UC Santa Barbara

UC Santa Barbara Previously Published Works

Title

An adaptive Bayesian inversion for upper-mantle structure using surface waves and scattered body waves

Permalink

https://escholarship.org/uc/item/4tv9d52h

Journal

Geophysical Journal International, 214(1)

ISSN

0956-540X

Authors

Eilon, Zachary
Fischer, Karen M
Dalton, Colleen A

Publication Date

2018-07-01

DOI

10.1093/gji/ggy137

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NoDerivatives License, availalbe at https://creativecommons.org/licenses/by-nd/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Geophysical Journal International

Geophys. J. Int. (2018) **214,** 232–253
Advance Access publication 2018 April 03
GJI Seismology

doi: 10.1093/gji/ggy137

An adaptive Bayesian inversion for upper-mantle structure using surface waves and scattered body waves

Zachary Eilon,[1,2] Karen M. Fischer[1] and Colleen A. Dalton[1]

[1]*Department of Earth, Environmental, and Planetary Sciences, Brown University, Providence, RI 02912, USA. E-mail: eilon@ucsb.edu*
[2]*Department of Earth Science, University of California Santa Barbara, Santa Barbara, CA 93106, USA*

Accepted 2018 April 2. Received 2018 March 27; in original form 2017 August 9

SUMMARY

We present a methodology for 1-D imaging of upper-mantle structure using a Bayesian approach that incorporates a novel combination of seismic data types and an adaptive parametrization based on piecewise discontinuous splines. Our inversion algorithm lays the groundwork for improved seismic velocity models of the lithosphere and asthenosphere by harnessing the recent expansion of large seismic arrays and computational power alongside sophisticated data analysis. Careful processing of *P*- and *S*-wave arrivals isolates converted phases generated at velocity gradients between the mid-crust and 300 km depth. This data is allied with ambient noise and earthquake Rayleigh wave phase velocities to obtain detailed V_S and V_P velocity models. Synthetic tests demonstrate that converted phases are necessary to accurately constrain velocity gradients, and *S–p* phases are particularly important for resolving mantle structure, while surface waves are necessary for capturing absolute velocities. We apply the method to several stations in the northwest and north-central United States, finding that the imaged structure improves upon existing models by sharpening the vertical resolution of absolute velocity profiles, offering robust uncertainty estimates, and revealing mid-lithospheric velocity gradients indicative of thermochemical cratonic layering. This flexible method holds promise for increasingly detailed understanding of the upper mantle.

Key words: Composition and structure of the mantle; North America; Inverse theory; Joint inversion; Seismic tomography; Cratons.

1 INTRODUCTION

With access to greater computational power and parallel processing, the geophysical community is engaged in the development of new statistical inverse techniques that permit increasingly sophisticated imaging of the Earth's interior. Several groups (e.g. Malinverno & Briggs 2004; Khan *et al.* 2011; Bodin *et al.* 2012b; Afonso *et al.* 2013b; Bodin *et al.* 2013; Drilleau *et al.* 2013; Shen *et al.* 2013a; Hawkins & Sambridge 2015; Bodin *et al.* 2016; Calò *et al.* 2016; Guo *et al.* 2016; Tork Qashqai *et al.* 2016; Shen & Ritzwoller 2016; Burdick & Lekic 2017) have developed Markov Chain Monte Carlo frameworks that ally Bayesian statistics with rapid forward calculations to generate probabilistic Earth models that explain seismic data. Increasingly, these inverse computations involve multiple data types with complementarily distinct sensitivities, such as the combination of surface wave phase velocities and body-wave-scattered phases. We offer a new contribution within this field, distinguished by a novel parametrization that uses adaptive splines, a robust method for refining body wave data, and the inclusion of *S*-wave-scattered phases that add sensitivity to mantle velocity gradients. We demonstrate that our technique is consistently able to capture velocity gradients within and at the base of the seismic lithosphere, while maintaining parsimony and smoothness of Earth models. We illustrate the method using stations that sample the North American continent, from the tectonically active U.S. north-west to the interior of the Wyoming and Superior cratons.

Accurate seismic imaging is paramount for our understanding of the processes that form, reshape and modify long-lived continental plates, as well as those governing the genesis and maturation of oceanic plates. Broad-band seismograph data from the continents improves each year, as equipment developments and wider instrumental coverage complement the longevity of 'backbone' installations. We are poised on the cusp of a sea-change in broad-band oceanic data, with ocean-bottom seismometer deployments rapidly escalating in quality and distribution. The proliferation of data, together with the increased computational resources becoming routinely available to geoscientists, demands novel methodologies to upgrade models of Earth structure.

Methodological progress is driven by scientific questions regarding the composition, thermal structure, strain history and distribution of melt in the upper mantle and crust. Seismic velocity is a highly useful proxy for interior physical properties (e.g. Dalton & Faul 2010; Jackson & Faul 2010; Priestley & McKenzie 2013; Abers *et al.* 2014; Hansen *et al.* 2015; Afonso *et al.* 2016) and

© The Author(s) 2018. Published by Oxford University Press on behalf of The Royal Astronomical Society.

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

large arrays are permitting increasingly sophisticated tomographic coverage of the plates (e.g. Obrebski *et al.* 2011; Yuan *et al.* 2011; French *et al.* 2013; Foster *et al.* 2014; Lin *et al.* 2014; Afonso *et al.* 2016; Shen & Ritzwoller 2016; Burdick & Lekic 2017). Most tomographic methods lack resolving power to discern sharp boundaries, but substantial gradients in velocity at the base of, or within, tectonic plates are well documented (e.g. Kind *et al.* 2012; Levander & Miller 2012; Hopper *et al.* 2014; Hansen *et al.* 2015; Hopper & Fischer 2015). Improved constraints on the magnitude, width and absolute velocities of these seismic discontinuities are essential for parsing the respective influences of temperature, composition, anisotropy and ponded melt (e.g. Fischer *et al.* 2010; Karato 2012; Rychert *et al.* 2012; Afonso *et al.* 2013a; Rader *et al.* 2015; Selway *et al.* 2015). Short-wavelength vertical changes in velocity likely cannot arise from gradual temperature gradients [notwithstanding possible contributions from thermally activated anelastic processes (Karato 2012; Olugboji *et al.* 2016)]; the correlation between strong, shallow velocity gradients and recent volcanism is indirect evidence for the importance of melt (e.g. Schmerr 2012, Hopper *et al.*, in preparation). Absolute velocities are strongly sensitive to thermal structure and the presence of melt, and help discriminate mid-lithospheric discontinuities (MLDs, which connote exogenous lithologies, isolated sills or deformation zones) from the lithosphere–asthenosphere boundary (LAB, generally considered to be the base of the rheologically coherent plate; e.g. Eaton *et al.* 2009; Fischer *et al.* 2010; Hansen *et al.* 2015).

To image the lithosphere-scale structure in fine details, we harness a combination of surface wave and body wave data (Fig. 1; *cf.* Özalaybey *et al.* 1997; Julia *et al.* 2000; Bodin *et al.* 2012b; Shen *et al.* 2013a). Surface wave phase velocities contain information about absolute values of V_S and V_P within the Earth, at the expense of poor vertical (and horizontal) resolution. High-frequency ambient noise data are sensitive to upper and mid crust, while earthquake surface waves constrain upper-mantle velocity structure. Body wave converted phases are generated at strong impedance contrasts within the Earth; *P–s* phases primarily help ascertain Moho depth and *S–p* converted phases provide robust, if lower resolution, information about velocity gradients at the base of the crust and within the mantle. However, these converted phases offer little information about absolute velocities. *S–p* data have an advantage over their *P–s* counterparts in their sensitivity to broader velocity gradients and in their avoidance of crustal reverberations that can occlude sub-Moho *P–s* conversions. We find the novel combination of *S–p* data and surface wave phase velocities to be particularly effective for constraining upper-mantle vertical velocity gradients.

Traditional tomography methods, primarily relying on least-squares inversions to derive a single optimal solution, offer efficiency but several limitations. Trade-offs between parameters like V_P/V_S ratio, crustal thickness and vertically varying velocity limit the resolution of linear inverse calculations (Menke 2017), and may bias our tectonophysical interpretations. Surface-wave-derived velocity maps are susceptible to artefacts due to the intrinsically coarse vertical resolution of the data. Regularization, required by the ill-posedness of mixed-determined problems, has a large effect on least-squares inversions. However, at best, choice of regularization rests upon poorly known estimates of data error; at worst, it is arbitrarily user-defined. Bayesian inversion algorithms address these shortcomings by generating an ensemble of likely models that fit the data, revealing how parameters covary and obviating the need for regularization (e.g. Mosegaard & Tarantola 1995; Mosegaard & Sambridge 2002; Malinverno & Briggs 2004; Bodin *et al.* 2012a). A particular advantage of Bayesian algorithms is their ability to straightforwardly incorporate multiple data types (*cf.* Afonso *et al.* 2013a,b, 2016; Guo *et al.* 2016; Tork Qashqai *et al.* 2016), such as the combination of surface and body wave data types we use. If the forward problem is computationally inexpensive, an arbitrarily large number of data sets can be brought to bear on the inversion.

Parametrization of the inverse problem dictates certain intrinsic constraints upon the solution. Coarse parametrizations offer robustness but lose the ability to resolve fine features. Fine parametrizations can capture short-wavelength structure, but may yield unstable solutions or computational impracticality. Many Bayesian algorithms to date parametrize $V(z)$ in terms of layered structure (e.g. using Voronoi cells; Bodin *et al.* 2013, 2016; Calò *et al.* 2016). The drawback to this approach is that it may exaggerate velocity discontinuities, including artefacts created by assumptions about the structure such as crustal thickness, a topic explored by Roy & Romanowicz (2017). We contend that adaptive, semi-smooth parametrizations are the optimal choice to capture the discontinuous smooth structure of the real Earth (e.g. Shen *et al.* 2013b; Hawkins & Sambridge 2015).

2 METHODS

2.1 Framework for a Bayesian inversion

The overall goal of this study is to implement an inverse algorithm that can capture detailed upper-mantle structure. We seek to recover models that are 'realistic' (as determined by conformity to prior expectations) and that resolve both coarse and fine-scaled structure through fitting multiple data types that have complementary sensitivity to distinct model features or depth ranges. Bayesian inverse methods offer these attributes, together with a posterior estimation of model uncertainty and trade-offs as well as the possibility for autonomous data weighting.

The framework for a Bayesian probabilistic approach to inverse problems has been laid out in detail in several recent papers (e.g. Mosegaard & Tarantola 1995; Malinverno 2002; Mosegaard & Sambridge 2002; Sambridge 2002; Sambridge *et al.* 2006; Bodin & Sambridge 2009; Bodin *et al.* 2012b, 2013; Shen *et al.* 2013b; Calò *et al.* 2016). Here, we briefly summarize the main points salient to our method.

We want to compute the (posterior) probability of a certain proposed model, given the observed data. From Bayes's theorem we have

$$\textbf{posterior} \propto \textbf{prior} \cdot \textbf{likelihood} \tag{1}$$

or

$$p(\mathbf{m}|\mathbf{d}_{\rm obs}) \propto p(\mathbf{m}) \cdot p(\mathbf{d}_{\rm obs}|\mathbf{m}), \tag{2}$$

where $p(\mathbf{m}|\mathbf{d}_{\rm obs})$ is the posterior probability of model ($\mathbf{m}$) given the observed data ($\mathbf{d}_{\rm obs}$). The prior probability ($p(\mathbf{m})$) is taken to be uniform within some bounds, and does not change between iterations. The likelihood function for data type i ($p(\mathbf{d}_i|\mathbf{m})$) is related to the misfit, Φ:

$$p(\mathbf{d}_i|\mathbf{m}) = \frac{1}{(\sqrt{2\pi}\,\sigma_i)^n} \times \exp\left\{-\frac{\Phi_i}{2\sigma_i^2}\right\}, \tag{3}$$

accounting for data error, σ_i, which we assume to be uncorrelated (i.e. the respective covariance matrices are $\sigma_i^2\,\mathbf{I}$). n is the degrees of freedom; for surface waves this is the number of frequencies measured, while for body waves this is estimated using the first zero-crossing of the data autocorrelation function (Silver & Chan 1991).

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

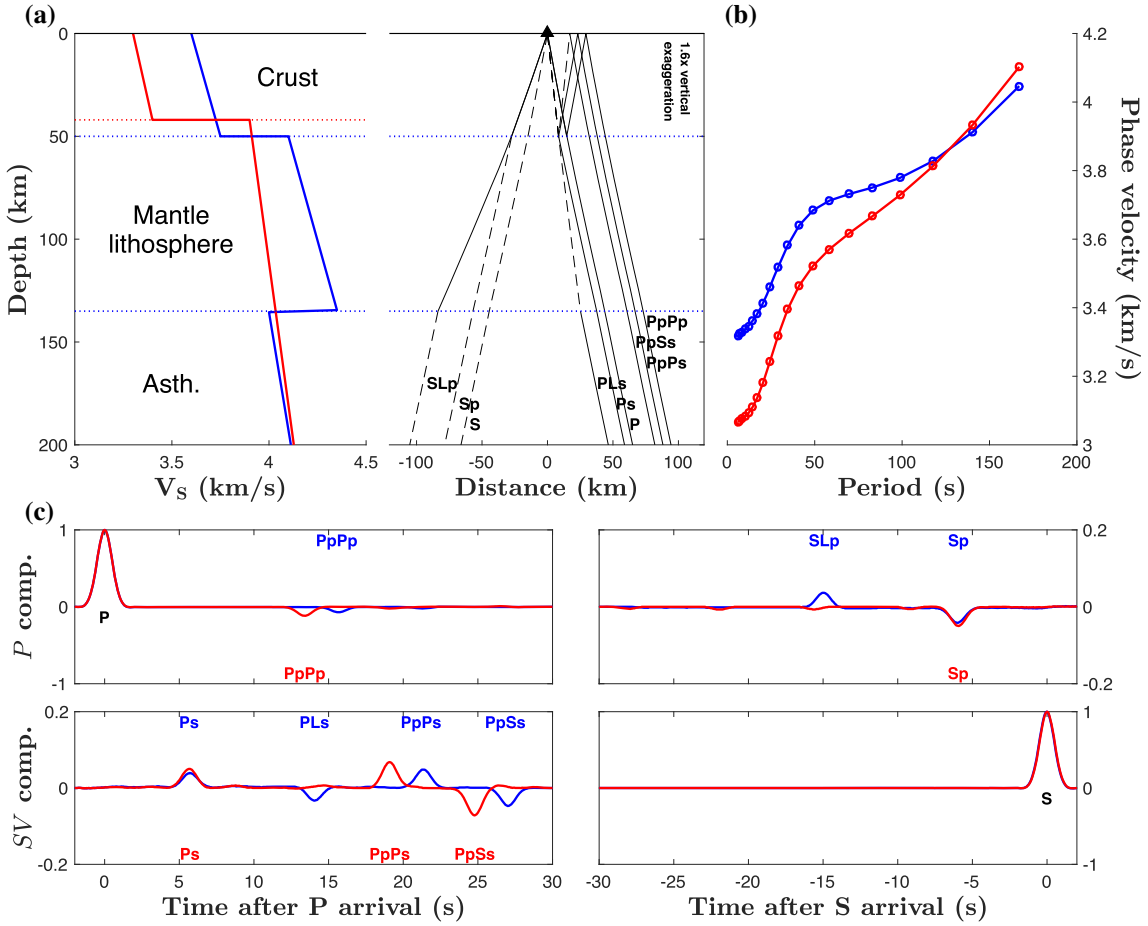

Figure 1. Summary of the data types, with two illustrative 1-D velocity models. (a) V_S profiles for the two models (blue model has $V_P/V_S = 1.68$, red model has $V_P/V_S = 1.8$) and ray paths for direct and converted phases passing through blue model. (b) Phase velocity dispersion curves for the two models. (c) *P–s* (left) and *S–p* (right) synthetic traces for the two 1-D models showing seismometer components rotated into *P* (top) and *SV* (bottom) coordinates, and with predicted arrivals indicated by phase labels.

The overall likelihood function for a given trial model is simply the product of misfits for each of the M data types:

$$p(\mathbf{d}_{\text{obs}}|\mathbf{m}) = \prod_{i=1}^{M} p(\mathbf{d}_i|\mathbf{m}). \tag{4}$$

Note that here we make the implicit assumption that data errors are uncorrelated across data types (*cf.* Khan *et al.* 2011; Drilleau *et al.* 2013; Shen *et al.* 2013b; Bodin *et al.* 2016; Calò *et al.* 2016; Roy & Romanowicz 2017). A more comprehensive characterization of seismic data error covariance is beyond the scope of this work. We calculate likelihood functions for body wave data (BW) and Rayleigh wave phase velocity dispersion curves (SW) from their respective misfit functions:

$$\Phi_{\text{BW}}(\mathbf{m}) = ||\mathbf{v}_p(t, \mathbf{m}) * \mathbf{H}(t) - \mathbf{h}_p(t, \mathbf{m}) * \mathbf{V}(t)||^2 \tag{5a}$$

$$\Phi_{\text{SW}}(\mathbf{m}) = ||\mathbf{C}(f) - \mathbf{c}_p(f, \mathbf{m})||^2, \tag{6}$$

where $\mathbf{H}$ and $\mathbf{V}$ denote the observed horizontal and vertical (respectively) components of body wave particle motion, $\mathbf{h}_p$ and $\mathbf{v}_p$ are the corresponding predicted data, $\mathbf{C}$ and $\mathbf{c}_p$ denote phase velocity observations and predictions, and the asterisk indicates convolution. The body wave misfit function makes use of the cross-convolution misfit (Menke & Levin 2003; Bodin *et al.* 2013) between vertical ($\mathbf{v}$) and horizontal ($\mathbf{h}$) seismograms, assuming that any converted energy due to seismic velocity gradients remains in the *P–SV* system. The surface wave misfit function is simply the difference in phase velocity, $\mathbf{c}$, at each frequency, f. Error distributions implicit in the forward calculations are assumed to be Gaussian.

2.1.1 *Monte Carlo Markov chains*

We employ a Monte Carlo Markov chain (MCMC) algorithm to explore the model space, seeking a family of adequately likely models that allows us to build a posterior likelihood distribution of model parameters (Fig. 2). This algorithm, which has been described in detail elsewhere (e.g. Hastings 1970; Mosegaard & Tarantola 1995; Malinverno 2002; Mosegaard & Sambridge 2002; Bodin *et al.* 2012a) entails the proposition of a random starting model, followed by sequential semi-random single-parameter perturbations in each iteration (k) that are accepted as new current models based on a Metropolis–Hastings acceptance criterion (Metropolis & Rosenbluth 1953; Hastings 1970). Supporting Information Table S1 defines the model parameters used, their prior bounds and their probabilities of perturbation in any iteration. All models are required to conform to *a priori* bounds, as well as to a set of conditions that enforce expectations about Earth structure, such as positive V_S increases at the Moho (Table 1).

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

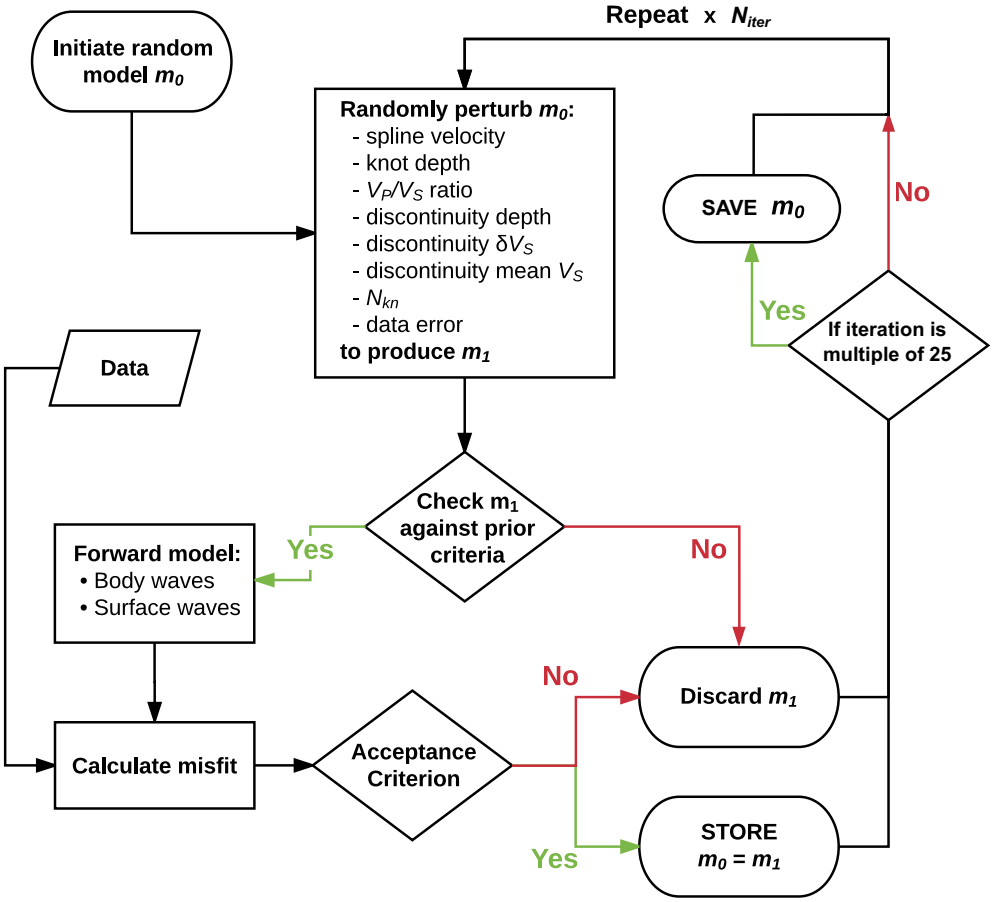

Figure 2. Algorithm for a single MCMC; several such chains are run in parallel.

Table 1. Model conditions.[a]

Conditions
V_S change at the Moho ≥ 0 km s^{-1}
V_S increase at the Moho ≤ 30 per cent
V_S within bounds (see the Supporting Information Table S1) in each layer
Adjacent spline knots no closer than 3 km

[a]Conditions to which any new trial model must conform.

The first phase of the Markov Chain (generally <2000 iterations) is characterized by steeply increasing model likelihood and decreasing misfit. Following this 'burn in' period, the model settles into a quasi-stationary state of sampling from the family of models that have acceptably high likelihood. It is from this region that the posterior sampling occurs. So as to avoid recording codependent models, the current model is only saved every 25 iterations. The final posterior distribution is formed from a random subset of 2000 saved models among the post-'burn in' phase.

During the 'burn in' period, the model parameter sampling is susceptible to becoming confined within local minima. In order to prevent this, and to maximize the efficiency of the inversion, we allow larger changes between sequential models, which we implement with a decaying thermal parameter, τ [following the simulated annealing approach (Kirkpatrick *et al.* 1983)]. τ decays over a 'cool down' period of 1500 iterations: $\tau = 1 + 3\,\mathrm{erfc}\,(k/500)$. τ pre-multiplies the standard deviation of the Gaussian distribution from which we draw random model parameter perturbations (so attempted perturbations are generally larger if τ is large) and also pre-multiplies the likelihood of the trial model, so misfit increases are more readily accepted early in the MCMC.

Because of the possibility for any single Markov chain to become irrevocably stuck in a local minimum, we run several individual chains in parallel for each station at which we perform the inversion. Since the starting model in each chain is random, there is the possibility of beginning with, or arriving at, a model that produces inhomogeneous *S*-waves and *P*-waves when predicted *P–s* and *S–p* phases are calculated. If 20 iterations in a row produce imaginary seismic data (the hallmark of inhomogeneous waves in our code) that chain is abandoned and re-started from scratch. At the conclusion of all chains, we discard chains that appear to stick in local misfit minima, identified as those with a mean misfit in any data type that exceeds the average final misfit across all chains by a factor of ≥ 1.3, as well as any chains whose current model does not change for >500 iterations in a row. Posterior distributions from all accepted chains are summed and normalized, to create a final posterior. In addition to plotting the posterior models, we select a representative final model (and its uncertainty bounds) from the median (and 68th

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

or 95th percentile) velocity values at each depth, having discretized models onto a fine depth grid.

2.1.2 Hierarchical bayes

Following Malinverno & Briggs (2004), we allow the inversion to determine the standard errors of the data (σ_i) by including these so-called 'hyperparameters' (higher-order model parameters that control the likelihood) as free parameters. This removes user agency in terms of weighting data misfit; the inversion itself only fits each data type as closely as it determines is statistically appropriate. This approach recognizes the uncertainty in our estimates of data error (e.g. for a multistation phase velocity inversion).

We assume that errors are normally distributed and, for the sake of computational expense, that phase velocities have the same error at each period (*cf.* Bodin *et al.* 2016; Calò *et al.* 2016). In any iteration, random perturbations from a log-normal distribution may be made to the data errors. In this case, the misfit will not change (the predicted data is the same) but the likelihood will increase if

$$\frac{\delta\sigma}{\sigma_0}\left(\frac{\Phi}{\sigma_0^2}-n\right)>0, \tag{7}$$

where σ_0 is the current value, and $\delta\sigma$ is the perturbation, of the data error. It follows that the inversion will drive the estimated data error towards $\sqrt{\Phi/n}$, which is the empirical root-mean-squared data error (see the Appendix).

2.2 Parametrization

2.2.1 Piecewise continuous splines

The velocity model is defined using piecewise continuous splines (Fig. 3). This choice is motivated by our goal of capturing relevant, realistic Earth structure while achieving sufficient parsimony to allow exhaustive Monte Carlo sampling. Drawing on previous work (Drilleau *et al.* 2013; Shen & Ritzwoller 2016), we parametrize 1-D V_S structure beneath each station with three layers (sediments, crystalline crust and mantle), each with vertically varying velocity. We solve for the thickness of the sedimentary and crustal layers as independent parameters.

In the sediment layer, two values (top and bottom) define the shear velocities, with a linear increase over the layer width. Within the crust, M_c cubic B-splines define the shear velocity structure, which is constrained to be monotonically increasing with depth. This assumption is easily relaxed in our algorithm, allowing for exploration of more complex crustal structure in future applications. Within the mantle, shear velocity is controlled by M_m cubic B-splines which span the layer to a maximum depth of z_{max}. The number of knots ($N_{kn} = M_c + M_m + 2$) that define the spline basis is varied during the inversion, as is their position (within the layer boundaries). The prior distribution for the number of knots in each layer is the Jeffrey's distribution.

Each of these model parameters (Supporting Information Table S1) is perturbed independently, except for velocities defining the discontinuities between lithologic layers (sediment to basement or crust to mantle), where we apply perturbations either to the average shear velocity at the discontinuity or to the magnitude of the shear velocity increase across the lithologic boundary. Tests demonstrated that this approach yielded markedly more efficient fitting of body wave converted phases than separately varying the two velocities that bound a given discontinuity. This is because single perturbations to V_S on only one side of a discontinuity often increase the misfit, creating an impediment to altering discontinuity properties. The depth of these sediment/crust and crust/mantle discontinuities is also permitted to change, in which case the basis splines and velocity coefficients are re-computed (by fitting to an interpolated or extrapolated V_S profile, depending on the layer shrinking or growing). If a discontinuity perturbation is large enough to move the discontinuity 'through' an intralayer knot, then the two knots swap places (i.e. the discontinuity moves to the depth of the intermediate knot, and the knot moves to the proposed depth of the discontinuity). This tends to slightly bias the algorithm against accepting models at the extremes of the prior distribution of discontinuity depth, but not to the extent that the posteriors are substantially skewed (Supporting Information Fig. S9).

2.2.2 Adding, removing or moving a spline

A primary goal of our approach is to capture steep velocity gradients within, or at the base of, the lithosphere (MLDs and LABs). These gradients require a greater density of spline knots in the regions of large second derivatives in $V_S(z)$. We also anticipate regions with gentle velocity gradients; these regions can be well defined by sparse splines. We use an adaptive parametrization, in which spline knots in the crust and mantle can move, be inserted, or disappear (with some limits, see Supporting Information Table S1).

Spline knots are moved by perturbing their depth by a random value drawn from a zero-centred Gaussian distribution. In this case, the velocity coefficients associated with each spline do not change. If a knot moves past the position (in depth) of another knot, then the velocity coefficients do not re-order (so the splines effectively swap order). Knots are not allowed to move beyond the bounds of their lithological layer (sediment, crust or mantle).

If a knot is removed, all splines are re-calculated, and their velocity coefficients are re-computed by optimally fitting the V_S profile with the new spline basis. In this case a parsimony term of $N_{kn}/(N_{kn}-1)$ pre-multiplies the final likelihood function. Thus, models with fewer splines tend to be accepted unless the coarser basis leads to significantly diminished model fit.

To add a knot, we select a random depth with uniform probability between the layer bounds. A new set of splines is calculated, including the new knot, and velocity coefficients are computed by fitting the existing V_S profile. Since this interpolation will provide an exact fit to the existing $V_S(z)$, we also perturb the velocity of the spline(s) closest to the new knot by some random δV drawn from a zero-centred Gaussian distribution. If a single spline accounts for >40 per cent of the weighting value at this depth, then only this spline's coefficient is modified. If two splines account for >40 per cent of the weighting, then they are each modified by $+\delta V/2$ and $-\delta V/2$, respectively. In this way, the addition of a knot tends to sharpen the velocity gradient near that depth. For knot addition, a parsimony term of $N_{kn}/(N_{kn}+1)$ pre-multiplies the final likelihood function; unless the more dense model fits the data better, it is unlikely to be accepted.

The re-computation of splines with each birth and death step means that it is not possible to write analytical expressions for the proposal (the probability of a certain change to the model) associated with these two steps. As a result, we define our prior empirically as the distribution sampled by our algorithm when the likelihood ratio is set to unity (Supporting Information S1; Mosegaard & Tarantola 1995; Agostinetti & Malinverno 2010). Evaluation of

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

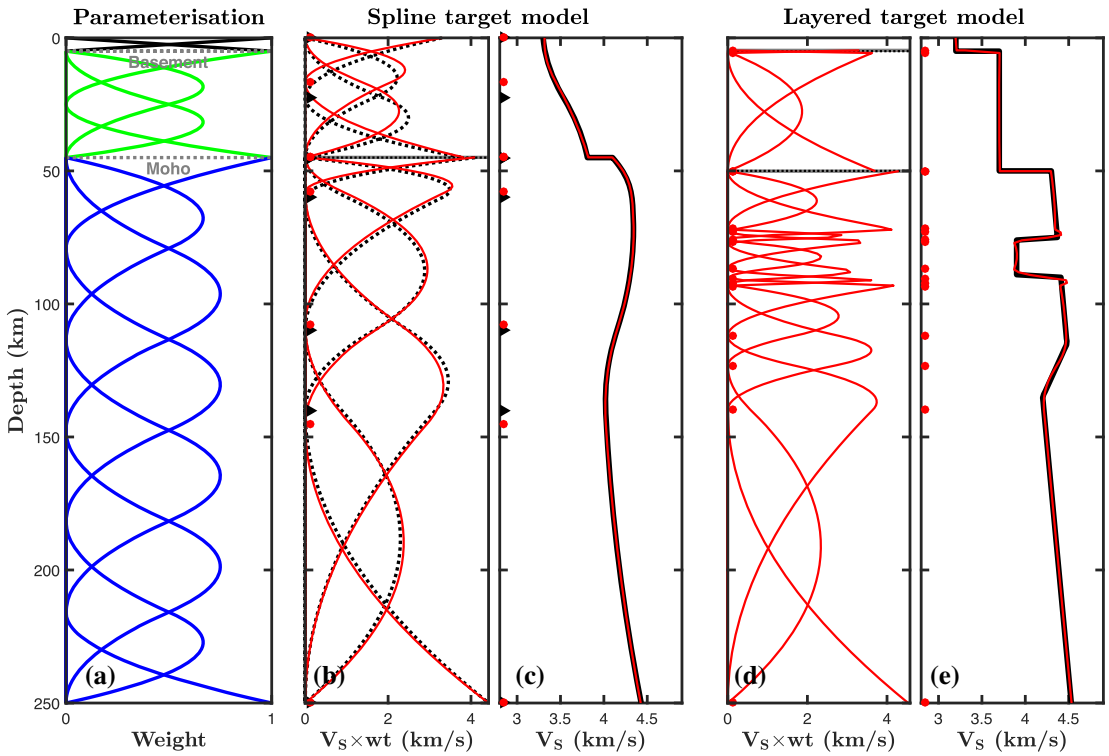

Figure 3. Adaptive discontinuous spline parametrization. (a) Example spline model, with discontinuities shown. (b,c) Results of fitting to a spline target model, with weighted splines shown at left and $V_S(z)$ at right. (d,e) Results of fitting a layered target model. Red (black) curves represent final (target) model. Red circles (black triangles) represent final (target) spline knot locations. Note: Unlike the data inversion, we use for the real modelling, (b,c) and (d,e) model fits were generated using an error term that penalized the difference between current and target $V_S(z)$ and $V_P(z)$; no data were calculated for these models. The final models shown here are not a pdf of the posterior, but the last model in each chain (which was almost invariant after the initial burn in).

this empirical prior (Supporting Information Fig. S9) shows that the algorithm does not bias the number of knots in each layer, which conform to a Jeffrey's distribution. However, the pdf of locations of knots in the regions adjacent to the boundaries is affected by our algorithm. So too, to a lesser extent, are the Moho depths and mantle velocity distributions, which somewhat undersample the extremes of their theoretical priors. The former arises from the fact that knots swap positions rather than move into different layers. The latter results form the way that birth and death steps are executed. Importantly, the algorithm efficiently samples the full model parameter space, introducing minimal bias. The empirical capture of our prior permits us to confidently identify robust features in the posterior distributions obtained once data constraints are added.

2.2.3 Conversion to layers

Although we have chosen to base our model parametrization on discontinuous splines, our forward model codes require inputs of discrete layered models. We convert our piecewise continuous model into a layered model according to the following algorithm:

(1) Lithological boundaries must be layer boundaries

(2) Regions with $V_S(z)$ gradients exceeding 0.5 s^{-1} are selected as layer boundaries, where the boundary depth is in the centre of the gradient.

(3) Regions with $V_S(z)$ gradients less than 0.0001 s^{-1} are selected as constant velocity layers.

(4) Regions with intermediate gradients are discretized into the minimum number of constant-thickness layers required to ensure no interlayer velocity jumps greater than 0.05 km s^{-1}.

This typically results in discrete models with ∼50 ± 5 layers (compared to typically ∼10–20 splines). These models are created solely for the forward problem, and are not stored.

2.3 The forward problem

2.3.1 Density and V_P scaling

The foregoing discussion defined the $V_S(z)$ model, which is converted to V_P and ρ using scaling factors. V_P/V_S and ρ/V_S in the sediments are defined by empirical observations of crustal rocks, as is ρ/V_S in the crust (Brocher 2005). ρ/V_S in the mantle is computed using a linear empirical scaling for lherzolite computed using the mineral properties calculator of Abers & Hacker (2016), as a function of depth. The mantle V_P/V_S ratio at each depth is fixed to its value in the *ak135* 1-D mantle model (Kennett *et al.* 1995). The V_P/V_S value in the crust is a free parameter in the inversion.

2.3.2 Body waves

To model the body wave data, we calculate synthetic Z-, R- and T-component seismograms using a propagator matrix approach (Keith & Crampin 1977a,b,c), optimized for parallel computing. At each layer, boundary (see Section 2.2.3) reflection and transmission coefficients are calculated, allowing for multiples and *P–SV* conversions. The ray parameter used is the data average. The source is a 1 s Gaussian pulse but, since we use the cross-convolution misfit (eq. 5a), both synthetic and actual source time functions can be moved to the front of the misfit term. These terms, therefore, end up scaling

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

the observed-minus-predicted impulse response, but if the model structure matches the true structure then the misfit will go to zero even if the source time functions differ (Menke & Levin 2003). This calculation is (on average) the most computationally demanding aspect of the forward problem, with each forward simulation taking ∼0.1 s per body wave arrival.

2.3.3 Surface waves

Surface wave phase velocities are calculated using the MINEOS code from CIG (Computational Infrastructure for Geodynamics) (Masters *et al.* 2007). This code requires a whole-Earth 1-D model, whereas our model extends only to 300 km depth. We construct whole-Earth models by using PREM (Dziewonski & Anderson 1981) at all depths >400 km and our model for depths 0–300 km. In the depth range 300–400 km, our model values of V_P, V_S and ρ linearly grade into PREM values. To ensure a smooth transition between our model and PREM, we do not permit any knots (other than the basal knot) deeper than 250 km depth, such that the bottom 50 km of our models cannot contain sharp velocity gradients. Q_κ and Q_μ values are taken from PREM, and in all the models presented in this manuscript we ignore radial anisotropy. Our data are minimally sensitive to the whole-Earth model used; even at the longest periods (180 s), less than 9 per cent of the Rayleigh wave V_S kernels extend deeper than 300 km, and periods ≤100 s have essentially no sensitivity beyond this depth.

Since MINEOS is a time-consuming calculation (typically >30 s per run), it is computationally impossible to re-run in every iteration. Instead, we use MINEOS to compute velocity perturbation kernels, $K_i(r)$, for each of the periods in our dataset for five parameters $M_i = V_{SV}$, V_{SH}, V_{PV}, V_{PH}, ρ. Phase velocity changes are then calculated by integrating the kernels over Earth radius:

$$\frac{\delta c}{c} = \int_0^{R_\oplus} \left[\sum_i \cdot K_i(r) \cdot \frac{\delta M_i}{M_i}(r) \right] \mathrm{d}r \tag{8}$$

at each frequency.

We tested the adequacy of the perturbation kernel approach by computing the error between exact and kernel-derived phase velocities as a function of the norm of the shear velocity perturbations. The further the current model deviates from the model used to construct the perturbation kernels, the less precise we expect the kernel-derived phase velocities to be.

For three periods (20, 62 and 176 s) we ran 10 chains of models, initially generating random models from which we calculated kernels at the outset of each chain. The starting models were randomly perturbed along each chain, and we computed exact and kernel-derived phase velocities at each iteration in order to determine the error in the kernel-derived phase velocities. For this test, we accepted all models and raised the thermal parameter to three, so this approach overestimated the error buildup in the true inversion.

For model perturbations with $\Delta = \|\delta V_{SV}/V_{SV}\|_2 + \|\delta V_{SH}/V_{SH}\|_2$ less than 0.5 per cent, the kernels yielded phase velocities that were insignificantly (<1 per cent or <0.04 km s^{-1}) different from the precise values from MINEOS at all frequencies (Fig. 4). At 20 s period, errors above the data uncertainty of ∼0.05 km s^{-1} (≈1.5 per cent) began to accrue in 8 per cent of cases that had Δ between 0.5 and 1 per cent, and in 27 per cent of chains for norms between 1 and 1.5 per cent. These errors are even smaller at longer periods.

Based on these tests, during our inversions we compute the perturbation norm at each iteration, and augment a counter by one each iteration for which $\Delta \leq 0.5$ per cent, by five for each iteration that $0.5 < \Delta \leq 1.0$ per cent, and by 10 for each iteration that $0.5 < \Delta \leq 1.5$ per cent. The full MINEOS calculation and new perturbation kernels are computed from the current model when the counter exceeds 200, or if Δ exceeds 1.5 per cent.

2.4 Data processing

2.4.1 Body wave stacking procedure

We seek to construct robust stacks of body wave arrivals that include *S–p* and *P–s* phases at individual stations. The computational expense of the forward problem requires that we compare synthetics to just a few high-quality observations. The automated process of producing these data stacks is illustrated for station ECSD (Fig. 5).

First, we use a cluster analysis algorithm to define groups of earthquakes from the recording lifetime of the stations, iteratively selecting clusters of events within windows of 30° in backazimuth and 20° in distance. For each cluster with ≥25 events, we rotate components to the $P-SV$ system using a free-surface transform with $V_S = 3.1$ km s^{-1} and $V_P = 5.58$ km s^{-1} (see Section 4.1), and apply a 100 s high-pass causal Butterworth filter before normalizing polarities and amplitudes using the parent seismogram channel. The filter is applied forwards in time for *S* arrivals, and backwards in time for *P* arrivals; this approach minimized the influence of sidelobes from the main phases impinging on the converted phases. Arrival polarities are flipped to the same sign using a three-stage algorithm that applies matched maxima, cross-correlation and sign-change criteria to estimate the polarity of the main arrival. If the three criteria yield no clear polarity, the trace is discarded. We align traces using cross-correlation (VanDecar & Crosson 1990; having narrow-band filtered the traces to 0.02–2 Hz for *P* and 0.05–1 Hz for *S*, both with two-pole acausal Butterworth filters), discarding any with a signal-to-noise less than three or a final cross-correlation coefficient less than 0.75. Prior to stacking, each individual daughter trace is migrated to a common ray parameter (taken to be the average ray parameter of the remaining arrivals) using a local velocity model (Section 4.1).

We sum traces to form a preliminary stack, before cross-correlating the individual traces against the preliminary stack, now discarding any with a cross-correlation coefficient less than 0.9. The re-aligned traces are normalized and summed to form the final stack. This processing procedure is somewhat laborious, but results in very high-quality stacked data (Fig. 5).

By stacking over multiple events, variations in moment rate function and radiation pattern deconstructively interfere, leaving us with seismograms that approximate the shallow scattering Green's function (convolved with some averaged source term) for that backazimuth and ray parameter. Given our assumption of isotropy, we can compare the same predicted data with observed data from each of the backazimuthal clusters within a given distance window. The predicted data is calculated using the mean ray parameter of all traces remaining in the final stacks for each distance window. We also conduct tests using synthetics that show how this stacking procedure handles the effect of varying source time function and ray parameter (Section 3.2).

Body waves offer sensitivity to strong velocity gradients in the shallow Earth due to the conversion of energy from the main phase to secondary phases with distinct propagation velocity (V_P versus V_S). In order to isolate these converted phases arriving after (before) the primary *P*(*S*)-wave, both predicted and observed data series are

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

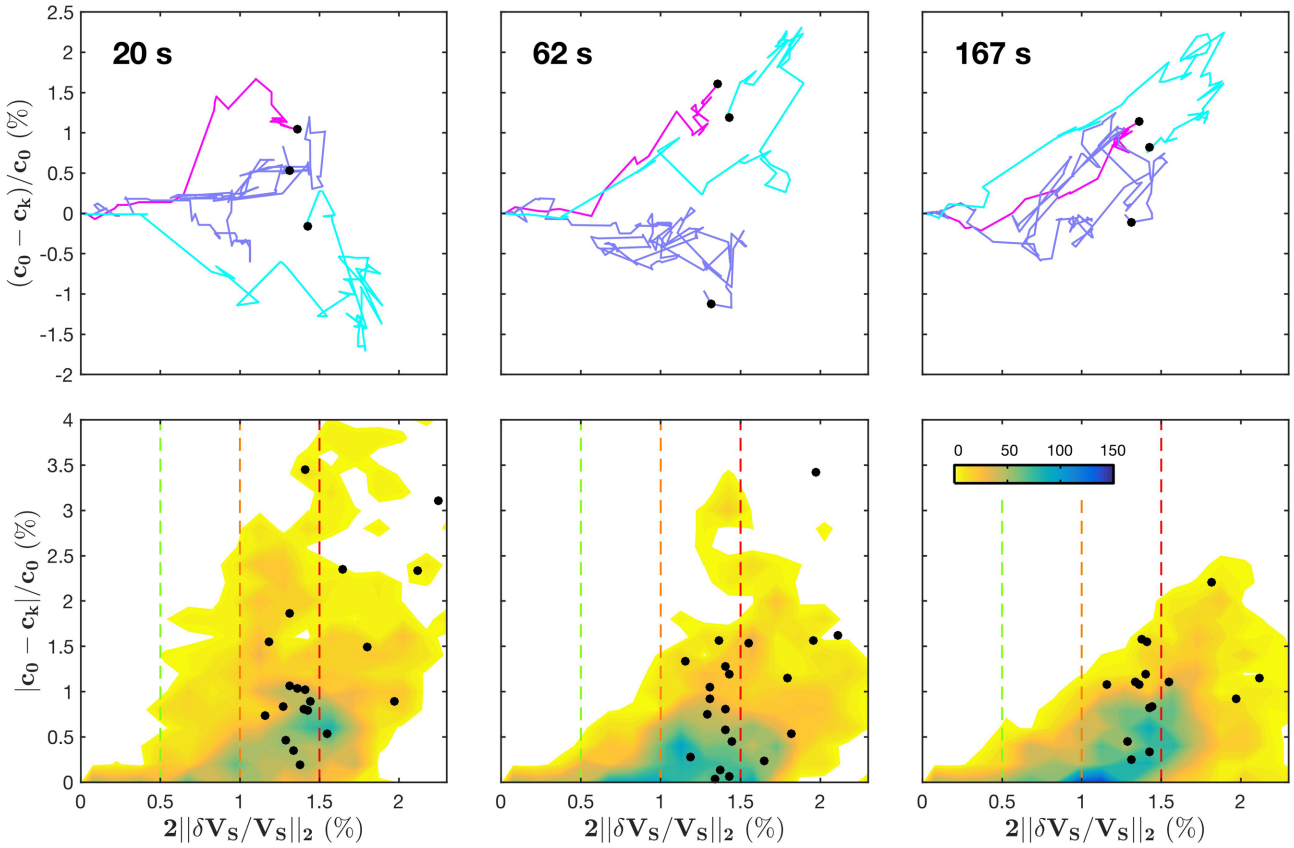

Figure 4. Tests of perturbation kernels, showing error in phase velocity as a function of fractional perturbations from the starting model. Error is measured as the fractional difference in phase velocity between exact (c_0) and kernel-derived (c_k) values. Model perturbation is computed as the norm of the shear velocity fractional perturbations, $2||\delta \mathbf{V}_S/\mathbf{V}_S||_2$. Columns show periods of 20, 62 and 167 s. Top row shows random walks of three chains of perturbations, each with a different random starting model. Bottom row shows a density surface of absolute error versus perturbation for all models visited along 24 random chains. Black dots represent the final states, when a new set of kernels would be computed. Green, orange, and red lines represent the thresholds for increasing a counter towards re-calculating kernels (Section 2.3.3).

windowed from 2(30) s before the main arrival to 25(2) s after the main arrival, with a 1 s taper.

If there were no velocity discontinuities in the Earth, we would expect this stacking process to produce a single Gaussian spike on the parent component (*P* for compressional waves and *SV* for shear waves) and no energy on the daughter component (*SV* for compressional waves and *P* for shear waves). The Moho, a positive velocity gradient (PVG), should produce a positive (negative) amplitude pulse of converted energy on the daughter component following (ahead of) the main arrival for *P*-(*S*-) waves. Negative velocity gradients (NVGs) deeper in the mantle, such as the LAB, will appear as a negative (positive) pulse of converted energy on the daughter component even further behind (ahead of) the main *P*-(*S*-) wave arrival (Fig. 1c).

2.4.2 Surface wave dataset

Within the continental US, we use the Rayleigh wave phase velocities of Bao *et al.* (2016), computed using an automated technique based on Helmholtz tomography (Jin & Gaherty 2015), at nine periods from 25 to 180 s. Across the region of the Wyoming craton, we use more detailed surface wave phase velocities maps from Dave & Li (2016), who employ a least-squares two-plane wave method to calculate phase velocities at 18 periods from 20 to 167 s. The dispersion curve we use is obtained by 2-D interpolation of the phase velocity map at the location of each station. In order to improve our sensitivity to shallow structure, we add ambient noise phase velocity results between 8 and 32 s from the model of Shen & Ritzwoller (2016). At the locations of our stations, these ambient noise data differ by less than 0.4 per cent (averaged over 8 to 32 s periods) from an independently collected dataset (Schmandt *et al.* 2015), bolstering our confidence in these data. We linearly average the phase velocities from the ambient noise and earthquake dispersion curves in their overlapping period range (20–32 s).

2.5 Misfit and likelihood

2.5.1 Body wave data

Mancinelli *et al.* (2017) have recently emphasized the importance of frequency dependency for recovery of scattered structure (*cf.* van der Meijde *et al.* 2003; Lekic & Fischer 2017). Based on these results, we evaluate body wave data misfit in two frequency ranges: 0.02–0.5 and 0.02–0.33 Hz for *P*-waves and 0.02–0.5 and 0.02–0.2 Hz for *S*-waves. We filter both observed and synthetic data within these ranges, and compute cross-convolution misfit for each filter band separately. This is mathematically equivalent to up-weighting the fit

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

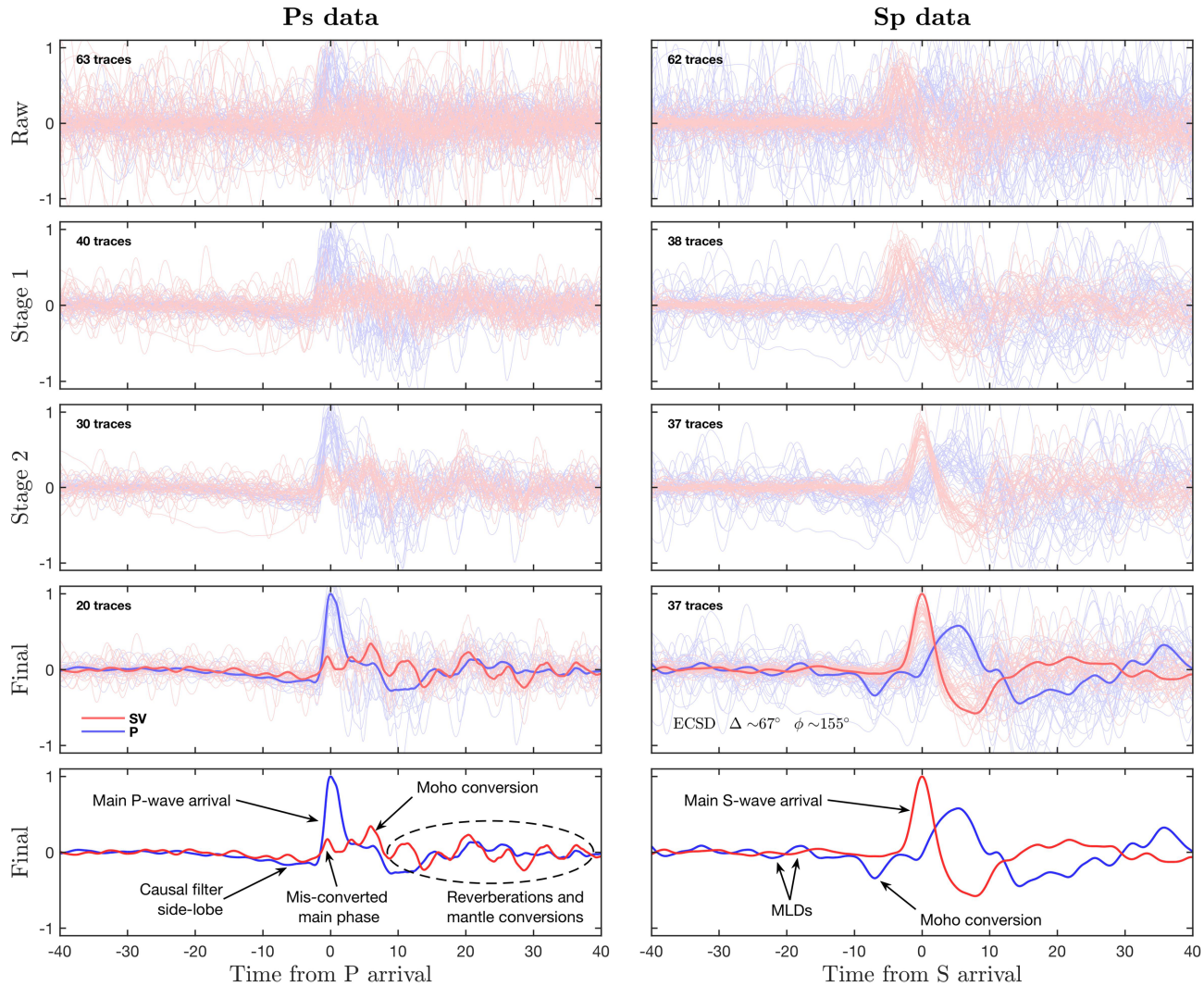

Figure 5. Example of body wave data processing at station ECSD for a cluster of 62 earthquakes shallower than 100 km depth, between 59.9° and 74.7° distance and 149.5° and 157.5° azimuth. 'Raw' data has been rotated to the *P–SV* coordinate frame with a free surface transform matrix. A 100 s high-pass filter has been applied. Stage 1: arrivals with parent component SNR less than three have been discarded, and polarities have been corrected. Stage 2: waveforms aligned by cross-correlation of parent component; low cross-correlation coefficient (<0.75) traces discarded; traces migrated to common ray parameter. Final stage: waveforms aligned by cross-correlation with the stack of Stage 2 data; low cross-correlation coefficient (<0.90) traces discarded. Light lines show all traces remaining at each stage, where red is the *SV* component and blue is the *P* component. Bold lines show final stacks formed by taking the mean of all 'Final stage' traces, after scaling each *P*, *SV* trace pair by a value that normalizes maximum *SV* amplitude to 1.0. In this plot, all daughter traces are multiplied by a factor of 2.5 so that they can be more clearly seen, but this factor is obviously not applied when the 'Final' stacks are employed as data in the inversion.

to longer periods. The total body wave likelihood is given by

$$\ln p(\mathbf{d}_{\rm BW}|\mathbf{m}) = A - \sum_{f}^{\rm low,high} \frac{1}{\kappa} \sum_{j=1}^{\kappa} \left(n_j \ln \sigma_{fj} + \frac{\Phi_{fj}}{2\sigma_{fj}^2} \right), \qquad (9)$$

where we have taken the logarithm to avoid computation errors that arise as p becomes very large. A is a constant [equal to $-(n/2)\ln(2\pi)$] that results from the normalization of the likelihood term and that does not change between models; κ is the number of earthquake clusters (treating *P*- and *S*-wave clusters separately); f is the frequency band. Φ is the cross-convolution misfit (eq. 5a) between observed and predicted data. The competing effect of data uncertainty on the likelihood is evident from the term inside the brackets: for given misfit, higher data noise reduces the likelihood through the term on the left, while it increases likelihood by penalizing misfit less through the term on the right (see the Appendix).

2.5.2 *Surface waves*

For phase velocity measurements, we weight the misfit at each frequency by the fraction of that frequency's V_S sensitivity kernel that is within the model (i.e. at depths shallower than the maximum model depth, $z_{\rm max}$). These misfit weights are normalized to an average value of unity (e.g. the 8 s misfit is up-weighted to 103 per cent while the 180 s misfit is down-weighted to 78 per cent).

3 SYNTHETIC TESTS

3.1 Adaptive splines

Before applying the inversion technique to real or synthetic data, we tested the ability of the adaptive parametrization to recover a target velocity profile. For these tests—unlike the actual inversion—no data were calculated or compared. Rather, the misfit term was taken

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

to be the norm of the difference between the target and the current velocity profiles ($\Phi = \|V_S(z)_i - V_S(z)_{true}\|_2 + \|V_P(z)_i - V_P(z)_{true}\|_2$). This exercise evaluated the optimal capability of the inversion prior to the computation of data (which opens up null model parameter spaces due to non-uniqueness).

Two target models were tested (Fig. 3). The first (A) was a model created using splines (Fig. 3b and c). The second (B) was composed of linear velocity gradients, including abrupt discontinuities in the mantle (Fig. 3d and e). In both cases, the MCMC was initiated with a random starting model, and iterated 10 000 times.

In both cases, our inversion was able to fit the target models extremely well. Final model A almost perfectly matched the location and weight of splines used to construct target model A, despite the chain having started with many more splines and with random knot locations. For model B, the fit was excellent, despite splines being highly non-ideal for capturing this sort of target structure. In both models, the splines were dense in regions with large second derivatives in the velocity profiles. Spline knots were sparse or absent in regions with constant velocity gradients. These tests demonstrated that our parametrization is capable of resolving both sharp and broad velocity gradients in the mantle, as well as the successful implementation of parsimony and transdimensionality.

3.2 Tests of body wave stacking procedure

The body wave data used in the inversions was produced by stacking numerous earthquake arrivals at the stations of interest so that small amplitude converted phases rise above the noise (Section 2.4.1). Differences in source magnitude, moment rate function, ray parameter and back azimuth (assuming 3-D structure) complicate this stacking procedure by introducing waveform heterogeneity and move-out interference. We tested the consequence of these factors by creating an artificial dataset of body wave arrivals from 60 random distances in the range of 30°–75°, as well as 30 random distances in the range of 60°–75° to provide more *S–p* phases at incidence angles that do not result in inhomogeneous converted phases. Synthetic data were forward modelled by propagating a 1 s Gaussian pulse through the velocity model, before convolving the output with a 20 s source time function that comprised three triangular wavelets with random height, width and relative timing. After normalization to unit power, Gaussian noise with $\sigma = 0.005$ was added. Finally, the waveforms were stacked to produce *P–s* and *S–p* data series (Fig. 6). The resultant stacked phases were markedly broader than the input Gaussian pulse, and had the form of a two-sided exponential function. The stacked phases were centred at the predicted times for an incident wave with the stack's mean ray parameter.

These stacked data were then inverted jointly with surface wave phase velocities using our inversion algorithm (using average ray parameters from each stack). The inversion did an excellent job of fitting the body wave data (Fig.6), and successfully recovered Moho depth, crustal V_P/V_S ratio and overall $V_S(z)$ structure.

We also performed the exercise of making the synthetic data stacks with and without migrating the data to a common ray parameter, in this case using the (known) input velocity model. In theory, the migration corrects for destructive interference between daughter-component phases that result from move out. Surprisingly, this step did not make a large difference to the quality of the stacks; the primary conversions became slightly sharper, but the crustal multiples were more deteriorated. We infer that if our sources are randomly distributed in distance, move out interference is largely accounted for by our stacking procedure. Nonetheless, because of the modest improvement in definition of primary conversions, our final data stack processing includes a migration step.

3.3 Sensitivities and resolution of different data types

Surface waves provide constraints on absolute velocities in the Earth, with limited vertical resolution. Teleseismic body wave arrivals encode information about strong velocity gradients in the sub-surface, but offer no constraints on absolute velocity structure. We conducted a series of synthetic inversions using each of these data types, individually and then in combination (Table 2, Fig. 7 and Supporting Information Figs S1–S4). These tests demonstrated the ability of our algorithm to recover a range of realistic Earth structure, including varying Moho depth (30 and 45 km) and LAB gradient (5 per cent $\Delta V_S/V_S$ over 0, 10 and 40 km depth). We refer to these different inversions with concatenations of the data involved, where 'SW' indicates Rayleigh wave phase velocities, 'Ps' indicates *P–s* body waves and 'Sp' indicates *S–p* body waves.

Synthetic data were calculated using same approach as applied for the forward modelling. Body wave data was computed for ray parameters of 7.385 and 6.148 s deg^{-1} for *P*-waves and 11.724 s deg^{-1} for *S*-waves, using a 2-s Gaussian source wavelet. *S*-wave synthetic time-series were long enough to reflect the upper ∼225 km of model structure. *P*-wave synthetic time-series varied in length, as described below. For the majority of the tests, Gaussian noise was added with standard deviation roughly equal to 1/5 of the maximum amplitude on the daughter component. This yielded a signal-to-noise ratio (SNR) lower than that of the real data (by a factor of ∼2 for *S–p* arrivals and a factor of ∼3 for *P–s* arrivals). SNR was calculated as the ratio of the maximum converted arrival amplitude on the daughter component to 2× the RMS of the pre-arrival noise on that component. We conducted a final test, using all data types, where we reduced the input noise to match observed values. Phase velocities were computed for 24 surface wave periods log-spaced from 6 to 167 s, and Gaussian noise was added with standard deviation of 0.015 km s^{-1}.

Surface waves alone did a reasonable job of recovering the average velocity profiles throughout, but poorly constrained Moho depth, and were not at all sensitive to deeper mantle discontinuities (Fig. 7 and Supporting Information Fig. S1–S4). By contrast, a combination of *P–s* and *S–p* body wave data types did an excellent job of capturing the magnitude of sharp discontinuities, and a reasonable job of capturing their depth (although this traded-off with V_P/V_S ratio, not shown), at the expense of poor fits to absolute velocities. The lack of sensitivity to absolute velocities is shown by the multiple sub-parallel likelihood profiles in the Ps+Sp panels of Fig. 7; these represent results from different chains that arbitrarily selected an absolute velocity early on and remained in that local misfit minimum.

Individually combining *P–s* and *S–p* data with surface waves provided surprisingly different improvements in fit (Fig. 7). Both body wave data types improved the models' recovery of Moho depth, in terms of both misfit and posterior error bounds (see Table 2). However, we found that only *S–p* conversions robustly provide sensitivity to deep, broad velocity gradients.

Our initial *P–s* and surface wave tests (PsSW panels) utilized 25 s of *P*-wave signal following the primary arrival. These inversions did a decent job of recovering the Moho, but for the most part did a very poor job of recovering deeper discontinuities due to interference from the Moho multiples. These multiples masked conversions from deeper velocity gradients and were often mistaken

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

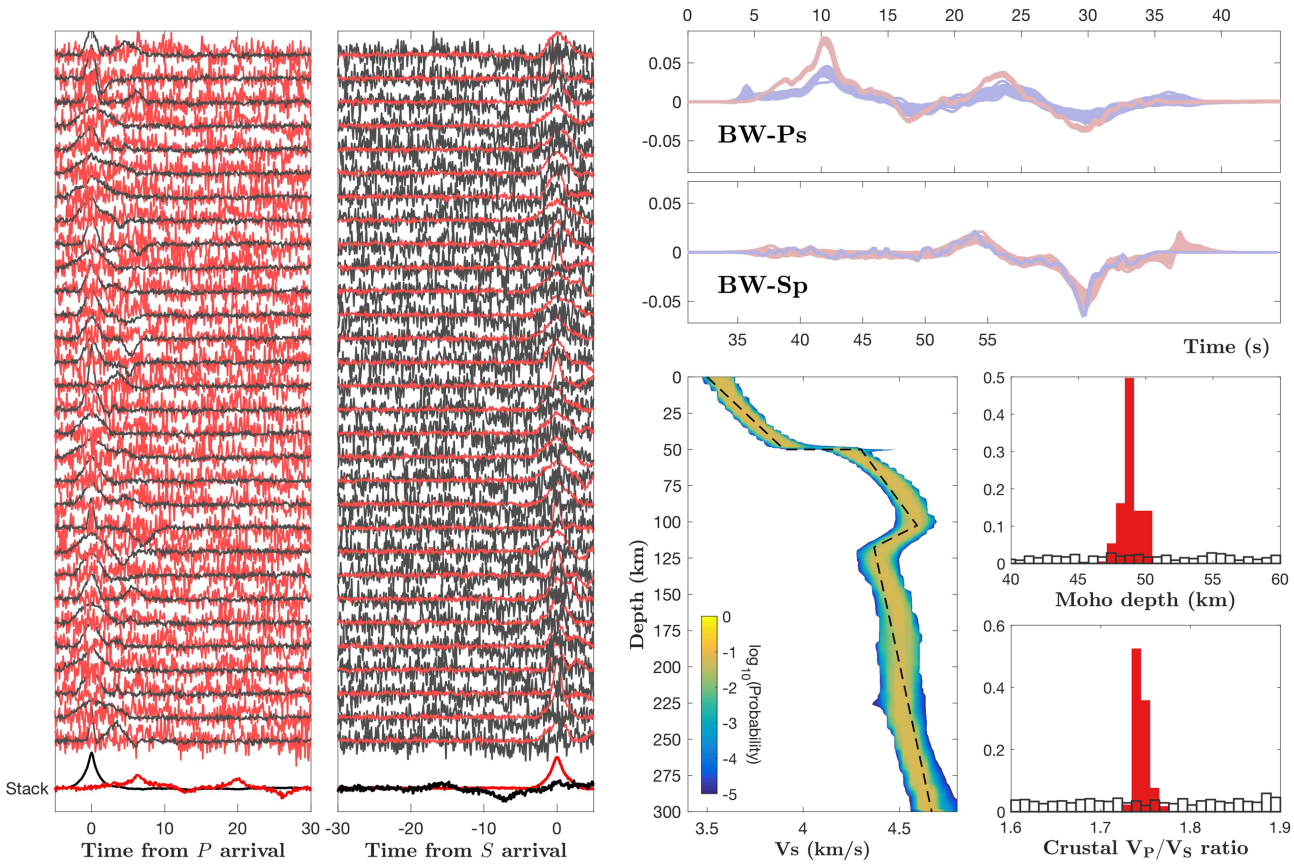

Figure 6. Example of synthetic test using stacked data. Left two panels show noisy synthetic body wave arrivals generated for random distances (30°–75°) and source–time functions (showing a subset of 300 total traces). Bottom traces are stacks of the random data. Black: *P*-component, red: *SV*-component, daughter components shown with 5× exaggeration. Top right: cross-convolution fits for 500 accepted models (red is $H_{obs}*v_{try}$, blue is $h_{try}*V_{obs}$) for *Ps* and *Sp* data types. Bottom right: histograms of prior (white) and posterior (red) from the suite of accepted models and $V_S(z)$ probability density function (true model shown as dashed black line).

Table 2. Moho and upper-mantle recovery for different data types (LAB width 10 km)∗.

Moho	Data	Z_{Moh} (km)	ΔV_S (per cent)	V_P/V_S	$V_S^{60\,km}$ (km s^{-1})
30	SW	40.2 ± 7.6	12.9 ± 10.4	1.74 ± 0.13	4.59 ± 0.13
	PsSp	30.3 ± 1.9	9.2 ± 8.5	1.79 ± 0.07	4.25 ± 0.19
	PsSW	29.7 ± 0.5	10.0 ± 8.5	1.80 ± 0.05	4.42 ± 0.07
	SpSW	30.9 ± 0.9	13.4 ± 9.9	1.78 ± 0.02	4.41 ± 0.04
	PsSpSW	30.0 ± 1.3	13.3 ± 4.1	1.80 ± 0.02	4.38 ± 0.03
45	SW	46.9 ± 18.3	8.7 ± 14.4	1.81 ± 0.07	4.39 ± 0.30
	PsSp	43.6 ± 1.5	7.5 ± 9.8	1.81 ± 0.03	4.33 ± 0.18
	PsSW	42.2 ± 7.0	4.1 ± 8.5	1.77 ± 0.13	4.37 ± 0.12
	SpSW	44.6 ± 1.7	8.5 ± 6.2	1.81 ± 0.02	4.38 ± 0.05
	PsSpSW	45.6 ± 1.4	9.7 ± 7.4	1.79 ± 0.02	4.36 ± 0.04

∗Columns are: (i) Moho depth (km); (ii) Data types used in the inversion; (iii) Moho depth recovered by inversion with 2σ error bounds from Gaussian fit to posterior histogram; (iv) Fractional velocity change across the Moho (true value is 10 per cent); (v) Crustal V_P/V_S value (true value is 1.8); (vi) V_S at 60 km depth (true value is 4.38 km s^{-1} for both models).

by the inversion algorithm as converted energy themselves, leading to models that include spurious velocity gradients at depth. Moreover, since some multiple energy was erroneously fit by spurious mantle velocity gradients, the ability of these inversions to correctly recover the Moho itself was deteriorated. On the basis of these tests, we were extremely cautious of incorporating late-arriving *P*-wave energy, and in all other tests presented (i.e. PsSp and PsSpSW) used only the first 10 s of *P*-wave synthetic time-series—long enough to constrain the upper ∼100 km of the model but not to contain crustal reverberations. The short *P*–*s* time-series do not constrain deeper mantle structure, but offer good Moho recovery.

The *S*–*p* data, which do not suffer from the problem of crustal multiples, markedly improved the recovery of mantle velocity gradients. The 10-km-wide LAB gradients were perfectly recovered, while the 40 km gradients were also well imaged, with only a small amount of smoothing at the changes in gradient. The 0-km-wide LAB gradient was smeared over 10 km depth, which is likely the maximum resolution for this sort of feature. The SpSW inversions also recovered both deeper (45 km depth) and shallower (30 km

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

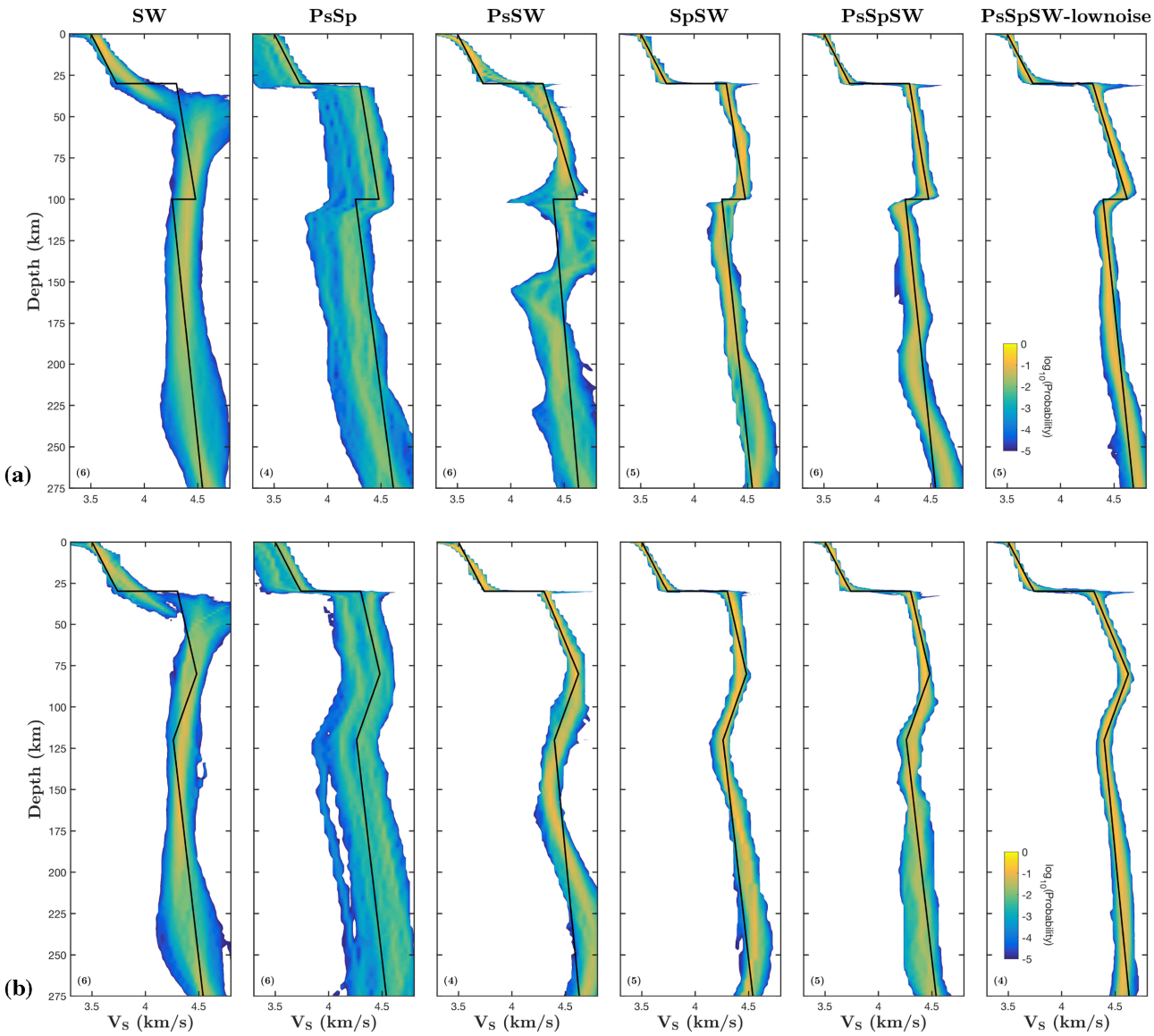

Figure 7. $V_S(z)$ probability density functions recovered using different combinations of data types, including *Ps* body waves, *Sp* body waves, and surface wave (*SW*) phase velocities. The input model (black line) has a 10 per cent $\Delta V_S/V_S$ Moho at 30 km depth and a 5 per cent 'LAB' negative velocity gradient at 100 km depth. We tested cases with an LAB width of (a) 0 km and (b) 40 km. Other test cases are shown in Supporting Information Figs S1–S4. The final panel shows the result using synthetics that had 'low' noise, comparable to the real data noise. The number of successful chains contributing to each posterior is shown in brackets at lower left.

depth) Moho discontinuities, with tight constraints both in depth and absolute velocity changes (Fig. 7 and Supporting Information Fig. S1–S4).

The combination of all three data types typically recovered the input structure successfully. However, adding *P–s* data only significantly improved retrieval of the target model in a subset of cases (Fig. 7). In other cases it had neutral impact (Supporting Information Figs S1–S4). The limiting factor for these tests appears to have been the level of noise in the data. Tests that used all three data types to fit low-noise synthetic data (with SNR matching our observed data, Section 4) did an excellent job of recovering target model structure, including the shape and depth of velocity discontinuities within the mantle. Comparison of low-noise and higher-noise PsSpSW tests is instructive. The former shows that the inversion is highly capable of recovering complex structure, given good enough data. The latter shows that moderate stochastic noise can systematically bias the recovered model to a small degree, but that our algorithm succeeds in exploring parameter space sufficiently for the true model to always lie within the posterior pdf.

These tests, therefore, help proof our posterior model errors. The $p = 0.05$ error bounds on the model parameters shown in Table 2 include the true values 95 per cent of the time (38 of the 40 estimated ranges), indicating the posterior uncertainty estimates are accurate.

For the model with a 30 km Moho and a 10-km-wide LAB gradient, we conducted an additional test with a 1.7-km-thick sedimentary layer above the crystalline crust. The recovery of sediment structure was good: estimated sediment thickness was 1.65 ± 0.10, and top and bottom velocities of 2.94 ± 0.09 and 3.09 ± 0.25 km s^{-1} were very close to the target values of 2.9 and 3.1 km s^{-1}, respectively (Supporting Information Fig. S14). This test demonstrates that the seismic properties of the sediments do not, in principle,

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

lie in the null space of the data. However, this shallow structure is controlled exclusively by short-period surface waves, which have noisy measurements. We do not permit a sediment layer in the other inversions included in this manuscript. Properly accounting for this layer may require inclusion of additional data types that provide shallow sensitivity, such as surface wave ellipticity (Li *et al.* 2016).

4 APPLICATIONS TO REAL DATA

In the interior of the continental US, numerous recent studies have leveraged dense datasets to reveal widespread but laterally discontinuous NVGs from 65 to 200 km depth and strong contrasts between the eastern and western US in terms of mantle discontinuity character (e.g. Levander & Miller 2012; Foster *et al.* 2013; Hopper *et al.* 2014; Lekic & Fischer 2014; Hansen *et al.* 2015; Hopper & Fischer 2015; Kind *et al.* 2015; Ford *et al.* 2016), crustal thickness and crust and mantle seismic velocities (e.g. Yuan *et al.* 2011; Lin *et al.* 2014; Porritt *et al.* 2014; Schmandt *et al.* 2015; Shen & Ritzwoller 2016; Burdick *et al.* 2017), gravity (*e.g.*, Schmandt *et al.* 2015), crustal structure (Lin *et al.* 2014), attenuation (Bao *et al.* 2016) and anisotropy (Lin & Schmandt 2014; Buehler & Shearer 2017). Our inversion technique is designed to accurately capture this lithospheric structure, particularly any sharp velocity gradients, in order to shed light on the processes that formed and continue to modify the continents. Thus, having proofed the method with synthetic data, we applied our inversion to data from long-lived stations that span the tectonic transition across the north west to north central continental US (Fig. 8): ECSD, EYMN, WVOR (all network US) and RSSD (network IU).

In the west, WVOR sits on the high lava plains (HLP), the site of bimodal volcanism since the mid-Miocene (Jordan *et al.* 2004), with comparably thin (<35 km) crust (Eagar *et al.* 2011; Levander & Miller 2012) and abnormally hot upper mantle, as inferred from low seismic velocities (Wagner *et al.* 2010) and resistivity (Meqbel *et al.* 2014). RSSD is on the re-worked Wyoming craton, an Archean province with high flexural strength (Lowry & Smith 1994), ≥50 km thick crust (Hansen & Dueker 2009; Ford *et al.* 2016) and thick seismic lithosphere (Humphreys *et al.* 2015; Ford *et al.* 2016) that sutured onto the Canadian shield at ∼1.8 Ga (Whitmeyer & Karlstrom 2007). Xenolith and seismological evidence point to pervasive modifications of the Wyoming craton during and since the Laramide Orogeny, with arguments for wholesale removal (and perhaps subsequent re-thickening) of the lower lithosphere (Carlson *et al.* 2004; Hearn 2004; Humphreys *et al.* 2015; Dave & Li 2016). ECSD and EYMN are situated within the Archean-age Superior craton, which has low heat flow (Humphreys 2009), 45–50 km thick crust (Ford *et al.* 2016; Shen & Ritzwoller 2016) and among the thickest seismic lithosphere observed worldwide (Yuan & Romanowicz 2010). Aside from a limited rifting event ∼1.1 Ga, this long-lived tectonic province has remained largely unaffected by regional tectonics since >2 Ga (Percival *et al.* 2006; Whitmeyer & Karlstrom 2007).

4.1 Complications using real data

The real data inversion presented several additional challenges compared to the synthetics. A major source of difficulty was the rotation of the data into *P*−*SV* components. It was not possible to obtain surface V_P and V_S values that did a perfect job removing the primary arrival on the daughter component on all traces (Section 2.4.1); the stacked data, therefore, have non-zero energy on the daughter at 0 s. This feature can introduce artefacts into the model (e.g. by causing the inversion to insert a low-velocity sedimentary layer that produces a large conversion almost co-temporally with the main pulse). To address this problem, we used the same surface V_P and V_S values to rotate the synthetic data from *Z*–*R* to *P*–*SV* as were used to rotate the actual data. This resulted in equally mis-rotated daughter components with energy at $t = 0$ s on the synthetic daughter, self-consistent with the real data. We also windowed the daughter trace to start 1 s after the maximum on the parent trace (using a 1.5 s Hanning taper). This effectively cut out most of the mis-rotated direct-phase energy on the daughter component, without affecting the converted energy from crustal and mantle velocity gradients.

3-D heterogeneity in the Earth further complicated the real data. Stacked data arriving from different backazimuthal bins revealed systematically different shallow structure as a function of direction from the station (Supporting Information Fig. S5). This may arise from dipping structures, anisotropy, or sharp lateral gradients in seismic properties near the station. While these phenomena are moderately problematic for stacking direct converted phases, the much larger move-out of crustal *P*-wave reverberations (*PpPp*, *PpPs*, and *PpSs*; Fig. 1) means that small lateral heterogeneities can cause these phases to destructively interfere in the data stack. The observed waveforms were often not well fit by forward-modelled synthetics, which suffer from no such move-out interference or backazimuthal variation.

To ameliorate this problem, we restricted the data in each stack to a backazimuthal window of ≤30° and a distance range of ≤20°. Within this window, we migrate data to the average ray parameter, assuming a single *P*–*s* or *S*–*p* conversion, and using the velocity model of Shen & Ritzwoller (2016) grading into AK135 (Kennett *et al.* 1995). Due to a combination of noisy data and the effect of our migration (which will exaggerate discrepancies in move-out of any reverberations) the stacked *P*-waveforms rarely contained clear crustal multiples. The fact that these multiples are evident in our synthetic stacks (Fig. 6) despite also migrating that data to common ray parameter perhaps indicates that their absence in the real data is largely attributable to strong lateral heterogeneity or crustal attenuation. Our pre-stacking migration step should help accentuate converted *S*–*p* phases in the data, which arrive before the primary phase (and well before crustal multiples). However, we note that these phases may be more vulnerable to lateral heterogeneity due to their sampling of a wider region beneath the stations. To fully address this issue will require an inversion that accounts for 3-D structure, which is beyond the scope of the present study.

Our synthetic tests indicated that that while *P*–*s* data add constraints on crustal structure and thickness, they are not as effective as *S*–*p* data at resolving deeper structure, and in fact can introduce spurious structure through mis-fit crustal multiples (Section 3.3, Fig. 7). Given this performance for (moderately noisy) synthetic data, we are skeptical that the real *P*–*s* data (Fig. 5) contain useful information deeper than the Moho. On the basis of these findings, we windowed the *P*-wave arrival to end 10 s after the main pulse, cutting out crustal multiples, and allowing the *P*–*s* data primarily to constrain the Moho depth.

Finally, recognizing that *S*−*p* stacks from different backazimuths use data that samples points hundreds of kilometres apart, for this study we only use the *S*−*p* stack at each station that has the most data. We use *P*−*s* data from the same approximate backazimuth as the *S*−*p* stacks. Capturing short-wavelength lateral velocity heterogeneities will require array analysis, beyond the scope of this manuscript.

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

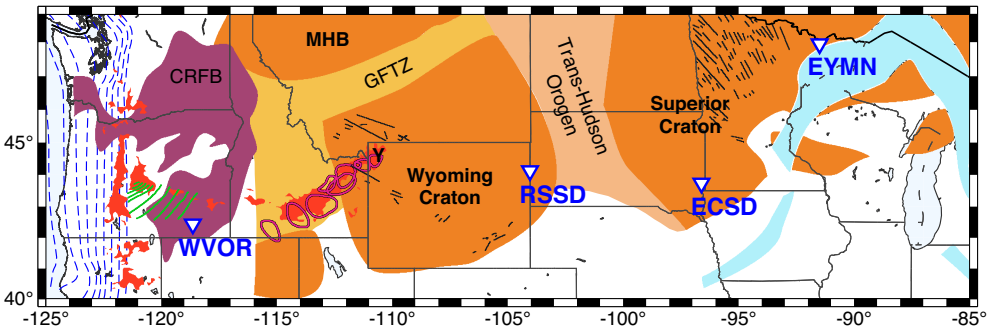

Figure 8. Tectonic map showing seismic stations in their geologic context. Cenozoic volcanism depicted in red; Snake River Plain volcanism shown with outlined polygons; high lava plains NW progression shown by 1 Ma green contours for last 10 Myr; black lines are Proterozoic dyke swarms; blue dashed lines are 10 km Juan de Fuca slab contours to 80 km depth; orange shading is Archean crust; blue shading is failed mid-continent rift. MHB is the Medicine Hat Block, CRFB is the Columbia River Flood Basalt; GFTZ is the Great Falls Tectonic Zone; Y is the present-day Yellowstone hotspot location.

4.2 ECSD station example

Figs 9 and 10 show results of an example PsSpSW inversion at station ECSD, on the South Dakota-Minnesota border, within the Superior craton interior. The stacked data for these inversions (Supporting Information Fig. S5) showed a clear positive conversion centred ∼7 s before the main *S*-wave arrival. This conversion, arising most likely from a relatively deep Moho discontinuity, was also evident on the four *P*-wave stacks as a broad pulse of positive energy on the *SV* component following the main arrival. Earlier in the *S*-wave time-series, more muted pulses of energy on the daughter component indicate conversions from velocity gradients deeper in the mantle.

For ECSD inversions, models quickly evolved towards stable values of likelihood and misfit by approximately 2000 iterations. The estimates of the number of model splines (not shown), data error and misfit (Fig. 10) were well described by stationary processes following the burn in, demonstrating the appropriateness of sampling from this likelihood space to derive the posterior model ensemble. Uncertainty in Moho depth was evident from modest variations in this parameter during and between model chains, but the final posterior distribution was well-peaked (Fig. 11). The algorithm responded to strong velocity gradients in the mid-lithosphere by tending to add more spline knots in this region – the pdf of knot locations (Fig. 10) had a broad swell centred at the depth of the mantle velocity gradients.

The inversion was clearly successful in deriving estimates of data error from the empirical fit of each data type; the values of σ closely track the RMS misfit for each data type. σ values for different individual *P–s* data stacks (from different earthquake clusters) differed slightly and were not very stable over the course of the inversion. This indicates that despite being unable to simultaneously fit all stacks perfectly, the algorithm did not select a single data stack to preferentially fit at the expense of mis-fitting the others. This was probably because the empirical data fits (Fig. 9) were all comparably good and, while the different stacks differed in detail, their signature of the Moho conversion was very similar, as it was between *P–s* and *S–p* data. The *S–p* data were crucial for the observation of deeper mantle gradients. Note also that fits to the lower-frequency body waveforms were comparable to the higher frequency data, but emphasize the coarser features in the data. Finally, the fit to Rayleigh wave phase velocities was excellent, as it was consistently for all inversions we performed.

4.3 Structure revealed by different data types

Similar to our synthetic tests (Section 3.3), we performed inversions at station ECSD with a variety of different combinations of data, to evaluate the structure constrained by each data type (Fig. 11).

The surface wave data alone returned very similar average $V_S(z)$ models across the different chains, indicating the robustness of this data set. However, these models evinced a large amount of scatter about the mean and poor constraints on the Moho. The *S*-wave data alone yielded more tightly confined model posteriors in each chain, but showed significant differences between chains due to trade-offs between Moho depth, absolute velocity, velocity jump and crustal V_P/V_S ratio. All the *S*-wave-derived models included marked velocity gradients within the mantle at depths likely corresponding to the mid-lithosphere (based on absolute velocities). *P*-wave data alone provided very little information, primarily constraining the Moho depth, with similar trade-offs to those for *S*-waves mentioned above.

We, therefore, show the model returned by joint inversions of *P*-waves and Rayleigh waves. As expected, this model had similar average velocities to the surface-wave-only profile. Surprisingly, this model did not have markedly better resolved crustal structure as determined by how strongly peaked the Moho depth and V_P/V_S posterior distributions were ($\sigma_{\text{Moho}} = 2.7$ km and $\sigma_{V_P/V_S} = 0.041$) compared to the surface-wave-only model ($\sigma_{\text{Moho}} = 3.9$ km and $\sigma_{V_P/V_S} = 0.024$). The V_S structure we recover at this station is similar to the result of Shen & Ritzwoller (2016), who employed very similar data types (*P* receiver functions and surface wave group and phase speeds).

Our preferred model, using all three datasets, has average velocities very similar to the SW or PsSW models [and to the model of Shen & Ritzwoller (2016)]. However, our model contains several additional features of interest, including a large velocity jump across the Moho, and two low velocity layers, one centered at 80 km depth and the other at 140 km depth. These low velocity layers are too thin to be resolved by surface waves alone, which have broad kernels at this depth. The negative gradients at the upper boundaries of these layers agree with the timing and polarity of body wave conversions in this region from a prior study with *S–p* data (Hopper & Fischer 2015) and the deeper phase was also seen in *P–s* data (Ford *et al.* 2016). Moho depth in the ECSD model (50.6 ± 5.5 km) is deeper than seen by Shen & Ritzwoller (2016) but is comparable to Moho depths in Schmandt *et al.* (2015) and Ford *et al.* (2016). The Moho velocity increase of ∼0.5 km s^{-1} is slightly larger than seen by Shen & Ritzwoller (2016), but consistent with the 0.55–0.8 km s^{-1}

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

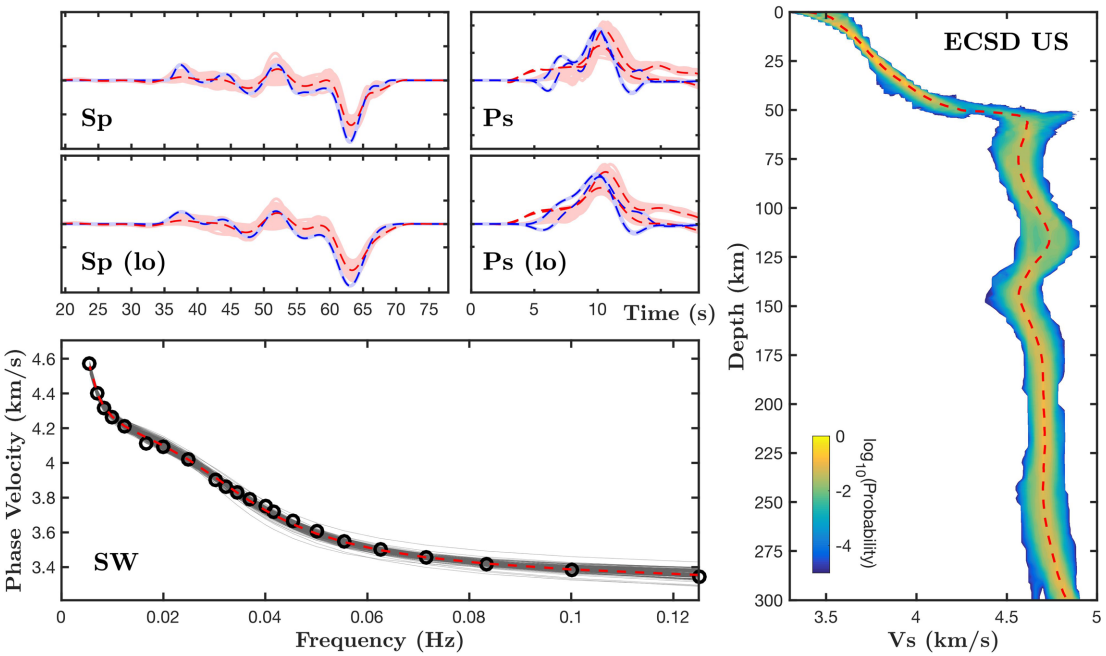

Figure 9. Example of data fit for a one-chain ECSD inversion using all three data types (Section 4.2). Top left: cross-convolution misfits for two different period bands (only 100 s high-pass filter and 'lo', which also includes low-pass filter at 3 s for *P* arrivals and 5 s for *S* arrivals), showing ($\mathbf{v}_p * \mathbf{H_{obs}}$) in red, ($\mathbf{h_p} * \mathbf{V_{obs}}$) in blue. An ideal fit would result in red and blue traces perfectly agreeing. We plot a subset of successful trial models' data in light colours and data from the final model with dark dashed lines. Each set of traces corresponds to a different earthquake data stack (Section 4.1). Bottom left: Rayleigh wave phase velocities, with observations plotted as circles, data from successful trial models in grey, and the final model in red dashed. Right: probability map from posterior ensemble of models, with final model in red. The single-chain result shown here differs slightly from the final multichain result in Figs 11 and 12.

difference between upper mantle and lower crust in the models of Schmandt *et al.* (2015) at this location. The average mantle velocities at this station appear typical for cratonic interiors (Fig. 12), but it is difficult to discern any deep NVG that might represent the LAB. The SpSW model was very similar to the PsSpSW model, and so is not shown here.

This example also illustrates the power of the Bayesian approach for understanding parameter covariance. Note that within the 120–150 km depth region that we have identified as a low-velocity zone, there is no point at which the maximum 2σ bound is less than the minimum 2σ bound for velocities above and below this feature. In other words, using the posterior velocity profile alone we could not reject the null hypothesis of constant or linearly increasing velocity throughout this depth range. However, when we consider parameter covariance within the posterior ensemble of models, we observe that V_S at ∼115 km depth exceeds V_S at ∼135 km depth in 82 per cent of the ensemble of models (Fig. 11). This makes a highly persuasive case that this low-velocity feature is required by the inversion.

Finally, the average fit to Rayleigh wave phase velocities was similar between the final model inverted using just the surface wave data (SW, RMS misfit 0.010 km s^{-1}) and the model inverted using all data types (PsSpSW, RMS misfit 0.011 km s^{-1}), despite the two velocity models having marked differences. This result nicely demonstrates the non-uniqueness of shear velocity models derived from surface wave phase velocities alone, as a result of the broad sensitivity kernels for this wave type. The low value of the data error in both cases—equal to ≈0.4 per cent—indicates how well fit this data type is in all our inversions.

4.4 Stations WVOR, RSSD and EYMN

Station WVOR, within the HLPs region of southeastern Oregon, evinces particularly low velocities almost immediately beneath the Moho. Based on the low mantle velocities at this station, the strong negative velocity gradient from ∼45–75 km depth [also observed in *S–p* CCP stacks across the HLPs (Hopper *et al.* 2014)] represents the LAB. Beneath this depth, the velocity profile is consistent with that of convecting mantle with an adiabatic thermal profile, as seen in the youngest portion of the oceans. It appears that the asthenosphere comes to within ∼80 km of the surface in this part of the continent. Particularly low velocities at ∼75 km depth may be indicative of trapped melt. The 39.8 ± 4.5 km Moho depth and crustal V_P/V_S ratio of 1.63 ± 0.08 (2σ bounds) that we estimate at this station are at the extreme bounds of a detailed *P–s* receiver function study in this area that inferred crustal thickness of ∼34 km and V_P/V_S ∼1.76 from $H-\kappa$ stacking analysis (Eagar *et al.* 2011). Some trade-off between these parameters is evident. Such a low V_P/V_S ratio may require a quartz-rich crustal mineralogy (Christensen 1996).

Station RSSD, sitting at the eastern edge of the Wyoming craton close to the Wyoming-South Dakota border, has a Moho depth of 47.8 ± 4.1 km and crustal V_P/V_S ratio of 1.86 ± 0.14. However, we hesitate to overinterpret the latter, as the posterior distribution shows that it is not well constrained. The average mantle velocity is consistent with expected values for continental interiors, but in detail has several features of note. There are three regions of strong velocity gradients: a negative velocity gradient from the Moho down to 80 km depth, a PVG from ∼120–170 km depth, and a broad negative velocity gradient (likely corresponding to the LAB) below 200 km depth. The shallow NVG has been previously imaged by receiver function studies (Foster *et al.* 2013; Hopper & Fischer 2015; Ford *et al.* 2016), although these studies estimated the depth

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

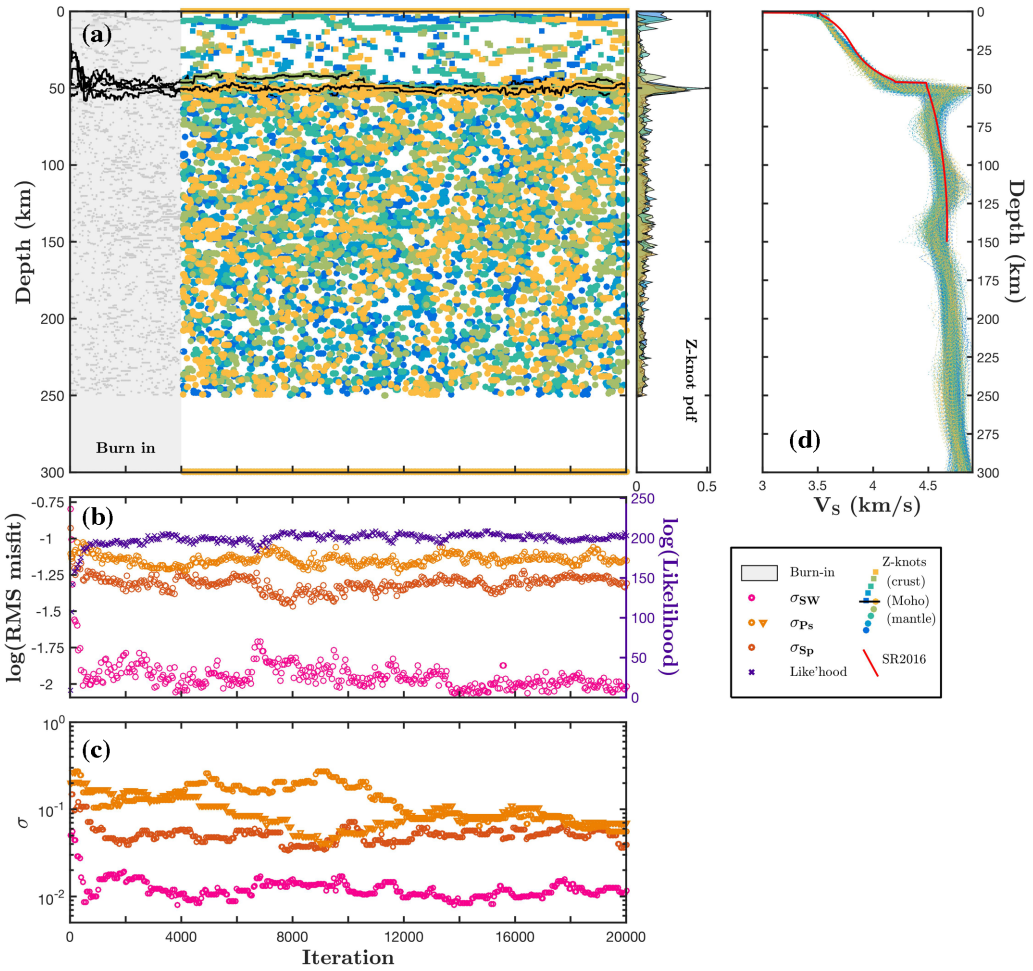

Figure 10. Example of an inversion at station ECSD using all data types, showing the change in spline knot locations, data error, likelihood and hyperparameters plotted against iteration for all accepted chains, as well as a sampling of accepted models. (a) Knot locations in the crust and mantle; grey points are all saved models, coloured points denote models after the burn-in phase of each chain (each parallel chain has its own colour). At right is a pdf of knot location. (b) RMS error of each data type (left axis), as well as the logarithm of the likelihood of that model. (c) Estimated data uncertainty for each data type. In (a), (b), and (c) only 1/15 models are shown. (d) $V_S(z)$ models from each chain (showing a subset of only 250 from each chain). Red profile: velocity model of Shen & Ritzwoller (2016) at this location, for reference.

of this feature to be somewhat greater (ranging from 70–100 km). The 120–160 km PVG has not been seen in prior studies and seems to represent a boundary between unusually slow upper lithosphere and relatively fast lower lithosphere. Individual *P*-wave stacks at this station reveal complex crustal structure, with intracrustal discontinuities noted previously (Ford *et al.* 2016). This latter structure is not clearly captured in our final model.

Station EYMN, in northeast Minnesota, sits within the interior of the Superior craton, close to the ∼1.1 Ga mid-continent rift. This station has thinner crust than the other cratonic stations in our data set, with a 40.6 ± 4.2 km deep Moho. This value is modestly shallower than previous estimates but not discrepant within uncertainty [*cf.* ∼44 km for Schmandt *et al.* (2015) and Shen & Ritzwoller (2016)]. If this crustal thickness was modified by the ancient rifting episode, no obvious signature of re-working is evident in the mantle velocity profile, which has uniformly high (4.65–4.85 km s^{-1}) and monotonically increasing shear velocities from the Moho down to 185 km depth. A broad negative velocity gradient from 185–250 km depth is indicative of a thermal LAB. The Moho velocity contrast of ∼0.5 km s^{-1} is greater than at our other stations, but well within the range of <0.7–1.3 km s^{-1} seen in the North American craton from active source studies (Snelson *et al.* 1998). The crustal V_P/V_S ratio of 1.66 ± 0.20 is more poorly constrained than other cratonic stations, with consequently greater uncertainty in Moho depth at this location.

5 DISCUSSION

5.1 Methodological improvements

Our results demonstrate the merit of utilizing multiple data types alongside a flexible parametrization to achieve detailed understanding of lithospheric structure. Our implementation of transdimensional, adaptive, piecewise-discontinuous splines holds significant promise as a tool for representing Earth structure. The seismic models we recover have well resolved crustal structure, Moho depth and lithospheric mantle velocities, as determined by strongly peaked posterior distributions for these properties. The posterior distributions offer robust uncertainty estimates and information on parameter covariance and null spaces (the term for model space within

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

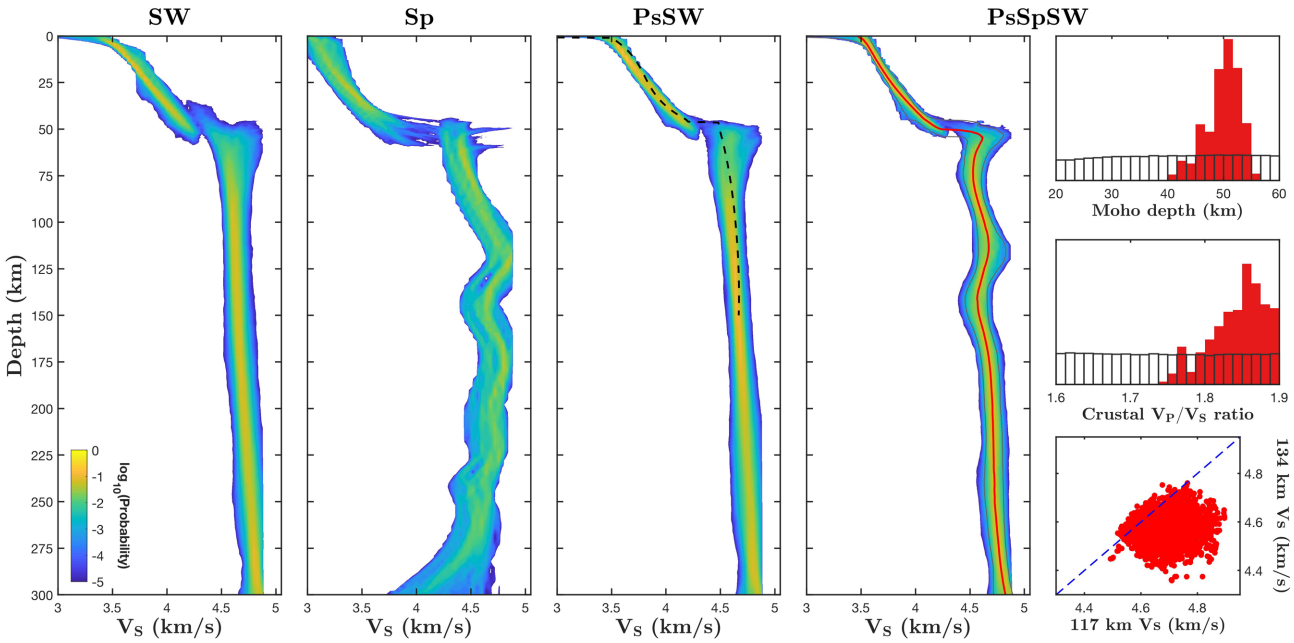

Figure 11. $V_S(z)$ models at station ECSD inverted using different combinations of data types. Left panels show probability density functions from different inversions (abbreviations as in Fig. 7), with preferred model in red, showing estimated 2σ bounds (grey lines) from a Gaussian fit at each depth. Black dashed line shows model of Shen & Ritzwoller (2016), who used similar data types, at this location. Right column shows histograms and samplings of the PsSpSW model posterior (red) and prior (white). Apparent clipping of the Sp and PsSW posterior is caused by the prior bound for mantle velocities of 4.9 km s^{-1}.

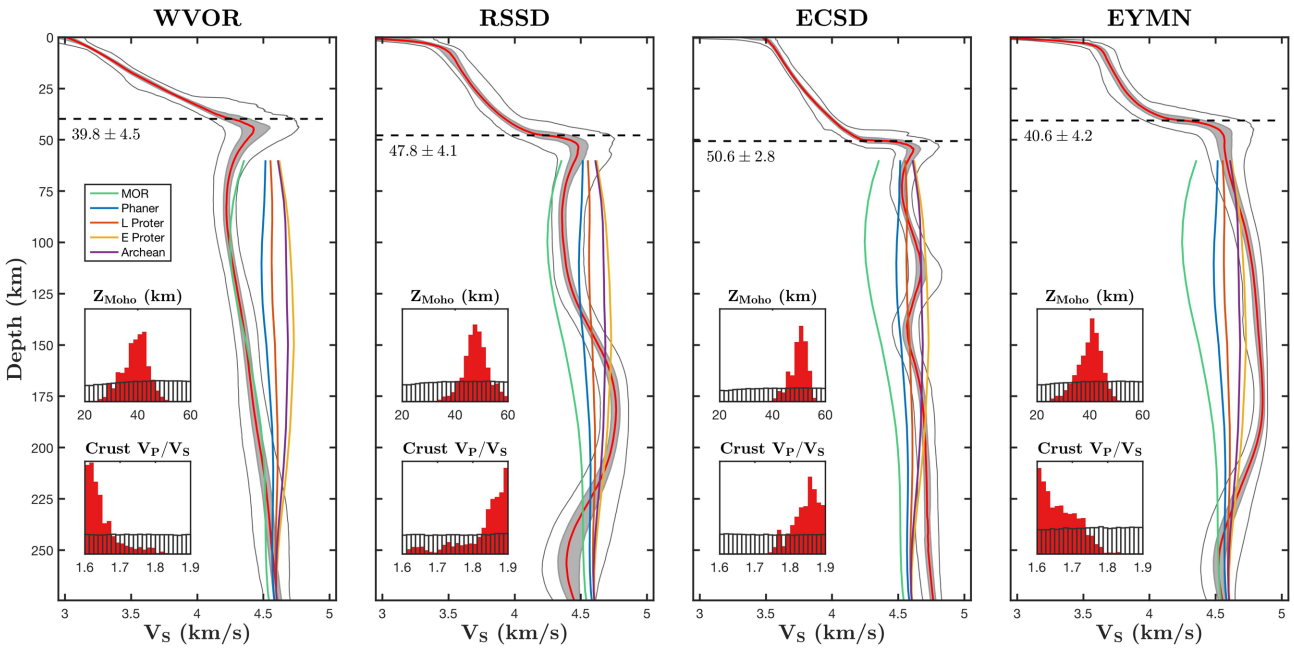

Figure 12. Final $V_S(z)$ models at four example stations across the north west continental US. Red lines: average model, grey lines: 2σ bounds. Insets show posterior (red) versus prior (white) estimates for Moho depth and crustal V_P/V_S ratio. Overlaid lines show shear velocity from global model SEMum2 averaged by age of the continental crust (Phanerozoic, Late Proterozoic, Early-Mid Proterozoic, Archean) and beneath 0–25 Myr seafloor.

which parameters can vary without affecting the data). The observed velocity profiles provide detailed resolution while maintaining model parsimony and realistic mantle velocity gradients. The importance of *S–p* data for observing deep velocity gradients is evident from inversions of both synthetic and real data. By estimating posterior uncertainty of each data type, we show that both *S–p* and surface wave phase velocity data are highly valuable for determining mantle structure, while *P–s* conversions sharpen sensitivity to crustal thickness. Our results are largely consistent with previous studies in the region in terms of crustal and mantle discontinuities (e.g. Hansen & Dueker 2009; Eagar *et al.* 2011; Levander & Miller 2012; Foster *et al.* 2013; Hopper *et al.* 2014; Lekic & Fischer 2014; Hopper & Fischer 2015; Ford *et al.* 2016; Shen & Ritzwoller 2016) and mantle velocity structure (Wagner *et al.* 2010; Obrebski *et al.* 2011; Humphreys *et al.* 2015; Shen & Ritzwoller 2016) but combine these two observations self-consistently to deliver Earth

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

models that allow us to examine imaged structure in full cognizance of parameter trade-offs.

5.2 Lithospheric structure across the northwestern to north-central US

Velocity models differ substantially between the stations we highlight as examples in this study. Fig. 12 shows these stations arranged from west to east. Several trends are clear. Average crustal velocity increases inward with distance from the edge of the cratonic lithosphere, and the gradient of velocities in the crust—arguably a proxy for heat flow—decreases. The average mantle velocity also increases along this transect. Consistent trends to the velocity gradients within, or at the base of, the mantle lithosphere are difficult to ascertain from this small sample of stations. However, qualitatively speaking there is a clear contrast between a sharp basal LAB for WVOR compared to generally higher velocities over the upper 200 km of mantle for cratonic stations, with a broad negative velocity gradient from 200–250 km depth that may represent a thermal LAB.

Station WVOR has by far the lowest upper-mantle velocities, consistent with an extremely thin mantle lithosphere comprising ∼40 km of crust above ∼20 km of mantle lithosphere with inverted velocity gradient. The strong, shallow LAB observed here is characteristic of the tectonically active western US (Fischer *et al.* 2010). This region is a good candidate for basal lithospheric removal via melt impounding or infiltration (Havlin *et al.* 2013). Extremely low upper-mantle velocities in this spot have been observed previously in surface (Wagner *et al.* 2010) and body-wave (Obrebski *et al.* 2011) studies and likely require melt. This agrees with the model of Meqbel *et al.* (2014), who use magnetotelluric data to argue for a shallow (∼55 km depth) LAB in this region overlying a highly conductive layer.

By contrast, within the Wyoming craton Meqbel *et al.* (2014) find that high resistivity persists down to approximately 250 km depth, in agreement with our much deeper inferred LAB at station RSSD. The most striking aspect to the velocity profile we recover at this station is the combination of unexpectedly low velocities from 60–125 km depth, and particularly high velocities from 150–200 km depth. This structure qualitatively agrees with Dalton *et al.* (2017), who found that the optimal match for cratonic lithosphere phase velocities was obtained using a xenolith-determined thermal model that was modified to include a 5 per cent velocity reduction from 60–80 km depth and a 5 per cent velocity increase from 200–250 km. Within the resolution of long-period surface waves, our velocity model is surprisingly similar to that which they propose. One explanation for the velocity reduction is the emplacement of volatile-rich mineral assemblages due to metasomatism and/or melt infiltration, ponding and crystallization (Hansen *et al.* 2015; Rader *et al.* 2015; Selway *et al.* 2015), consistent with evidence from xenolith geochemistry (Carlson *et al.* 2004; Hearn 2004), and potentially related to Farallon slab subduction during the Laramide orogeny. The high velocities at depth could represent an underthrust oceanic plateau (Humphreys *et al.* 2015) or refertilized mantle with a high garnet and pyroxene content that elevates velocities (Kopylova & Russell 2000; Hopper & Fischer 2015).

The two low velocity layers observed in the mantle beneath ECSD, one centred at 80 km and the other at 140 km, could also represent volatile-rich phases emplaced by metasomatism and melting. The shallower layer would be consistent with melts that stalled and cooled just below a solidus for carbonated and hydrated peridotite (Dasgupta 2013; Hansen *et al.* 2015; Rader *et al.* 2015). Metasomatism may explain the deeper layer, or it could represent a layer formed at shallower depths that was underthrust beneath or underneath the lithosphere of the Superior craton, for example, during orogenesis (Hopper & Fischer 2015). Alternatively, this deeper layer could relate to the original lithospheric base, defined compositionally, beneath which the thermal portion of the craton subsequently developed (Yuan & Romanowicz 2010). In this case, the apparent velocity reduction may, in fact, represent a change in anisotropic fabric (Wirth & Long 2014).

EYMN lies on the flank of the Mid-continent rift (Fig. 8), and its crustal thickness (∼40 km) is comparable to values found at other rift-flank locations along the strike of the rift, as opposed to thicker crust found near the rift axis (∼55 km; e.g. Zhang *et al.* 2016). Signatures of ancient rifting in the mantle beneath EYMN are less clear, except to the extent that the absence of low velocity layers within the mantle lithosphere could represent the destruction of such layers by rift-related thermal anomalies. Expanding our technique to more stations in this neighbourhood will permit a more thorough examination of the lateral contiguity and average strength of these low velocity layers, in particular the deeper low velocity layer observed beneath ECSD, which does not seem to persist throughout the Superior craton, based on our analysis of station EYMN.

5.3 Future work

This study represents a proof-of-concept for our method, and a validation of the utility of converted *S–p* phases for constraining mantle structure. Future steps will entail the expansion of our dataset to all long-lived stations in this region. If possible, we will extend the analysis to short-lived stations (<5 yr deployments), if sufficient earthquake arrivals are available to form robust station stacks. At the expense of vulnerability to strong 3-D velocity variations, we will explore aggregation of data across 'mini-arrays' of 3+ stations within the dense TA network in order to improve signal-to-noise on the body wave stacks. At present, our assumption of 1-D structure beneath the stations is a source of data misfit; body wave stacks at some stations reveal significant backazimuthal variation (Supporting Information Fig. S7) that cannot be simultaneously fit by any 1-D velocity profile. We plan to employ automated waveform cluster analysis to separate backazimuthal bins for separate inversions as a first step towards fully 3-D implementation.

Our methodology makes it trivial to extend the inversion to include more data types, as long as the forward-problem is computationally efficient. In particular, MINEOS can compute Rayleigh wave group velocities at no significant added expense. Inclusion of Rayleigh wave ellipticity (Li *et al.* 2016; Shen & Ritzwoller 2016) would improve constraints on crustal V_P/V_S ratios, which our posterior distributions reveal to be the most poorly resolved parameter within our models. In this inversion, we have solved only for isotropic structure, but our codes can handle radially anisotropic models at no extra expense—the addition of Love wave data would then allow us to determine anisotropic lithospheric structure that has been linked to chemical layering within the cratons (Yuan & Romanowicz 2010), albeit at the expense of longer run times necessitated by the addition of anisotropic parameters that expand the model space explored by the MCMCs.

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

6 CONCLUSIONS

We have developed a novel approach for obtaining upper-mantle velocity structure by simultaneous inversion of multiple complimentary data types. The use of adaptive discontinuous splines achieves a description of a realistic Earth model that is both parsimonious and detailed, allowing MCMCs to exhaustively explore a limited parameter space while capturing fine scale features we seek to understand. The data uncertainty is left as a free parameter for which the inverse algorithm solves; we show this tends towards the posterior RMS of the data fit. By pre-migrating stacked body wave data that is carefully aligned and quality controlled, converted phases originating at sharp velocity gradients are brought to the fore. In particular, *S–p* conversions are shown to be extremely powerful for the imaging of deep discontinuities. The joint inversion of these data with Rayleigh wave phase velocities reveals marked layering within continental cratons. In the tectonically active northwestern US, we find much thinner lithosphere overlying very low mantle velocities that grade into an adiabatic velocity profile. The benefits of our Bayesian approach include robust parameter uncertainty estimates, quantification of trade-offs and the flexibility to modularly add additional data constraints as we continue to develop, and more widely implement, this technique.

ACKNOWLEDGEMENTS

This research was supported by the NSF CSEDI Program under award EAR-1361487 to KMF and CAD. Many thanks are owed to Thomas Bodin for discussions regarding the inversion algorithm and the convergence of data error estimates. Thanks to Riddhi Dave and Aibing Li, and to Weisen Shen and Fan-Chi Lin for making their phase velocity models available. Phase velocities from Bao *et al.* (2016), Dave & Li (2016), Shen & Ritzwoller (2016) and Schmandt *et al.* (2015) are available online or upon personal request to the authors. All body wave data used are available through the IRIS DMC (network codes US, IU). We thank two anonymous reviewers and Juan Carlos Afonso for their helpful comments. Thanks to Emily Hopper for geologic map data. Final models are available at Library ftp (TBD).

REFERENCES

Abers, G.A. & Hacker, B.R., 2016. A MATLAB toolbox and Excel workbook for calculating the densities, seismic wave speeds, and major element composition of minerals and rocks at pressure and temperature, *Geochem. Geophys. Geosyst.,* **17**(2), 616–624.

Abers, G.A., Fischer, K.M., Hirth, G., Wiens, D.A., Plank, T., Holtzman, B.K., McCarthy, C. & Gazel, E., 2014. Reconciling mantle attenuation-temperature relationships from seismology, petrology, and laboratory measurements, *Geochem. Geophys. Geosyst.,* **15,** 3521–3542.

Afonso, J.C., Fullea, J., Griffin, W.L., Yang, Y., Jones, A.G., D Connolly, J.A. & O'Reilly, S.Y., 2013a. 3-D multiobservable probabilistic inversion for the compositional and thermal structure of the lithosphere and upper mantle. I: a prioripetrological information and geophysical observables, *J. geophys. Res.,* **118**(5), 2586–2617.

Afonso, J.C., Fullea, J., Yang, Y., Connolly, J.A.D. & Jones, A.G., 2013b. 3-D multi-observable probabilistic inversion for the compositional and thermal structure of the lithosphere and upper mantle. II: General methodology and resolution analysis, *J. geophys. Res.,* **118**(4), 1650–1676.

Afonso, J.C., Rawlinson, N., Yang, Y., Schutt, D.L., Jones, A.G., Fullea, J. & Griffin, W.L., 2016. 3-D multiobservable probabilistic inversion for the compositional and thermal structure of the lithosphere and upper mantle: III. Thermochemical tomography in the Western-Central U.S., *J. geophys. Res.,* **121**(10), 7337–7370.

Agostinetti, N.P. & Malinverno, A., 2010. Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.,* **95**(2), 15303–15.

Bao, X., Dalton, C.A., Jin, G., Gaherty, J.B. & Shen, Y., 2016. Imaging Rayleigh wave attenuation with USArray, *Geophys. J. Int.,* **206,** 241–259.

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.,* **178**(3), 1411–1436.

Bodin, T., Sambridge, M., Rawlinson, N. & Arroucau, P., 2012a. Transdimensional tomography with unknown data noise, *Geophys. J. Int.,* **189**(3), 1536–1556.

Bodin, T., Sambridge, M., Tkalcic, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012b. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.,* **117**(B2), B02301–24.

Bodin, T., Yuan, H. & Romanowicz, B., 2013. Inversion of receiver functions without deconvolution—application to the Indian craton, *Geophys. J. Int.,* **196**(2), 1025–1033.

Bodin, T., Leiva, J., Romanowicz, B., Maupin, V. & Yuan, H., 2016. Imaging anisotropic layering with Bayesian inversion of multiple data types, *Geophys. J. Int.,* **206**(1), 605–629.

Brocher, T.M., 2005. Empirical relations between elastic wavespeeds and density in the Earth's crust, *Bull. seism. Soc. Am.,* **95**(6), 2081–2092.

Buehler, J.S. & Shearer, P.M., 2017. Uppermost mantle seismic velocity structure beneath USArray, *J. geophys. Res.,* **122**(1), 436–448.

Burdick, S. & Lekic, V., 2017. Velocity variations and uncertainty from transdimensional P-wave tomography of North America, *Geophys. J. Int.,* **209**(2), 1337–1351.

Burdick, S. *et al.*, 2017. Model update May 2016: upper-mantle heterogeneity beneath North America from travel-time tomography with global and USArray data, *Seismol. Res. Lett.,* **88**(2A), 319–325.

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 planet. Sci. Lett.,* **449,** 164–175.

Carlson, R.W., Irving, A.J., Schulze, D.J. & , .Hearn, B.C., Jr, 2004. Timing of Precambrian melt depletion and Phanerozoic refertilization events in the lithospheric mantle of the Wyoming Craton and adjacent Central Plains Orogen, *Lithos,* **77**(1-4), 453–472.

Christensen, N.I., 1996. Poisson's ratio and crustal seismology, *J. geophys. Res.,* **101**(B2), 3139–3156.

Dalton, C.A. & Faul, U.H., 2010. The oceanic and cratonic upper mantle: Clues from joint interpretation of global velocity and attenuation models, *Lithos,* **120**(1-2), 160–172.

Dalton, C.A., Bao, X. & Ma, Z., 2017. The thermal structure of cratonic lithosphere from global Rayleigh wave attenuation, *Earth planet. Sci. Lett.,* **457,** 250–262. doi:10.1016/j.epsl.2016.10.014.

Dasgupta, R., 2013. Ingassing, storage, and outgassing of terrestrial carbon through geologic time,, *Rev. Mineral. Geochem.,* **75** (1), 183–229.

Dave, R. & Li, A., 2016. Destruction of the Wyoming craton: seismic evidence and geodynamic processes, *Geology,* **44**(11), 883–886.

Drilleau, M., Beucler, E., Mocquet, A., Verhoeven, O., Moebs, G., Burgos, G., Montagner, J.P. & Vacher, P., 2013. A Bayesian approach to infer radial models of temperature and anisotropy in the transition zone from surface wave dispersion curves, *Geophys. J. Int.,* **195**(2), 1165–1183.

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

Eagar, K.C., Fouch, M.J., James, D.E. & Carlson, R.W., 2011. Crustal structure beneath the high lava plains of eastern Oregon and surrounding regions from receiver function analysis, *J. geophys. Res.,* **116**(B2), 15303–15318.

Eaton, D.W., Darbyshire, F., Evans, R.L., Grütter, H., Jones, A.G. & Yuan, X., 2009. The elusive lithosphere–asthenosphere boundary (LAB) beneath cratons, *Lithos,* **109**(1-2), 1–22.

Fischer, K.M., Ford, H.A., Abt, D.L. & Rychert, C.A., 2010. The Lithosphere-Asthenosphere Boundary, *Annu. Rev. Earth Planet. Sci.,* **38**(1), 551–575.

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, *J. geophys. Res.,* **121**(9), 6675–6697.

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

Foster, A., Ekström, G. & Nettles, M., 2014. Surface wave phase velocities of the Western United States from a two-station method, *Geophys. J. Int.,* **196**(2), 1189–1206.

Foster, K., Dueker, K., Schmandt, B. & Yuan, H., 2013. A sharp cratonic lithosphere–asthenosphere boundary beneath the American Midwest and its relation to mantle flow, *Earth planet. Sci. Lett.,* **402,** 1–8.

French, S., Lekic, V. & Romanowicz, B., 2013. Waveform tomography reveals channeled flow at the base of the oceanic asthenosphere, *Science,* **342**(6155), 224–227.

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 Res.,* **37,** 252–265, doi:10.1016/j.gr.2016.07.002.

Hansen, S. & Dueker, K., 2009. P-and S-wave receiver function images of crustal imbrication beneath the Cheyenne Belt in southeast Wyoming, *Bull. seism. Soc. Am.,* **99**(3), 1953–1961.

Hansen, S.M., Dueker, K. & Schmandt, B., 2015. Thermal classification of lithospheric discontinuities beneath USArray, *Earth planet. Sci. Lett.,* **431,** 36–47, doi:10.1016/j.epsl.2015.09.009.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika,* **57**(1), 97–109.

Havlin, C., Parmentier, E.M. & Hirth, G., 2013. Dike propagation driven by melt accumulation at the lithosphere–asthenosphere boundary, *Earth planet. Sci. Lett.,* **376,** 20–28.

Hawkins, R. & Sambridge, M., 2015. Geophysical imaging using trans-dimensional trees, *Geophys. J. Int.,* **203,** 972–1000.

Hearn, B.C., 2004. The Homestead kimberlite, central Montana, USA: mineralogy, xenocrysts, and upper-mantle xenoliths, *Lithos,* **77**(1–4), 473–491.

Hopper, E. & Fischer, K.M., 2015. The meaning of midlithospheric discontinuities: A case study in the northern US craton, *Geochem. Geophys. Geosyst.,* **16**(12), 4057–4083.

Hopper, E., Ford, H.A., Fischer, K.M., Lekić, V. & Fouch, M.J., 2014. The lithosphere–asthenosphere boundary and the tectonic and magmatic history of the northwestern United States, *Earth planet. Sci. Lett.,* **402,** 69–81.

Humphreys, E., 2009. Relation of flat subduction to magmatism and deformation in the western United States, *Geol. Soc. Am. Memo.,* **204,** 85–98.

Humphreys, E.D., Schmandt, B., Bezada, M.J. & Perry-Houts, J., 2015. Recent craton growth by slab stacking beneath Wyoming, *Earth planet. Sci. Lett.,* **429,** 170–180.

Jackson, I. & Faul, U.H., 2010. Grainsize-sensitive viscoelastic relaxation in olivine: Towards a robust laboratory-based model for seismological application, *Phys. Earth planet. Inter.,* **183**(1-2), 151–163.

Jin, G. & Gaherty, J.B., 2015. Surface wave phase-velocity tomography based on multichannel cross-correlation, *Geophys. J. Int.,* **201**(3), 1383–1398.

Jordan, B.T., Grunder, A.L., Duncan, R.A. & Deino, A.L., 2004. Geochronology of age-progressive volcanism of the Oregon High Lava Plains: Implications for the plume interpretation of Yellowstone, *J. geophys. Res.,* **109**(B10), 225–19.

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

Karato, S., 2012. On the origin of the asthenosphere, *Earth Planet. Sci. Lett.,* **321-322,** 95–103.

Keith, C.M. & Crampin, S., 1977a. Seismic body waves in anisotropic media: synthetic seismograms, *Geophys. J. Int.,* **49**(1), 225–243.

Keith, C.M. & Crampin, S., 1977b. Seismic body waves in anisotropic media: Propagation through a layer, *Geophys. J. Int.,* **49**(1), 209–223.

Keith, C.M. & Crampin, S., 1977c. Seismic body waves in anisotropic media: Reflection and refraction at a plane interface, *Geophys. J. Int.,* **49**(1), 181–208.

Kennett, B. L.N., Engdahl, E.R. & Buland, R.P., 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.,* **122**(1), 108–124.

Khan, A., Zunino, A. & Deschamps, F., 2011. The thermo-chemical and physical structure beneath the North American continent from Bayesian inversion of surface-wave phase velocities, *J. geophys. Res.,* **116,** B09304.

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

Kind, R., Yuan, X., Mechie, J. & Sodoudi, F., 2015. Structure of the upper mantle in the north-western and central United States from USArray S-receiver functions, *Solid Earth,* **6**(3), 957–970.

Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P., 1983. Optimization by simulated annealing, *Science,* **220**(4598), 671–680.

Kopylova, M.G. & Russell, J.K., 2000. Chemical stratification of cratonic lithosphere: constraints from the Northern Slave craton, Canada, *Earth planet. Sci. Lett.,* **181,** 71–87.

Lekic, V. & Fischer, K.M., 2014. Contrasting lithospheric signatures across the western United States revealed by Sp receiver functions, *Earth planet. Sci. Lett.,* **402,** 90–98.

Lekic, V. & Fischer, K.M., 2017. Interpreting spatially stacked Sp receiver functions, *Geophys. J. Int.,* **210**(2), 874–886.

Levander, A. & Miller, M.S., 2012. Evolutionary aspects of lithosphere discontinuity structure in the western U.S., *Geochem. Geophys. Geosyst.,* **13**(7), Q0AK07, doi:10.1029/2012GC004056 .

Li, G., Chen, H., Niu, F., Guo, Z. & Yang, Y., 2016. Measurement of Rayleigh wave ellipticity and its application to the joint inversion of high-resolution S wave velocity structure beneath northeast China, *J. geophys. Res.,* **121,** 864–880.

Lin, F.-C. & Schmandt, B., 2014. Upper crustal azimuthal anisotropy across the contiguous U.S. determined by Rayleigh wave ellipticity, *Geophys. Res. Lett.,* **41**(23), 8301–8307.

Lin, F.C., Tsai, V.C. & Schmandt, B., 2014. 3-D crustal structure of the western United States: application of Rayleigh-wave ellipticity extracted from noise cross-correlations, *Geophys. J. Int.,* **198**(2), 656–670.

Lowry, A.R. & Smith, R.B., 1994. Flexural rigidity of the Basin and Range-Colorado Plateau-Rocky Mountain transition from coherence analysis of gravity and topography, *J. geophys. Res.,* **99**(B10), 20123–20140.

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

Mancinelli, N.J., , , Fischer, K.M. & Dalton, C.A., 2017. How sharp is the cratonic lithosphere-asthenosphere transition? *Geophys. Res. Lett.,* doi:10.1002/2017GL074518

Masters, G., Barmine, M.P. & Kientz, S., 2007. Mineos user's manual, in *Computational Infrastructure for Geodynamics,* California Insitute of Technology.

Menke, W., 2017. The uniqueness of single data function, multiple model functions, inverse problems including the Rayleigh wave dispersion problem, *Pure appl. Geophys.,* **174**(4), 1699–1710.

Menke, W. & Levin, V., 2003. The cross-convolution method for interpreting SKSsplitting observations, with application to one and two-layer anisotropic earth models, *Geophys. J. Int.,* **154**(2), 379–392.

Meqbel, N.M., Egbert, G.D., Wannamaker, P.E., Kelbert, A. & Schultz, A., 2014. Deep electrical resistivity structure of the northwestern U.S. derived from 3-D inversion of USArray magnetotelluric data, *Earth planet. Sci. Lett.,* **402,** 290–304.

Metropolis, N. & Rosenbluth, A.W., 1953. Equation of state calculations by fast computing machines, *J. Chem. Phys.,* **21**(6), 1087–1092.

Mosegaard, K. & Sambridge, M., 2002. Monte Carlo analysis of inverse problems, *Inverse Probl.,* **18,** R29–R54.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.,* **100**(B7), 12 431–12 447.

Obrebski, M., Allen, R.M., Pollitz, F.F. & Hung, S.H., 2011. Lithosphere–asthenosphere interaction beneath the western United States from the joint inversion of body-wave traveltimes and surface-wave phase velocities, *Geophys. J. R. Astron. Soc.,* **185**(2), 1003–1021.

Olugboji, T.M., Park, J. & Karato, S., 2016. Nature of the seismic lithosphere-asthenosphere boundary within normal oceanic mantle from high-resolution receiver functions, *Geochem. Geophys. Geosyst.,* **17**(4), 1265–1282.

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

Özalaybey, S., Savage, M.K. & Sheehan, A.F., 1997. Shear-wave velocity structure in the northern Basin and Range province from the combined analysis of receiver functions and surface waves, *Bull. seism. Soc. Am.,* **87**(1), 183–199.

Percival, J.A., Sanborn-Barrie, M., Skulski, T., Stott, G.M., Helmstaedt, H. & White, D.J., 2006. Tectonic evolution of the western Superior Province from NATMAP and Lithoprobe studies, *Canad. J. Earth Sci.,* **43**(7), 1085–1117.

Porritt, R.W., Allen, R.M. & Pollitz, F.F., 2014. Seismic imaging east of the Rocky Mountains with USArray, *Earth planet. Sci. Lett.,* **402,** 16–25.

Priestley, K. & McKenzie, D., 2013. The relationship between shear wave velocity, temperature, attenuation and viscosity in the shallow part of the mantle, *Earth planet. Sci. Lett.,* **381,** 78–91.

Rader, E., Emry, E., Schmerr, N., Frost, D., Cheng, C., Menard, J., Yu, C.Q. & Geist, D., 2015. Characterization and Petrological Constraints of the Midlithospheric Discontinuity, *Geochem. Geophys. Geosyst.,* **16**(10), 3484–3504.

Roy, C. & Romanowicz, B., 2017. On the implications of a priori constraints in transdimensional Bayesian inversion for continental lithospheric layering, *J. geophys. Res.,* **122,** 10118–10131.

Rychert, C.A., Schmerr, N. & Harmon, N., 2012. The Pacific lithosphere-asthenosphere boundary: seismic imaging and anisotropic constraints from SS waveforms, *Geochem. Geophys. Geosyst.,* **13**(9).

Sambridge, M., 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.,* **40**(3), 238.

Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2006. Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.,* **167**(2), 528–542.

Schmandt, B., Lin, F.-C. & Karlstrom, K.E., 2015. Distinct crustal isostasy trends east and west of the Rocky Mountain Front, *Geophys. Res. Lett.,* **42**(23), 10290–10298.

Schmerr, N., 2012. The Gutenberg Discontinuity: Melt at the Lithosphere-Asthenosphere Boundary, *Science,* **335**(6075), 1480–1483.

Selway, K., Ford, H. & Kelemen, P., 2015. The seismic mid-lithosphere discontinuity, *Earth planet. Sci. Lett.,* **414,** 45–57.

Shen, W. & Ritzwoller, M.H., 2016. Crustal and uppermost mantle structure beneath the United States, *J. geophys. Res.,* **121,** 4306–4342.

Shen, W., Ritzwoller, M.H. & Schulte-Pelkum, V., 2013a. A 3-D model of the crust and uppermost mantle beneath the central and western US by joint inversion of receiver functions and surface wave dispersion, *J. geophys. Res.,* **118,** 1–15.

Shen, W., Ritzwoller, M.H., Schulte-Pelkum, V. & Lin, F.-C., 2013b. Joint inversion of surface wave dispersion and receiver functions: a Bayesian Monte-Carlo approach, *Geophys. J. Int.,* **192**(2), 807–836.

Silver, P.G. & Chan, W.W., 1991. Shear wave splitting and subcontinental mantle deformation, *J. geophys. Res.,* **96**(B10), 16429–16454.

Snelson, C.M., Henstock, T.J., Miller, G.R., Miller, K.C. & Levander, A., 1998. Crustal and uppermost mantle structure along the Deep Probe seismic profile, *Rocky Mt. Geol.,* **33**(2), 181–198.

Tork Qashqai, M., Carlos Afonso, J. & Yang, Y., 2016. The crustal structure of the Arizona transition zone and southern Colorado plateau from multi-observable probabilistic inversion, *Geochem. Geophys. Geosyst.,* **17**(11), 4308–4332.

van der Meijde, M., Marone, F. & Giardini, D., 2003. Seismic evidence for water deep in Earth's upper mantle, *Science,* **300,** 1556–1558.

VanDecar, J.C. & Crosson, R.S., 1990. Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares, *Bull. seism. Soc. Am.,* **80**(1), 150–169.

Wagner, L., Forsyth, D.W., Fouch, M.J. & James, D.E., 2010. Detailed three-dimensional shear wave velocity structure of the northwestern United States from Rayleigh wave tomography, *Earth planet. Sci. Lett.,* **299,** 273–284.

Whitmeyer, S.J. & Karlstrom, K.E., 2007. Tectonic model for the Proterozoic growth of North America, *Geosphere,* **3**(4), 220–259.

Wirth, E.A. & Long, M.D., 2014. A contrast in anisotropy across mid-lithospheric discontinuities beneath the central United States–A relic of craton formation, *Geology,* **42**(10), 851–854.

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

Yuan, H., Romanowicz, B., Fischer, K.M. & Abt, D.L., 2011. 3-D shear wave radially and azimuthally anisotropic velocity model of the North American upper mantle, *Geophys. J. Int.,* **184**(3), 1237–1260.

Zhang, H. *et al.*, 2016. Distinct crustal structure of the North American Mid-continent Rift from Pwave receiver functions, *J. geophys. Res.,* **121**(11), 8136–8153.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. $V_S(z)$ probability density functions recovered using different combinations of data types, including *Ps* body waves, *Sp* body waves and surface wave (*SW*) phase velocities. The input model (black line) has a 10 per cent $\Delta V_S/V_S$ Moho at 30 km depth, and a 10-km-wide, 5 per cent negative velocity gradient at 100 km depth. The final panel shows the result using synthetics that had 'low' noise, comparable to the real data noise. Inversions performed with six parallel chains (Section 2.1.1); the number of successful chains contributing to each posterior is shown in brackets at lower left.

Figure S2. As for Fig. S1, but for target model with 45 km Moho depth, and 0-km-wide LAB gradient.

Figure S3. As for Fig. S1, but for target model with 45 km Moho depth, and 10-km-wide LAB gradient.

Figure S4. As for Fig. S1, but for target model with 45 km Moho depth, and 40-km-wide LAB gradient.

Figure S5. Normalized *P*-wave arrival stacks (top) and *S*-wave arrival stacks (bottom) recorded at station ECSD, aligned on the parent arrival. Blue: *P*-component, red: *SV*-component. The daughter component in each plot is scaled up by 3×. The number of events per stack is shown at lower left, and the arrivals in each stack are plotted as a function of backazimuth and great circle distance in the polar plot to right. Grey plots indicate traces from backazimuths that were not used in the inversion shown in the main article.

Figure S6. As for Fig. S5 but for station WVOR.

Figure S7. As for Fig. S5 but for station RSSD.

Figure S8. As for Fig. S5 but for station EYMN.

Figure S9. Test for perturbation algorithm bias, comparing theoretical prior (black lines) to the empirical prior (red bars) histograms, as described in Section S1. Blue curves show the theoretical values for the Jeffrey's distribution.

Figure S10. Comparisons of prior (white) and posterior (red) distributions for various model parameters at station ECSD. Velocities are in km s^{-1}.

Figure S11. As for Fig. S10 but for station WVOR.

Figure S12. As for Fig. S10 but for station RSSD.

Figure S13. As for Fig. S10 but for station EYMN.

Figure S14. Comparisons of prior (white) and posterior (red) distributions for various model parameters for a synthetic inversion using *P–s*, *S–p* and surface wave data. The values for the target model (with a 30-km-deep Moho and a 5 per cent mantle velocity decrease over 10 km centred at 100 km depth) are shown with blue dashed lines. This model contained a thin sediment layer above the crystalline crust. Velocities are in km s^{-1}.

Figure S15. Results of a synthetic test to assess recovery of a negative velocity gradient within the lithosphere. This inversion was performed using *P–s*, *S–p* and Rayleigh wave synthetic data, with Gaussian noise added [*cf.* (Section 2.1.1)] and four parallel chains. The posterior model ensemble is shown in the heat-map, with a histogram of discontinuity depths to right (sediments were not included in the target model, nor permitted in the inversion. Left plots

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018

show the prior (black bars) versus posterior (red bars) distributions for Moho depths, as well as comparisons of velocities at different depths within the models that populate the final ensemble. Clearly, the negative velocity gradient at the 'LAB' is well captured, as is the negative gradient within the lithosphere.

Figure S16. Surface wave phase velocity fits for the four stations highlighted in this paper. Observed values are plotted as cyan circles. Red lines are dispersion curves computed for the median models at each station (i.e. the red lines in Fig. 12). Box and whiskers plots show the distribution of phase velocities predicted by all of the models in the ensemble at each station (sampling from all chains). The whiskers denote maximum/minimum values, while the boxes (which are in most cases so thin that they cannot be discerned) show the two-sigma bounds of phase velocity predicted by accepted models. Note that these inversions also incorporated *P–s* and *S–p* data (not shown).

Figure S17. Results of synthetic tests comparing recovered structure when the input noise is Gaussian random noise or a 'realistic' noise time-series. The 'realistic' noise is computed by taking the power spectrum of the −95 to −25 s window prior to the main arrival on Ps and Sp components for station RSSD (selected arbitrarily) and re-constituting this noise, having randomized the phase. The noise RMS for both 'real' and Gaussian cases is scaled to the amplitude of the observed noise using the signal-to-noise ratio of the primary arrival on the parent component. The phase velocities have Gaussian noise with a standard deviation of 0.015 km s^{-1}. The target model, with a Moho depth of 45 km, a 0-km-wide LAB at 120 km depth and a crustal V_P/V_S ratio of 1.8 is shown in blue, while the final model is depicted as in Fig. 12.

Table S1. Perturbations and bounds of inversion parameters.

Table S2. Probabilities of altering inverse parameters.

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: DATA ERROR CONVERGENCE

Here we demonstrate briefly how the random perturbations to the data error will lead to convergence on the appropriate data misfit. Consider the derivative of the likelihood function (eq. 3) with respect to σ:

$$\frac{\partial\, p(\mathbf{d}_i|\mathbf{m})}{\partial\, \sigma_j} = \frac{\partial}{\partial\, \sigma_j}\left(\frac{1}{(\sqrt{2\pi}\,\sigma_i)^n}\exp\left\{-\frac{\Phi_i}{2\sigma_i^2}\right\}\right) \tag{A1}$$

$$= \frac{\delta_{ij}}{(\sqrt{2\pi}\,\sigma_i)^n}\exp\left\{-\frac{\Phi_i}{2\sigma_i^2}\right\}\left[\Phi_i\,\sigma_j^{-3} - n_i\,\sigma_j^{-1}\right] \tag{A2}$$

at the maximum likelihood, this derivative is zero, so

$$0 = \left[\Phi_i\,\sigma_i^{-3} - n_i\,\sigma_i^{-1}\right] \tag{A3}$$

$$\sigma_i = \sqrt{\frac{\Phi_i}{n_i}}. \tag{A4}$$

The optimal value of σ is thus the RMS of the data misfit. Over the course of the inversion, we therefore expect σ to converge towards the true data error.

Now consider a perturbation to σ, comparing current value '0' to trial value 'prop'. The log-likelihood ratio is

$$\ln\left(\frac{p_{\rm prop}}{p_0}\right) = \ln\left(\frac{\sigma_{\rm prop}^{-n}\exp(-\Phi/2\sigma_{\rm prop}^2)}{\sigma_0^{-n}\exp(-\Phi/2\sigma_0^2)}\right) \tag{A5}$$

$$= n\ln\left(\frac{\sigma_0}{\sigma_{\rm prop}}\right) + \frac{\Phi}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_{\rm prop}^2}\right). \tag{A6}$$

These two terms work in opposite directions. If $\sigma_{\rm prop} > \sigma_0$ then the first term of this log-likelihood function is negative, while the second term is positive. If $\sigma_{\rm prop} < \sigma_0$, then the opposite is true. A change in σ is definitely accepted if the likelihood increases, that is,

$$\frac{\Phi}{2\sigma_0^2}\left[1-\gamma^{-2}\right] - n\ln\gamma > 0, \tag{A7}$$

where $\gamma = \sigma_{\rm prop}/\sigma_0$.

We are generating proposal σ values by perturbation of existing values: $\sigma_{\rm prop} = \sigma_0 + \delta\sigma$, where $|\delta\sigma/\sigma_0| \ll 1$. Thus $\gamma = 1 + x$, where $x = \delta\sigma/\sigma_0$.

Inserting this into eq. (A7), we accept the model if

$$\frac{\Phi}{2\sigma_0^2}\left[1-[1+x]^{-2}\right] - n\ln[1+x] > 0. \tag{A8}$$

Since $x \ll 1$, we can apply Taylor expansions and discard terms in x^2 and higher, giving

$$x\left(\frac{\Phi}{\sigma_0^2} - n\right) > 0 \tag{A9}$$

as the acceptance criterion.

Recognizing that $\sqrt{\Phi/n} =$ RMS, our acceptance criterion is equivalent to

Accept larger σ ($x > 0$) if $\sigma_0 <$ RMS

Accept smaller σ ($x < 0$) if $\sigma_0 >$ RMS.

Thus, the acceptance criterion drives the estimated uncertainty towards the empirical data error at that stage of the inversion.

Downloaded from https://academic.oup.com/gji/article-abstract/214/1/232/4959626
by University of California-SB user
on 30 May 2018