geophysical Journal International,* 133**,** 615-629.

Gok, R., Pasyanos, M.E. & Zor, E. (2007), Lithospheric structure of the continent-continent collision zone: eastern Turkey, *Geophysical Journal International,* 169**,** 1079-1088.

Gorman, A. R., Clowes, R. M., Ellis, R. M. et al., (2002), Deep Probe: Imaging the roots of western North America, *Can. Journ. Earth Sci.,* 39(3)**,** 375-398.

Hetenyi, G. & Bus, Z., (2007). Shear wave velocity and crustal thickness in the Pannonian Basin from receiver function inversions at four permanent stations in Hungary, *Journal of Seismology,* 11**,** 405-414.

Horspool, N.A., Savage, M.K. & Bannister, S. (2006), Implications for intraplate volcanism and back-arc deformation in northwestern New Zealand, from joint inversion of receiver functions and surface waves, *Geophysical Journal International,* 166**,** 1466-1483.

Jones, C. H., and R. A. Phinney (1998), Seismic structure of the lithosphere from teleseismic converted arrivals observed at small arrays in the southern Sierra Nevada and vicinity, California, *J. Geophys. Res.*, *103*(B5), 10,065–10,090, doi:10.1029/97JB03540.

Julia, J., Ammon, C.J. & Herrmann, R.B. (2003), Lithospheric structure of the Arabian Shield from the joint inversion of receiver functions and surface-wave group velocities, *Tectonophysics,* 371**,** 1-21.

Julia, J., Ammon, C.J., Herrmann, R.B. & Correig, A.M. (2000), Joint inversion of receiver function and surface wave dispersion observations, *Geophysical Journal International,* 143**,** 99-112.

Kanamori, H. and D. Anderson (1977), Importance of physical dispersion in surface wave and free oscillation problems : Review, *Revs. Geophys. Space Phys*., 15(1):105-112.

Khan, A., Zunino, A. and Deschamps, F. (2011) The thermo-chemical and physical structure beneath the North American continent from Bayesian inversion of surface-wave phase velocities , J. Geophys. Res., 116, B09304, doi:10.1029/2011JB008380.

Langston, C. A. (1979), Structure Under Mount Rainier, Washington, Inferred From Teleseismic Body Waves, *J. Geophys. Res.*, *84*(B9), 4749–4762, doi:10.1029/JB084iB09p04749.

Last, R. J., A. A. Nyblade, C. A. Langston, and T. J. Owens (1997), Crustal structure of the East African Plateau from receiver functions and Rayleigh wave phase velocities, *J. Geophys. Res.*, *102*(B11), 24,469–24,483, doi:10.1029/97JB02156.

Lawrence, J.F. & Wiens, D.A., (2004). Combined receiver-function and surface wave phase-velocity inversion using a niching genetic algorithm: Application to Patagonia, *Bulletin of the Seismological Society of America,* 94**,** 977-987.

Levander, A., B. Schmandt, M. S. Miller, K. Liu, K. E. Karlstrom, R. S. Crow, C. T. A. Lee, and E. D. Humphreys (2011), Continuing Colorado plateau uplift by delamination-style convective lithospheric downwelling, *Nature*, *472*(7344), 461-U540.

Levin, V. & Park, J., (1997). P-SH conversions in a flat-layered medium with anisotropy of arbitrary orientation, *Geophysical Journal International,* 131**,** 253-266.

Levin, V., N.M. Shapiro, J. Park, and M.H. Ritzwoller (2002), Seismic evidence for catastrophic slab loss beneath Kamchatka, *Nature*, 418, 763-767.

Levshin, A.L., M.H. Ritzwoller, and N.M. Shapiro (2005), The use of crustal higher modes to constrain crustal structure across Central Asia, *Geophys. J. Int.*, 160, 961-972.

Ligorria, J. P., and C. J. Ammon (1999), Iterative deconvolution and receiver-function estimation, *Bulletin of the Seismological Society of America*, *89*(5), 1395-1400.

Lin, F.C., Moschetti, M.P. & Ritzwoller, M.H., (2008). Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps, *Geophysical Journal International,* 173**,** 281-298.

Lin, F.C., M.H. Ritzwoller, Y. Yang, M.P. Moschetti, and M.J. Fouch (2011), Complex and variable crustal and uppermost mantle seismic anisotropy in the western United States, *Nature Geoscience*, Vol 4, Issue 1, 55-61.

Lin, F.C., Ritzwoller, M.H. & Snieder, R., (2009). Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array, *Geophysical Journal International,* 177**,** 1091-1110.

Lin, F.C. & Ritzwoller, M.H., 2011. Helmholtz surface wave tomography for isotropic and azimuthally anisotropic structure, *Geophysical Journal International,* 186**,** 1104-1120.

Lin, F.C.,V. Tsai, and M.H. Ritzwoller (2012a), The local amplification of surface waves: A new observable to constrain elastic velocities, density, and anelastic attenuation, J. Geophys. Res., in press.

Lin, F.C.,B. Schmandt, and V.C. Tsai (2012b), Joint inversion of Rayleigh wave phase velocity and ellipticity using USArray: constraining velocity and density structure in the upper crust, submitted to Geophys. Res. Letts.

Liu, Q.Y., Li, Y., Chen, J.H., van der Hilst, R.D., Guo, B.A., Wang, J., Qi, S.H. & Li, S.C., (2010). Joint inversion of receiver function and ambient noise based on Bayesian theory, *Chinese Journal of Geophysics-Chinese Edition,* 53**,** 2603-2612.

Lowry, A.R. & Perez-Gussinye, M., (2011). The role of crustal quartz in controlling Cordilleran deformation, *Nature,* 471**,** 353-+.

Lucente, F. P., N. Piana Agostinetti, M. Moro, G. Selvaggi, and M. Di Bona (2005), Possible fault plane in a seismic gap area of the southern Apennines (Italy) revealed by receiver function analysis, *J. Geophys. Res.*, *110*, B04307, doi:10.1029/2004JB003187.

Maraschini, M. & Foti, S., 2010. A Monte Carlo multimodal inversion of surface waves, *Geophysical Journal International,* 182**,** 1557-1566.

Megnin, C., and B. Romanowicz (2000), The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms, *Geophysical Journal International*, *143*(3), 709-728.

Molnar, S., S.E. Dosso and J.F. Cassidy, (2010). Bayesian inversion of microtremor array dispersion data in southwestern British Columbia, *Geophys. J. Int.,* 183, 923-940.

Mooney, W. D., and M. K. Kaban (2010), The North American upper mantle: Density, composition, and evolution, *J. Geophys. Res.*, *115*, B12424, doi:10.1029/2010JB000866.

Moorkamp, M., A. G. Jones, and S. Fishwick (2010), Joint inversion of receiver functions, surface wave dispersion, and magnetotelluric data, *J. Geophys. Res.*, *115*, B04318, doi:10.1029/2009JB006369.

Moschetti, M.P., M.H. Ritzwoller, and F.C. Lin (2010), Seismic evidence for widespread crustal deformation caused by extension in the western USA, *Nature*, 464, Number 7290, 885-889.

Moschetti, M. P., M. H. Ritzwoller, and N. M. Shapiro (2007), Surface wave tomography of the western United States from ambient seismic noise: Rayleigh wave group velocity maps, *Geochem. Geophys. Geosyst.*, *8*, Q08010, doi:10.1029/2007GC001655.

Moschetti, M. P., M. H. Ritzwoller, F. C. Lin, and Y. Yang (2010a), Crustal shear wave velocity structure of the western United States inferred from ambient seismic noise and earthquake data, *Journal of Geophysical Research-Solid Earth*, *115*.

Moschetti, M.P., M.H. Ritzwoller, F.C. Lin, and Y. Yang, Crustal shear velocity structure of the western US inferred from amient noise and earthquake data (2010b), *J. Geophys. Res.*, 115, B10306, doi:10.1029/2010JB007448.

Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, *100*(B7), 12,431–12,447, doi:10.1029/94JB03097.

Mosegaard K. and Sambridge M. 2002. Monte Carlo analysis of inverse problems. *Inverse Problems* **18**, 29–54.

Nicholson, T., Bostock, M. & Cassidy, J.F., (2005). New constraints on subduction zone structure in northern Cascadia, *Geophysical Journal International,* 161**,** 849-859.

Ozalaybey, S., Savage, M.K., Sheehan, A.F., Louie, J.N. & Brune, J.N., (1997). Shear-wave velocity structure in the northern Basin and Range province from the combined analysis of receiver functions and surface waves, *Bulletin of the Seismological Society of America,* 87**,** 183-199.

Pasyanos, M.E., Tkalcic, H., Gok, R., Al-Enezi, A. & Rodgers, A.J., (2007). Seismic structure of Kuwait, *Geophysical Journal International,* 170**,** 299-312.

Piana Agostinetti, N., F. P. Lucente, G. Selvaggi, and M. Di Bona (2002), Crustal Structure and Moho Geometry beneath the Northern Apennines (Italy), *Geophys. Res. Lett.*, *29*(20), 1999,

doi:10.1029/2002GL015109.

Piana Agostinetti, N. & Chiarabba, C., 2008. Seismic structure beneath Mt Vesuvius from receiver function analysis and local earthquakes tomography: evidences for location and geometry of the magma chamber, *Geophysical Journal International,* 175**,** 1298-1308.

Piana Agostinetti, N. & Malinverno, A., 2010. Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophysical Journal International,* 181**,** 858-872.

Pollitz, F.F. Observations and interpretation of fundamental mode Rayleigh wavefields recorded by the Transportable Array (USArray) (2008), *Geophys. J. Int.,* **173**,189-204.

Pollitz, F. F. & Snoke, J. A., 2010, Rayleigh-wave phase-velocity maps and three dimensional shear velocity structure of the western US from local non-plane surface wave tomography. *Geophys. J. Int.*, **180**, 1153–1169.

Ritzwoller, M. H., N. M. Shapiro, A. L. Levshin, and G. M. Leahy (2001), Crustal and upper mantle structure beneath Antarctica and surrounding oceans, *J. Geophys. Res.*, *106*(B12), 30,645–30,670, doi:10.1029/2001JB000179.

Ritzwoller, M.H., N.M. Shapiro, and G.M. Leahy, A resolved mantle anomaly as the cause of the Australian-Antarctic Discordance (2003), *J. Geophys. Res.*, 108, no. B12, 2559, doi:10.1029/2003JB002522.

Ritzwoller, M.H., N.M. Shapiro, S. Zhong (2004), Cooling history of the Pacific lithosphere, *Earth Planet. Sci. Letts.*, 226, 69-84.

Ritzwoller, M.H., F.C. Lin, and W. Shen, Ambient noise tomography with a large seismic array, *Compte Rendus Geoscience*, (2011), 13 pages, doi:10.1016/j.crte.2011.03.007.

Roy, M., T. H. Jordan, and J. Pederson (2009), Colorado Plateau magmatism and uplift by warming of heterogeneous lithosphere, *Nature*, *459*(7249), 978-U102.

Salah, M.K., Chang, S.J. & Fonseca, J., 2011. Crustal structure beneath the Lower Tagus Valley, southwestern Iberia using joint analysis of teleseismic receiver functions and surface-wave dispersion, *Geophysical Journal International,* 184**,** 919-933.

Sambridge, M., 1999a. Geophysical inversion with a neighbourhood algorithm - I. Searching a

parameter space, *Geophysical Journal International,* 138**,** 479-494.

Sambridge, M., 1999b. Geophysical inversion with a neighbourhood algorithm - II. Appraising the ensemble, *Geophysical Journal International,* 138**,** 727-746.

Sambridge M. 2001. Finding acceptable models in nonlinear inverse problems using a neighbourhood algorithm. *Inverse Problems* **17**, 387–403.

Sambridge M. and Mosegaard K. 2002. Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics* **40**, 1–29

Savage, M. K. (1998), Lower crustal anisotropy or dipping boundaries? Effects on receiver functions and a case study in New Zealand, *J. Geophys. Res.*,*103*, 15069-15087.

Schulte-Pelkum, V. (2012), 2-D structure and anisotropy from harmonic out-of-plane receiver functions, in prep.

Schultz, C.A., Myers, S.C., Hipp, J. & Young, C.J., 1999. Nonstationary Bayesian kriging: a predictive technique to generate spatial corrections for seismic detection, location and identification, *Physics of the Earth and Planetary Interiors,* 113**,** 321-338.

Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science,* 307**,** 1615-1618.

Shapiro, N.M. & Ritzwoller, M.H., 2002. Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophysical Journal International,* 151**,** 88-105.

Shapiro, N. M., M. H. Ritzwoller, and E. R. Engdahl (2008), Structural context of the great Sumatra-Andaman Islands earthquake, *Geophys. Res. Lett.*, *35*, L05301, doi:10.1029/2008GL033381.

Sheehan, A. F., C. H. Jones, M. K. Savage, S. Ozalaybey, and J. M. Schneider (1997), Contrasting lithospheric structure between the Colorado Plateau and Great Basin: Initial results from Colorado Plateau - Great Basin PASSCAL Experiment, *Geophys. Res. Lett.*, *24*(21), 2609–2612, doi:10.1029/97GL02782.

Shibutani, T., M. Sambridge, and B. Kennett (1996), Genetic algorithm inversion for receiver functions with application to crust and uppermost mantle structure beneath eastern

Australia, *Geophys. Res. Lett.*, *23*(14), 1829–1832, doi:10.1029/96GL01671.

Socco, L.V. and D. Boiero, Improved Monte Carlo inversion of surface wave data (2008), *Geophys. Prospecting*, 56, 357-371.

Tarantola, A., and B. Valette (1982), Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys.*, *20*(2), 219–232, doi:10.1029/RG020i002p00219.

Tkalčić, H., M. E. Pasyanos, A. J. Rodgers, R. Gök, W. R. Walter, and A. Al-Amri (2006), A multistep approach for joint modeling of surface wave dispersion and teleseismic receiver functions: Implications for lithospheric structure of the Arabian Peninsula, *J. Geophys. Res.*,*111*, B11311, doi:10.1029/2005JB004130.

Tokam, A.P.K., Tabod, C.T., Nyblade, A.A., Julia, J., Wiens, D.A. & Pasyanos, M.E., 2010. Structure of the crust beneath Cameroon, West Africa, from the joint inversion of Rayleigh wave group velocities and receiver functions, *Geophysical Journal International,* 183**,** 1061-1076.

Villasenor, A., M.H. Ritzwoller, A.L. Levshin, M.P. Barmin, E.R. Engdahl, W. Spakman, and J. Trampert (2001), Shear velocity structure of Central Eurasia from inversion of surface wave velocities, *Phys. Earth Planet. Int.*, 123(2-4), 169 - 184.

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

Vinnik, L.P., Aleshin, I.M., Kaban, M.K., Kiselev, S.G., Kosarev, G.L., Oreshin, S.I. & Reigber, C., 2006. Crust and mantle of the Tien Shan from data of the receiver function tomography, *Izvestiya-Physics of the Solid Earth,* 42**,** 639-651.

Vinnik, L.P., Reigber, C., Aleshin, I.M., Kosarev, G.L., Kaban, M.K., Oreshin, S.I. & Roecker, S.W., 2004. Receiver function tomography of the central Tien Shan, *Earth and Planetary Science Letters,* 225**,** 131-146.

Wilson, C. K., Jones, C. H., and Gilbert, H. J., 2003. Single-chamber silicic magma system inferred from shear wave discontinuities of the crust and uppermost mantle, Coso geothermal area, California, *J. Geophys. Res.,* 108(B5)**,** 2226.

Yang, Y. & Forsyth (2006), D. Regional tomographic inversion of the amplitude and phase of

Rayleigh waves with 2-D sensitivity kernels, *Geophys. J. Int.*, **166**, 1148–1160.

Yang, Y.J., Ritzwoller, M.H., Levshin, A.L. & Shapiro, N.M., 2007. Ambient noise rayleigh wave tomography across Europe, *Geophysical Journal International,* 168**,** 259-274.

Yang, Y., and M. H. Ritzwoller (2008), Teleseismic surface wave tomography in the western U.S. using the Transportable Array component of USArray, *Geophys. Res. Lett.*, *35*, L04308, doi:10.1029/2007GL032278.

Yang, Y., A. Li, and M.H. Ritzwoller (2008a), Crustal and uppermost mantle structure in southern Africa revealed from ambient noise and teleseismic tomography, *Geophys. J. Int.*, doi:10.1111/j.1365-246X.2008.03779.x.

Yang, Y., M. H. Ritzwoller, F.-C. Lin, M. P. Moschetti, and N. M. Shapiro (2008b), Structure of the crust and uppermost mantle beneath the western United States revealed by ambient noise and earthquake tomography, *J. Geophys. Res.*, *113*, B12310, doi:10.1029/2008JB005833.

Yang, Y., M. H. Ritzwoller, and C. H. Jones (2011), Crustal structure determined from ambient noise tomography near the magmatic centers of the Coso region, southeastern California, *Geochem. Geophys. Geosyst.*, *12*, Q02009, doi:10.1029/2010GC003362.

Yang, Y., M. H. Ritzwoller, Y. Zheng, W. Shen, A. L. Levshin, and Z. Xie (2012), A synoptic view of the distribution and connectivity of the mid-crustal low velocity zone beneath Tibet, *J. Geophys. Res.*, *117*, B04303, doi:10.1029/2011JB008810.

Yao, H.J., van der Hilst, R.D. & de Hoop, M.V., 2006. Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis - I. Phase velocity maps, *Geophysical Journal International,* 166**,** 732-744.

Yeck, W., Sheehan, A. F. and Schulte-Pelkum, V., Iterative H-κ stacking to obtain accurate crustal thicknesses beneath sedimentary basins, 2012, in prep.

Yoo, H.J., Herrmann, R.B., Cho, K.H. & Lee, K., 2007. Imaging the three-dimensional crust of the Korean Peninsula by joint inversion of surface-wave dispersion and teleseismic receiver functions, *Bulletin of the Seismological Society of America,* 97**,** 1002-1011.

Yoshizawa, K. and B.L.N. Kennett (2002), Non-linear waveform inversion for surface waves with a neighbourhood algorithm – application to multimode dispersion measurements,

Geophys. .Int. 149, 118-133.

Zhao, L.S., Sen, M.K., Stoffa, P. & Frohlich, C., 1996. Application of very fast simulated annealing to the determination of the crustal structure beneath Tibet, *Geophysical Journal International,* 125**,** 355-370.

Zheng, X. F., Z. X. Yao, J. H. Liang, and J. Zheng (2010), The role played and opportunities provided by IGP DMC of China National Seismic Network in Wenchuan earthquake disaster relief and researches, *Bull. Seismol. Soc. Am*., 100(5B), 2866-2872.

Zheng, Y., W. Shen, L. Zhou, Y. Yang, Z. Xie, and M. H. Ritzwoller (2011), Crust and uppermost mantle beneath the North China Craton, northeastern China, and the Sea of Japan from ambient noise tomography, *J. Geophys. Res.*, *116*, B12312, doi:10.1029/2011JB008637.

Zhou, L., J. Xie, W. Shen, Y. Zheng, Y. Yang. H. Shi, and M.H. Ritzwoller (2012), The structure of the crust and uppermost mantle beneath South China from ambient noise and earthquake tomography, *Geophys. J. Int.*, doi: 10.1111/j.1365-246X.2012.05423.x.

Zhu, L., and H. Kanamori (2000), Moho depth variation in southern California from teleseismic receiver functions, *J. Geophys. Res.*, *105*(B2), 2969–2980, doi:10.1029/1999JB900322.

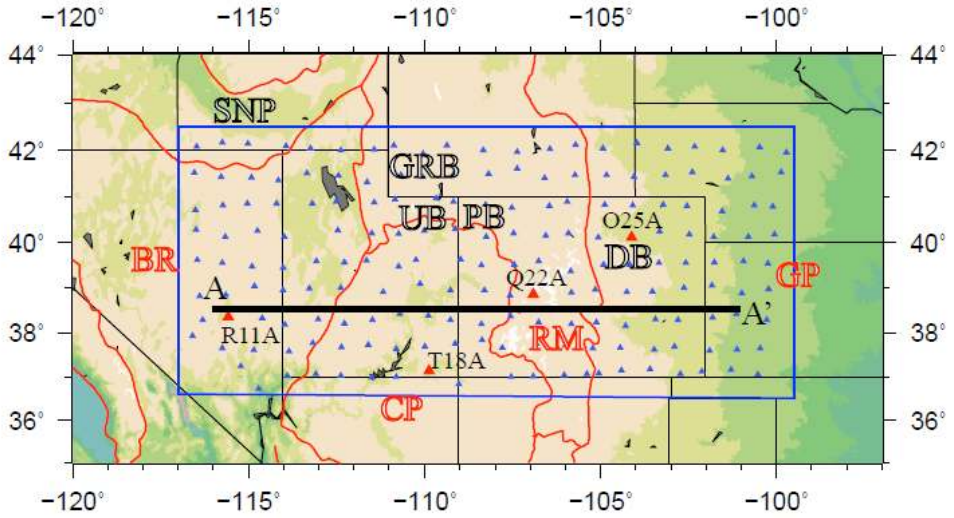

Figure 1. Stations of the Earthscope USArrayTransportable Array(TA) used in this study are shown with blue triangles. The main geological provinces are outlined with red contours and titled with red abbreviations (the Basin and Range (BR) province, the Colorado Plateau (CP), the Rocky Mountains (RM), 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 northeastern 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 GP.

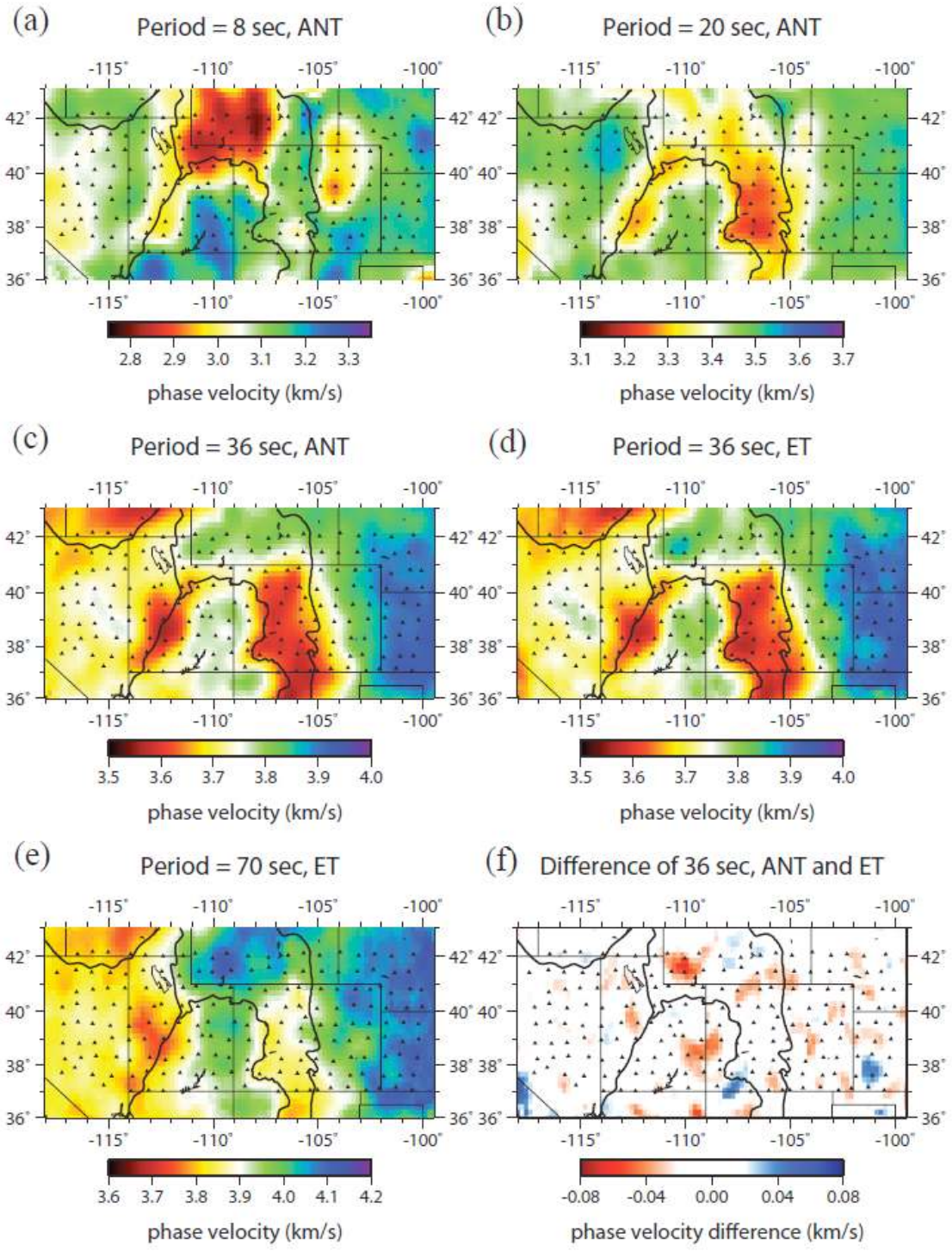

Figure 2. (a-c) Phase velocity maps at periods of 8, 20 and 36 sec from ambient noise eikonal tomography (ANT). (d-e) Phase velocity maps at 36 and 70 sec from earthquake Helmholtz tomography (ET). (e) Difference between the ANT and EQ results at 36 sec period presented in (c) and (d).

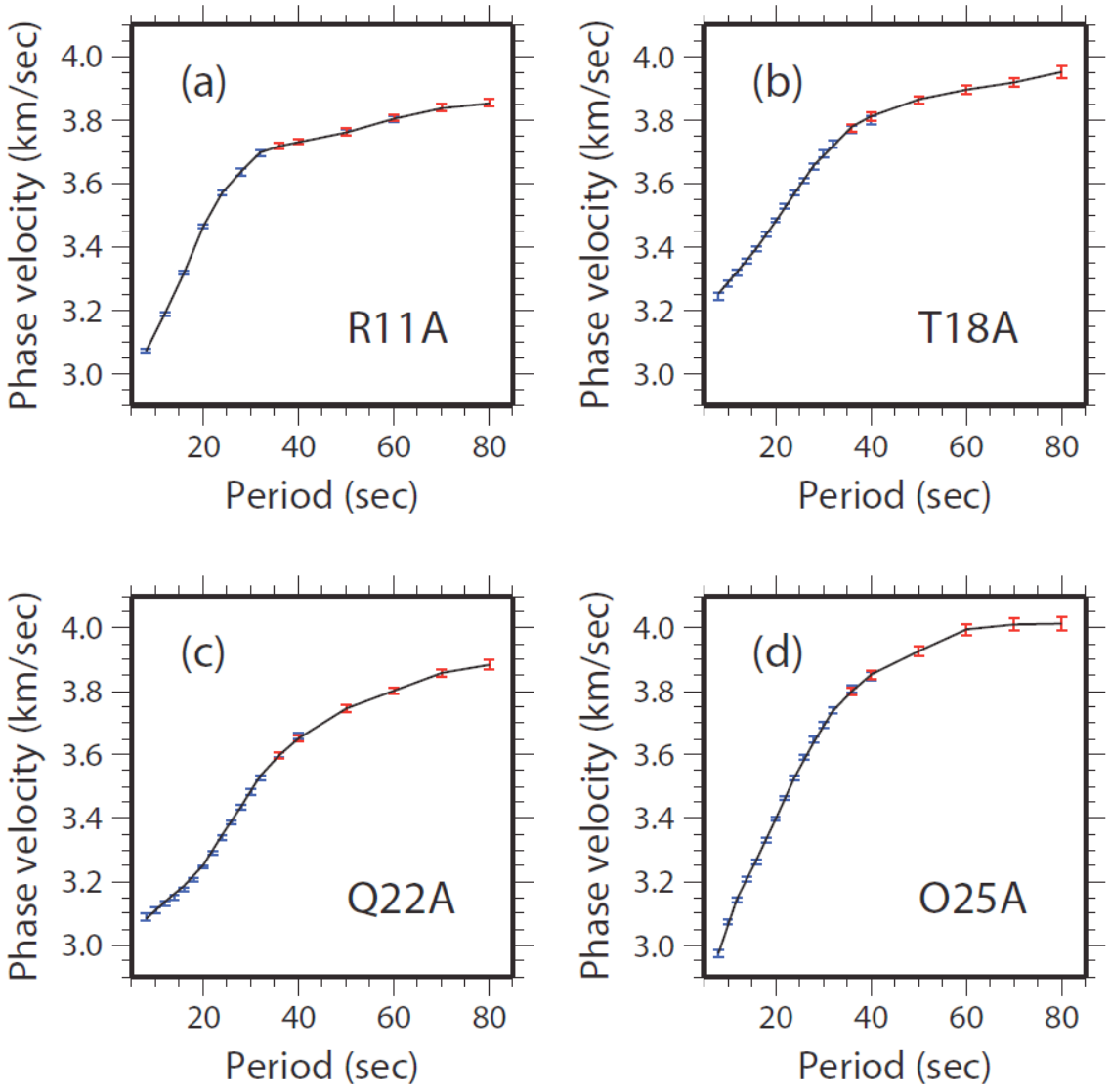

Figure 3. Phase velocity curves taken from the maps presented in Fig. 2 (and similar maps at the intervening periods) at the locations of stations R11A, T18A, Q22A and O25A, respectively. Blue symbols are uncertainties from the ambient noise tomography maps, red symbols are uncertainties for earthquake tomography maps, and the line is the joint dispersion curve from the two kinds of measurements.

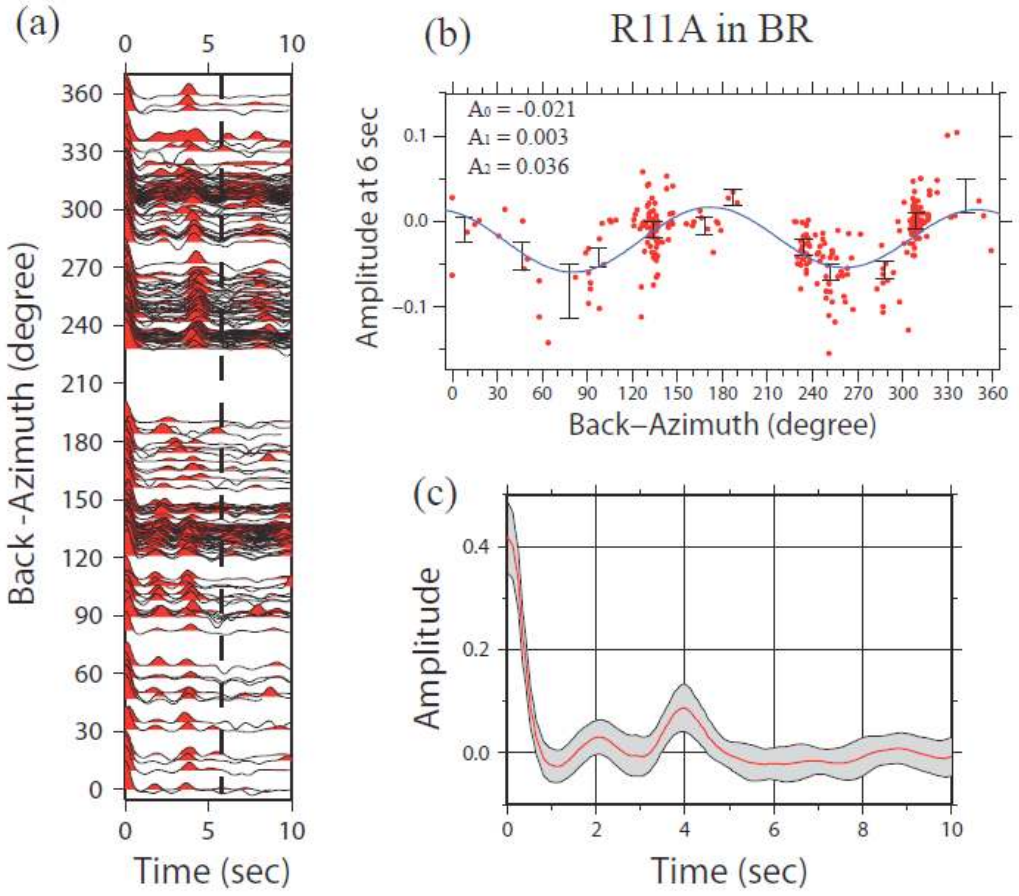

Figure 4. (a) Raw receiver function (RF) waveforms for station R11A in the BR are plotted vs. back-azimuth, 6 sec time time is identified with the dashed line. (b) Red dots are RF amplitude measured at 6 sec time with azimuthally binned amplitudes plotted with black error-bars. The harmonic fitting result is shown with the blue curve and the amplitudes of coefficients A_0, A_1, and A_2 (eq. (1)) are shown. (c) The A_0 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 (eqn. (1)) .

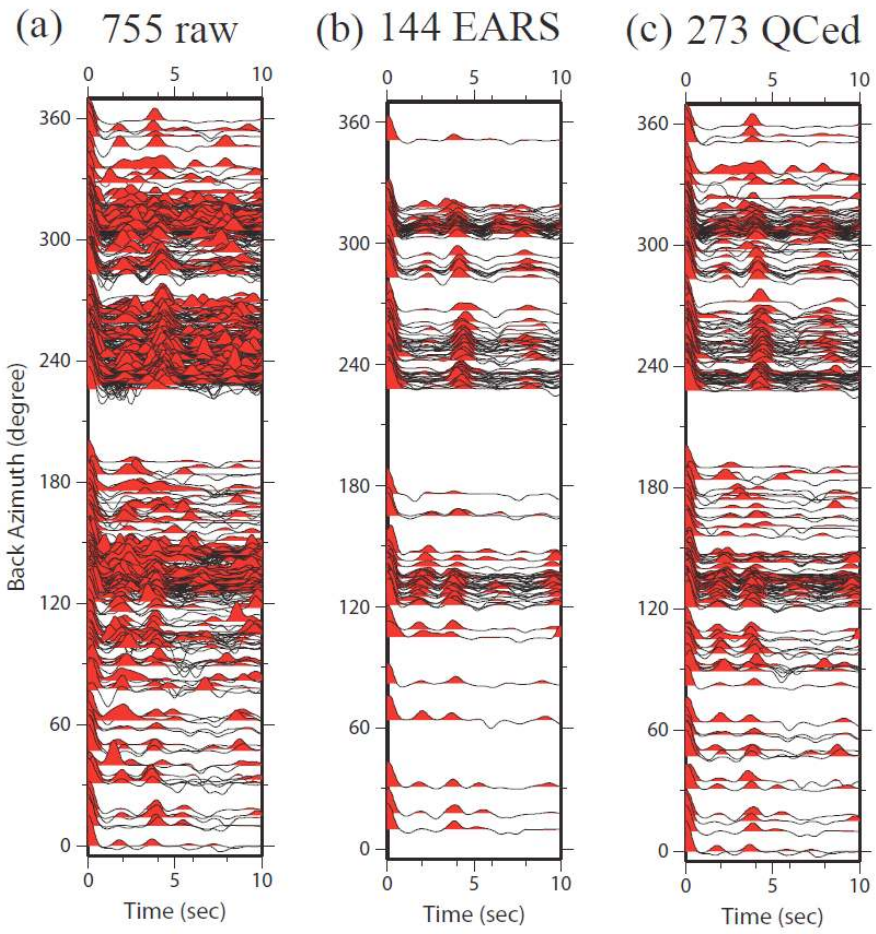

Figure 5. (a) The 377 non-quality controlled, slowness-corrected receiver functions (RFs) are plotted along back-azimuth for TA station R11A. (b) The same as (a), but for the 144 RFs in (a) that have been quality controlled by EARS. (c) The 273 quality controlled RFs including the 144 RFs from EARS and an additional 129 RFs that are consistent with them.

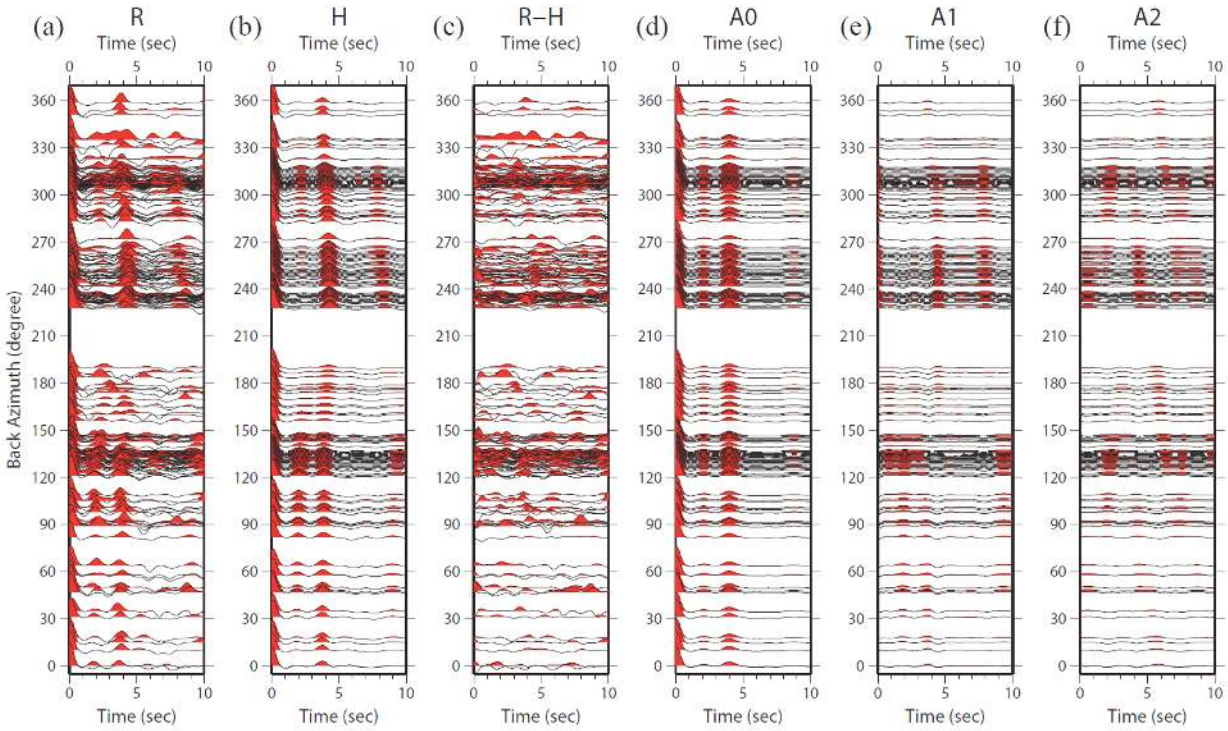

Figure 6. (a) QCed RF are plotted along back-azimuth for station R11A. (b) The reconstructed RF from Harmonic Stripping. (c) The difference between (a) and (b), which determines the uncertainties measured for the RF waveform. (d-f) The 3 components from harmonic stripping.

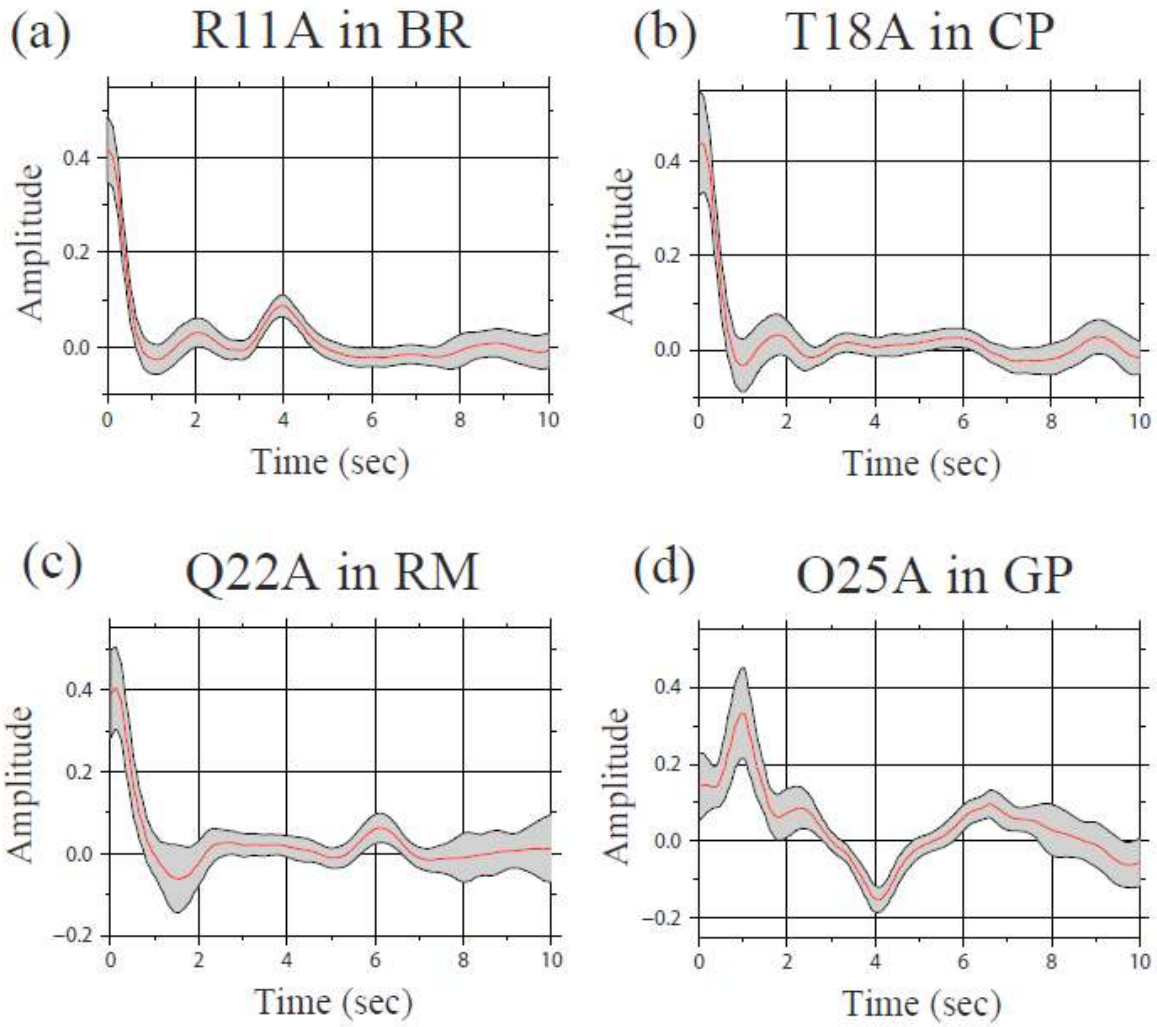

Figure 7: The A_0 component receiver function (red curve) and uncertainty (gray corridor) for four stations whose locations are identified in Fig. 1. The uncertainty is the rms of the residual remaining after the harmonic fitting (eqn. (1)) to the azimuthally dependent RFs, reduced by a factor of two between 3 and 8 sec.

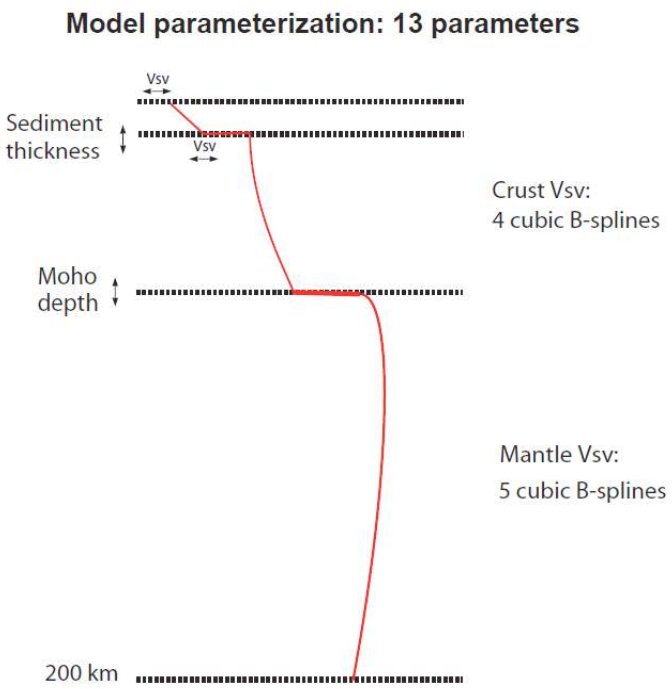

Figure 8. Model parameterization illustrating the 13 model parameters used in the Monte Carlo sampling of model space.

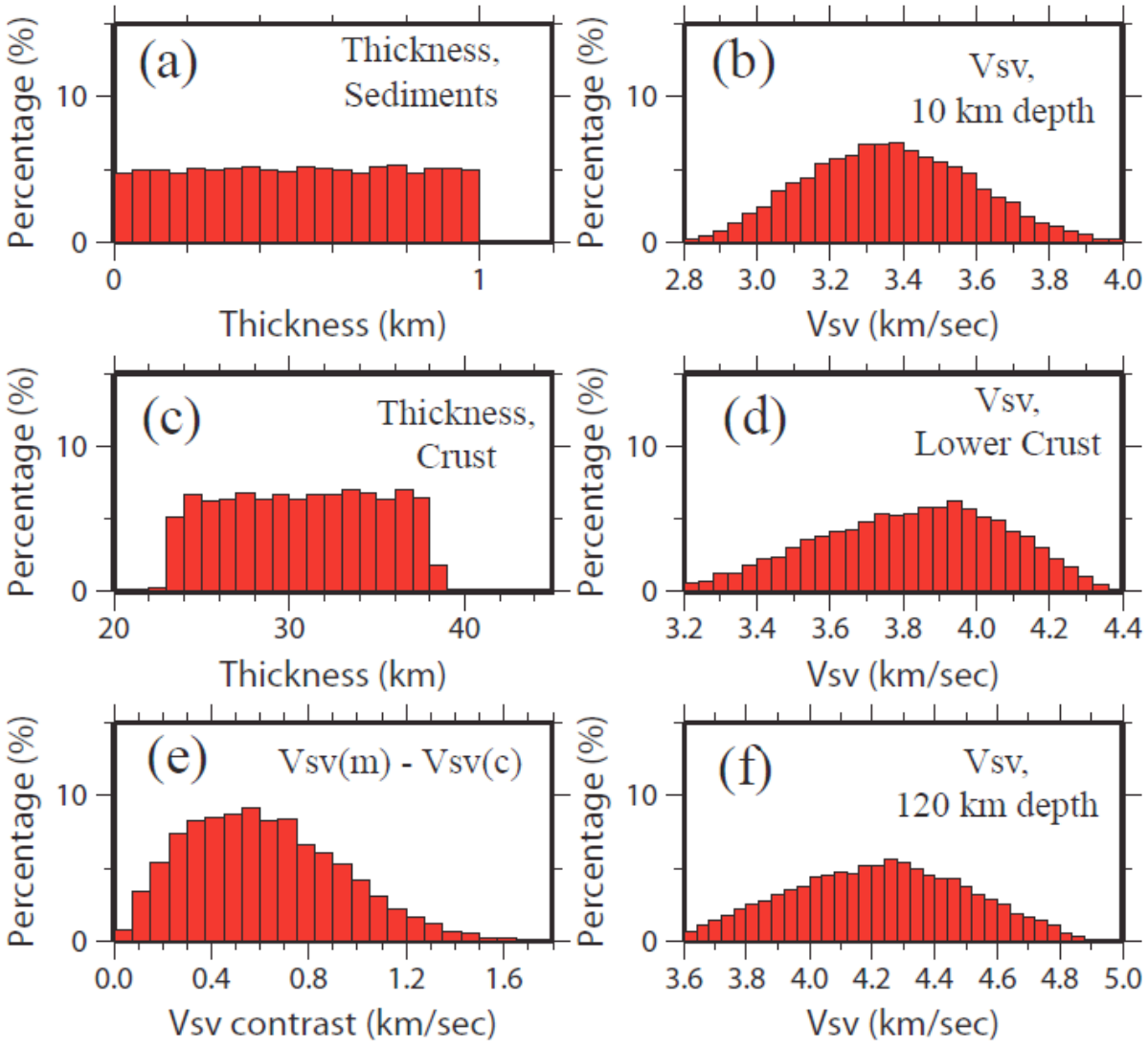

Figure 9. Prior distribution for several of the model parameters at the location of TA station R11A: (a) sediment thickness, (b) Vsv at 10 km depth, (c) crustal thickness, (d) Vsv in the lower crust , (e) velocity contrast from 4 km above to 4 km below Moho (mantle – crust), and (f) Vsv at 120 km depth.

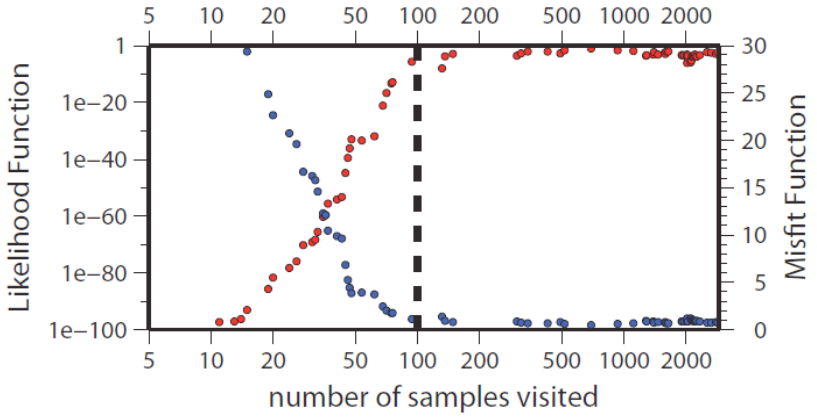

Figure 10. Convergence of sampling the posterior distribution of models when only surface wave dispersion data at the location of station R11A are used. Models are provisionally accepted according to the criterion in eq. (9), and then replaced by other models as the procedure evolved. Red dots are the likelihood function (eq. (5)) for each model that is accepted during the Metropolis sampling of model space, while blue dots are the rms-misfit (eqn. (10)). The convergence point is at ~100 models (dashed line).

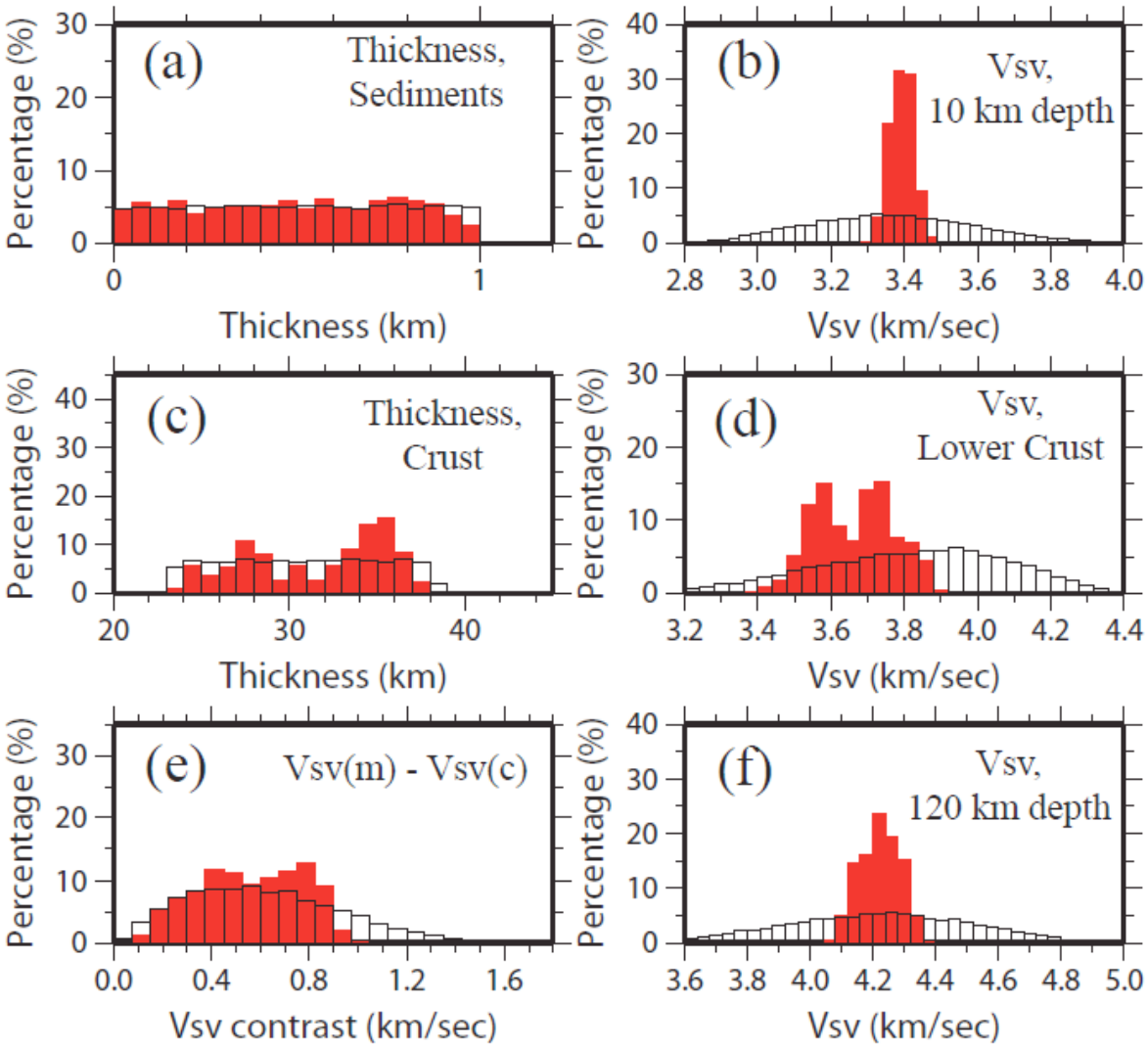

Figure 11. (a)-(f) The same as Fig. 9 for TA station R11A, but plotted here are the posterior distributions after Monte Carlo sampling using surface wave data alone (models fit the surface wave data). The prior distributions from Fig. 9 are plotted as write histograms in the background. Surface wave 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.

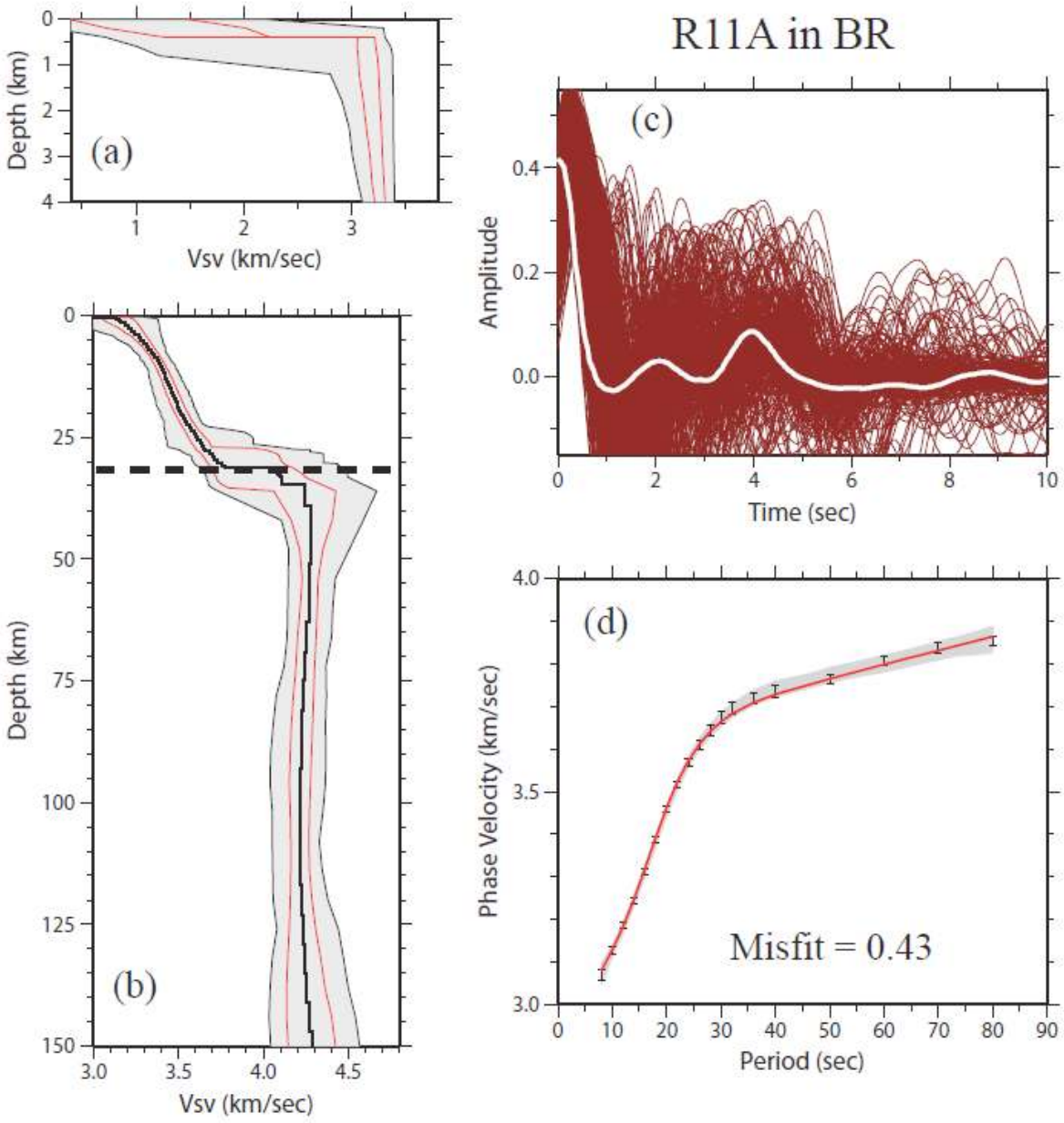

Figure 12. (a-b) The model ensemble at the location of TA station R11A in the Basin and Range province resulting from the Monte Carlo fit to surface wave data alone is shown in full width (black corridor with gray fill), 1 σ width (red corridor) and the average model (black curve). Moho is identified as a dashed line at ~ 32 km. (c) The observed RF (white line) is plotted with predicted receiver functions (red lines) computed from all of the models accepted (eq. (11)) by Monte Carlo sampling, showing that RFs are not not well fit, on average, by models constrained by surface wave data alone. (d) The observed Rayleigh wave phase velocity dispersion curve data (black error bars) are plotted with predicted surface wave phase velocity curves computed from all accepted models (eq. (11), gray lines). The red curve is the predicted phase velocity curve from the best fitting model. The reduced rms-misft (eqn. (10)) of the best fitting model is 0.43.

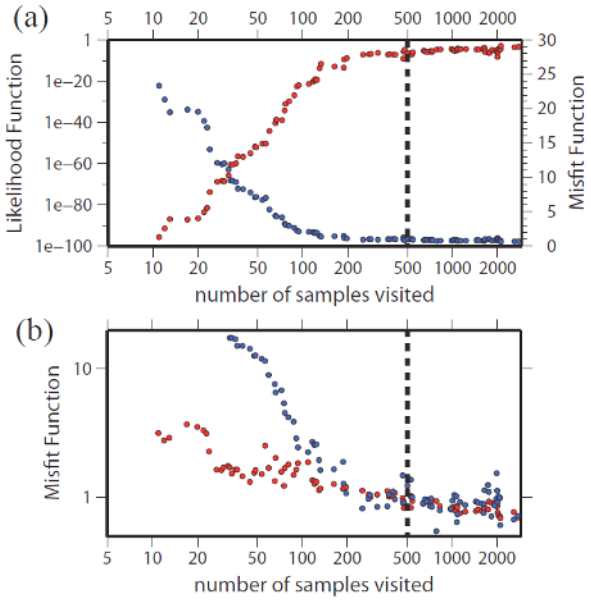

Figure 13. (a) The joint likelihood function (eqn. (16), red dots) and the joint misfit function (eqn. (17), blue dots) for TA station R11A as a function of number of models sampled. (b) Misfit for each individual data set in the same search as (a). Blue dots are for surface wave phase velocity data and red dots are for the RF data. When the model converges to maximize the likelihood functional, both misfits typically converge to < 1. In both plots, the location of 500 sampled models is identified with a dashed line, where the fit approaches convergence.

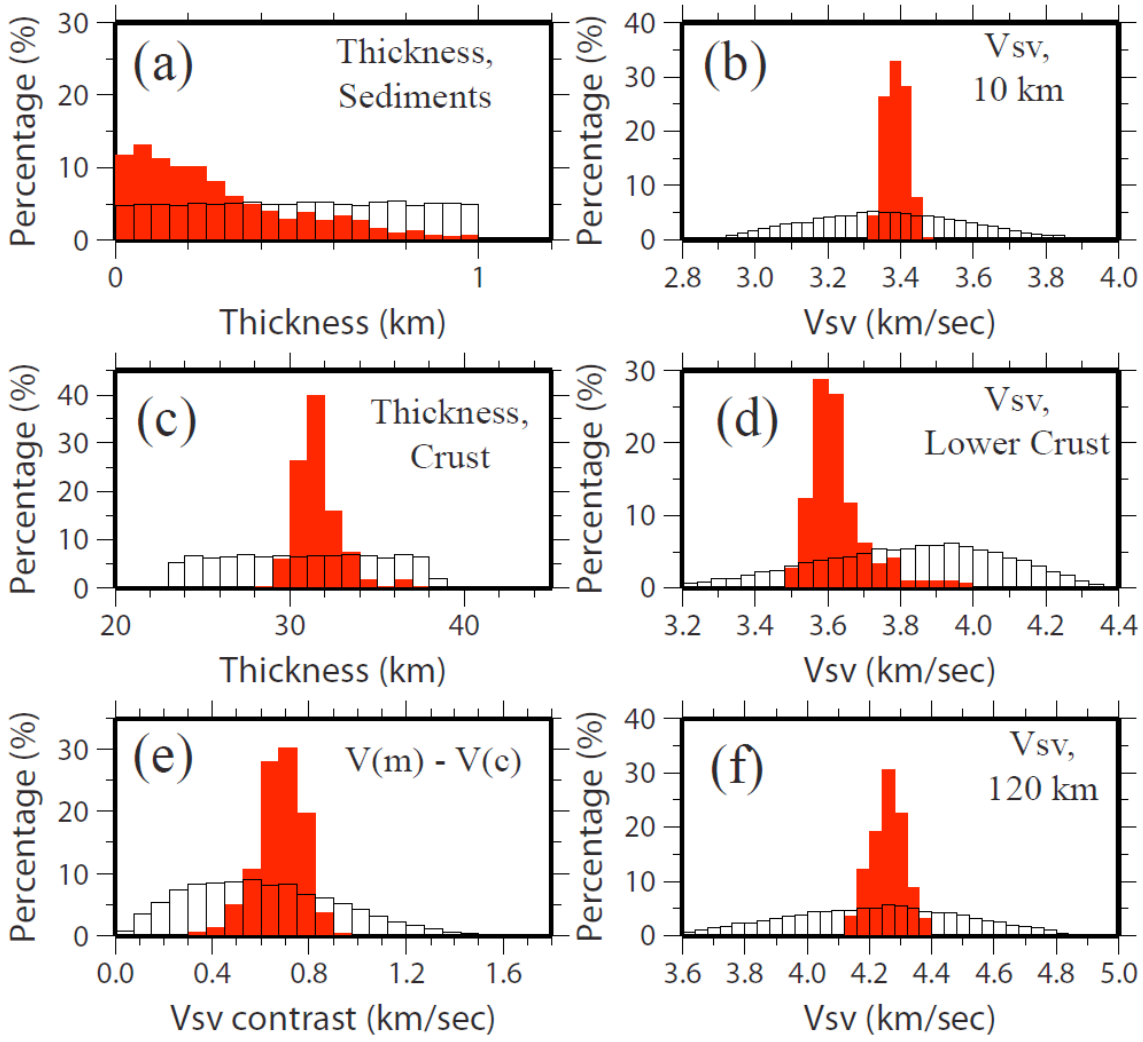

Figure 14. (a-f) The same as Fig. 11, but for the posterior distribution resulting from the joint application of surface wave phase velocities and receiver functions. Note the sharpening of distributions of parameters near the Moho in (c) – (e) compared to the distributions resulting from surface wave data alone (Fig. 11).

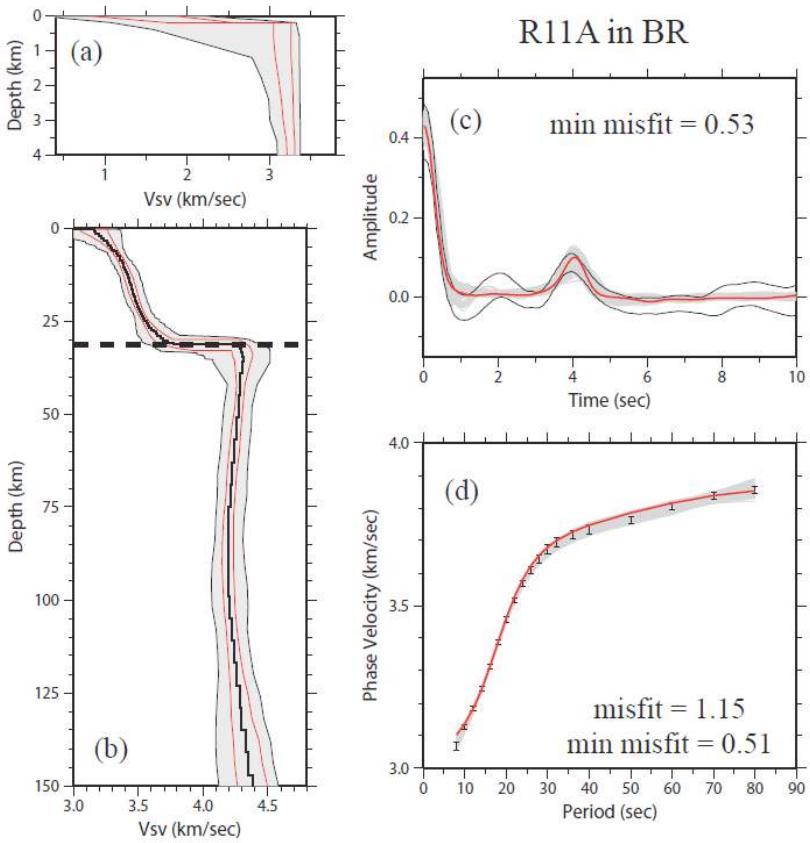

Figure 15. (a-b): The same as Fig. 12, but for results from the joint use of surface wave phase velocities and RFs at TA station R11A in the Basin and Range province. (c) The synthetic receiver functions from the accepted model ensemble are plotted with gray lines, with the best fitting receiver function shown as the red curve. The parallel black lines are the estimated uncertainty of the RF. A clear P-to-S converstion near 4 sec period is seen on this RF, necessitating a large velocity jump at a shallow Moho. (d) The surface wave dispersion from the model that best predicts the receiver function in (c) is shown with the red curve and has a misfit of 1.15. The predicted surface wave dispersion curves from all accepted models are plotted with gray lines. Of those, the best fitting one has a misfit of 0.51

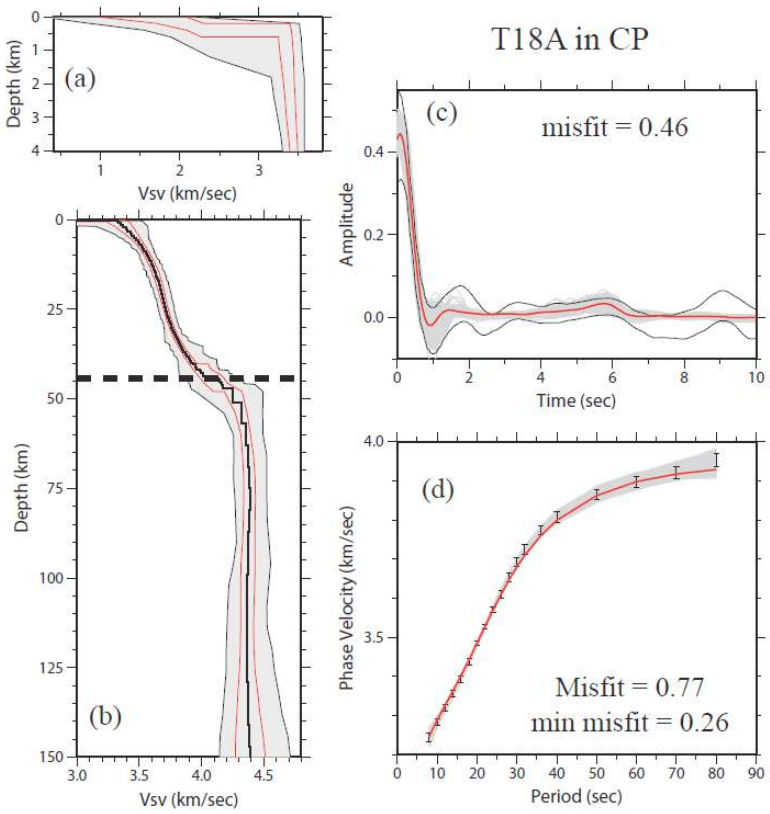

Figure 16. The same as Fig. 15, but for the result at TA station T18A located in the Colorado Plateau. In (c), misfit is for the model that fits the RF best. No clear P-to-S Moho conversion is seen on the RF, requiring a gradient Moho. In (d), misfit is for the model that fits the RF best and min misfit is for the model that fits the phase velocities best.

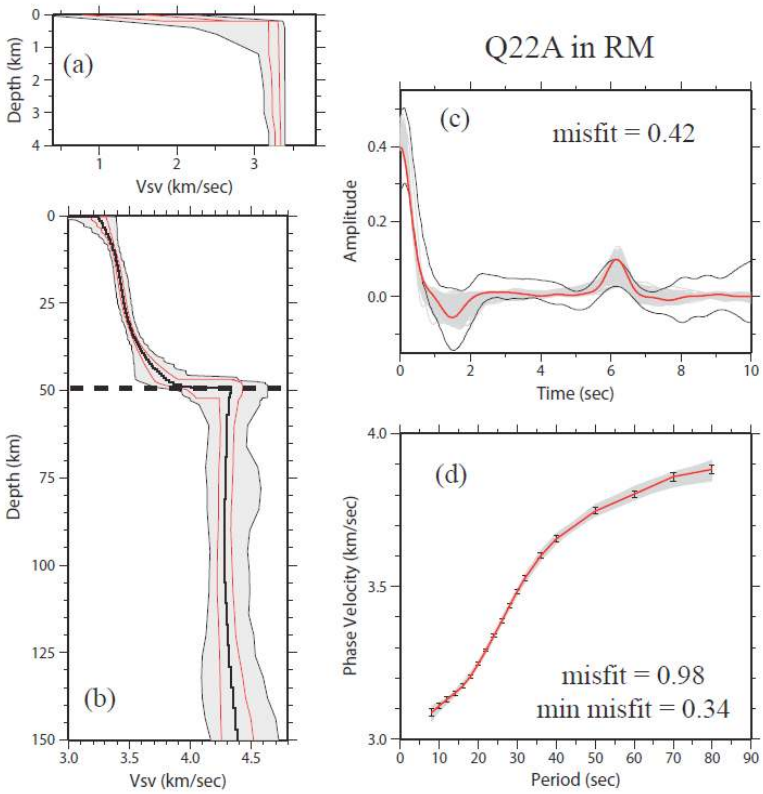

Figure 17. The same as Fig. 15, but for the result at TA station Q22A located in the Colorado Rocky Mountains. A clear P-to-S conversion is seen near 6 sec, requiring a large velocity jump at a deeper Moho than in Fig. 15.

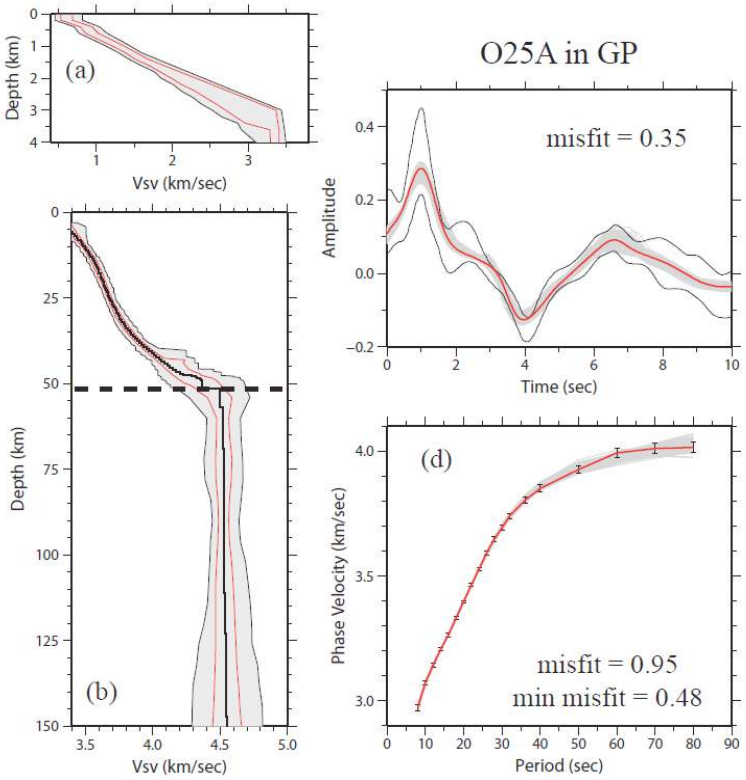

Figure 18. The same as Fig. 15, but for the result at TA station O25A located in the Great Plains. Strong sedimentary reverberations dominate the RF adding uncertainty in the location of and velocity jump at the Moho.

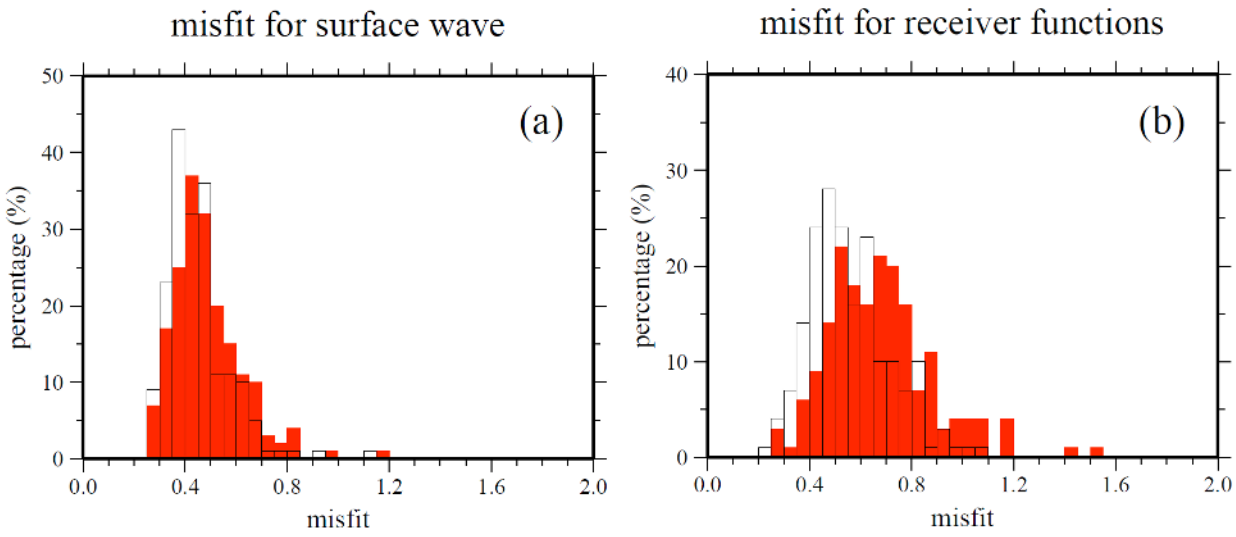

Figure 19. (a) Misfit histogram showing the misfits to surface wave data over the grid covering the area of study. The white histogram is for the model that best fits the surface wave data at each point and the red histogram is for the model that jointly best fits both surface wave and RF data. At most grid points, 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 RFs. The white histogram is for the model that best fits the RF data at each point and the red histogram is for the model that jointly best fits both data sets. At most grid points, the jointly best fitting models fit the RF data only slightly worse than the model that fits the RF data best.

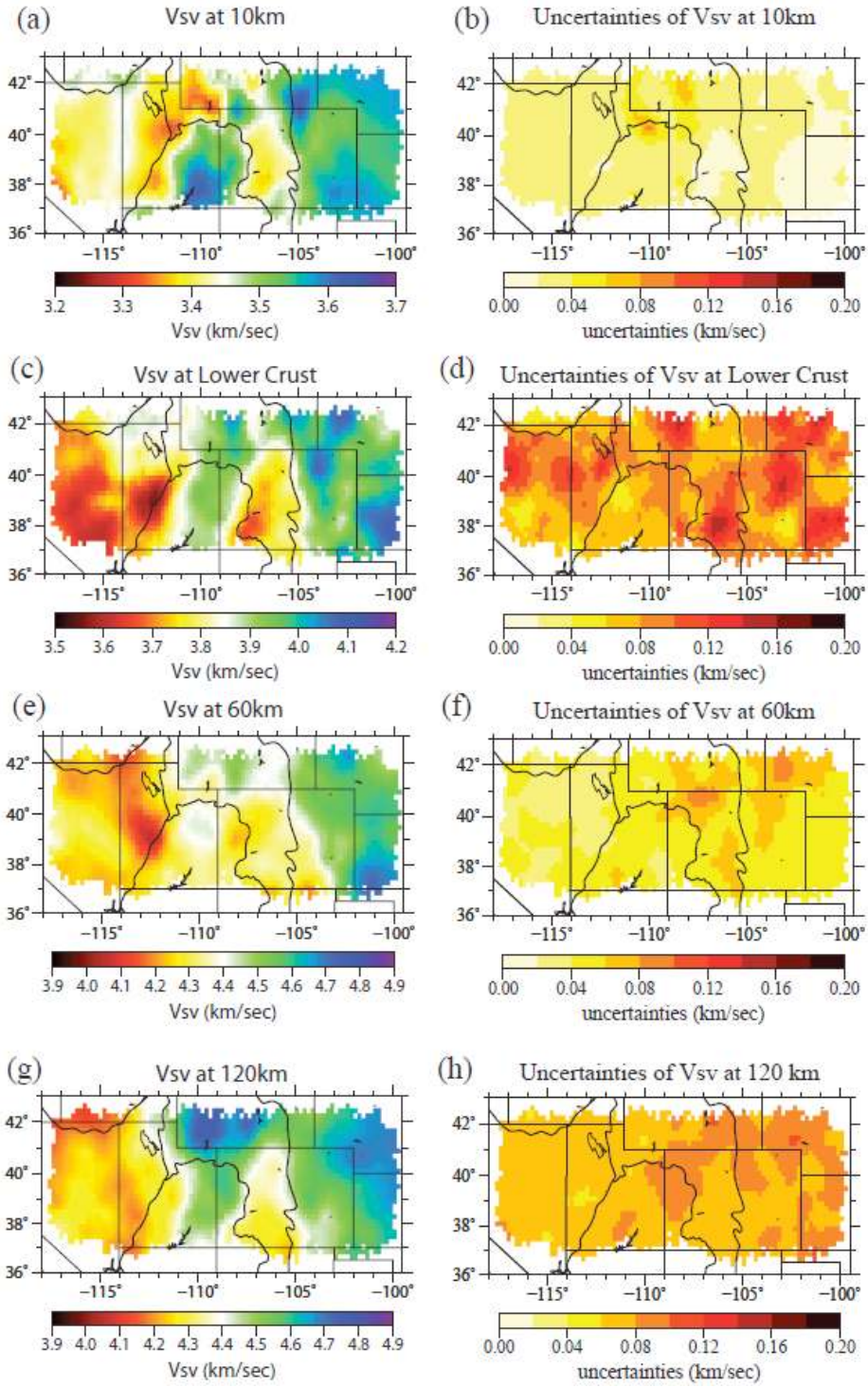

Figure 20: Map views of Vsv at different depths (left) with uncertainties (right). (a-b) Vsv at 10 km depth. (c-d) Vsv in the lower crust, 4 km above Moho. (e-f) Vsv at 60 km depth. (g-h) Vsv at 120 km depth.

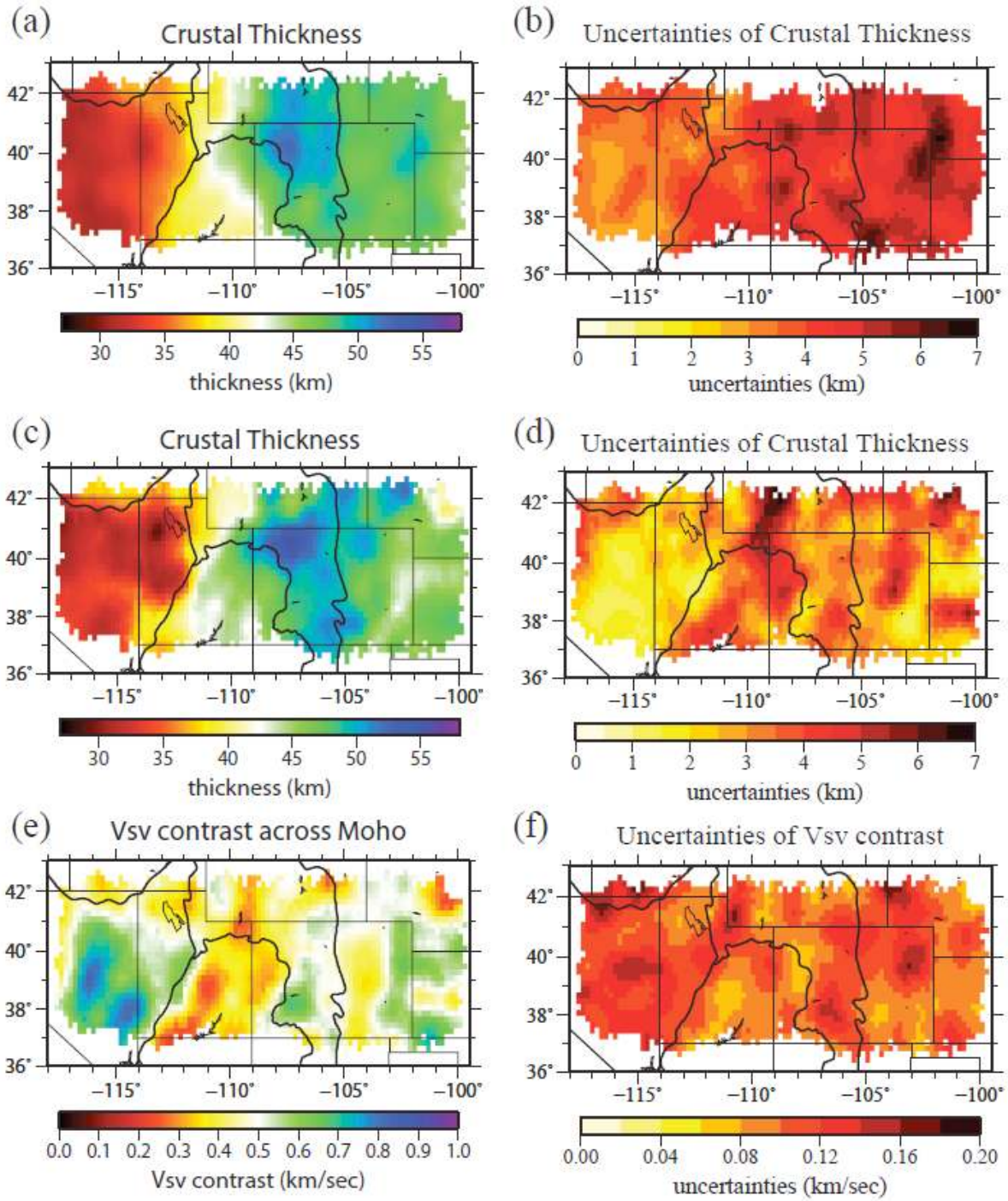

Figure 21. (a)-(b) Crustal thickness and its uncertainty resulting from use of surface wave data alone. (c)-(d) Crustal thickness and its uncertainty resulting from the joint use of surface wave and RF data. (e)-(f) Vsv contrast from 4 km below to 4 km above Moho and its uncertainty from the joint use of surface wave and RF data.

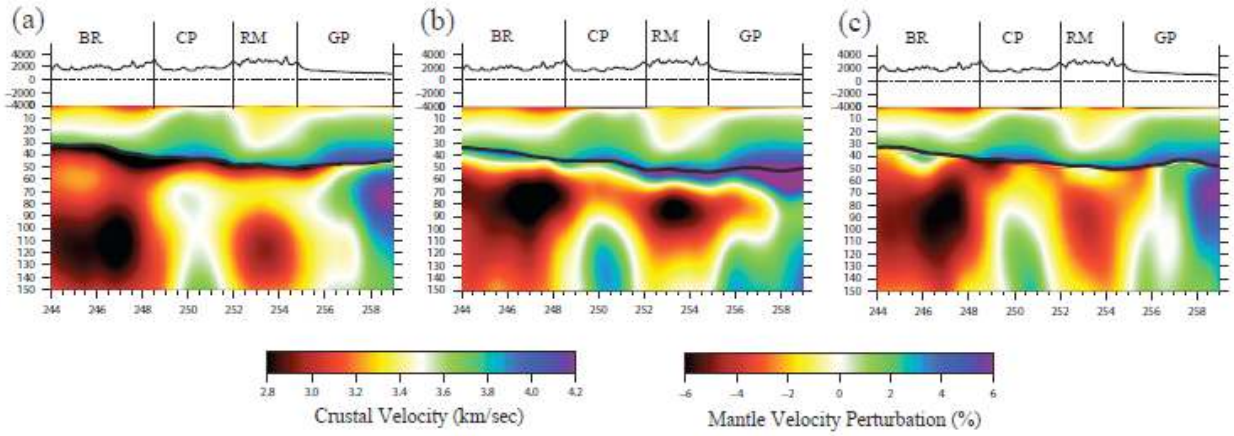

Figure 22. (a) Vsv model along transect AA' in Fig. 1 constructed using surface wave data alone with a positive gradient constraint in the uppermost mantle. (b) The same as (a), but with a negative constraint on the uppermost mantle velocity gradient. (c) Vsv model from the joint use of both surface wave and RF data with no constraint on the Vsv gradient in the upper-most mantle. Crustal structure is presented in absolute shear wave speed but mantle structure is presented as the percent perturbation relative to 4.4 km/s. Geological provinces are presented with abbreviations (Fig. 1) overlying surface topography. Mantle velocity perturbation is taken relative to 4.4 km/s.

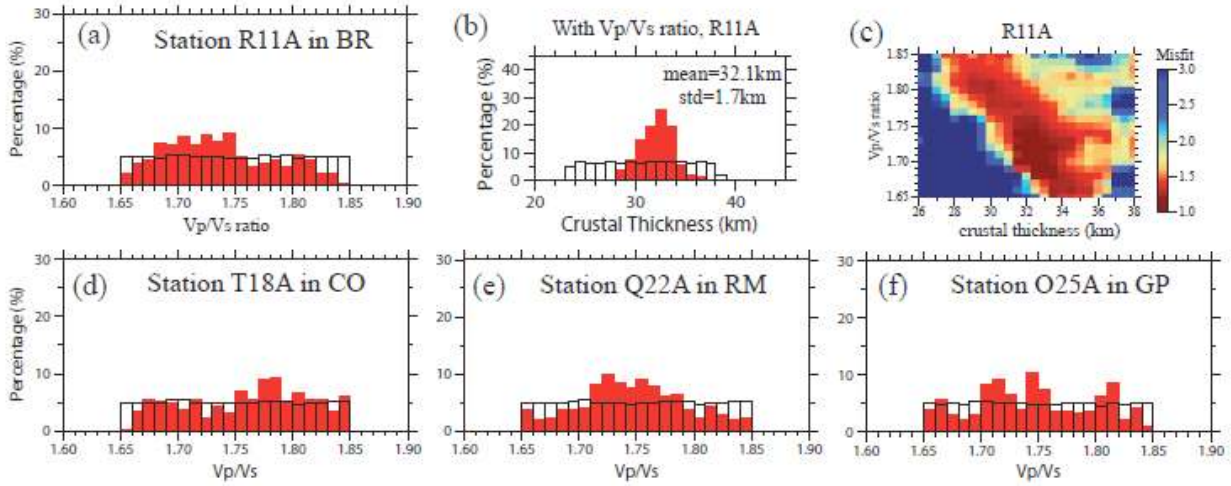

Figure 23: (a) Posterior distribution of Vp/Vs ratio when it is included as a parameter in the joint inversion. The prior distribution is shown with the unfilled box outlined in black. (b) The posterior distribution of crustal thickness for station R11A from the inversion that includes Vp/Vs as a free parameter. (c) The joint misfit (χ_{joint}) is plotted as a function of crustal thickness and Vp/Vs ratio. Trade-off between crustal thickness and the Vp/Vs ratio is clear. (d)-(f) The same as (a), but for stations T18A, Q22A and O25A, respectively.

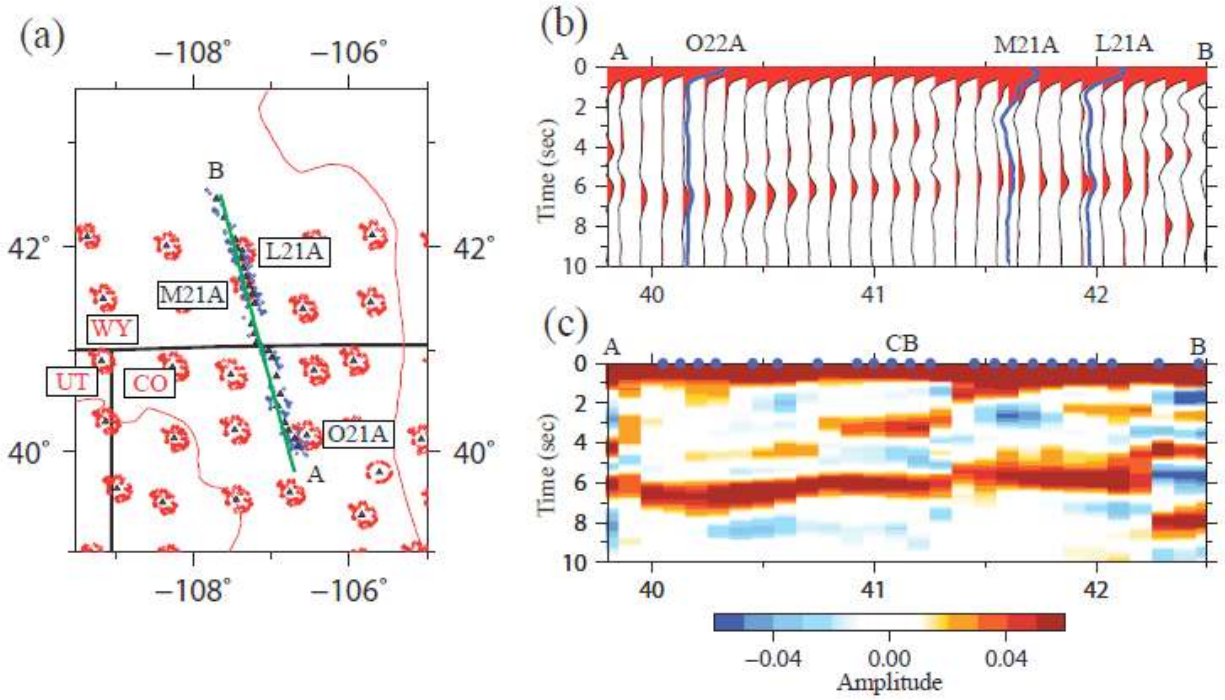

Figure 24: (a) Piercing points of the incident P-wave at the Moho for TA stations are shown with red dots. The piercing points for CD-ROM stations are shown with blue dots. The green line is the transect AB along the CD-ROM line from Colorado to Wyoming. Three TA stations near the transect are identified with names (L21A, M21A and O21A from north to south). State boundaries are outlined with black lines and red lines are geological provinces (Fig. 1). (b) Moho conversion point (MCP) stacked RFs are illustrated with red waveforms along transect AB in (a). For comparison, single-station processed RFs ($A_0(t)$) are shown with blue waveform for the three stations shown in (a). (c) The smoothed image of the red RFs in (b). Blue dots indicate the location of the CD-ROM stations. The location of the Cheyenne Belt is marked as CB.