AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Solid Earth

RESEARCH ARTICLE
10.1002/2017JB014968

Key Points:
- Fixing crustal discontinuities can lead to erroneous results in modeling upper mantle structure
- Treating the V_p/V_s as an unknown does not improve results and cannot prevent the appearance of spurious layers
- We confirm the presence of a sharp MLD at 70 km atop a ~50 km thick low shear velocity layer in the middle of the NA craton

Supporting Information:
- Supporting Information S1

Correspondence to:
C. Roy,
corinna.roy@berkeley.edu

Citation:
Roy, C., & Romanowicz, B. A. (2017). On the implications of a priori constraints in transdimensional Bayesian inversion for continental lithospheric layering. *Journal of Geophysical Research: Solid Earth*, *122*, 10,118–10,131. https://doi.org/10.1002/2017JB014968

Received 13 SEP 2017
Accepted 14 OCT 2017
Accepted article online 19 OCT 2017
Published online 13 DEC 2017

©2017. American Geophysical Union. All Rights Reserved.

On the Implications of A Priori Constraints in Transdimensional Bayesian Inversion for Continental Lithospheric Layering

C. Roy[1] and B. A. Romanowicz[1,2,3]

[1]Berkeley Seismological Laboratory, University of California, Berkeley, CA, USA, [2]Institut de Physique du Globe de Paris, Paris, France, [3]Physics of the Earth's Interior, Collège de France, Paris, France

Abstract Monte Carlo methods are powerful approaches to solve nonlinear problems and are becoming very popular in Earth sciences. One reason being that, at first glance, no constraints or explicit regularization of model parameters are required. At second glance, one might realize that regularization is done through a prior. The choice of this prior, however, is subjective, and with its choice, unintended or undesired extra information can be injected into the problem. The principal criticism of Bayesian methods is that the prior can be "tuned" in order to get the expected solution. Consequently, detractors of the Bayesian method could easily argue that the solution is influenced by the form of the prior distribution, which choice is subjective. Hence, models obtained with Monte Carlo methods are still highly debated. Here we investigate the influence of a priori constraints (i.e., fixed crustal discontinuities) on the posterior probability distributions of estimated parameters, that is, vertical polarized shear velocity V_{SV} and radial anisotropy ξ, in a transdimensional Bayesian inversion for continental lithospheric structure. We follow upon the work of Calò et al. (2016), who jointly inverted converted phases (*P* to *S*) without deconvolution and surface wave dispersion data, to obtain 1-D radial anisotropic shear wave velocity profiles in the North American craton. We aim at verifying whether the strong lithospheric layering found in the stable part of the craton is robust with respect to artifacts that might be caused by the methodology used. We test the hypothesis that the observed midlithospheric discontinuities result from (1) fixed crustal discontinuities in the reference model and (2) a fixed V_p/V_s ratio. The synthetic tests on two Earth models show that a fixed V_p/V_s ratio does not introduce artificial layering, even if the assumed value is slightly wrong. This is an important finding for real data inversion where the true value is not always available or accurate. However, fixing crustal discontinuities can lead to the introduction of spurious layering, and this is not recommended. Additionally, allowing the V_p/V_s ratio to vary does not help preventing that. Applying the modified approach resulting from these tests to two stations (FRB and FCC) in the North American craton, we confirm the presence of at least one midlithospheric low-velocity layer. We also confirm the difficulty of consistently detecting the lithosphere-asthenosphere boundary in the craton.

1. Introduction

The construction of crustal and uppermost mantle seismic velocity models over extended continental regions is critical for our understanding of the formation of continents and the thermal and compositional structure of the lithosphere. Of particular interest in this context is the presence and depths of discontinuities within the lithosphere and the lateral variations thereof, as well as the sharpness of the lithosphere-asthenosphere boundary (LAB).

The presence of layering within the continental lithosphere of the North American craton has been known for a long time (Hales, 1969) and confirmed in many recent studies: for example, Abt et al. (2010) used *Ps* and *Sp* receiver functions (RFs) at 93 permanent seismic stations on the North American craton, while Rychert et al. (2010) exploited only *Ps*, and Kumar et al. (2012) only *Sp* RFs but for a larger number of U.S. array stations. On the other hand, Ford et al. (2016) focused on the layering in Wyoming and the Superior Provinces using anisotropic *Ps* RFs. In contrast, Yuan and Romanowicz (2010) combined long-period seismic waveforms and *SKS* splitting data in full waveform inversion and revealed lithospheric layering manifested by a rapid change in the fast axis direction of anisotropy.

Recently, Calò et al. (2016) imaged discontinuities in the upper mantle of the North American craton by modeling surface wave dispersion data and converted *Ps* phases using a Markov chain Monte Carlo transdimensional

method (Malinverno, 2002), while Bodin et al. (2016) added *SKS* waveforms to this approach. They presented evidence for strong lithospheric layering in the stable part of the North American craton, marked by at least two midlithospheric discontinuities (MLDs) at intermediate depths. However, they used an averaged V_p/V_s ratio provided by other RF studies and assumed this value to be constant in the crust and upper mantle. Given this strong assumption, the question came up, whether the observed MLDs are real features or artifacts resulting from the assumptions made, such as effects of Moho multiples due to erroneous constraints on crustal discontinuity depths imposed in the prior of V_{SV}.

Indeed, Monte Carlo approaches are becoming more popular than iterative linearized inversion techniques, to solve nonlinear problems such as those encountered in studies of Earth's seismic structure. Both approaches combine a priori information about the model, the data, and their errors. In the iterative linearized inversion approach, the problem is solved via optimization of a data misfit function. The choice of a starting model combined with necessary regularization may constrain the solution to a local minimum of the objective function. In contrast, in Monte Carlo-based approaches, this problem is avoided in that the solution is obtained by generating an ensemble of candidate models, described by probabilistic sampling of a posterior probability density function.

Monte Carlo methods are widely used in seismology to investigate crust and upper mantle structure: for example, Gouveia and Scales (1998) applied them for inversion of surface seismic field data and take into account near-surface effects, uncertainty in amplitude scaling, and model parametrization errors. Chen et al. (2007) took advantage of them for modeling a 1-D reservoir in the Gulf of Mexico by joint inversion of seismic and electromagnetic data. Arnold and Townend (2007) estimated tectonic stress using Monte Carlo methods, which also became popular for the inversion of receiver functions (Agostinetti & Malinverno, 2010; Lucente et al., 2005). For the first time, Käufl et al. (2013) performed a feasibility study of a probabilistic 3-D full waveform inversion for the Australian continent. Young et al. (2013) imaged the 3-D shear velocity structure of the Tasmanian crust using a Bayesian transdimensional inversion. On the other hand, Pachhai et al. (2014) explored ultralow velocity zone parameters at the base of the mantle under the Philippine Sea, while Wang et al. (2015) inverted simultaneously for fault slip and stress drop and Mustać and Tkalčić (2016) for moment tensors of point sources using Monte Carlo methods. Quite recently, Wirth et al. (2017) used a Markov chain Monte Carlo (MCMC) with Gibbs sampling method for the interpretation of radial and transverse component receiver functions. An overview of applications to other fields in Earth science is given by Sambridge and Mosegaard (2002).

The Monte Carlo approach also presents some drawbacks: The Bayesian formulation allows to account for prior knowledge if that information can be expressed as a probability distribution $p(m)$. However, the choice of the prior is subjective, and with its choice, unintended or undesired extra information can be injected into the problem. The principal criticism of the Bayesian approach is that users often "tune" the prior in order to get the expected solution (Bodin et al., 2012; Scales & Snieder, 1997; Scales & Tenorio, 2001). Consequently, detractors of the Bayesian method could easily argue that the solution is influenced by the form of the prior distribution, the choice of which is subjective. Finally, theoretical errors are not always easily quantifiable.

Here we follow up on the work of Calò et al. (2016) and investigate the robustness of the lithospheric layering found in that study. The objective is to explore how different assumptions influence the posterior distributions and to confirm or disprove the MLDs reported by Calò et al. (2016). To investigate this question, we extend the multiobservable probabilistic inversion method of Calò et al. (2016) inverting jointly Love and Rayleigh wave dispersion data and converted phases (*P* to *S*) for V_{SV}, and radial anisotropy parameter $\xi = (V_{SH}/V_{SV})^2$. In addition, we include the V_p/V_s ratio as an unknown parameter in the crust and upper mantle. Previously, the latter parameter was included in multiobservable probabilistic inversion studies of the crust: for the crustal structure of the Arizona transition zone (Tork Qashqai et al., 2016), below South China (Shan et al., 2014), and North China by Guo et al. (2016). In section 4 we present the results of a series of synthetic tests, where we evaluate the impact of several assumptions in the inversion: a fixed or variable V_p/V_s ratio, fixed crustal discontinuities in the reference model, and a combination of both assumptions. We perform these tests with noise added to the synthetic data. We apply parallel tempering (Sambridge, 2014) to ensure fast convergence. Then we perform the same analysis for real data, as described in section 5, for one station (FRB), already presented in Calò et al. (2016) and another station (FCC) of the Canadian National Seismograph Network located in the North American craton. We consider a second station to bring out regional variations in the lithospheric layering of the North American craton.

2. Multiobservable Probabilistic Inversion Method

2.1. Bayesian Inference

In a Bayesian approach all information is described in probabilistic terms to calculate the posterior probability distribution, which describes the probability of having a discontinuous model m for the observed data d. The posterior is defined through Bayes' theorem (Jaynes, 2003) that combines prior knowledge about the model, that is, the prior probability $p(m)$ and a likelihood function $p(d|m)$, that is, the probability of observing the data, for a given particular model:

$$p(m|d) \propto p(d|m)\, p(m), \tag{1}$$

where m is the model vector and d the data vector. The posterior represents how our prior knowledge of the model parameters is improved by the data: If they are the same, the data add no new information. In our problem, the number of layers in the model is an unknown and a random variable itself; therefore, the posterior becomes a transdimensional function (Green, 2003).

We use the reversible jump Markov chain Monte Carlo (rj-MCMC) algorithm (Green, 1995, 2003) to sample the posterior. The rj-MCMC algorithm is an iterative method, where models are generated in a chain. The starting point of this chain is chosen randomly, and each model is a randomly perturbed version of the previous one. Samples in the first part of the chain, called burn-in period, are discarded, because the random walk is assumed to be stationary after a certain number of sampled models. We follow the implementation presented in Bodin et al. (2012, 2016) for joint inversion of RFs without deconvolution and surface waves but expand the algorithm for the case of an unknown V_p/V_s ratio. In the following sections, we define the likelihood and the prior.

2.2. The Likelihood Function

The likelihood function involves the difference between the model simulation and observed data. It determines how well a given model with a particular set of parameter values can reproduce the observed data. The larger the value of the likelihood, the closer the model is to the observed data. Assuming that the measurement errors are independent, the likelihood function for the surface waves is given by (Tarantola, 2005)

$$p(d|m) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} * e^{\left(\frac{-\phi(m)}{2\sigma^2}\right)}, \tag{2}$$

where n is the number of data points and $\phi(m)$ the misfit function. The measurement data errors σ are unknown and estimated jointly with the model parameters. This approach is also referred to as hierarchical Bayes (Malinverno & Briggs, 2004; Rosas-Carbajal et al., 2013). The likelihood increases with decreasing data misfit and the proposed model is more likely to explain the data. The misfit function $\phi(m)$ for the surface waves is simply given by the difference between observed and modeled dispersion curves, that is, $\phi(m) = \sum_i \left(d_i^{\text{obs}} - d_i^{\text{est}}\right)^2$.

We use the cross-convolution misfit function for the converted phases as introduced by Bodin et al. (2014):

$$\phi(m,t) = v(t,m) * H(t)_{\text{obs}} - h(t,m) * V(t)_{\text{obs}}, \tag{3}$$

where $V_{\text{obs}}(t)$ and $H_{\text{obs}}(t)$ are the observed vertical and radial component of the seismograms and $v(t,m)$ and $h(t,m)$ are the vertical and radial impulse response functions calculated for the model m. For real data, v and h are a stack of waveforms coming from similar distances and back azimuths. The major advantage of this approach compared to the standard RF method is the avoidance of the numerically unstable deconvolution process and related errors. It removes the source and distance effects in the seismograms and allows us to isolate the effect of the structure below the recording station. In a rigorous Bayesian framework, this definition of the likelihood for the converted phases (equation (3)) is not exact (Dettmer et al., 2015), because it describes the distribution of residuals rather than the distribution of the data. The exact likelihood for the residuals can be found in Bodin et al. (2016). Nevertheless, this approximation of the likelihood by a distribution of residuals was also used by Stähler and Sigloch (2014) for Bayesian moment tensor inversion.

2.3. The Prior

In a Bayesian framework, the role of regularization is played by the prior distribution, which limits the space of possible Earth models by giving higher probability to those that agree more closely with prior knowledge. In our problem, the prior is what we think is reasonable for shear wave velocity and radial anisotropy according

to previous studies. An important point in our approach is that we do not put any constraints on the number of layers in the model; indeed, it is an unknown and a random variable itself. Therefore, the posterior becomes a transdimensional function (Green, 2003). Assuming that the parameters we want to invert for are independent, the prior is then the product of the priors of each parameter. We follow here the prior distribution of Bodin et al. (2016), but we also allow for a variable V_p/V_s ratio and change the prior distribution for the model m with k layers accordingly to

$$p(m|k) = \underbrace{p(k)}_{\substack{\text{prior}\\ \text{number of}\\ \text{layers}}} \underbrace{p(z|k)}_{\substack{\text{prior}\\ \text{depth of}\\ \text{layers}}} \underbrace{p\left(V_s|k\right)}_{\text{prior } V_s} \underbrace{p\left(V_p/V_s|k\right)}_{\text{prior } V_p/V_s \text{ ratio}} \underbrace{p(\xi|k)}_{\substack{\text{prior radial}\\ \text{anisotropy}}}. \tag{4}$$

The priors for V_p/V_s, ξ, and V_{SV} are uniform distributions over the range of possible values and integrate to one. Assuming that each component V_p/V_{s_i} of the vector V_p/V_s is uniformly distributed over the interval $[V_p/V_{s_{\min}}, V_p/V_{s_{\max}}]$, the prior distribution for the V_p/V_s ratio is given by

$$p\left(V_p/V_{s_i}|k\right) = \frac{1}{\Delta V_p/V_s}, \tag{5}$$

where $\Delta V_p/V_s = V_p/V_{s_{\max}} - V_p/V_{s_{\min}}$.

3. Data

Images of the anisotropic crust and upper mantle structure are usually obtained by analyzing seismic observables such as surface wave dispersion curves, *SKS* wave splitting, and converted phases. Since they sample the Earth differently in terms of wave propagation, they are sensitive to various length scales leading to different resolution properties. Converted phases are sensitive to relative changes in *S* velocity and can resolve the depth of upper mantle discontinuities well, whereas surface waves can be used to constrain volumetric heterogeneity and anisotropy. These data sets are often used and interpreted separately and result in models that are frequently incompatible.

To overcome this limitation, we jointly invert surface wave dispersion data of Love and Rayleigh waves, together with high-frequency body wave converted phases (*P* to *S* converted phases). The dispersion data set consists of azimuthally averaged phase velocities for Love and Rayleigh waves in the period range 16–150 s (Shapiro & Ritzwoller, 2002) and group velocities in the period range 25–250 s (Ekström, 2011).

4. Synthetic Tests

In this section, we describe synthetic tests performed to answer the following four questions:

1. How does a variable or fixed V_p/V_s ratio influence the posterior probability distributions for the velocity of vertically polarized shear waves (V_{SV}) and radial anisotropy, as described by the parameter $\xi = (V_{SH}/V_{SV})^2$?
2. Does a fixed V_p/V_s ratio influence the posteriors toward an erroneous solution?
3. If this is the case, can it be prevented when the V_p/V_s ratio is an unknown?
4. Does fixing the depth of the Moho in the reference model lead to the appearance of artificial layers in depth?

To answer the above questions, we computed synthetic data for two known, synthetic upper mantle Earth models. The first model is a simple 1-D velocity model with only one crustal layer. The second model consists of several anisotropic layers in the upper mantle with alternating high and low velocities. We calculated synthetic body wave converted phases (*P* to *S* converted phases) using the reflectivity scheme of Levin and Park (1998), which returns the impulse response of an incoming planar wave to a stack of anisotropic layers. The impulse response was then convolved with a box car function to form a synthetic waveform. We acknowledge that these synthetic waveforms are far from being realistic, but the aim here is to show the ability of our approach to recover the synthetic model in an ideal case. The length of the time window of 51 s is long enough to include the converted phase, as well as the later arriving Moho multiples for an average Moho depth of 40 km.

Surface wave dispersion curves were computed by normal mode summation in a spherical Earth (Takeuchi & Saito, 1972) for periods between 25 and 250 s for both Love and Rayleigh waves, although group velocity data are available for periods of 16–150 s. Eventually, we added random white Gaussian noise, that is, random noise, to the synthetic data after their computation. We acknowledge that real noise is correlated, and in our study the noise is uncorrelated.

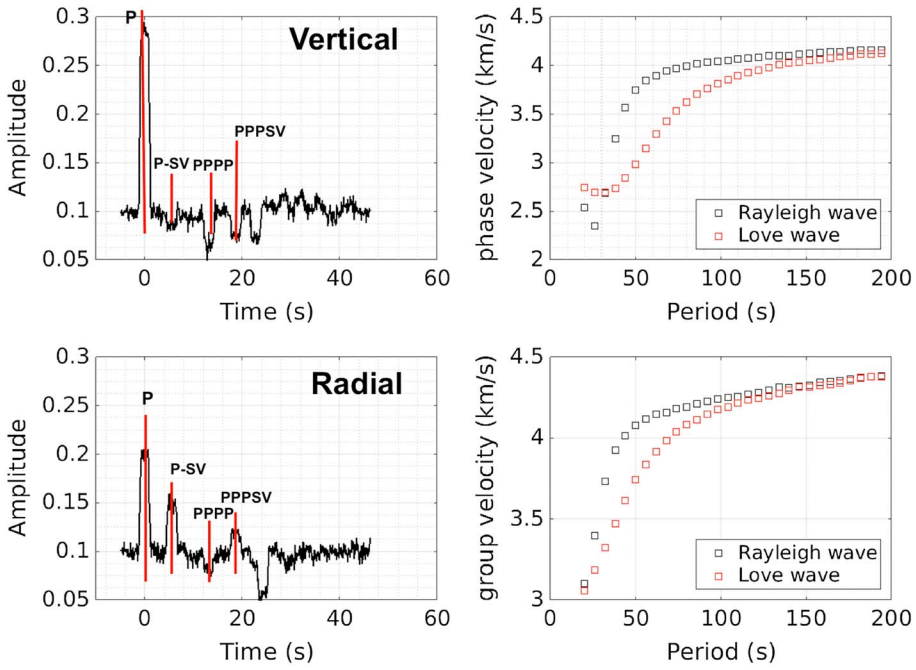


Figure 1. (Left) Synthetic body waves for an incoming planar *P* wave (slowness = 0.075 km s^{-1}) generated by a point source in time, traversing the layered Earth model shown as black line in Figure 2. Approximately 3–5% of *P* arrival amplitude was added to the synthetics as Gaussian white noise. (Right) Dispersion curves for Love and Rayleigh waves of Ekström (2011) and Shapiro and Ritzwoller (2002) (with 0.1 km/s noise added).

We assumed a vertical transversely isotropic model, where the radial anisotropy is described by $\xi = \left(V_{SH}/V_{SV}\right)^2$. The other three parameters of radial anisotropy and density are parametrized in terms of Voigt average isotropic *S* and *P* velocity using the empirical scaling laws of Montagner and Anderson (1989). We then applied the rj-MCMC algorithm with parallel tempering (Sambridge, 2014) to obtain the posterior probability density distributions for shear wave velocity V_{SV} and radial anisotropy ξ.

4.1. Simple Model

The first synthetic model has only one discontinuity at 40 km depth, representing the Moho discontinuity over an infinite half-space. The shear wave velocity increases from 3.2 km/s to 4.7 km/s at this discontinuity. Crust and mantle are both anisotropic, with $\xi = 0.8$. The synthetic converted phase and dispersion curves with ~3–5% white noise are shown in Figure 1. The waveforms contain the *P* to *S* converted phases, as well as the later arriving Moho multiples. We performed three different tests using this synthetic data set:

1. First, we applied the rj-MCMC to the synthetics with ~3–5% Gaussian white noise and inverted for V_{SV} and ξ, assuming a fixed V_p/V_s ratio and do not fix the Moho depth in the reference model.
2. Using the data with noise, we allowed the V_p/V_s ratio to vary in the rj-MCMC, and we do not fix the Moho depth.
3. This is the same as (2) but we fix the Moho discontinuity in the reference model at a wrong depth, to investigate its effect on the posterior distribution.
4. This is the same as (3) but with a variable V_p/V_s ratio.

We define our model $V_{SV} = V_0 + \delta V_{SV}$, where V_0 is a layered reference model that remains fixed in the inversion and δV_{SV} is a perturbation we want to invert for. This is illustrated in Figure S1 in the supporting information that shows V_0 with one and δV_{SV} with two layers and a half space, and consequently, in this case, V_{SV} will have three layers. The prior distribution of δV_{SV} is uniformly distributed over the range $[\delta V_{min} \delta V_{max}]$. The simple reference model that we used for the inversion of synthetics is summarized in Table 1. When we fix the Moho discontinuity in Test 3 or other crustal discontinuities in the later described tests, that means that we fix them in the reference model V_0.

The posterior solutions for V_{SV} and radial anisotropy ξ of these tests are shown in Figure 2. For all parameters (e.g., number of layers and noise parameters), we assumed uniform prior distributions in the rj-MCMC

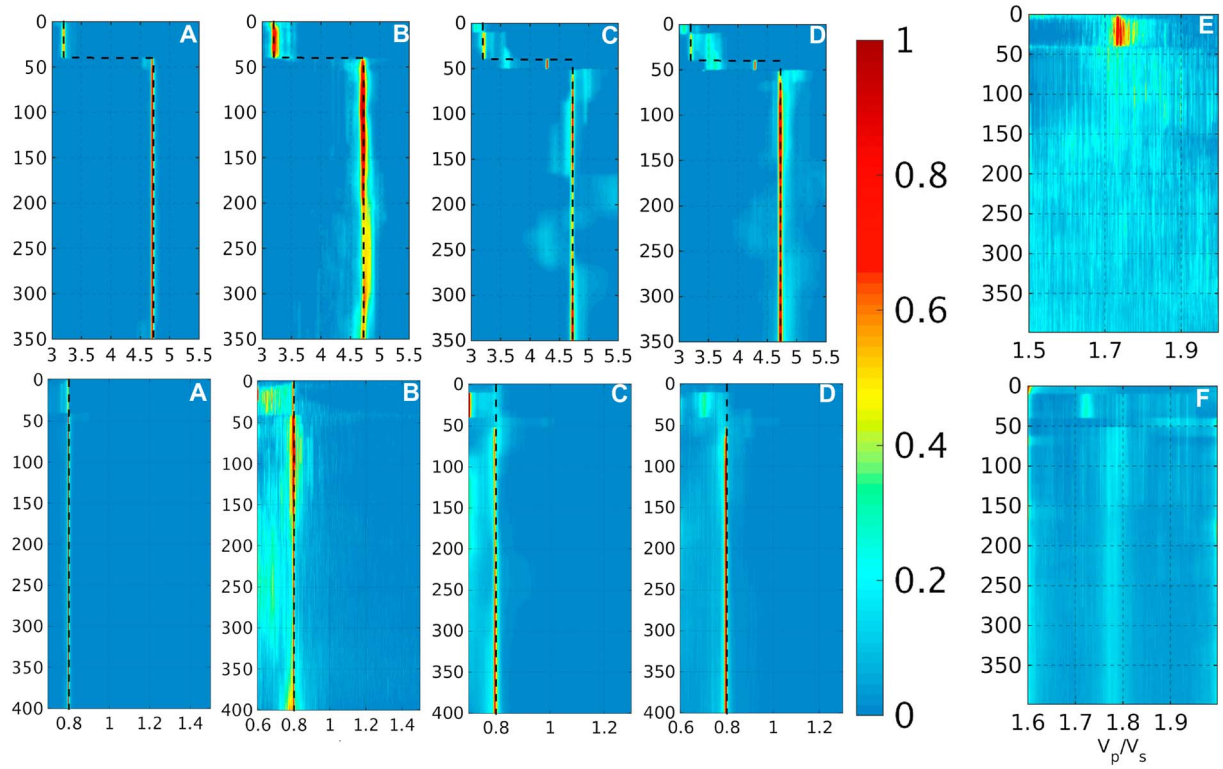


Figure 2. Posterior probability distributions for (top row) V_{SV} and (bottom row) ξ and for the V_p/V_s ratio using the synthetic data in Figure 1 with random Gaussian noise. The black dotted line represents the true model. (a) (Test 1) Fixed V_p/V_s ratio in the rj-MCMC. (b) (Test 2) The V_p/V_s ratio is treated as an unknown variable in the rj-MCMC. (c) (Test 3) The V_p/V_s ratio is fixed and a Moho at the wrong depth is implemented in the prior distribution for V_{SV}. (d) (Test 4) The V_p/V_s ratio is treated as an unknown variable, and a Moho at the wrong depth is implemented in the prior distribution for V_{SV}. (e and f) Probability distribution of the V_p/V_s ratio corresponding to the results shown in Figures 2b and 2d (Test 3) and in 2c and 2f (Test 4).

with bounds set to 3–50 for the number of layers, 0–0.05 m/s and 0–4 km/s for the noise in the converted phases and the surface waves, 0.6–1.5 for ξ, and 2.7–5.7 km/s for V_{SV}. In the parallel tempering algorithm, we used four chains per processor at different temperatures and allowed exchanges at all temperature levels. An example of the uniform priors compared to their posterior distribution for Test 1 is shown in Figure S4 for the number of layers and the noise hyperparameter.

In the first simulation, the V_p/V_s ratio in the rj-MCMC was fixed to the true value used to generate the synthetics. The true model represented by the black line is very well recovered in both, V_{SV} and ξ (Figure 2a) and the number of layers, as well as the misfits and the noise parameters are converged (Figure 3).

The variable V_p/V_s ratio in Test 2 broadens the posterior probability distribution for both, V_{SV} and ξ, but the true model is still recovered and the parameters are converged. In Figures 2c and 2d we clearly see that fixed crustal discontinuities at a wrong depth introduce artificial layers between 150 km and 250 km depth,and that a variable V_p/V_s ratio can help to reduce this effect a little but not prevent it. However, our computations show little sensitivity to the V_p/V_s ratio (Figures 2e and 2f), mainly in the crust. The best result is Figure 2a, because the probability distribution is well centered around the true model and does not broaden with depth.

Figure 3 displays the average convergence of the misfit in both data sets, as well as the convergence of the noise parameters and the number of layers for the four tests. Clearly, all four parameters are converged and stable and do not vary significantly with sampled models any more in all tests. In particular, the number of layers is stable over 5,000 sampled models (Test 1) and 20,000 (Test 4).

Table 1
Reference Model Kept Fixed in the Inversion for the Synthetic Data

Depth (km)	V_{SV} (km/s)	ξ	V_p/V_s
0	4.2	1	1.77
1,149	6.61	1	1.77

Note. In later simulations we added crustal discontinuities to this model.

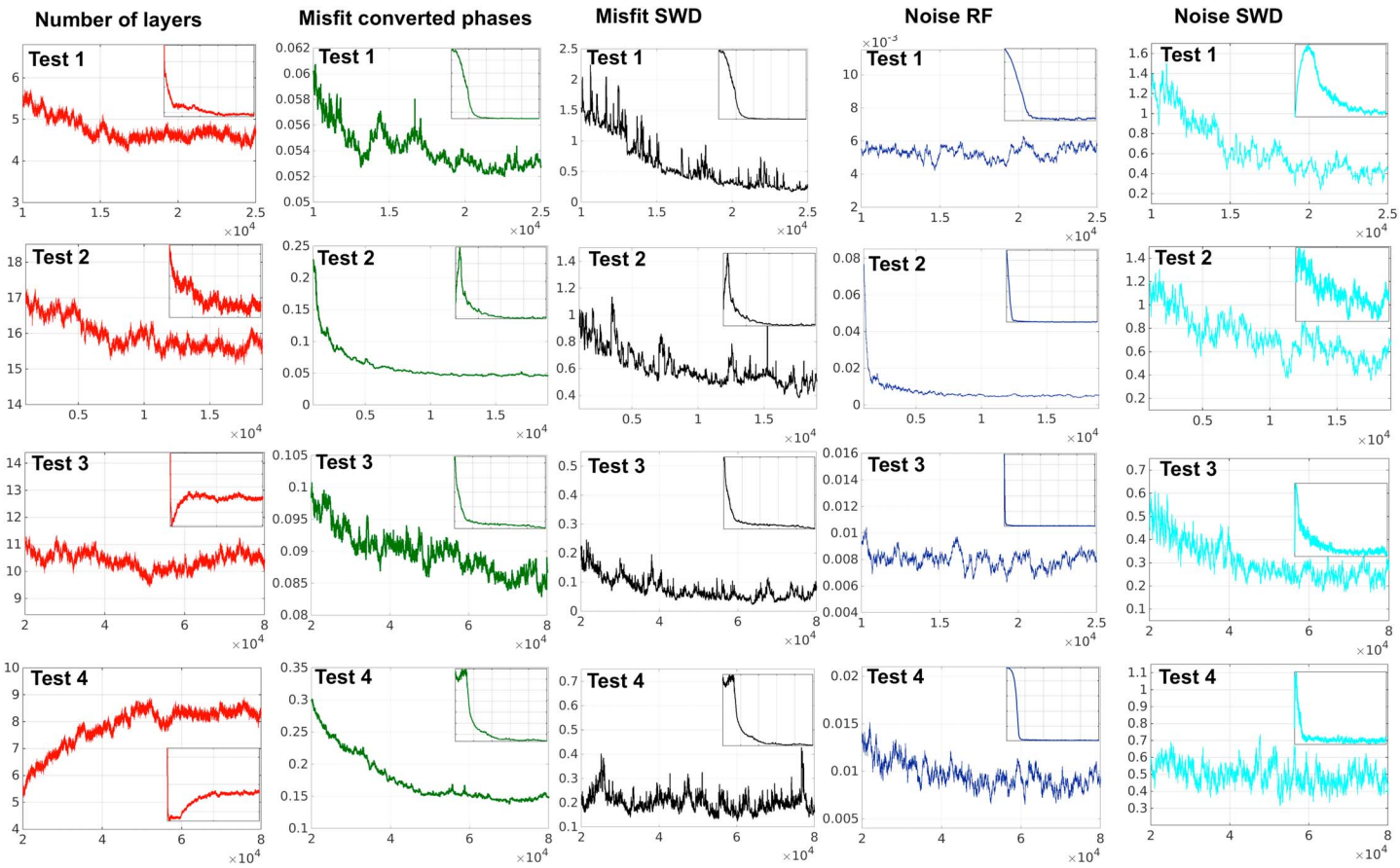


Figure 3. Convergence of the average number of layers, the noise parameters and the misfit as function of sampled models for the computations in Figure 2 after the burn-in phase. The inlay shows the parameters for all sampled models.

4.2. Complex Model

The previous results on a relatively simple synthetic model show that the true model can be recovered even if noise is added to the synthetics and the Moho is fixed at the wrong depth. However, the previous model is very simple and far from reality. We now assess a more complex one with eight layers, seven of which are anisotropic. The V_p/V_s ratio has the same value for all depths. The V_{SV} profile chosen is similar to the upper mantle structure found by Calò et al. (2016) under stations on the North American craton, that is, FRB, with several layers between 100 and 200 km depth, where they found MLDs. The values for the radial anisotropy here are not related to any known structure in the upper mantle of the North American craton and are chosen arbitrarily. It also has been shown by Calò et al. (2016) that the radial anisotropy changes at interfaces, although this change was not clear everywhere. Here the radial anisotropy varies at every discontinuity in order to see how well this can be recovered. We assume uniform prior distribution for all parameters, with the same bounds as for the simple model, as we added the same amount of noise to the synthetics. We applied parallel tempering and allowed exchanges between all temperature levels.

We performed the same four simulations as previously for the simple model but added a fifth, where we fixed the V_p/V_S ratio to a slightly too high value. The results of these tests are summarized in Figure 4.

The best result is clearly Figure 4a, were we fixed the V_p/V_s ratio to the true value and we did not fix any crustal discontinuities. We observe here again that a variable V_p/V_s ratio introduces more uncertainty and results in loss of resolution with depth. Clearly, we can see the artificial layers introduced between 50 and 100 km depth by fixing the crustal discontinuity at the wrong depth (Figures 4c and 4d), and a variable V_p/V_s ratio cannot prevent this. The radial anisotropy is not resolved well in any case, and we are only able to recover the true model in the crust.

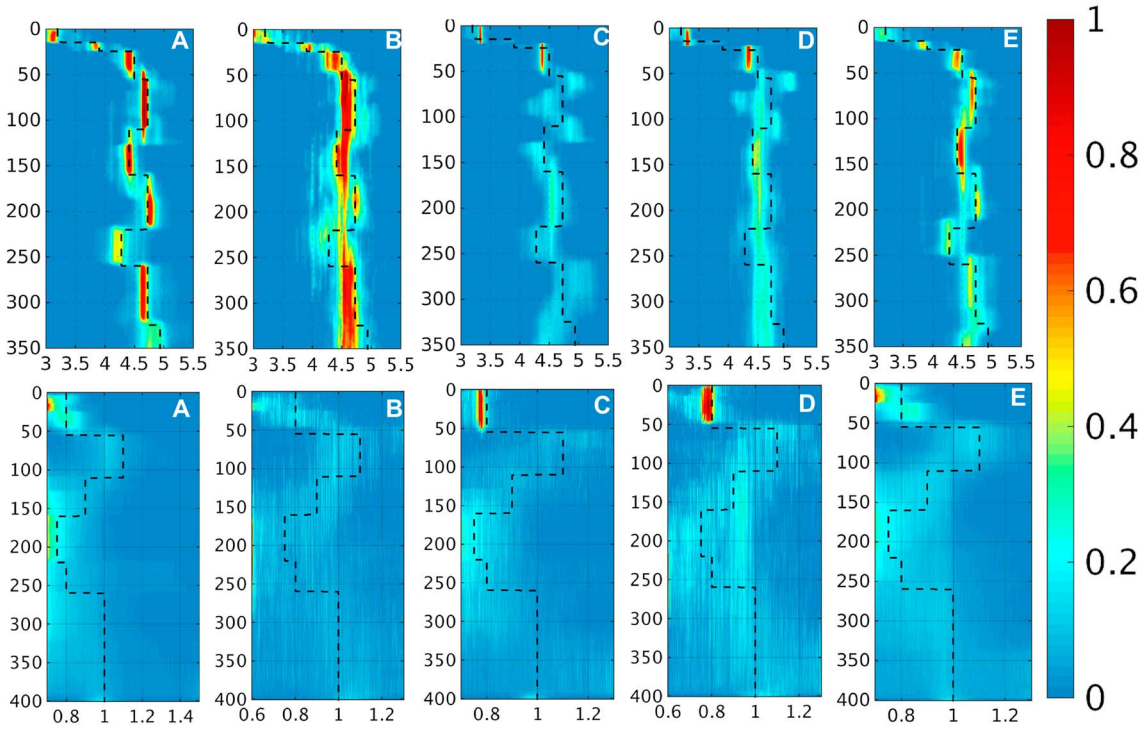


Figure 4. Posterior probability distributions for (top row) V_{SV} and (bottom row) ξ ratio using synthetic data with random Gaussian noise. The black dotted line represents the true model. (a) (Test 1) Fixed V_p/V_s ratio in the rj-MCMC. (b) (Test 2) The V_p/V_s ratio is treated as an unknown variable in the rj-MCMC. (c) (Test 3) The V_p/V_s ratio is fixed, and a Moho at the wrong depth is implemented in the prior distribution for V_{SV}. (d) (Test 4) The V_p/V_s ratio is treated as an unknown variable, and a Moho at the wrong depth is implemented in the prior distribution for V_{SV}. (e) Test(5) The V_p/V_s ratio is fixed to a wrong value.

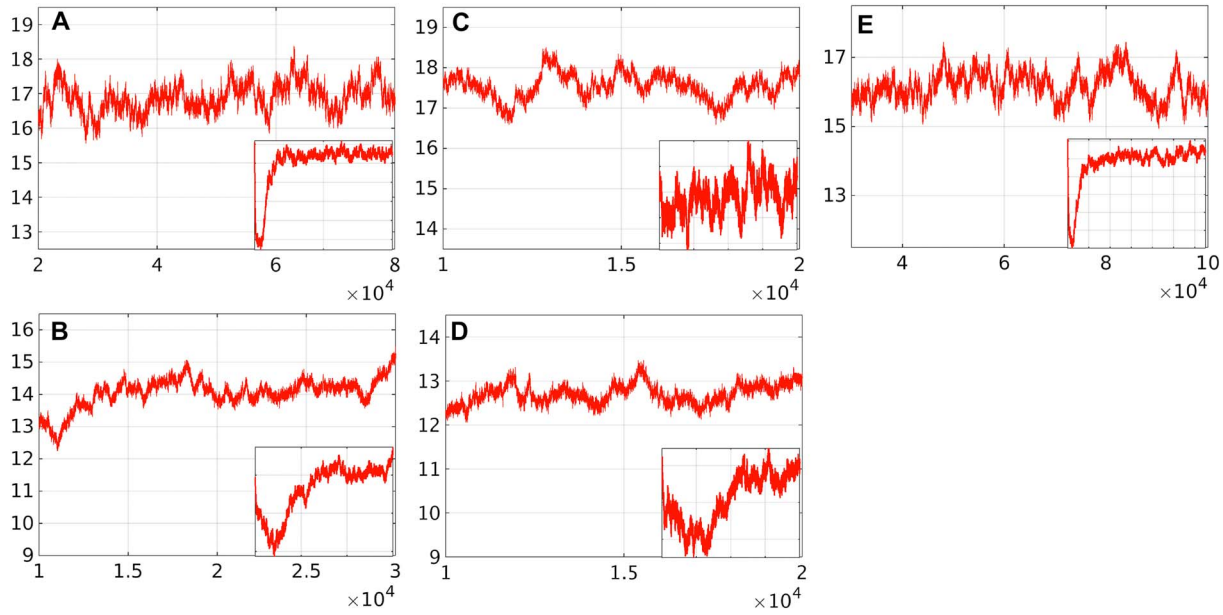


Figure 5. (a–e) Convergence of the average number of layers as a function of sampled models for the simulations in Figure 4 after the burn-in phase. The inlay shows the parameters for all sampled models. Note that Figures 5b–5d display the convergence after two consecutive runs.

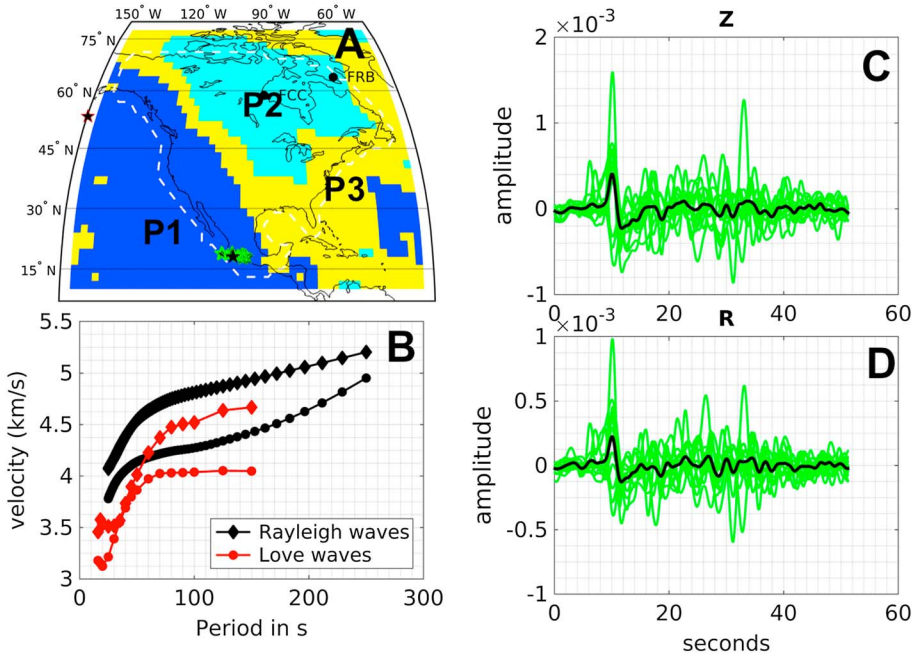


Figure 6. Data at station FCC used in the inversion shown in Figure 7. (a) Station map modified from Calò et al. (2016). The North American continent is divided into three regions by applying k-means cluster analysis to SEMUCB-WM1 (French & Romanowicz, 2014). The light blue color indicates the craton region P2, according to the nomenclature of Calò et al. (2016). Green stars mark the location of earthquakes for which the filtered records at station FCC are shown in Figures 6c and 6d used in Figure 7. The green stars mark the locations of earthquakes used for the waveform stack in Figures 6a and 6b. The red star is the average location of the earthquakes used at station FRB. (b) Group and phase velocities for Love and Rayleigh waves at FCC. (c and d) Vertical and radial component of data recorded at FCC (green) and stack (black).

When the V_p/V_s ratio is fixed to a wrong value (Figure 4e), the true model is still recovered, although the posterior distribution is broader and less well confined around the true model than the computations with the correct value (Figure 4a). This means that the true value of the V_p/V_s ratio does not need to be exactly known for real data inversion. The number of layers is stable and converged in all cases, also in the case with crustal discontinuities fixed at the wrong depth (Figure 5).

5. Real Data

5.1. Station FCC

Figure 6 shows surface wave dispersion data and body wave stacks at station FCC, Canada, used to obtain the results in Figure 7. In the station stack of the waveforms in Figures 6b and 6c, 30 recordings of earthquakes at the coast of South America with an average back azimuth of 195° were stacked to reduce the noise. Care must be taken that the epicenters of the stacked waveforms are close together to avoid incoherent stacking of the waveforms. The phase and group velocities for the surface waves are provided by Shapiro and Ritzwoller (2002) and Ekström (2011), respectively.

We applied parallel tempering with changes between all temperature levels and performed four computations, with variable or fixed V_p/V_s ratio and fixed crustal discontinuities. The results are summarized in Figure 7.

In all four cases, we see the Moho at 34 km depth, followed by discontinuities at 77 km (MLD) and around 150 km (LAB) depth. The second layer at 150 km depth is less clear in the inversion, where the V_p/V_s ratio is allowed to vary in the rj-MCMC (Figure 7c). This structure found in the upper 150 km is similar to what Calò et al. (2016) found at the station FRB. When the Moho is fixed (Figures 7b and 7d), at 34 km (Table S3), no additional layers appear in depth, the result is similar to the one obtained without fixed crustal discontinuities, and the MLD at 77 km is still visible. That no additional layers appear indicates that the Moho was put at the correct depth; however, all deeper layers disappear. The layering is also visible as changes of radial anisotropy at the depth of the layers; this is best resolved in the crust. The shear wave velocities found here are very

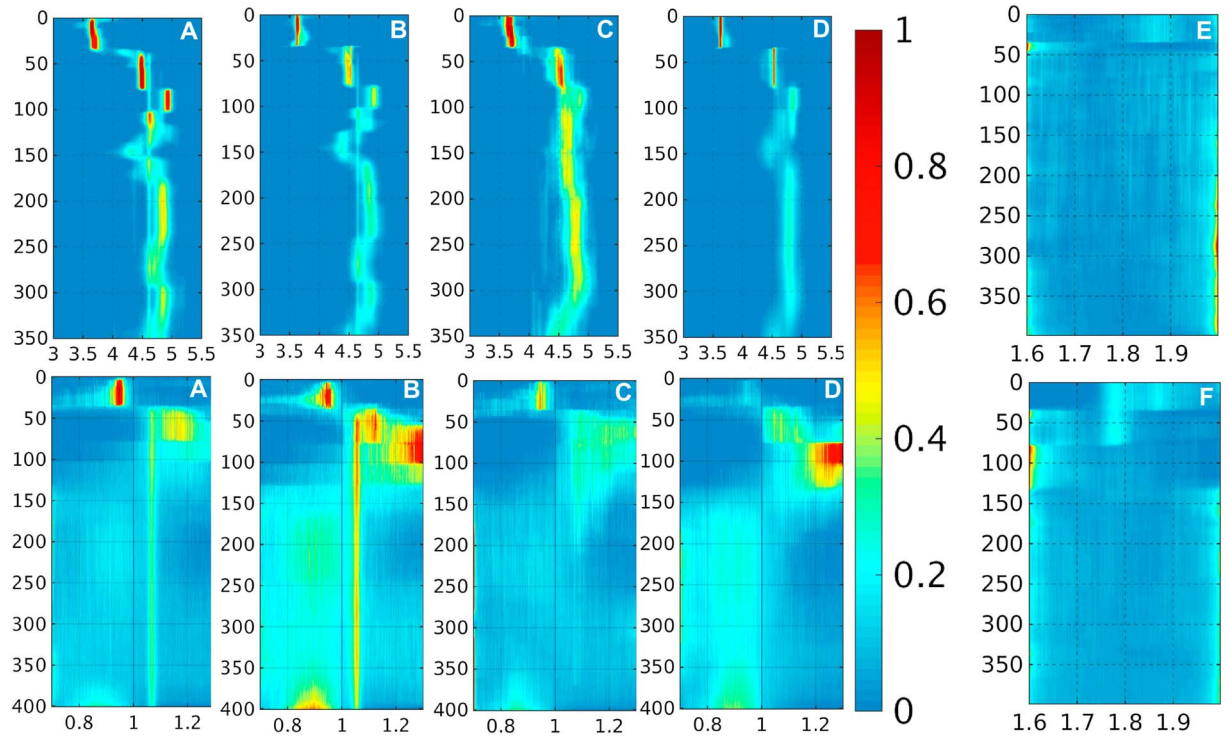


Figure 7. Posterior probability distributions for (top row) V_{SV} and (bottom row) ξ and for the V_p/V_s ratio at station FCC. (a and b) Fixed V_p/V_s ratio, without and with fixed crustal discontinuities, respectively. (c and d) Variable V_p/V_s ratio, without and with fixed crustal discontinuities, respectively. (e and f) Probability distribution of the V_p/V_s ratio corresponding to the results shown in Figures 7b and 7d.

high at ~100–150 km, around 4.8 km/s at 100 km. This is hard to explain by a standard craton geotherm and peridotite composition but corroborates results found from tomography (French & Romanowicz, 2014). Here the LAB may be the discontinuity at 150 km depth and it may be real, since it is most stable when no crustal discontinuities are fixed but visible in the other inversions.

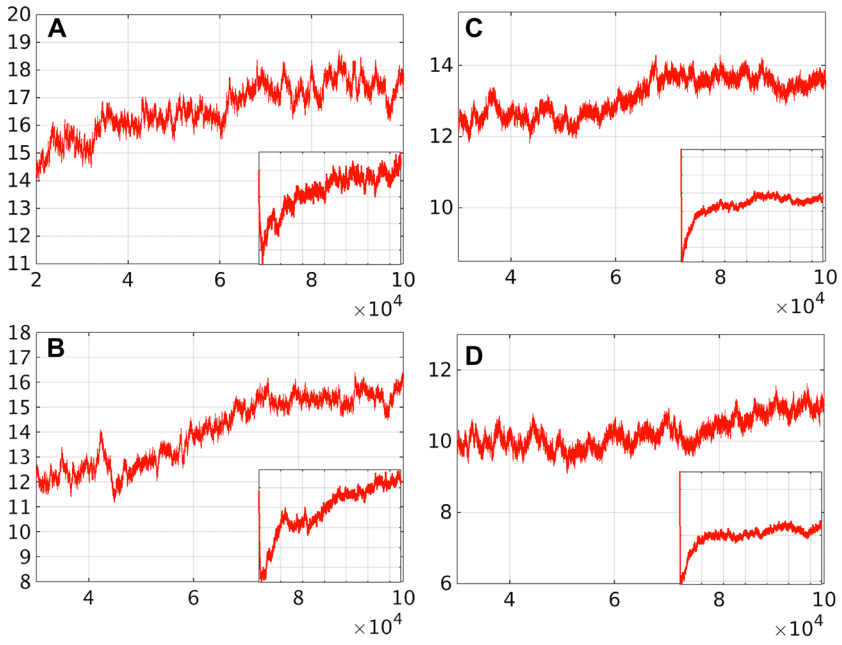


Figure 8. Convergence of the average number of layers as function of sampled models for the real data in Figure 6 after the burn-in phase. The inlay shows the parameters for all sampled models.

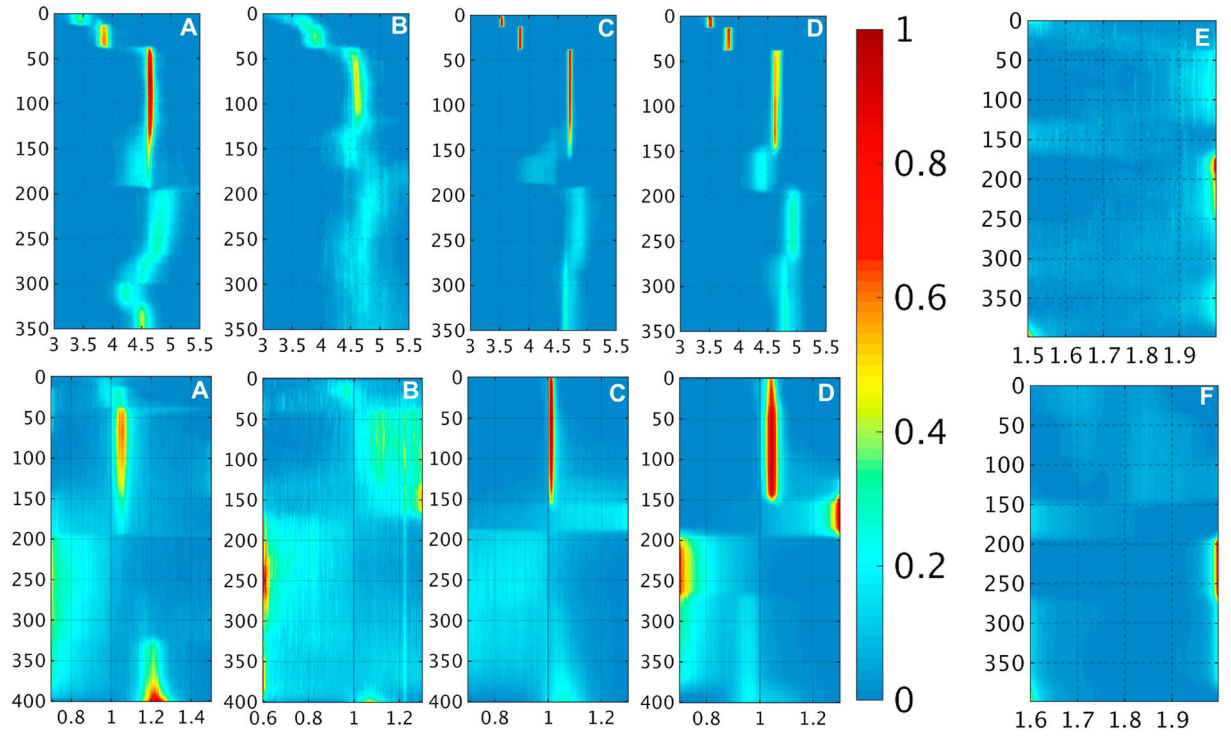


Figure 9. Posterior probability distributions for (top row) V_{SV} and (bottom row) ξ and for the V_p/V_s ratio at station FRB. (a and b) Results with fixed and variable V_p/V_s ratio, respectively. (c and d) Results obtained with fixed Conrad and Moho discontinuity and fixed or variable V_p/V_s ratio, respectively. (e and f) Probability distribution of the V_p/V_s ratio corresponding to the results shown in Figures 9b and 9d.

The posterior distribution of the V_p/V_s ratio (Figures 7e and 7f) is very broad and indicates little sensitivity of the data to it. The posterior distribution of V_{SV} (Figure 7c) is blurred, and layers are not as sharp as in Figure 7a with a fixed V_p/V_s ratio. The average number of layers is stable and converged (Figure 8), also in the case with fixed crustal discontinuities (B&D).

5.2. Station FRB

We performed the same four modeling exercises as for FCC, at station FRB. Here we stacked body waves arriving from back azimuth 220°. We applied parallel tempering with allowed changes between all temperature levels and performed four computations, with variable or fixed V_p/V_s ratio and fixed crustal discontinuities. The results are summarized in Figure 9.

The Moho here is at 34 km depth, but we do not observe any MLD around 70 km as for FCC. We see a clear decrease in V_{SV} at 150 km depth in all four calculations, too, which is smeared when we allow the V_p/V_s ratio to vary (Figure 9b) but still visible when we fix crustal discontinuities (Figures 9c and 9d). This indicates that we fixed the Moho at the right depth, and this discontinuity is no artifact of later arriving Moho multiples. At this depth, the radial anisotropy changes, too, from larger than 1 to smaller than 1, although the exact value is not well resolved and we observe a small change in the V_p/V_s as well.

The Moho depth and the decrease in V_{SV} agree well with the results of Calò et al. (2016). However, they observe a MLD at 70 km depth, two more between 150 and 250 km, and a LAB at 250 km, while we do not see any. These discontinuities might be artifacts of the fixed crustal discontinuities or lack of convergence of the result. The results presented here are converged, as displayed in Figure 10 by the number of layers. We note that the number of layers is very high when we fix crustal discontinuities (Figure 10b), but this is compensated by a variable V_p/V_s ratio in Figure 10d. The large number in Figure 10b, however, might indicate a poor data fit. The variable V_p/V_s ratio does not introduce much change in our results (Figure 10b); therefore, we conclude that the artificial layers in Calò et al. (2016) are not due to a fixed V_p/V_s ratio.

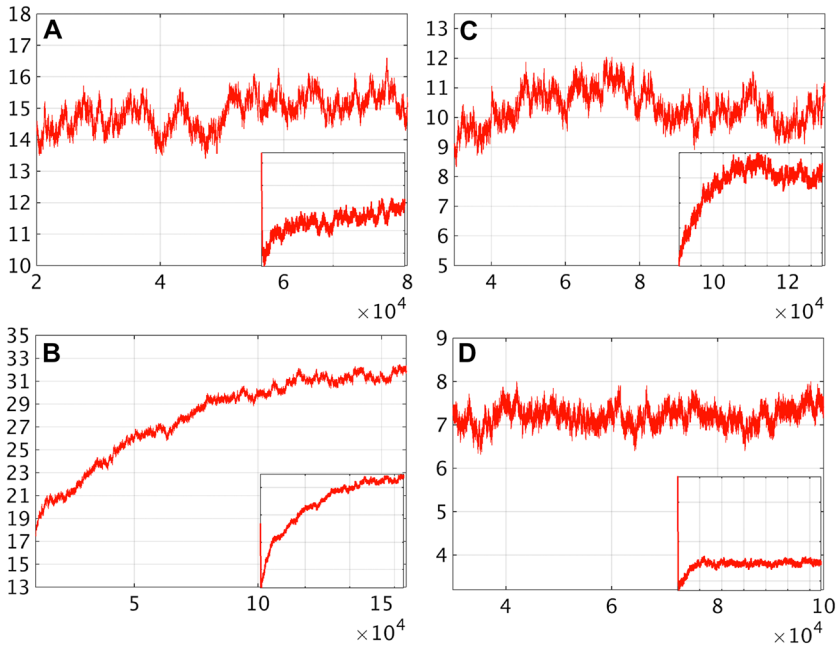


Figure 10. Convergence of the average number of layers as function of sampled models for the results in Figure 9 after the burn-in phase. The inlay shows the number of layers for all sampled models.

6. Discussion and Conclusions

We evaluated a 1-D Bayesian transdimensional Monte Carlo approach on synthetics and real data to explore its ability to constrain depth variations of radial anisotropy ξ and vertical shear wave velocity V_{SV}. We inverted converted phases (*P* to *S*) and surface wave dispersion data simultaneously. In the model parametrization, the number of layers and V_{SV} and ξ are treated as unknowns and are only constrained by the data. Additionally, the noise in the two data sets is also unknown and inverted for simultaneously. Compared to previous studies of Calò et al. (2016), the V_p/V_s ratio was treated as random variable in the rj-MCMC and we applied parallel tempering. We reviewed critically the effects of a variable V_p/V_s ratio and the effect of fixed crustal discontinuities on the posterior. We found that allowing a variable V_p/V_s ratio does not improve results and cannot prevent the appearance of spurious layers in depth, which are meant to fit the later arriving Moho multiples. In the synthetic tests we found that we can still recover the true model with a slightly wrong V_p/V_s ratio, even in the case with noise in the data. We conclude that the V_p/V_s ratio does not need to be known very accurately for real data. Introducing fixed crustal discontinuities in the reference model leads to spurious layers, when put at the wrong depth, and is not recommended.

We come to the conclusion that a model with many thin layers of similar velocity is indistinguishable from a model with constant velocity. Applying these insights to real data, we can confirm the overall structure found by Calò et al. (2016) but it seems that the MLDs observed by Calò et al. (2016) between 150 and 200 km depth may be artifacts due to the fixed crustal discontinuities. Additionally, contrary to Calò et al. (2016), our posterior distribution for V_{SV} is not bimodal between 200 and 250 km depth, and we clearly see the layering better in the radial anisotropy at discontinuities. At station FCC, we observe a similar structure as at FRB but with a shallower Moho and an MLD at 70 km depth.

For upper mantle inversions using a Bayesian approach, we recommend avoiding the introduction of discontinuities in the crust at specific depths and recommend fixing the V_p/V_s to a value given by independent crustal studies. If a value is not available, however, the V_p/V_s can be used as a random variable in the rj-MCMC, without impacting the results significantly. An alternative method, not tried here, is to treat V_p/V_s as unknown but independent of depth, as done by Quijano et al. (2016) for attenuation in a seabed sediment.

In conclusion, we confirm the presence of a sharp MLD around 70 km atop a ~50 km thick low shear velocity layer in the middle of the North American craton, but other MLDs are not as robust. The LAB may be detected

at around 150 km at FCC, which is in good agreement with recent tomographic studies that show thick lithosphere around Hudson Bay (e.g., Schaeffer & Lebedev, 2014; Yuan et al., 2014). It may be deeper at FRB; however, the results are inconclusive.

Acknowledgments
This research used the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley, the Los Alamos National Laboratory computational facility, and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant ACI-1053575. Seismic data are from the IRIS DMS (http://ds.iris.edu/ds/nodes/dmc/) and the Canadian National Seismograph Network (http://www.earthquakescanada.nrcan.gc.ca). The research was supported by NSF EAR grant 1460205 and by UC Lab Fees Collaboration grant 12-LR-236345. We thank J. Dettmer and M. Sambridge for helpful suggestions which improved the quality of the manuscript and Thomas Bodin for helpful advice.

References

Abt, D. L., Fischer, K. M., French, S. W., Ford, H. A., Yuan, H., & Romanowicz, B. (2010). North American lithospheric discontinuity structure imaged by *Ps* and *Sp* receiver functions. *Journal of Geophysical Research*, *115*, B09301. https://doi.org/10.1029/2009JB006914

Agostinetti, N. P., & Malinverno, A. (2010). Receiver function inversion by trans-dimensional Monte Carlo sampling. *Geophysical Journal International*, *181*(2), 858–872. https://doi.org/10.1111/j.1365-246X.2010.04530.x

Arnold, R., & Townend, J. (2007). A Bayesian approach to estimating tectonic stress from seismological data. *Geophysical Journal International*, *170*(3), 1336–1356. https://doi.org/10.1111/j.1365-246X.2007.03485.x

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K., & Rawlinson, N. (2012). Transdimensional inversion of receiver functions and surface wave dispersion. *Journal of Geophysical Research*, *117*, B02301. https://doi.org/10.1029/2011JB008560

Bodin, T., Yuan, H., & Romanowicz, B. (2014). Inversion of receiver functions without deconvolution—Application to the Indian craton. *Geophysical Journal International*, *196*(2), 1025–1033. https://doi.org/10.1093/gji/ggt431

Bodin, T., Leiva, J., Romanowicz, B., Maupin, V., & Yuan, H. (2016). Imaging anisotropic layering with Bayesian inversion of multiple data types. *Geophysical Journal International*, *206*, 605–629. https://doi.org/10.1093/gji/ggw124

Calò, M., Bodin, T., & Romanowicz, B. (2016). Layered structure in the upper mantle across North America from joint inversion of long and short period seismic data. *Earth and Planetary Science Letters*, *449*, 164–175. https://doi.org/10.1016/j.epsl.2016.05.054

Chen, J., Hoversten, G. M., Vasco, D., Rubin, Y., & Hou, Z. (2007). A Bayesian model for gas saturation estimation using marine seismic AVA and CSEM data. *Geophysics*, *72*(2), WA85–WA95. https://doi.org/10.1190/1.2435082

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

Dettmer, J., Dosso, S. E., Bodin, T., Stipčević, J., & Cummins, P. R. (2015). Direct-seismogram inversion for receiver-side structure with uncertain source–time functions. *Geophysical Journal International*, *203*(2), 1373–1387.

Ekström, G. (2011). A global model of Love and Rayleigh surface wave dispersion and anisotropy, 25–250 s. *Geophysical Journal International*, *187*(3), 1668–1686.

Ford, H. A., Long, M. D., & Wirth, E. A. (2016). Midlithospheric discontinuities and complex anisotropic layering in the mantle lithosphere beneath the Wyoming and Superior Provinces. *Journal of Geophysical Research: Solid Earth*, *121*, 6675–6697. https://doi.org/10.1002/2016JB012978

French, S., & Romanowicz, B. (2014). Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography. *Geophysical Journal International*, *199*(3), 1303–1327.

Gouveia, W. P., & Scales, J. A. (1998). Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis. *Journal of Geophysical Research*, *103*(B2), 2759–2779. https://doi.org/10.1029/97JB02933

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*(4), 711–732.

Green, P. J. (2003). Trans-dimensional Markov chain Monte Carlo. *Oxford Statistical Science Series*, *27*, 179–198.

Guo, Z., Afonso, J. C., Qashqai, M. T., Yang, Y., & Chen, Y. J. (2016). Thermochemical structure of the North China craton from multi-observable probabilistic inversion: Extent and causes of cratonic lithosphere modification. *Gondwana Research*, *37*, 252–265. https://doi.org/10.1016/j.gr.2016.07.002

Hales, A. (1969). A seismic discontinuity in the lithosphere. *Earth and Planetary Science Letters*, *7*(1), 44–46.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. New York: Cambridge University Press.

Käufl, P., Fichtner, A., & Igel, H. (2013). Probabilistic full waveform inversion based on tectonic regionalization—Development and application to the Australian upper mantle. *Geophysical Journal International*, *193*(1), 437–451.

Kumar, P., Kind, R., Yuan, X., & Mechie, J. (2012). USArray receiver function images of the lithosphere-asthenosphere boundary. *Seismological Research Letters*, *83*(3), 486–491.

Levin, V., & Park, J. (1998). *P-SH* conversions in layered media with hexagonally symmetric anisotropy: A cookbook. *Geodynamics of Lithosphere and Earth's Mantle*, *151*, 669–697.

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

Malinverno, A. (2002). Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophysical Journal International*, *151*(3), 675–688.

Malinverno, A., & Briggs, V. A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics*, *69*(4), 1005–1016.

Montagner, J.-P., & Anderson, D. L. (1989). Petrological constraints on seismic anisotropy. *Physics of the Earth and planetary Interiors*, *54*(1), 82–105.

Mustać, M., & Tkalčić, H. (2016). Point source moment tensor inversion through a Bayesian hierarchical model. *Geophysical Journal International*, *204*(1), 311–323.

Pachhai, S., Tkalčić, H., & Dettmer, J. (2014). Bayesian inference for ultralow velocity zones in the Earth's lowermost mantle: Complex ULVZ beneath the east of the Philippines. *Journal of Geophysical Research: Solid Earth*, *119*, 8346–8365. https://doi.org/10.1002/2014JB011067

Quijano, J. E., Dosso, S. E., Dettmer, J., & Holland, C. (2016). Geoacoustic inversion for the seabed transition layer using a Bernstein polynomial model. *The Journal of the Acoustical Society of America*, *140*(6), 4073–4084.

Rosas-Carbajal, M., Linde, N., Kalscheuer, T., & Vrugt, J. A. (2013). Two-dimensional probabilistic inversion of plane-wave electromagnetic data: Methodology, model constraints and joint inversion with electrical resistivity data. *Geophysical Journal International*, *196*(3), 1508–1524.

Rychert, C. A., Shearer, P. M., & Fischer, K. M. (2010). Scattered wave imaging of the lithosphere–asthenosphere boundary. *Lithos*, *120*(1), 173–185.

Sambridge, M. (2014). A Parallel Tempering algorithm for probabilistic sampling and multimodal optimization. *Geophysical Journal International*, *196*(1), 357–374. https://doi.org/10.1093/gji/ggt342

Sambridge, M., & Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, *40*(3), 1009. https://doi.org/10.1029/2000RG000089

Scales, J. A., & Snieder, R. (1997). To Bayes or not to Bayes? *Geophysics*, *62*(4), 1045–1046.

Scales, J. A., & Tenorio, L. (2001). Prior information and uncertainty in inverse problems. *Geophysics*, *66*(2), 389–397. https://doi.org/10.1190/1.1444930

Schaeffer, A., & Lebedev, S. (2014). Imaging the North American continent using waveform inversion of global and USArray data. *Earth and Planetary Science Letters*, *402*, 26–41.

Shan, B., Afonso, J. C., Yang, Y., Grose, C. J., Zheng, Y., Xiong, X., & Zhou, L. (2014). The thermochemical structure of the lithosphere and upper mantle beneath south China: Results from multiobservable probabilistic inversion. *Journal of Geophysical Research: Solid Earth*, *119*, 8417–8441. https://doi.org/10.1002/2014JB011412

Shapiro, N., & Ritzwoller, M. (2002). Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle. *Geophysical Journal International*, *151*(1), 88–105.

Stähler, S., & Sigloch, K. (2014). Fully probabilistic seismic source inversion—Part 1: Efficient parameterisation. *Solid Earth*, *5*(2), 1055.

Takeuchi, H., & Saito, M. (1972). Seismic surface waves. *Methods in Computational Physics*, *11*, 217–295.

Tarantola, A. (2005). Inverse problem theory and methods for model parameter estimation (chapter 1, pp. 1–40). Society for Industrial and Applied Mathematics.

Tork Qashqai, M., Carlos Afonso, J., & Yang, Y. (2016). The crustal structure of the Arizona transition zone and southern Colorado Plateau from multiobservable probabilistic inversion. *Geochemistry, Geophysics, Geosystems*, *17*, 4308–4332.

Wang, L., Zöller, G., & Hainzl, S. (2015). Joint determination of slip and stress drop in a Bayesian inversion approach: A case study for the 2010 M8.8 Maule earthquake. *Pure and Applied Geophysics*, *172*(2), 375–388. https://doi.org/10.1007/s00024-014-0868-x

Wirth, E. A., Long, M. D., & Moriarty, J. C. (2017). A Markov chain Monte Carlo with Gibbs sampling approach to anisotropic receiver function forward modeling. *Geophysical Journal International*, *208*(1), 10–23.

Young, M. K., Rawlinson, N., & Bodin, T. (2013). Transdimensional inversion of ambient seismic noise for 3D shear velocity structure of the Tasmanian crust. *Geophysics*, *78*(3), WB49–WB62. https://doi.org/10.1190/geo2012-0356.1

Yuan, H., & Romanowicz, B. (2010). Lithospheric layering in the North American craton. *Nature*, *466*(7310), 1063–1068.

Yuan, H., French, S., Cupillard, P., & Romanowicz, B. (2014). Lithospheric expression of geological units in central and eastern North America from full waveform tomography. *Earth and Planetary Science Letters*, *402*, 176–186.