Direct-seismogram inversion for receiver-side structure with uncertain source-time functions

Jan Dettmer, Stan E. Dosso, Thomas Bodin, Josip Stipcevic, Phil R. Cummins

Geophysical Journal International

Geophys. J. Int. (2015) **203,** 1373–1387
GJI Seismology

doi: 10.1093/gji/ggv375

Direct-seismogram inversion for receiver-side structure with uncertain source–time functions

Jan Dettmer,[1,2] Stan E. Dosso,[1] Thomas Bodin,[3,4] Josip Stipčević[2,5] and Phil R. Cummins[2]

[1] *School of Earth and Ocean Sciences, University of Victoria, Victoria BC, Canada. E-mail: jand@uvic.ca*
[2] *Research School of Earth Sciences, Australian National University, Canberra ACT, Australia*
[3] *Berkeley Seismological Laboratory,* 215 *McCone Hall, UC Berkeley, Berkeley CA* 94720-4760, *USA*
[4] *Laboratoire de Géologie de Lyon, École Normale Superieure de Lyon, Université de Lyon*-1, *CNRS, F*-69364 *Lyon Cedex* 07, *France*
[5] *Department of Geophysics, Faculty of Science, University of Zagreb, Zagreb, Croatia*

Accepted 2015 September 7. Received 2015 September 4; in original form 2014 December 17

SUMMARY

This paper presents direct-seismogram inversion (DSI) for receiver-side structure which treats the source signal incident from below (the effective source–time function—STF) as a vector of unknown parameters in a Bayesian framework. As a result, the DSI method developed here does not require deconvolution by observed seismogram components as typically applied in receiver-function inversion and avoids the problematic issue of choosing subjective tuning parameters in this deconvolution. This results in more meaningful inversion results and uncertainty estimation compared to classic receiver-function inversion. A rigorous derivation is presented of the likelihood function required for unbiased inversion results. The STF is efficiently inferred by a maximum-likelihood closed-form expression that does not require deconvolution by noisy waveforms. Rather, deconvolution is only by predicted impulse responses for the unknown environment (considered to be a 1-D, horizontally stratified medium). For a given realization of the parameter vector which describes the medium below the station, data predictions are computed as the convolution of the impulse response and the maximum-likelihood source estimate for that medium. Therefore, the assumption of a Gaussian pulse with specified parameters, typical for the prediction of receiver functions, is not required. Directly inverting seismogram components has important consequences for the noise on the data. Since the signal processing does not require filtering and deconvolution, data errors are less correlated and more straightforward to model than those for receiver functions. This results in better inversion results (parameter values and uncertainties), since assumptions made in the derivation of the likelihood function are more likely to be met by the inversion process. The DSI method is demonstrated for simulated waveforms and then applied to data for station Hyderabad on the Indian craton. The measured data are inverted with both the new DSI and traditional receiver-function inversion. All inversions are carried out for a trans-dimensional model that treats the number of layers in the model as unknown. Results for DSI are consistent with previous studies for the same location. The DSI has clear advantages in trans-dimensional inversion. Uncertainty estimates appear more realistic (larger) in both model complexity (number of layers) and in terms of seismic velocity profiles. Receiver-function inversion results in more complex profiles (highly-layered structure) and suggests unreasonably small uncertainties. This effect is likely also significant when the parametrization is considered to be fixed but exacerbated for the trans-dimensional model: If hierarchical errors are poorly estimated, trans-dimensional models overestimate the structure which produces unfavourable results for the receiver-function inversion.

Key words: Inverse theory; Probability distributions; Computational seismology; Statistical seismology.

1 INTRODUCTION

The study of converted teleseismic waves is one of the most important methods to infer receiver-side seismic structure of the Earth's crust and upper mantle. The coda of *P* and *S* waves contains large numbers of converted phases and their multiples which can be used to infer the shear wave velocity (V_s) structure and, to a lesser extent, the compressional wave velocity (V_p) structure underneath a receiver. Since the signals are complex superpositions of effects related to the source, source–region, receiver–region and propagation path, initial efforts were directed at separating the source and path effects from the local receiver-side effects (e.g. Phinney 1964; Burdick & Langston 1977; Vinnik 1977; Langston 1979). This separation is achieved by deconvolving the vertical (D_v) seismogram component from the radial (D_r) or transverse which produces a receiver function (RF). The RF can be interpreted as an approximation of the response of local structure (a layer stack below the receiver) to a plane wave incident from below (Bostock 2007). Early RF work was based on forward modelling or linearized inversion which both were widely applied (e.g. Owens *et al.* 1984; Owens 1987; Ammon *et al.* 1990; Cassidy 1992). However, strong nonlinearities can cause highly non-unique solutions and care must be taken in the interpretation of results (Ammon *et al.* 1990; Cassidy 1992).

Over the last decade several nonlinear inversion examples have shown the potential to infer complex and meaningful crust and upper-mantle models from RF waveforms (e.g. Sambridge 1999; Shapiro & Ritzwoller 2002; Frederiksen *et al.* 2003; Kiselev *et al.* 2008; Agostinetti & Malinverno 2010; Stipčević *et al.* 2011; Bodin *et al.* 2012; Shen *et al.* 2012; Brillon *et al.* 2013). In particular, joint inversion with other data types has been important in improving constraints on V_p and V_s structure. For example, surface wave dispersion (SWD) data provide information about the absolute value of velocities (Julia *et al.* 2000; Tkalčić *et al.* 2006) and can break strong parameter correlations (Bodin *et al.* 2012), such that the joint inversion result is significantly better resolved than results for either data set considered individually. Similarly, shear wave RF data can contribute additional constraints on the V_p and V_s, reducing uncertainties (Vinnik *et al.* 2004), and teleseismic traveltime residuals can constrain the V_p structure (Vinnik *et al.* 2006; Kiselev *et al.* 2008).

Much work is also directed at using RF waveforms to image earth structure by directly interpreting RFs and migrating data for several stations to infer 2-D structure along profiles (Bostock & Rondenay 1999; Bostock 2002; Bostock *et al.* 2002; Kind *et al.* 2002; Audet *et al.* 2009; Kind *et al.* 2012). While RF imaging has led to significant tectonic insights, migration depends on seismic velocity information and care must be taken in the interpretation with respect to multiples. In addition, uncertainties of inferred parameters may be difficult to estimate due to strong model assumptions.

All such inference techniques require removing the effects of the source–time function (STF) from the seismograms which produces the RF. This is commonly achieved by deconvolution of one component from another (Ammon 1991; Ligorria & Ammon 1999). Since deconvolution can be considered an ill-posed linear inverse problem where noise causes numerical instability, the data processing typically includes some form of filtering and stabilisation. Both the filtering and stabilisation require the subjective choice of control parameters and the effect of such choices on the estimated seismic velocity profiles is difficult to quantify. This is particularly problematic for studies that infer parameter uncertainty estimates (e.g. Bayesian inversion) which can depend strongly on tuning parameters. In addition, the interpretation of fine structure from RF results may be strongly impacted by tuning parameters.

To address deconvolution issues, Bodin *et al.* (2014) proposed a cross-convolution misfit function that does not involve any deconvolution (Menke & Levin 2003) and, conceptually, would have a minimum at the true model parameters (for noise-free data without any theory error). Similar approaches are also applied in blind deconvolution (Royer *et al.* 2012). Bodin *et al.* (2014) treated the exponential of the negative of the cross-convolution misfit as a likelihood function and applied it in probabilistic inversion. However, a proper likelihood function for probabilistic inference must be derived from an assumption about the distribution of data errors (e.g. Gaussian distributed) and residuals must be defined as a difference between observed and predicted data vectors (Gelman *et al.* 2003). This is not the case for a cross-convolution misfit, which, to our knowledge, cannot be applied in a Bayesian formulation. Note that a similar approach (basing a likelihood function on conveniently-defined residuals rather than on the residual error distribution) is also used by Stähler & Sigloch (2014) for moment tensor inversion.

This work represents a novel Bayesian inversion for receiver-side structure that does not require deconvolution of observed seismogram components, thereby avoiding the problem of choosing subjective stabilisation parameters. The method considers the STF as unknown and provides rigorous derivation of a likelihood function which is required for unbiased inversion results (not the case in the other approaches discussed above). Therefore, noise on the seismograms (including theory error) is properly translated to parameter uncertainties. Importantly, considering the STF as unknown results in more meaningful uncertainty estimation that does not depend on subjectively chosen tuning parameters. The inversion accounts for the limited knowledge about the STF in the uncertainty estimates for seismic velocities and provides a probabilistic estimate for the effective STF incident from below.

2 PROCESSING OF RECEIVER FUNCTIONS

At a receiver at the Earth's surface, the seismogram D for a time window after the *P*-wave arrival can be considered as the convolution of an effective STF $s(t)$ with the impulse response E of the local structure

$$\begin{aligned} D_v(t) &= s(t) * E_v(t), \\ D_r(t) &= s(t) * E_r(t), \\ D_t(t) &= s(t) * E_t(t), \end{aligned} \tag{1}$$

where D_i are the vertical, radial and transverse components ($i \in (v, r, t)$) and $*$ indicates time-domain convolution. The impulse responses E_i is only for the receiver-side structure and assumes an impulsive plane wave incident from below (Bostock 2007). While the following development can be extended to the transverse component in a straightforward manner, $D_t(t)$ is not considered here to improve clarity in the notation. This treatment also assumes the instrument response to be either removed, considered as part of $s(t)$, or included in the impulse response (i.e. convolve the instrument response with the impulse response).

The effective STF is considered to be the waveform due to a distant earthquake (~30°–90° epicentral distance) that is incident from below at the bottom of the deepest layer (for a given time window). Hence, $s(t)$ includes the effects of earthquake rupture as well as effects due to complexities in the source region and due to

propagation along a path from the source region to the deepest layer beneath the receiver location. The shallow receiver-side structure is considered to be represented by the unknown plane-wave impulse responses $E_v(t)$ and $E_r(t)$. In general, $s(t)$ is unknown, can have complex shape and is dependent on the particular earthquake. To infer the local receiver-side structure, the unknown $s(t)$ needs to be separated from the unknown $E_i(t)$ (Burdick & Langston 1977; Vinnik 1977; Ammon 1991). Traditionally, this has been carried out by deconvolving one component from the other which gives a waveform referred to as a (P) RF (Ammon 1991). Since the radial component shows the strongest signal for P-to-S conversions, radial P RFs are most commonly considered. However, in recent years S RFs and transverse RFs have received significant attention (Farra & Vinnik 2000; Kiselev *et al.* 2008; Miller & Piana Agostinetti 2012). Since observed seismograms exhibit stochastic noise, deconvolution requires stabilisation. For example, water-level stabilisation (Clayton & Wiggins 1976) can be applied to the seismogram spectra to avoid numerical instability and can preserve amplitudes. Following Ammon (1991) and Cassidy (1992), RF processing is typically carried out in the frequency domain

$$H(\omega) = E_r(\omega)E_v^*(\omega)G(\omega)/\Phi(\omega), \tag{2}$$

where ω is angular frequency and $*$ denotes the complex conjugate. The denominator Φ includes the vertical-component spectrum and a numerical water-level stabilisation c

$$\Phi(\omega) = \max\left\{E_v(\omega)E_v^*(\omega),\, c\, \max_{\omega'}\left\{E_v(\omega')E_v^*(\omega')\right\}\right\}. \tag{3}$$

The Gaussian filter $G(\omega)$ is given by

$$G(\omega) = \xi \exp\left(-\omega^2/4a^2\right), \tag{4}$$

where ξ normalizes the filter and a is a parameter to control the filter width. Amplitudes of the RF are preserved by normalizing with

$$A(\omega) = E_v(\omega)E_v^*(\omega)G(\omega)/\Phi(\omega). \tag{5}$$

The RF is typically analysed in the time domain. A significant practical challenge is setting the two tuning parameters c and a which can both affect the resulting waveform significantly. The values for these parameters are typically set by a trial-and-error approach and can cause significant subjectivity. Further, both parameters intrinsically trade off resolvable structure and stability. Most commonly, practitioners apply the smallest water level that produces stable results (judged subjectively and requiring experience) and apply a filter width a such that the desired frequencies are present in the signal (also subjectively trading off resolution and stability). Typical values are $a = 2.5$ rad s^{-1} for crustal studies and $a = 1$ rad s^{-1} for studies also considering the upper mantle. In Bayesian inversion, the effect of uncertain c and a parameters on inversion results is difficult to quantify and limits useful applications. Other processing approaches, such as iterative deconvolution (Ligorria & Ammon 1999) may provide some advantages but suffer from the same fundamental requirement for tuning parameters. A recently developed Bayesian approach to deconvolution (Kolb & Lekic 2014) shows promise to address the issue of tuning parameters but is not considered here due to the high computational cost.

As another concern, the Gaussian filter and deconvolution processes change the noise characteristics of the data and can cause significant challenges for Bayesian parameter and uncertainty estimation. While Bayesian sampling has been applied to RFs (Bodin *et al.* 2012; Shen *et al.* 2012; Brillon *et al.* 2013), the Bayesian formulation fundamentally relies on assuming a statistical distribution form for the residual errors (difference between observation and prediction). The assumed residual distribution is used to derive the likelihood function which quantifies how well data are fit by the model. To obtain meaningful inversion results with that likelihood function, the actual residual errors must be reasonably consistent with the assumption. Most commonly, residuals are considered to be a combination of measurement error and theory error (due to an imperfect model (Tarantola 2005)) and assumed to be Gaussian distributed with zero mean and covariance $\mathbf{C}_d$. While Bayesian methods can estimate $\mathbf{C}_d$ from the data (referred to as hierarchical models/estimation), limitations exist since the time series to estimate correlations from is of finite length and correlations can be of similar length.

In comparison to seismogram traces, which exhibit clear stochastic noise, the noise on highly processed RF waveforms can be extremely correlated and difficult to quantify. Bodin *et al.* (2012) and Dettmer *et al.* (2012) proposed hierarchical estimation of off-diagonal (covariance) terms in $\mathbf{C}_d$ (assuming a simple covariance matrix parametrization) to address the strong correlations. However, the simple types of correlation (e.g. correlations decay exponentially with lag) have often failed to produce meaningful results for RF inversion in practice.

3 LIKELIHOOD FUNCTION FOR DIRECT-SEISMOGRAM INVERSION

This section develops a novel Bayesian approach to invert time-domain waveforms (seismogram components) directly by treating the STF as uncertain (model dependent) in the inversion. The work is motivated by Bodin *et al.* (2014); however, we consider the STF to be unknown and estimate it as part of the inversion. In addition, we provide a rigorous derivation of a likelihood function for Bayesian inversion that results in a general inversion method. Importantly, assuming $s(t)$ as unknown results in probabilistically estimating $s(t)$ as part of the inverse problem and accounts for the uncertainty due to the fact that $s(t)$ is unknown. Therefore, the method has the potential to be relevant to the problem of estimating the STF of earthquakes as well as being potentially applicable to the inversion of deep-earth phases (Flanagan & Shearer 1998; Menke & Levin 2003; Chambers *et al.* 2005; Idehara 2011; Pachhai *et al.* 2014).

Bayesian inversion treats unknown parameters as random variables and uses probabilities to express the degree of belief that a parameter value reflects reality. Parameter values and uncertainties are inferred from the posterior probability density which combines prior information (independent of the data) and data information (through the likelihood function) to specify the state of knowledge about the parameters. Bayesian inversion requires formulating a model which includes a choice of physical theory, an appropriate parametrization and a statistical characterisation of residual errors, which together capture the response of the system under study. The likelihood function of the model parameters originates from the statistical residual-error distribution and quantifies how well predictions fit the data. Model selection (discriminating between possible models) is an important component of Bayesian inversion and is addressed here by trans-dimensional (trans-D) models (the number of layers in the 1D Earth model is considered to be unknown). Formulating the likelihood function for the DSI with unknown STF is the focus of this section. For a brief overview of trans-D Bayesian inversion, the reader is referred to Appendix A.

In the following, a likelihood is derived to invert $D_v(t)$ and $D_r(t)$ (eq. 1) jointly while assuming that both model parameters $\mathbf{x}$ and the parametrized STF are unknown. Assuming that residual errors for

$D_v(t)$ and $D_r(t)$ are independent Gaussian distributed with zero mean and covariance matrices $\sigma_v^2\mathbf{I}$ and $\sigma_r^2\mathbf{I}$, respectively, the likelihood function is given by (Tarantola 2005)

$$\begin{aligned} L(\mathbf{x}) &= \exp\left[-\sum_i\left[\frac{N_i}{2}\log(2\pi)+N_i\log\sigma_i \right.\right. \\ &\qquad \left.\left. +\frac{1}{2\sigma_i^2}\sum_{j=1}^{N_i}\left((D_i)_j-\left(D_i(\mathbf{x})\right)_j\right)^2\right]\right] \\ &\propto \exp\left[-\sum_i\left[N_i\log\sigma_i+\frac{1}{2\sigma_i^2}\sum_{j=1}^{N_i}\left((D_i)_j-\left(D_i(\mathbf{x})\right)_j\right)^2\right]\right] \\ &= \exp\left[\phi(\mathbf{x})\right], \end{aligned} \tag{6}$$

where $i \in \{v, r\}$, $N_i = N$ is the number of time samples (same number of samples on each component) and $D_i(\mathbf{x})$ is the predicted seismogram component for the model parameters $\mathbf{x}$. Considering N discreet time samples, and substituting eq. (1) into eq. (6), the log-likelihood function ϕ can be written as

$$\begin{aligned} \phi(\mathbf{x}) &= -\sum_{j=1}^{N}\left[N\log\sigma_v+\left((D_v)_j-(D_v(\mathbf{x}))_j\right)^2/2\sigma_v^2 \right. \\ &\qquad \left. +N\log\sigma_r+\left((D_r)_j-(D_r(\mathbf{x}))_j\right)^2/2\sigma_r^2\right] \\ &= -\sum_{j=1}^{N}\left[N\log\sigma_v+\left((D_v)_j-\sum_{k=1}^{N_s}s_k(E_v(\mathbf{x}))_{j-k}\right)^2/2\sigma_v^2 \right. \\ &\qquad \left. +N\log\sigma_r+\left((D_r)_j-\sum_{k=1}^{N_s}s_k(E_r(\mathbf{x}))_{j-k}\right)^2/2\sigma_r^2\right], \end{aligned} \tag{7}$$

where s_k are N_s unknown source parameters that give the STF. In eq. (7), both s (the source) and $\mathbf{x}$ (the receiver-side structure and data error statistics) are unknown. In the following, we show how to estimate s efficiently as part of a Bayesian sampling algorithm for $\mathbf{x}$.

Since s is unknown, eq. (7) could be solved by Bayesian sampling, resulting in N_s unknown STF parameters. However, a much more efficient approach is to consider an analytic maximum likelihood estimate of s (for a given parameter vector $\mathbf{x}$) by taking the partial derivative of ϕ with respect to the source $\partial\phi/\partial s_l = 0$. Solving for s gives (see Appendix B for details)

$$s(\mathbf{x}) = \frac{D_v * E_v(\mathbf{x})/\sigma_v^2 + D_r * E_r(\mathbf{x})/\sigma_r^2}{E_v(\mathbf{x}) * E_v(\mathbf{x})/\sigma_v^2 + E_r(\mathbf{x}) * E_r(\mathbf{x})/\sigma_r^2}, \tag{8}$$

where division indicates time-domain deconvolution. Eq. (8) is a maximum-likelihood source estimate for a given set of earth parameters $\mathbf{x}$. Note that deconvolution here involves the predicted earth impulse responses E_v and E_r rather than measured seismograms as in standard RF analysis. The predicted impulse responses are noise free and unlikely to cause numerical instability. For the cases considered here, $N_s < N$ (i.e. the STF is of shorter duration than the seismograms) so that the problem is over-determined with $2N$ data and N_s unknowns. However, this work does not address the issue of an optimal STF length N_s or the potential to regularize the estimation of s. Such approaches can potentially further stabilize the inversion for larger N_s. To obtain s, consider the time domain convolution of vectors u (the denominator in eq. 8) and s to give vector w (the numerator in eq. 8) of lengths N_u, N_s and $N_w = N_u + N_s - 1$, respectively,

$$w = u * s = Us, \tag{9}$$

where U is a Toeplitz matrix of dimension N_w by N_s whose non-zero rows are given by

$$U = \begin{bmatrix} u_1 & 0 & \cdots & & \\ u_2 & u_1 & \ddots & & \\ u_3 & u_2 & \ddots & & \\ \vdots & \vdots & \ddots & & \\ u_{N_u} & u_{N_u-1} & \ddots & & \\ 0 & u_{N_u} & \ddots & & \\ 0 & 0 & \ddots & & \\ \vdots & \vdots & \ddots & & \\ 0 & 0 & \cdots & u_{N_u} & u_{N_u-1} \\ 0 & 0 & 0 & 0 & u_{N_u} \end{bmatrix}. \tag{10}$$

Therefore, the deconvolution of w by u can be expressed as a discrete linear inverse problem where eq. (9) is solved for s. Since this problem is overdetermined, no exact solution exists and the linear least-squares approach is to minimize $||w - Us||_2$ which gives a solution $s = (U^\top U)^{-1}U^\top w$. This solution is well defined provided U has no small singular values, which is often the case if $N > N_s$. Here, s is solved for by the Lapack algorithm DGELSS (Anderson *et al.* 1999) which computes the minimum-norm solution to $||w - Us||_2$. To avoid instability, singular values smaller than 10^{-12} are omitted in DGELSS. This stabilisation was sufficient in all inversion work in this paper. To obtain the source estimate, the least squares solution to eq. (9) is applied for the deconvolution in eq. (8).

Eqs (7) and (8) can be used in a Bayesian inversion algorithm to treat the environment parametrization, the STF and the noise standard deviations σ_v and σ_r as unknown (Fig. 1). Hence, $\mathbf{x}$, s, σ_v and σ_r are all estimated probabilistically and uncertainties reflect the inability to specify precise values for these unknowns in the inversion results (the posterior density). In particular, $\mathbf{x}$, σ_v and σ_r are explicitly sampled with a Metropolis-Hastings-Green algorithm

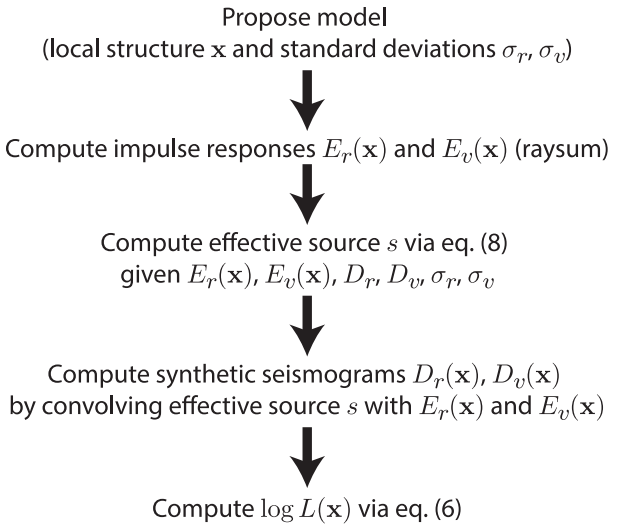

Figure 1. Schematic of evaluating the likelihood function in direct seismogram inversion.

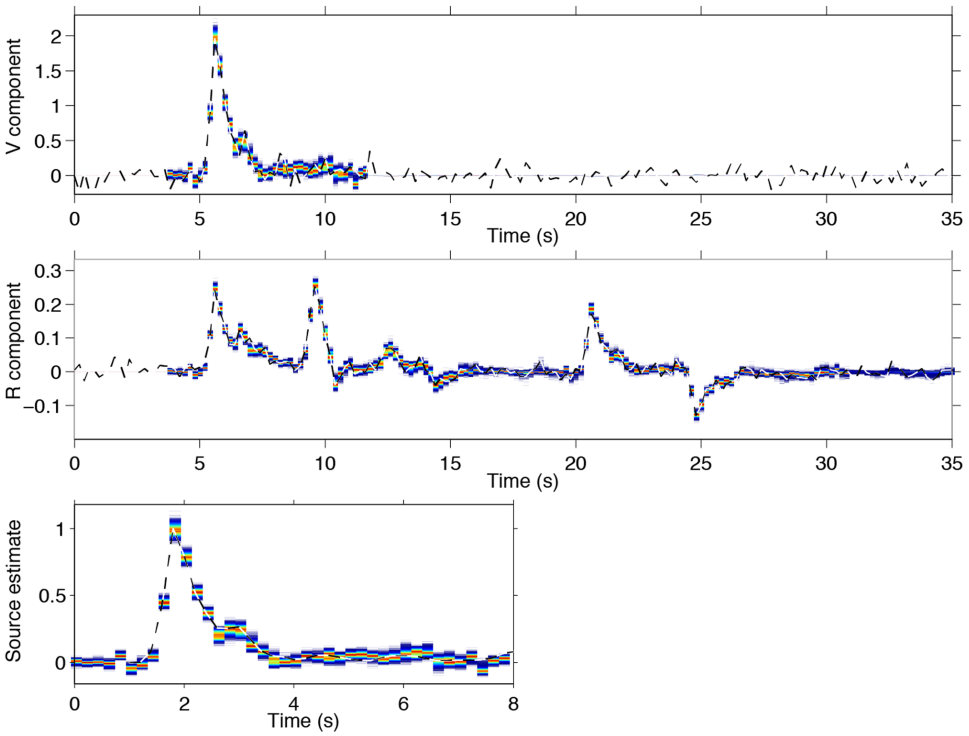

Figure 2. Noisy simulated waveforms (dashed) for (a) vertical and (b) radial components. The range of data predictions produced by the Bayesian sampling is shown as a density for each time sample (colour). Both waveforms are fit well in the inversion. (c) The true STF (dashed) and the probabilistic STF estimate (colour) closely agree.

(Brooks *et al.* 2011), while s is implicitly sampled as a maximum-likelihood estimate for each realisation of $\mathbf{x}$ (Fig. 1). This method is a fully Bayesian approach to the source equalisation problem without any requirement to find tuning parameters required in the processing of RFs. In addition, the expressions in eqs (6)–(8) are straightforward to generalize and can be applied to any number and combinations of receiver components.

In this paper, time-domain stacked seismograms are considered to improve the signal to noise ratio and to better meet the assumption of Gaussian distributed errors. While stacking cancels noise, it inevitably results in a loss (due to averaging) of the distinct information in the various seismograms and it may be advantageous (at higher computational cost) to consider all waveforms jointly (the total likelihood function then becomes the product over the likelihood for each event). However, stacking also provides the advantage that the stacked seismograms from several events are more likely to exhibit Gaussian data errors (by the central limit theorem). Stacks are obtained by following the approach of Shearer (1991) and Kumar *et al.* (2010). In particular, only events of similar magnitude are considered ($M_w \approx 6$) and events are aligned on the P arrival peak and sign reversal is applied when the P arrival amplitude is negative. Formal normal move out correction is not considered here. Rather, normal move out is addressed by limiting the events to a small range of ray parameters (see Supporting Information Fig. S1 for details).

Although not considered in this paper, the application of this STF estimation approach to dense arrays appears feasible and may result in great benefits by reducing the requirement to stack over multiple earthquake events. As long as several stations of an array share reasonably similar receiver-side structure, such stations can be treated together in eq. (8) by including a sum over their components. This case could provide better constraints on the structure underneath the array and also has potential to provide estimates of high-quality STFs for individual events. Larger numbers of stations would result in a better constrained problem (more data with independent noise for the same number of unknowns).

4 INVERSION RESULTS

This section first considers results for simulated waveforms which are generated by convolution of the impulse response for a layered earth model with a complex STF. Then, observed waveforms for 350 events at the HYB station are time-domain stacked and inverted with the new method presented in this paper. Those results are compared to inversion of stacked RF for the same data.

4.1 Simulation

Synthetic seismograms are computed for a model that consists of four isotropic, homogeneous layers over an isotropic half space. The method of Frederiksen & Bostock (2000) is adapted to obtain the impulse response for this layer stack. Next, an effective STF (see Fig. 2) is convolved with the impulse response to produce radial and vertical components. Gaussian noise (zero mean) is added with standard deviations for radial and vertical components of $\sigma_r = 0.012$ and $\sigma_v = 0.1$, respectively. The different noise levels are chosen to examine the hierarchical estimation of different noise levels for each component via eqs (6) and (8).

The model is parametrized in terms of perturbations from a background model (Table 1). This parametrization has the advantage of permitting changes in prior as a function of depth which is otherwise not possible in trans-D partition models. The prior is chosen to be ± 1.2 km s^{-1} for V_s and ± 0.1 for V_p/V_s. Density is treated with an empirical relationship (Gardner *et al.* 1974).

Table 1. Background model used for the inversion of simulated data.

Depth (km)	V_s (km s^{-1})	V_p/V_s
0.0	2.5	1.8
15.0	3.0	1.8
115.0	4.0	1.82

The inversions were carried out on a computer cluster using 36 CPU cores. Each computer core simulates a particular Markov chain and chains interact in terms of parallel tempering steps to increase sampling efficiency (for details on the sampling algorithm, see Dettmer *et al.* 2013). Convergence was judged by visual inspection of chain histories and marginal densities of the first and last quarter of the sample (Dettmer *et al.* 2013). When no significant difference is visible, the inversion is stopped and considered to have converged. Fig. 2 shows the simulated noisy seismograms and the density of all data predictions. To obtain the density of data predictions, a normalized histogram of all data predictions for models in the posterior sample is computed at each time sample. This data-prediction density visualizes the range of data predictions that are supported by the posterior. The range of predictions in Fig. 2 fits the simulated data well. Computed in the same way, the range of STFs sampled by the posterior agrees closely with the true STF that was used to generate the synthetic seismograms (Fig. 2). Fig. 3 considers the marginal densities for the noise standard deviations. Both estimates are consistent with the true values, indicating that the data are fit to noise level (no over or under fitting).

The main inversion results are presented in Fig. 4 in terms of interface probability as a function of depth and marginal profile densities for V_s and V_p/V_s ratio. The positions of interfaces agree closely with the true model for all major interfaces. The minor discontinuity at 50 km depth is just resolved by the data. In addition, the V_s values are well resolved throughout the model and agree closely with the true model. The V_p/V_s ratio is not well resolved. While some limited sensitivity exists above the large discontinuity at 25 km depth, the parameter is essentially undetermined below that depth. The limited sensitivity to the V_p/V_s ratio is consistent with the fact that most information in these data is due to *P*-to-*S* conversions at the major interfaces. Hence, the data exhibit little sensitivity to V_p structure.

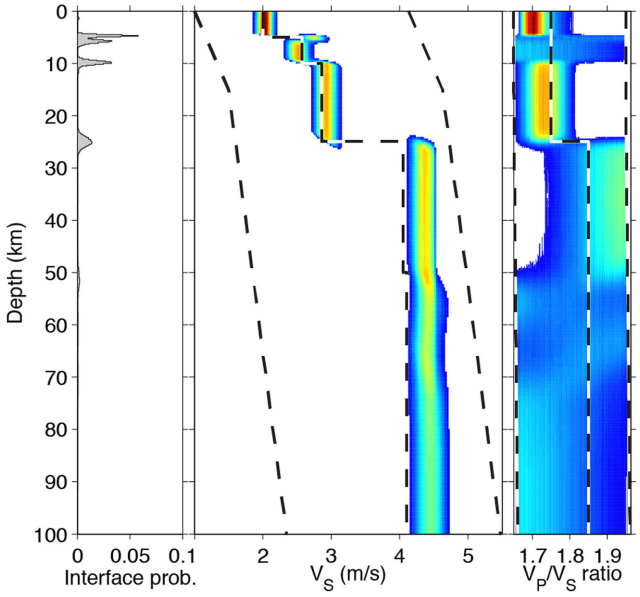

Figure 4. True earth model (dashed) and inversion results in terms of interface probability as a function of depth and profile marginal densities for V_s and V_p/V_s ratio. The extent of the uniform prior bounds are also shown (dashed).

4.2 Application to the Hyderabad station, Indian craton

This section presents inversion results for the joint inversion of vertical and radial seismogram components and surface SWD data (Ekström 2011) for station Hyderabad (HYB), India. The station is located on the eastern Dharwar craton that formed at 2.5 Gyr and stabilized in the early Proterozoic (Kiselev *et al.* 2008). The same data have been considered previously (Bodin *et al.* 2014) albeit with a different stacking method and fewer long periods in the SWD data. Since this study only considers an isotropic, homogeneous and

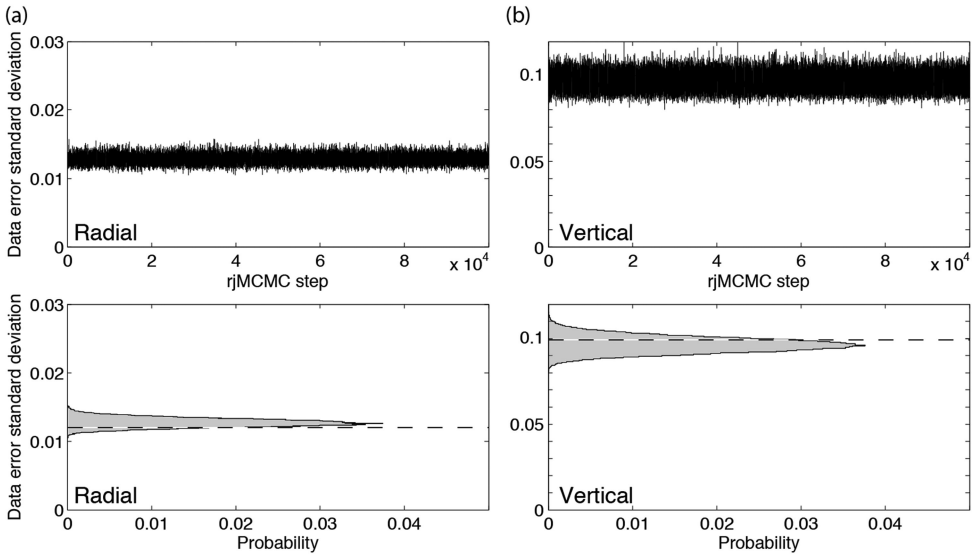

Figure 3. True error standard deviations (dashed) for (a) vertical and (b) radial components and inversion results in terms of chain history (top) and marginal densities (bottom).

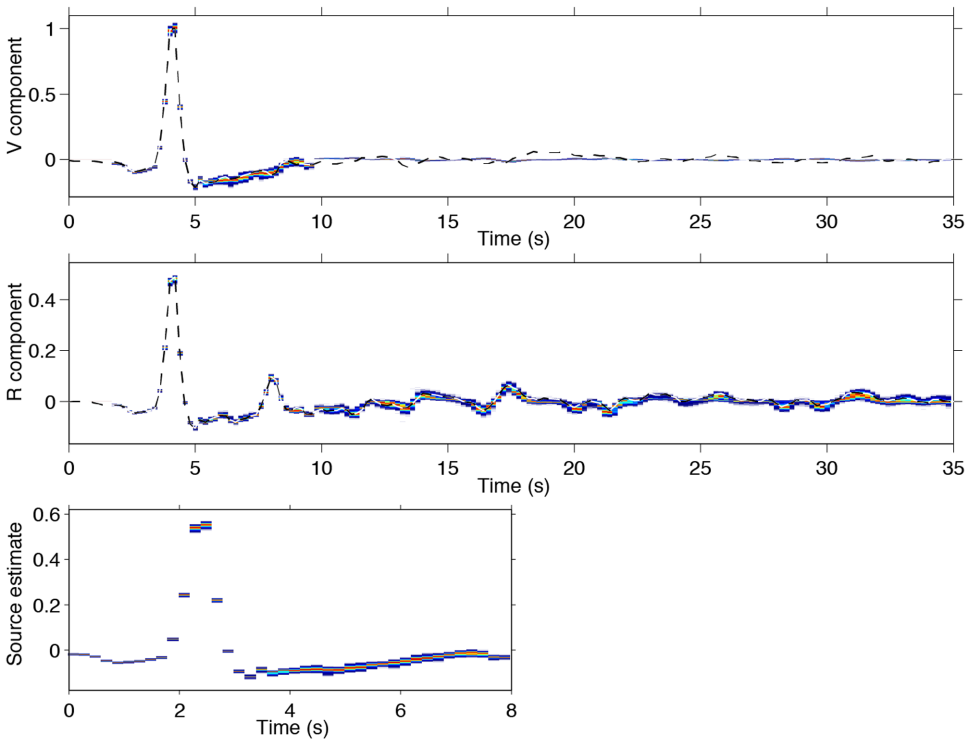

Figure 5. Observed (stacked) waveforms at station HYB (dashed) in terms of (a) vertical and (b) radial components. The range of data predictions produced by the joint Bayesian inversion is shown as a density for each time sample (colour). Both waveforms are fit well in the inversion. (c) The probabilistic STF estimate (colour) for the time-stacked waveforms.

horizontally stratified model for the medium below the receiver, data are limited to backazimuths of 260°–310° and slowness values between 0.056 and see Supporting Information Fig. S1). Transverse components are not considered due to the same limitations in the model. Multiples of *P* waves (*PmP*) may also have an effect on the DSI method when they arrive beyond the extent of the parametrized STF. However, the time-stacked waveforms for HYD do not show evidence of a clear *PmP* (the first would be expected at ∼14 s). This may be due to weakly dipping structure that cause azimuthal smearing which is averaged over in the stacking (Lombardi *et al.* 2008) and/or averaging over epicentral distances (while aligning on the *P* arrival). Both effects are not straightforward to model and would cause a substantial increase in computational cost. Therefore, the results in the work do not include *PmP* multiples in the synthetic seismograms. To obtain the data in Fig. 5, 350 earthquakes of $M_w \approx 6$ that occurred between 1997 and 2007 were stacked in the time domain. The stacking follows the approach of Shearer (1991) and Kumar *et al.* (2010). In particular, events are aligned on the *P* arrival peak and sign reversal is applied when the *P* arrival amplitude is negative. All traces are time windowed to a total length of 35 s starting ∼4 s before the P arrival. The events are then normalized to equal energy and rotated to vertical, radial and tangential components, and finally stacked (see Supporting Information Fig. S2 for all waveforms contributing to the stack). The sampling rate of all waveforms is 0.2 s. The resulting stack shows a short (∼2 s), strong signal on the vertical component for the *P* arrival (Fig. 5) which is consistent with the magnitude of the earthquakes considered ($M_w \approx$ 6 earthquakes typically cause ruptures of ∼10 km extent, limiting the STF duration to a few seconds, Wells & Coppersmith (1994)). The SWD data (Fig. 6) are extracted from Ekström (2011) and span periods from 25 to 250 s. To simplify the treatment of the residual

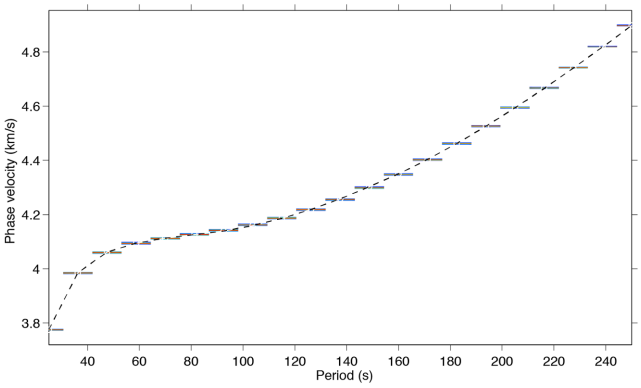

Figure 6. Observed SWD data (dashed) for the region near station HYB (Ekström 2011). The range of data predictions produced by the joint inversion is shown as a density for each time sample (colour). The SWD data are fit well in the inversion.

error statistics, resampling to even spacing in periods was applied for a total of 21 points between 25 and 250 s.

The earth model is parametrized in terms of perturbations from a background model (Table 2) and prior bounds are set to ± 1 km s^{-1} for V_s and ± 0.1 for V_p/V_s. Since the SWD data extend to periods of 250 s, they are sensitive to much greater depths than considered feasible for this joint inversion. Hence, the model for the seismogram data assumes a homogeneous half space below 350 km depth, and the model for SWD data assumes a layered half space that is based on a simplified reference model (see Table 2, Dziewonski & Anderson 1981). The layered half space is perturbed by the half-space value in the trans-D model for both V_s and V_p/V_s. In this way, the model avoids extending the trans-D domain to great

Table 2. Background model used for inversion of measured data. To 350 km depth, the values are applied to both RF and SWD data, below that, the values are given by PREM to better constrain the SWD data at greater depth.

Depth (km)	V_s (km s^{-1})	V_p/V_s
0.0	3.5	1.8
15.0	4.0	1.8
115.0	4.0	1.81
350.0	4.65	1.81
PREM		
371.0	4.75	1.863
400.0	4.77	1.867
471.0	5.14	1.848
571.0	5.43	1.843
600.0	5.52	1.840
670.0	5.57	1.843
771.0	6.24	1.774
871.0	6.31	1.781
971.0	6.38	1.789
1071.0	6.44	1.798

depth (aids efficiency) and also avoids biases in the results due to the greater sensitivity depth of the SWD data. Alternatively, the SWD data could be limited to shorter periods. However, due to the integrating nature of surface waves as a function of depth, a clear depth limitation is difficult to estimate and depends on the inversion result.

For the joint inversions 48 computer cores were employed and each core simulates from a Markov chain. Interactions between chains are in terms of parallel tempering steps and samples are recorded from 4 chains with $\beta = 1$ (see Appendix A). Fundamental-mode dispersion curves are predicted by normal-mode summation with the method of Saito (1988). The D_r and D_v components are considered to have noise that is independent from that of the SWD data. Hence, the likelihood function is defined by the product of likelihoods for the two data types and an additional parameter is required for the SWD error standard deviation. Note that no other weights to scale the relative importance of RF and other data types in the inversion are required in Bayesian inversion. The weight of each data type is intrinsically given by its error standard deviation. In particular, choosing a subjective weighting could substantially change the results and hierarchical estimation such as applied here should be preferred. Convergence of the posterior sampling was judged by comparing the first and last quarter of the sampling history in terms of marginal densities and chain histories for the logarithm of the likelihood value and the model index k (see eq. A1).

Due to the short P arrival (Fig. 5), we set s to have an extent of 8 s ($N_s = 40$). Fig. 5 shows that the D_r and D_v components are both well fit by the inversion and the STF estimate is mostly constrained by the vertical component. Since the fundamental source assumption for this method is a plane P wave incident from below, the P arrival is expected to dominate the D_v component for the ray parameters considered here. Fig. 7 shows marginal densities for the standard deviations of D_r and D_v components. The standard deviation is significantly larger for the D_v component, suggesting that some of the simplifications in the model (horizontal stratification, isotropic) cause some theory error. In addition, note that the standard deviations for D_v and D_r are approximately an order of magnitude larger than the pre-event noise level (see Supporting Information Fig. S4). Hence, theory error dominates for this inversion and it is crucial to consider hierarchical estimation of noise parameters to avoid over-fitting of the data. The fit of SWD data (Fig. 6) is very close, resulting in low values for the SWD residual standard deviation (Fig. 7). Such low residual estimates are likely due to the SWD curve being inferred via regularized inversion (Ekström 2011) rather than measured. However, these data are the best currently available and provide important constraints for the RF inversions.

The main inversion results are shown in Fig. 8 in terms of profile probability densities for V_s and V_p/V_s. The most prominent feature in the V_s density is the Moho, a strong discontinuity at 30 km depth where V_s jumps to ~4.55 km s^{-1}. Above that is a layered crust with a topmost layer of $V_s \sim 3.35$ km s^{-1} extending to 10 km

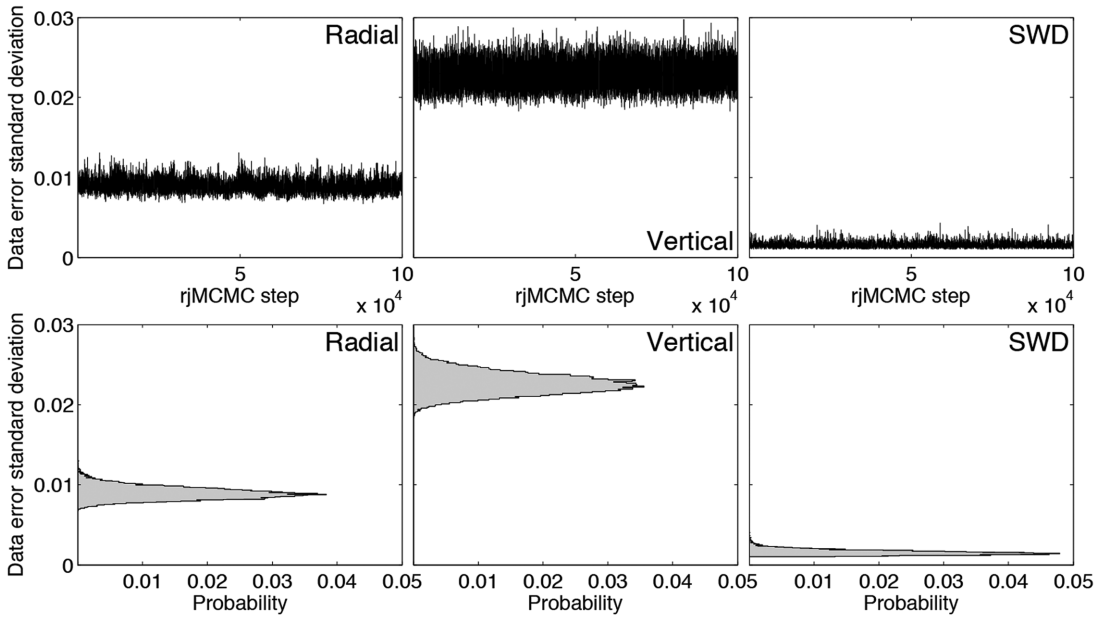

Figure 7. The standard deviation of the three independent data sets considered in the inversion of (a) radial and (b) vertical waveforms, and (c) SWD dispersion data. Note that no subjective choice is required to weight the individual data types.

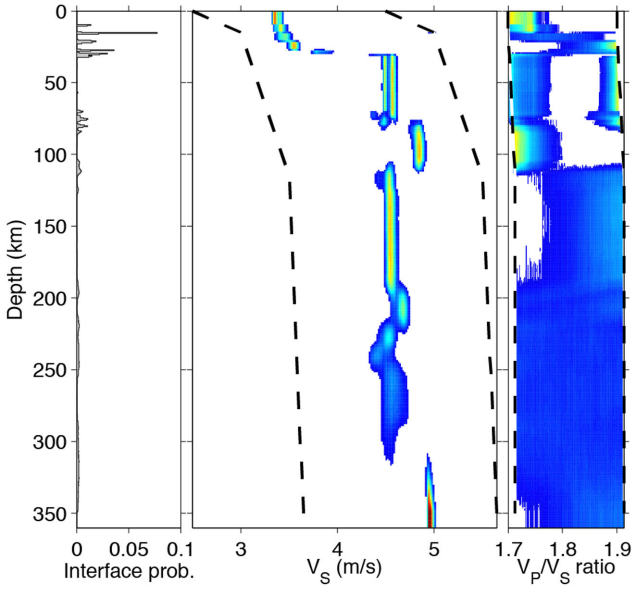

Figure 8. Inversion results (colour) in terms of profile marginal densities for station HYB and prior bounds (dashed).

depth. Between 10 and 30 km depth, the inversion results indicate a gradient with increasing velocity towards the Moho. However, it is possible that the crust has in fact a much simpler structure (two layers above the Moho) and that the gradient is due to the seismic data averaging over some depth (D_r and D_v due to band-limited signal, and SWD due to intrinsic depth integration). The inclusion of SWD data likely increases this averaging effect. Below the Moho, V_s is mostly around 4.55 km s^{-1}, which closely agrees with the results from joint RF (P and S) and teleseismic traveltime residuals in Kiselev *et al.* (2008).

The profile also exhibits strong discontinuities at 75 and 110 km depth. In the layer between these depths V_s increases to 4.8 km s^{-1}. Below 110 km depth, the structure returns to a value of ∼4.55 km s^{-1} until velocities decrease below 220 km. While the lithosphere-asthenosphere boundary is not clearly resolved, these results suggest that it likely occurs at depths below 220 km. Uncertainties at these depths are also significantly larger than in the more shallow parts of the profile, making inferences on structure more challenging. The larger uncertainties with respect to interface positions and velocities at these depths appear reasonable given the much smaller amplitudes in the waveforms to constrain these features and the lower sensitivity in the SWD data. Hence, the approach produces uncertainties that appear to be consistent with the type of data considered.

The results for the V_p/V_s ratio are highly uncertain. Below 100 km depth, the uncertainties comprise the full width of the prior bounds, meaning that the data have virtually no information about V_p within these bounds. At shallower depths, some sensitivity exists but the uncertainties are still high and exhibit multiple modes. It is unlikely that this is due to meaningful V_p structure. Rather, the parameter likely accounts for some of the theory error in the inversion which results in unreasonable V_p/V_s estimates. These results illustrates that if V_p structure is of interest, it is desirable to include additional data (e.g. shear wave RFs, teleseismic traveltime residuals) in joint inversions to provide information about P velocities. In addition to better constraining V_p, such additional data could also improve constraints on V_s by improving the V_p/V_s estimate. However, this additional complication is not considered here and may also require

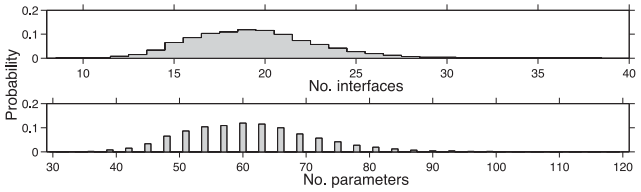

Figure 9. Complexity of the trans-D model in terms of the number of layers (top) and the number of parameters (bottom).

a more general model to reduce theory errors (e.g. by accounting for potential anisotropy and/or dipping interfaces).

Fig. 9 shows a marginal distribution of the model index k (the number of interfaces in the model). The distribution is fairly spread out with significant probability from 10 to 30 interfaces. This means that the uncertainty in terms of model choice (how many layers should be included in the inversion) is high and emphasizes the importance of using a trans-D model to analyse these data.

Bayesian parameter uncertainty estimates depend (potentially strongly) on how well the prior assumptions about the noise are met in the inversion. Fig. 10 considers residual error analysis to examine how well the assumptions used in the derivation of the likelihood function eq. (6) are met. In particular, the derivation assumed that data errors are Gaussian distributed and uncorrelated such that the covariance matrix is represented by only a standard deviation. Hence, the residual error should be both random (uncorrelated) and Gaussian distributed to be consistent with the formulation in eq. (6).

Note that while the seismogram noise is directly accessible (pre-event) it is often not a sufficient approximation of the residual errors in an inversion. The residual errors are a combination of measurement and theory errors, of which the theory errors are shown to dominate in this case (Fig. 7 and Supporting Information Fig. S4). While the pre-event noise is correlated, the magnitude of this noise is ∼10 times less than the residual errors in the inversion.

Randomness is examined here by comparing the density of auto-covariance functions of the residuals for all models in the posterior density to the autocovariance of a Gaussian pseudo random series. To obtain the density, data predictions are computed for all samples in the posterior, which are used to produce a large sample of residual errors. Then, the autocovariance function is computed for each series of residuals in the sample. Finally, the autocovariance functions are displayed in terms of their density as a function of lag (Figs 10a–c). Note that the results for the R component do not display any significant residual correlation (the main lobe of the density closely resembles a delta function). For the D_v component, the central peak is also very narrow but some residual correlation exists at larger lags (likely due to the later times in the waveform). These correlations are not likely to have significant impacts on the inversion results.

Similarly, histogram densities of standardized residuals (normalized by the standard deviation for each sample) are computed for the three data types and are compared to a Gaussian distribution with zero mean and unit standard deviation (Figs 10d–f). The residuals for both D_r and D_v components closely resemble a Gaussian distribution, which increases the confidence in the inversion results.

The auto-covariances for the SWD residuals indicate the strongest correlations. For this reason, the likelihood function for the SWD data included an autoregressive (AR) error model (Dettmer *et al.* 2012). However, given that the curve includes only 21 data, estimation of any noise parameters is challenging. Hence, the inclusion of the AR process had little effect on the inversion results. The SWD residual histograms are also highly uncertain due to the small

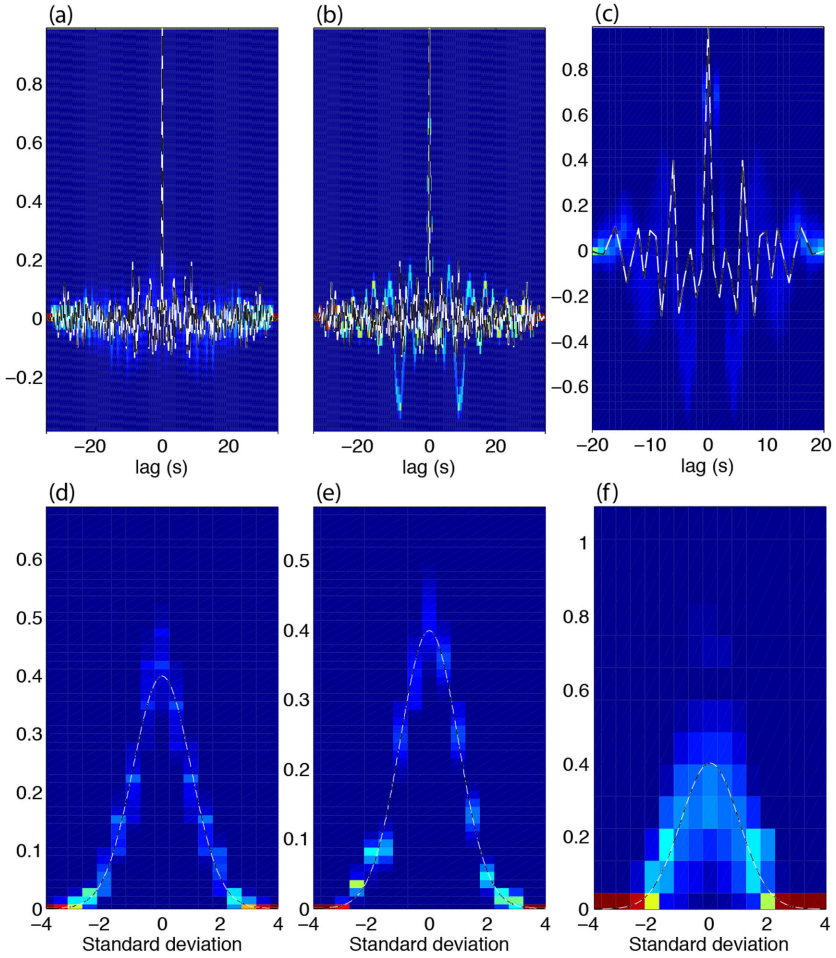

Figure 10. Residual analysis for HYB data in terms of autocovariance (a–c) and residual histogram (d–f) densities (colour) compared to simulated Gaussian (zero mean, unit standard deviation) densities (dashed). For (a, d) radial and (b, e) vertical components, and (c, f) SWD data.

number of data. These results suggest that these SWD are not ideally suited for Bayesian inversion but since no other data are available, are still included in the inversion as they provide important information in terms of absolute velocities. It is possible that the small errors on the data (rooted in the approach used to infer the SWD curve) result in underestimated uncertainties in the deeper part of the profile (where the SWD data provide most of the information).

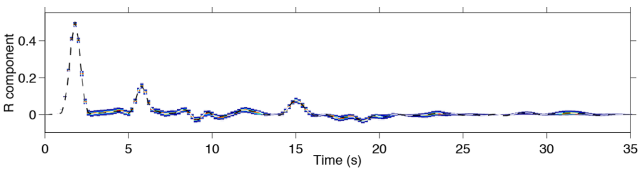

Figure 11. Observed RF (dashed) and range of data predictions (colour) for station HYB.

4.3 Comparison to receiver function inversion

This section presents RF inversion results for station HYB. The purpose of this comparison is to examine what impact the different formulation has on inversion results. Differences can be due to the following points: First, tuning parameters that are required for classic RF processing are problematic in uncertainty estimation since the estimates do not contain the effects of uncertainty in those tuning parameters. Second, the seismogram inversion treats the STF as unknown and uncertain, resulting in rigorous (and likely larger) uncertainty estimates. Third, avoiding deconvolution by noisy (observed) waveforms results in more data errors that are less correlated than filtered and deconvolved RF.

The RF was processed by applying water level deconvolution (Ammon 1991) to the events selected in Section 4.2. The processing for all events was performed with a water level of 0.005 and a Gaussian filter width of 2.5 rad s^{-1}. Both parameters were set subjectively in a trial-and-error approach based on visual inspection of many waveforms. Nonetheless, many of the events used in to generate the stacks in the previous section produced unstable RF results. Therefore, before stacking, an additional selection step (Tkalčić *et al.* 2006) based on the correlation coefficient at zero lag was applied to obtain a reasonable RF stack (Fig. 11). The RFs contributing to the final stack are required to have a 0.9 correlation with at least 5 other waveforms in the stack (see Supporting Information Fig. S3 for details on the selection process).

The inversion was carried out for a joint likelihood function that includes RF and SWD data. The seismogram inversion in the previous sections requires no specification of a Gaussian pulse (and hence avoids these parameters altogether) since predicted waveforms are based on the impulse responses convolved with the STF estimate s (eq. 8). The delay time was set to be 0 s, so that data predictions are relevant for the likelihood computation starting at

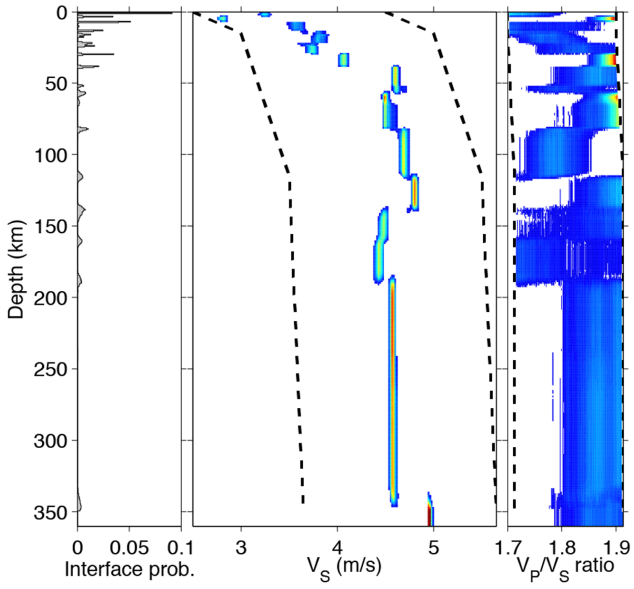

Figure 12. Receiver function inversion results (colour) in terms of profile marginal densities for station HYB and prior bounds (dashed). Note the much smaller (unrealistic) uncertainties than for the DSI result in Fig. 8.

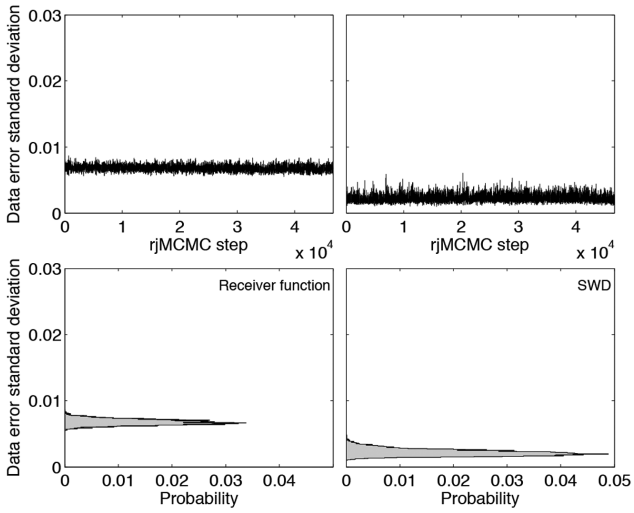

Figure 13. Estimated standard deviation for the RF and SWD data.

the peak of the *P* arrival (0 s). Otherwise, the algorithm and prior are identical to those applied in Section 4.2. The main inversion result is shown in Fig. 12 in terms of profile densities for V_s and V_p/V_s. The crustal structure in these results is complicated and shows clear signs of over-parametrization: V_s values of several layers jump back and forth. This problem is likely exacerbated by the trans-D model. The trans-D model estimates model complexity based on the hierarchical noise estimate (Fig. 13). The lower such noise estimates, the higher the model complexity. The problem in the RF is that the highly processed waveforms exhibits little noise and the statistical attributes of the noise are difficult to understand (it is not clear how deconvolution, water level and filtering affect the noise characteristics). Therefore, the noise is highly correlated and while estimating correlated noise in hierarchical models is possible (Bodin *et al.* 2012; Dettmer *et al.* 2012; Kolb & Lekic 2014), such estimation can be difficult and relies on model assumptions that may not be met for observed data. Hence, the seismogram inversion from the

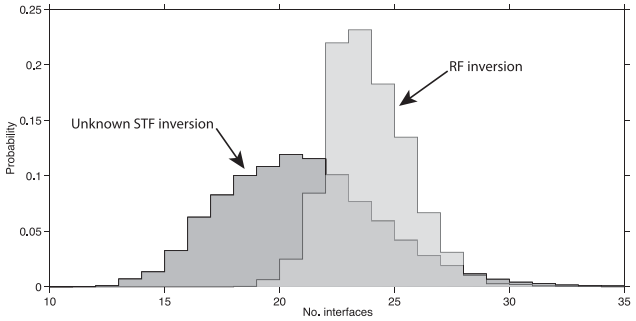

Figure 14. Comparison of model complexity of the trans-D model for the inversion with unknown STF (Section 4.2) and for the RF inversion (Section 4.3). Note that the uncertainty in the number of layers is much lower for the RF inversion. In addition, the RF inversion overestimates the number of layers which results in unreasonably complex shallow structure.

Section 4.2 appears advantageous, since it avoids signal processing steps that substantially change noise characteristics.

The problem of over-estimated complexity (i.e. over parametrization) is also visible when marginal distributions for the number of interfaces are considered (Fig. 14). The *k* marginal for the inversion with unknown STF is considerably wider than in the case of RF inversion. In addition, for RF inversion, *k* peaks at a higher value which suggest that the algorithm estimates higher complexity for the RF data in contrast to the D_r and D_v data. However, these effects are likely due to a lack in the ability to properly account for the effects of signal processing on noise.

The results in Fig. 12 also show little increase in uncertainty as a function of depth which is unlikely to be realistic. To compare the uncertainties between the two approaches, Fig. 15 displays the inversion parameters in terms of the density of layer nodes and perturbation values from the background profile (Table 2). A layer node is considered to be located at the top interface of a layer and defines the homogeneous region below until the next interface is encountered. To emphasize the difference in uncertainty (given the limited dynamic range of the colour scale), the base 10 logarithm of the density is displayed. The RF results have unreasonably small uncertainties in both V_s and V_p/V_s. These under-estimated uncertainties are likely due to the strong effects that signal processing (deconvolution and filtering) has on the noise in the seismograms. Since it is unclear how to account for such processing in the Bayesian framework efficiently (treating signal processing parameters as unknown appears to be computationally prohibitive today), rigorous uncertainty estimation is not possible.

5 CONCLUSION

This work considered a new approach to Bayesian inversion for receiver-side structure where the STF is treated as unknown. This substantially simplifies the data processing as it does not require any deconvolution or specification of tuning parameters. Rather, seismogram components are directly inverted by formulating a joint likelihood function for the seismogram components while assuming the STF to be unknown. The unknown STF is here estimated by a closed-form expression for the maximum likelihood STF given a realisation of the parameter vector. This expression is obtained by setting to zero the partial derivative of the likelihood function with respect to the unknown source parameters. Importantly, this approach accounts for the limited knowledge about the STF and accounts for this in the inversion results (producing larger, more realistic uncertainty estimates for V_s profiles).

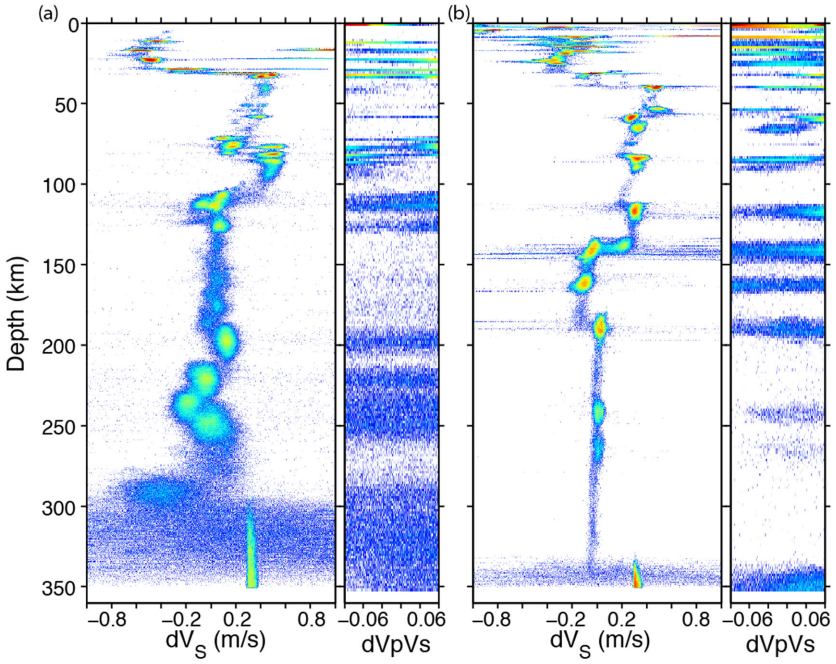

Figure 15. Comparison of inversion results for (a) DSI and (b) RF inversion. Results are shown in terms of the logarithm (base 10) of the perturbation position density. In (a), uncertainty increases with depth but the half-space is closely constrained due to the small errors on the long-period SWD data. In (b), uncertainty estimates are unrealistically small due to signal processing effects on data noise and since the RF inversion does not treat the STF as uncertain.

Since the inversion works on time-domain stacked seismograms and avoids filtering and deconvolution, the errors on the data are much more straightforward to understand and produce more meaningful results when used with trans-D inference. A significant issue with Bayesian inference is that the inferred model parameter uncertainty is closely linked to the noise on the data (which may include significant theory error). When signal processing steps (deconvolution, filtering) substantially change the noise characteristics, Bayesian algorithms may produce unreasonable uncertainty results. This problem is greatly exacerbated when the model parametrization is considered to be unknown as in trans-D models. We have shown that classic RF inversion may produce misleading results and that, in the examples considered here, the new approach produces much more reasonable inferences.

A fundamental assumption of the likelihood derivation as presented in this work is that the noise on the vertical, radial and transverse components is independent. This may not always be true for components of the same receiver. In principle, the likelihood function can be generalized to consider noise correlations between receiver components by including an appropriate covariance matrix. However, it may be more sensible to consider multiple receivers in an array. Therefore, we note that the approach is general and is straightforward to extend to any number and combination of seismogram components. In addition, the application is not limited to the inversion of receiver-side structure but can be potentially useful in inversion methods that consider other seismic phase, such as ScP or other phases that are sensitive to deep earth structure but require accounting for unknown STFs.

ACKNOWLEDGEMENTS

Inversions were performed on the Terrawulf III computational facility and a computer cluster operated by the authors at the University of Victoria. Terrawulf III is supported through the AuScope Australian Geophysical Observing System. AusScope is funded under the National Collaborative Research Infrastructure Strategy and the Education Investment Fund (EIF3), both Australian Commonwealth Government Programs. The computer facility at the University of Victoria is funded by the Canadian National Science and Engineering Research Council and by the United States Office of Naval Research. We would like to thank two anonymous reviewers and the editor for thorough reviews that substantially improved the quality of the manuscript. The research algorithm is available upon request.

REFERENCES

Agostinetti, N.P. & Malinverno, A., 2010. Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.,* **181,** 858–872.

Ammon, C.J., 1991. The isolation of receiver effects from teleseismic P waveforms, *Bull. seism. Soc. Am.,* **81**(6), 2504–2510.

Ammon, C.J., Randall, G.E. & Zandt, G., 1990. On the nonuniqueness of receiver function inversions, *J. geophys. Res.,* **95**(B10), 15303, doi:10.1029/JB095iB10p15303.

Anderson, E. *et al.*, 1999. *LAPACK Users' Guide,* 3rd edn, Society for Industrial and Applied Mathematics.

Audet, P., Bostock, M.G., Christensen, N.I. & Peacock, S.M., 2009. Seismic evidence for overpressured subducted oceanic crust and megathrust fault sealing, *Nature,* **457**(7225), 76–78.

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.,* **178,** 1411–1436.

Bodin, T., Yuan, H. & Romanowicz, B., 2014. Inversion of receiver functions without deconvolution–application to the Indian craton, *Geophys. J. Int.,* **196**(2), 1025–1033.

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

Bostock, M., 2002. Kirchhoff-approximate inversion of teleseismic wavefields, *Geophys. J. Int.,* **149**(3), 787–795.

Bostock, M.G., 2007. Theory and observations—teleseismic body-wave scattering and receiver-side structure, in *Treatise on Geophysics Volume 1: Seismology and the Structure of the Earth,* pp. 219–246, eds Romanowicz, B. & Dziewonski, A., Elsevier.

Bostock, M.G. & Rondenay, S., 1999. Migration of scattered teleseismic body waves, *Geophys. J. Int.,* **137,** 732–746.

Bostock, M.G., Hyndman, R.D., Rondenay, S. & Peacock, S.M., 2002. An inverted continental Moho and serpentinization of the forearc mantle, *Nature,* **417,** 536–538.

Brillon, C., Cassidy, J.F. & Dosso, S.E., 2013. Onshore/offshore structure of the Juan de Fuca Plate in Northern Cascadia from Bayesian receiver function inversion, *Bull. seism. Soc. Am.,* **103**(5), 2914–2920.

Brooks, S., Gelman, A., Jones, G. & Meng, X. (eds), 2011. *Handbook of Markov chain Monte Carlo,* pp. 1–592, Springer.

Burdick, L.J. & Langston, C.A., 1977. Modeling crustal structure through the use of converted phases in teleseismic body-wave forms, *Bull. seism. Soc. Am.,* **67**(3), 677–691.

Cassidy, J.F., 1992. Numerical experiments in broadband receiver functions, *Bull. seis Soc. Am.,* **82**(3), 1453–1474.

Chambers, K., Deuss, A. & Woodhouse, J.H., 2005. Reflectivity of the 410-km discontinuity from PP and SS precursors, *J. geophys. Res.,* **110**(B2), B02301, doi:10.1029/2004JB003345.

Clayton, R.W. & Wiggins, R.A., 1976. Source shape estimation and deconvolution of teleseismic bodywaves, *Geophys. J. R. astr. Soc.,* **47**(1), 151–177.

Dettmer, J. & Dosso, S.E., 2012. Trans-dimensional matched-field geoacoustic inversion with hierarchical error models and interacting Markov chains, *J. acoust. Soc. Am.,* **132,** 2239–2250.

Dettmer, J., Dosso, S.E. & Holland, C.W., 2010. Trans-dimensional geoacoustic inversion, *J. acoust. Soc. Am.,* **128,** 3393–3405.

Dettmer, J., Holland, C.W. & Dosso, S.E., 2013. Trans-dimensional uncertainty estimation for dispersive seabed sediments, *Geophysics,* **78,** WB63–WB76.

Dettmer, J., Molnar, S., Steininger, G.A., Dosso, S.E. & Cassidy, J.F., 2012. Trans-dimensional inversion of microtremor array dispersion data with hierarchical autoregressive error models, *Geophys. J. Int.,* **188,** 719–734.

Dettmer, J., Benavente, R., Cummins, P.R. & Sambridge, M., 2014. Trans-dimensional finite-fault inversion, *Geophys. J. Int.,* **199,** 735–751.

Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.,* **25**(4), 297–356.

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

Farra, V. & Vinnik, L.P., 2000. Upper mantle stratification by P and S receiver functions, *Geophys. J. Int.,* **141,** 699–712.

Flanagan, M.P. & Shearer, P.M., 1998. discontinuities by stacking SS precursors Bounce points, *J. geophys. Res.,* **103**(97), 2673–2692.

Frederiksen, A.W. & Bostock, M.G., 2000. Modelling teleseismic waves in dipping anisotropic structures, *Geophys. J. Int.,* **141**(2), 401–412.

Frederiksen, A.W., Folsom, H. & Zandt, G., 2003. Neighbourhood inversion of teleseismic Ps conversions for anisotropy and layer dip, *Geophys. J. Int.,* **155,** 200–212.

Gardner, G. H.F., Gardner, L.W. & Gregory, A.R., 1974. Formation velocity and density-the diagnostic basis for stratigraphic traps, *Geophysics,* **39**(6), 770–780.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B., 2003. *Bayesian Data Analysis,* Chapman and Hall/CRC.

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

Idehara, K., 2011. Structural heterogeneity of an ultra-low-velocity zone beneath the Philippine Islands: implications for core–mantle chemical interactions induced by massive partial melting at the bottom of the mantle, *Phys. Earth planet. Inter.,* **184,** 80–90.

Julia, J., Ammon, C.J., Herrmann, R.B. & Correig, A.M., 2000. Joint inversion of receiver function and surface wave dispersion observations, *Geophys. J. Int.,* **143,** 99–112.

Kind, R. *et al.*, 2002. Seismic images of the crust and upper mantle beneath Tibet: evidence for eurasian plate subduction references, *Science,* **298**(5596), 1–4.

Kind, R., Yuan, X. & Kumar, P., 2012. Seismic receiver functions and the lithosphereasthenosphere boundary, *Tectonophysics,* **536–537,** 25–43.

Kiselev, S., Vinnik, L., Oreshin, S., Gupta, S., Rai, S.S., Singh, A., Kumar, M.R. & Mohan, G., 2008. Lithosphere of the Dharwar craton by joint inversion of P and S receiver functions, *Geophys. J. Int.,* **173**(3), 1106–1118.

Kolb, J.M. & Lekic, V., 2014. Receiver function deconvolution using trans-dimensional hierarchical Bayesian inference, *Geophys. J. Int.,* **197**(3), 1719–1735.

Kumar, P., Kind, R. & Yuan, X., 2010. Receiver function summation without deconvolution, *Geophys. J. Int.,* **180**(3), 1223–1230.

Langston, C.A., 1979. Structure under Mount Rainier, Washington, inferred from teleseismic body waves, *J. geophys. Res.,* **84**(9), 4749–4762.

Ligorria, P. & Ammon, C.J., 1999. Iterative deconvolution and receiver-function estimation, *Bull. seism. Soc. Am.,* **89**(5), 1395–1400.

Lombardi, D., Braunmiller, J., Kissling, E. & Giardini, D., 2008. Moho depth and Poisson's ratio in the Western-Central Alps from receiver functions, *Geophys. J. Int.,* **173**(1), 249–264.

MacKay, D.J.C., 2003. *Information Theory, Inference, and Learning Algorithms,* pp. 343–386, Cambridge Univ. Press.

Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.,* **151,** 675–688.

Menke, W. & Levin, V., 2003. The cross-convolution method for interpreting SKS splitting observations, with application to one and two-layer anisotropic earth models, *Geophys. J. Int.,* **154,** 379–392.

Miller, M.S. & Piana Agostinetti, N., 2012. Insights into the evolution of the Italian lithospheric structure from S receiver function analysis, *Earth planet. Sci. Lett.,* **345–348,** 49–59.

Minsley, B.J., 2011. A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data, *Geophys. J. Int.,* **187,** 252–272.

Owens, T.J., 1987. Crustal structure of the Adirondacks determined from broadband teleseismic waveform modeling, *J. geophys. Res.,* **92**(B7), 6391–6401.

Owens, T.J., Zandt, G. & Taylor, S.R., 1984. Seismic evidence for an ancient rift beneath the Cumberland Plateau, Tennessee: a detailed analysis of broadband teleseismic P waveforms, *J. geophys. Res.,* **89,** 7783–7795.

Pachhai, S., Tkalčić, H. & Dettmer, J., 2014. Bayesian inference for ultra-low velocity zones in the Earth's lowermost mantle: multiple-layer ULVZ beneath the east of the Philippines, *J. geophys. Res.*, **119**, 8346–8365.

Phinney, R.A., 1964. Structure of the Earth's crust from spectral behavior of long-period body waves, *J. geophys. Res.,* **69**(14), 2997–3017.

Royer, A.A., Bostock, M.G. & Haber, E., 2012. Blind deconvolution of seismograms regularized via minimum support, *Inverse Probl.,* **28**(12), 125 010–125 027.

Saito, M., 1988. Disper80: a subroutine package for the calculation of seismic normal-mode solutions, in *Seismological Algorithms,* pp. 294–319, ed. Doornbos, D.J., Academic Press.

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

Sambridge, M., 2014. A parallel tempering algorithm for probabilistic sampling and multimodal optimization, *Geophys. J. Int.,* **196,** 357–374.

Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2006. Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.,* **167,** 528–542.

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

Shearer, P.M., 1991. Imaging global body wave phases by stacking long-period seismograms, *J. geophys. Res.,* **96,** 20 353–20 364.

Shen, W., Ritzwoller, M.H., Schulte-Pelkum, V. & Lin, F.-C., 2012. Joint inversion of surface wave dispersion and receiver functions: a Bayesian Monte-Carlo approach, *Geophys. J. Int.,* **192**(2), 807–836.

Stähler, S.C. & Sigloch, K., 2014. Fully probabilistic seismic source inversion—part 1: Efficient parameterisation, *Solid Earth,* **5,** 1055–1069.

Stipčević, J., Tkalčić, H., Herak, M., Markušić, S. & Herak, D., 2011. Crustal and uppermost mantle structure beneath the External Dinarides, Croatia,

determined from teleseismic receiver functions, *Geophys. J. Int.,* **185**(3), 1103–1119.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* pp. 1–57, SIAM.

Tkalčić, H., Pasyanos, M.E., Rodgers, A.J., Gök, R., Walter, W.R. & Al-Amri, A., 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**(B11), B11311, doi:10.1029/2005JB004130.

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

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 planet. Sci. Lett.,* **225**(1–2), 131–146.

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, *Izv. Phys. Solid Earth,* **42**(8), 639–651.

Wells, D.L. & Coppersmith, K.J., 1994. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. seism. Soc. Am.,* **84**(4), 974–1002.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this paper:

Figure S1. Slowness-azimuth pairs for all $M_w = 6$ events from 1997 to 2007 at station HYB and those that were selected for this study (red).

Figure S2. Seismograms for all chosen events (grey) after aligning on the P arrival and sign reversing. The stacked waveform (red) used in this study and one standard deviation bounds (green) are also shown.

Figure S3. Receiver functions for (a) all chosen events from Fig. 1 (grey) and those used for stacking this study (black). (b) The RFs used for stacking (grey) and The final stacked RF used in the inversion (black).

Figure S4. Pre-event noise of time-stacked radial and vertical components: The pre-event time series for (a) radial and (b) vertical components and the auto covariance functions (c, d) indicate some correlation. While the noise is correlated, the histograms in (e) and (f) suggest it is reasonably Gaussian and agrees well with a theoretical Gaussian (dashed). Note that this noise does not include any theory error due to limitations in the data prediction and/or model parametrization. Importantly, the standard deviation of the radial and vertical components is ∼0.0025 and ∼0.0042, respectively. These standard deviations are approximately one order of magnitude lower than the hierarchical estimates from the inversion, showing that theory error is dominant. Therefore, it is not sufficient to consider these noise esimates only. Rather, this work treats the noise as unknown and infers noise parameters as part of the inversion. (http://gji.oxfordjournals.org/lookup/suppl/doi:10.1093/gji/ggv375/-/DC1).

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

APPENDIX A: BAYESIAN INFERENCE WITH TRANS-DIMENSIONAL MODELS

Bayesian methods estimate parameter values and uncertainties by quantifying the posterior probability density which combines prior information (independent of the data) and data information (through the likelihood function) to specify the state of knowledge about the parameters. For completeness, the following section briefly reviews the Bayesian formulation as presented in our previous work (Dettmer *et al.* 2010, 2012, 2014). More complete treatment can be found elsewhere (MacKay 2003; Brooks *et al.* 2011). In particular, the following considers the parametrization in terms of the number of earth layers to be unknown which we refer to as a trans-D model (Green 1995; Malinverno 2002; Sambridge *et al.* 2006; Bodin & Sambridge 2009; Agostinetti & Malinverno 2010; Minsley 2011; Dettmer & Dosso 2012). Therefore, the inversion results account for the unknown parametrization in the posterior parameter and uncertainty estimates, which are deemed to be a more appropriate representation of the data and prior information.

Let $\mathbf{d}$ be a random variable of N observed data and $\mathcal{M}_k$ denote a group of models specifying particular choices of physical theory, model parametrization and error statistics, where $k \in \mathcal{K}$ and $\mathcal{K}$ is a countable set. In this particular application, k will index the number of interfaces in the parametrization. Let $\mathbf{m}_k$ be a random variable of M_k parameters representing a realisation of model $\mathcal{M}_k$. Green (1995) shows that Bayes' rule can be written for a Bayesian hierarchical model to include parameter k

$$p(k, \mathbf{m}_k|\mathbf{d}) = \frac{p(k)p(\mathbf{d}|k, \mathbf{m}_k)p(\mathbf{m}_k|k)}{\sum_{k' \in \mathcal{K}} \int p(k')p(\mathbf{d}|k', \mathbf{m}'_{k'})p(\mathbf{m}'_{k'}|k')d\mathbf{m}'_{k'}}, \tag{A1}$$

where $p(k)$ is the prior over the $\mathcal{K}$ models considered. The state variables $(k, \mathbf{m}_k)$ are of dimension $M_k + 1$ and the state space is trans-D and given by the union of all fixed-dimensional spaces in $\mathcal{K}$, that is, $\bigcup_{k \in \mathcal{K}}(\{k\} \times \mathbb{R}^{M_k})$. A Markov chain that samples this state space can be defined and converges to the trans-D posterior $p(k, \mathbf{m}_k|\mathbf{d})$. Note that the posterior probability density $p(k, \mathbf{m}_k|\mathbf{d})$ intrinsically addresses model selection and typical inferences about expectations do not require the computation of normalizing constants [the denominator in eq. (A1)].

The conditional probability $p(\mathbf{d}|k, \mathbf{m}_k)$ in eq. (A1) describes the residual-error statistics, where residual errors are defined as the difference of observed and predicted data. For observed (fixed) data, $p(\mathbf{d}|k, \mathbf{m}_k)$ is interpreted as the likelihood function $L(\mathbf{x})$, where $\mathbf{x} = (k, \mathbf{m}_k)$ is the parameter vector (seismic velocities, layering geometry, . . .). The likelihood function quantifies how well data are fit by predictions and is a crucial component in any Bayesian formulation. The derivation of the likelihood function is considered in detail in Section 3.

To obtain parameter and uncertainty inferences, the posterior density $p(k, \mathbf{m}_k|\mathbf{d})$ must be estimated via numerical integration ('sampling'). Due to the high dimension of the model space, care must be taken in choosing efficient algorithms for the sampling. Here Markov-chain Monte Carlo (MCMC) sampling is applied with the Metropolis-Hastings-Green (MHG) algorithm (Brooks *et al.* 2011). The MHG algorithm

simulates a Markov chain by proposing new states $\mathbf{x}' = (k', \mathbf{m}'_{k'})$ of the chain based on the current state $\mathbf{x}$ and a proposal distribution q centred on the current state. The proposed state is then accepted/rejected based on the acceptance probability

$$\alpha = \min\left[1, \frac{p(\mathbf{x}')}{p(\mathbf{x})}\left(\frac{L(\mathbf{x}')}{L(\mathbf{x})}\right)^{\beta}\frac{q(\mathbf{x}|\mathbf{x}')}{q(\mathbf{x}'|\mathbf{x})}|\mathbf{J}|\right], \tag{A2}$$

where the annealing parameter β is used for sampling efficiency when the algorithm is implemented with many Markov chains in parallel. Such parallel tempering can substantially increase sampling efficiency for trans-D models (Dettmer & Dosso 2012; Sambridge 2014).

APPENDIX B: MAXIMUM LIKELIHOOD SOURCE–TIME FUNCTION ESTIMATE

To obtain eq. (8) from eq. (7) requires solving $\partial\phi/\partial s_l = 0$ for s. The misfit ϕ can be written (omitting terms that do not depend on s)

$$\begin{aligned}\phi(\mathbf{x}) &\propto \sum_{j=1}^{N}\left((D_v)_j - \sum_{k=1}^{N_s} s_k (E_v(\mathbf{x}))_{j-k}\right)^2 \Big/ 2\sigma_v^2 + \left((D_r)_j - \sum_{k=1}^{N_s} s_k (E_r(\mathbf{x}))_{j-k}\right)^2 \Big/ 2\sigma_r^2 \\ &= \sum_{j=1}^{N}\left((D_v)_j^2 - 2(D_v)_j \sum_{k=1}^{N_s} s_k (E_v(\mathbf{x}))_{j-k} + \left(\sum_{k=1}^{N_s} s_k (E_v(\mathbf{x}))_{j-k}\right)^2\right)\Big/ 2\sigma_v^2 \\ &\quad + \left((D_r)_j^2 - 2(D_r)_j \sum_{k=1}^{N_s} s_k (E_r(\mathbf{x}))_{j-k} + \left(\sum_{k=1}^{N_s} s_k (E_r(\mathbf{x}))_{j-k}\right)^2\right)\Big/ 2\sigma_r^2 .\end{aligned} \tag{B1}$$

First we consider the maximum likelihood estimate for jth source term (omitting σ_v, σ_r, and $\mathbf{x}$ for simplicity)

$$\begin{aligned}0 = \partial\phi/\partial s_l &= -2\sum_{j=1}^{N}(D_v)_j (E_v)_{j-l} + 2\sum_{j=1}^{N}\left(\sum_{k=1}^{N_s} s_k (E_v)_{j-k}\right)(E_v)_{j-l} \\ &\quad - 2\sum_{j=1}^{N}(D_r)_j (E_r)_{j-l} + 2\sum_{j=1}^{N}\left(\sum_{k=1}^{N_s} s_k (E_r)_{j-k}\right)(E_r)_{j-l} .\end{aligned} \tag{B2}$$

Rearranging gives

$$\sum_{j=1}^{N}(D_v)_j (E_v)_{j-l} + \sum_{j=1}^{N}(D_r)_j (E_r)_{j-l} = \sum_{j=1}^{N}\left(\sum_{k=1}^{N_s} s_k (E_v)_{j-k}\right)(E_v)_{j-l} + \sum_{j=1}^{N}\left(\sum_{k=1}^{N_s} s_k (E_r)_{j-k}\right)(E_r)_{j-l}, \tag{B3}$$

which holds for all $l = 1, \ldots, N$ and can be written more compact in terms of convolutions (including σ_v and σ_r explicitly again)

$$D_v * E_v/\sigma_v^2 + D_r * E_r/\sigma_r^2 = (s * E_v) * E_v/\sigma_v^2 + (s * E_r) * E_r/\sigma_r^2 , \tag{B4}$$

which can be rearranged to give eq. (7) for s

$$s = \frac{D_v * E_v/\sigma_v^2 + D_r * E_r/\sigma_r^2}{E_v * E_v/\sigma_v^2 + E_r * E_r/\sigma_r^2} . \tag{B5}$$