Electronic Seismologist

BayesBay: A Versatile Bayesian Inversion Framework Written in Python

Fabrizio Magrini*[1], Jiawen He[1], and Malcolm Sambridge[1]

Abstract

The lack of versatile tools for Bayesian inference presents a significant challenge to researchers in geophysics, who often resort to developing bespoke codes to address specific classes of inverse problems. In this study, we present BayesBay, a Python package for generalized transdimensional and hierarchical Markov chain Monte Carlo sampling. Leveraging object-oriented programming principles, BayesBay facilitates the definition of Bayesian sampling problems across a range of applications. This includes joint inversions of multiple data sets with different forward functions and unknown noise properties, as well as complex parameterizations involving multiple parameters with unknown dimensionality and/or spatially varying priors. We illustrate BayesBay from both a technical and a practical perspective. The first two applications are common in geophysics: a 2D tomographic problem and a joint inversion for the 1D subsurface structure. The third involves partition modeling and requires a sophisticated parameterization with two nested levels of transdimensionality. In all cases, BayesBay recovers known solutions, highlighting its potential to address a broad range of inverse problems.

Cite this article as Magrini, F., J. He, and M. Sambridge (2025). BayesBay: A Versatile Bayesian Inversion Framework Written in Python, *Seismol. Res. Lett.* **96**, 2052–2064, doi: 10.1785/0220240275.

Introduction

Inferring the Earth's interior structure from observations collected at the surface relies on geophysical inversion. Inverse problems in geophysics have traditionally been tackled by searching for a single "optimal" Earth model (e.g., Parker, 1994; Rawlinson and Sambridge, 2003; Menke, 2018). This is despite the fact that the observations themselves often do not require, or even allow, a unique solution to exist (Sambridge *et al.*, 2013). In general, different geophysical observables are sensitive to different—and often complementary—properties of the Earth's interior, making their combined use well suited to mitigating the nonuniqueness of the inverse problem at hand (e.g., Vozoff and Jupp, 1975; Afonso *et al.*, 2016). Nevertheless, without a comprehensive understanding of noise characteristics, standard (i.e., deterministic) methods require assigning weights to different observables, a cumbersome choice that is often difficult to justify through robust criteria. Furthermore, the interpretation of single "best-fitting" models can be problematic, particularly when they are influenced by a priori regularization or incorrect noise assumptions (Sambridge *et al.*, 2013; Rawlinson *et al.*, 2014).

Over the past few decades, probabilistic methods have grown in popularity in the geosciences because they are capable of overcoming these recognized challenges (e.g., Sambridge *et al.*, 2013). Central to these approaches is Bayes' theorem,

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

where $p(a|b)$ denotes the conditional probability of a given b. In words, Bayes' theorem states that the posterior probability of the model given the data, $p(\mathbf{m}|\mathbf{d})$, is proportional to the likelihood, $p(\mathbf{d}|\mathbf{m})$, scaled by our prior knowledge of the model, $p(\mathbf{m})$.

An effective method for Bayesian sampling is reversible-jump Markov chain Monte Carlo (RJ-MCMC, Green, 1995, 2003), which allows for inference on both model parameters and dimensionality. Transdimensional approaches such as RJ-MCMC are often described as naturally parsimonious (Malinverno, 2002; Sambridge *et al.*, 2006): akin to Occam's razor, they inherently favor simpler (i.e., low dimensional) models over complex ones. This property makes them suited to overcoming the curse of dimensionality. Moreover, hierarchical Bayesian inference offers a principled approach to data fusion, allowing the data themselves to determine the noise properties and avoiding the need to weigh different data classes prior to inversion (e.g., Malinverno and Briggs, 2004; Bodin, Sambridge, Rawlinson, and Arroucau, 2012).

1. Research School of Earth Sciences, The Australian National University, Acton, Australia, https://orcid.org/0000-0003-2417-2686 (FM); https://orcid.org/0009-0007-7754-825X (JH); https://orcid.org/0000-0003-2858-3175 (MS)

*Corresponding author: fabrizio.magrini@anu.edu.au

© Seismological Society of America

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/96/3/2052/7099576/srl-2024275.1.pdf
by Australian National University user

The technical advantages outlined earlier have established Bayesian methods, especially RJ-MCMC, as the state-of-the-art methods for a wide range of inverse problems encountered in geophysics (e.g., Piana Agostinetti and Malinverno, 2010; Bodin, Sambridge, Tkalčić, *et al.*, 2012; Dettmer *et al.*, 2014; Dosso *et al.*, 2014; Mustać and Tkalčić, 2016; Hawkins *et al.*, 2018, 2019; Izquierdo *et al.*, 2020). Such widespread success has prompted the development of software libraries tailored to specific modeling problems. Python examples include *MTfit* (Pugh and White, 2018), focused on seismic moment tensor characterization; *BayHunter* (Dreiling and Tilmann, 2019), enabling the inversion of receiver functions and/or surface-wave dispersion measurements; and *COMPOSTI* (Niskanen and Lähivaara, 2023), addressing seismic waveform inversions.

Although these specialized packages offer valuable functionalities, their application is confined to a relatively narrow range of uses. On the other hand, more generic Bayesian inference libraries—such as *emcee* (Foreman-Mackey *et al.*, 2013) and *PyMC3* (Salvatier *et al.*, 2016)—also exist; these provide powerful modeling tools but lack support for transdimensional and hierarchical sampling as well as for sophisticated parameterizations (e.g., involving spatially varying priors), which are essential for addressing a broad range of inverse problems in the geosciences.

This study introduces BayesBay, a Python package for transdimensional and hierarchical MCMC sampling. Originating from the Common Framework for Inference project (see Data and Resources), BayesBay aims to bridge the gap between the specificity of specialized libraries and the versatility of generalized frameworks. This is achieved with a modular architecture that offers both ready-to-use (high-level) and customizable (low-level) features. We illustrate BayesBay functionalities in the following sections, both from a technical perspective and by applying it to three complex inverse problems.

Method

To sample the posterior distribution, BayesBay implements an RJ-MCMC method (Green, 1995, 2003), which is a generalization of the Metropolis–Hastings algorithm (Metropolis *et al.*, 1953; Hastings, 1970) allowing for transdimensional parameterizations. The algorithm starts from an initial model $\mathbf{m}$ and proposes a new model $\mathbf{m}'$ based on a perturbative approach. The new model is then accepted (in which case, $\mathbf{m} \leftarrow \mathbf{m}'$) with probability

$$\alpha(\mathbf{m}'|\mathbf{m}) = \underbrace{\frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})}}_{\text{Likelihood ratio}} \underbrace{\frac{p(\mathbf{m}')}{p(\mathbf{m})}}_{\text{Prior ratio}} \underbrace{\frac{q(\mathbf{m}|\mathbf{m}')}{q(\mathbf{m}'|\mathbf{m})}}_{\text{Proposal ratio}} |\mathbf{J}|, \quad (2)$$

where it is understood that $\alpha = \min(1,\alpha)$. In equation (2), the Jacobian $\mathbf{J}$ of the transformation from $\mathbf{m}$ to $\mathbf{m}'$ accounts for the volume change in the parameter space under the proposed transformation. Through the forward operator $\mathbf{g}$, the likelihood expresses how well a model explains the data, and reads

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{C}_d|}} \exp\left\{\frac{-\Phi(\mathbf{m})}{2}\right\}, \quad (3)$$

where n denotes the size of the data vector, $\mathbf{C}_d$ denotes the data covariance matrix, and

$$\Phi(\mathbf{m}) = [\mathbf{g}(\mathbf{m}) - \mathbf{d}]^T \mathbf{C}_d^{-1} [\mathbf{g}(\mathbf{m}) - \mathbf{d}] \quad (4)$$

is the Mahalanobis distance between observations and model predictions.

In MCMC methods, the process of proposing a new model and deciding whether to accept it is repeated many times to build a sequence of models $\mathcal{M} = \{\mathbf{m}_t\}$, where t denotes the Markov chain iteration. In practice, a burn-in period typically precedes the generation of $\mathcal{M}$ to allow convergence of the Markov chain to a stationary distribution. After the burn-in period is completed, the subsequent iterations are used to populate $\mathcal{M}$, providing an approximation to the posterior distribution that can be used to estimate the sought model parameters (for discussions about convergence and other related topics see, e.g., Cowles and Carlin, 1996; Green and Hastie, 2009).

It is evident from the aforementioned discussion that the acceptance probability must be computed at each Markov chain iteration. In BayesBay, we analytically evaluate the ratios in equation (2) for the types of perturbations most common in the geophysical literature (e.g., Bodin, Sambridge, Tkalčić, *et al.*, 2012; Piana Agostinetti *et al.*, 2015; Izquierdo *et al.*, 2023). This approach avoids the need to separately calculate all the numerators and denominators before taking their ratios, which enhances computational efficiency. The specific form of equation (2) depends on the class of proposed perturbation. In BayesBay, these include the perturbation of a model parameter, the addition (birth) or removal (death) of model parameters, and the perturbation of the data noise. Detailed mathematical expressions for each type of perturbation are provided in the Appendix.

BayesBay API and Workflow

BayesBay leverages object-oriented programming principles to ensure that each component of the inversion—such as the discretization, the physical properties to be inferred, and the data noise—is a self-contained unit. This design allows solution of a wide range of inverse problems and facilitates their parameterization. Moreover, it enables seamless integration of various forward solvers and data sets, promoting the joint use of multiple data types in geophysical inversions.

BayesBay provides accessibility and versatility through a modular architecture comprising 27 Python classes (Fig. 1). Table 1 summarizes the tutorials available in the online documentation, with useful links included in Data and Resources. Inspired by examples from statistics and geophysics, these tutorials illustrate practical applications of our library in posterior sampling problems. A typical workflow involves the following steps.

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/96/3/2052/7099576/srl-2024275.1.pdf
by Australian National University user

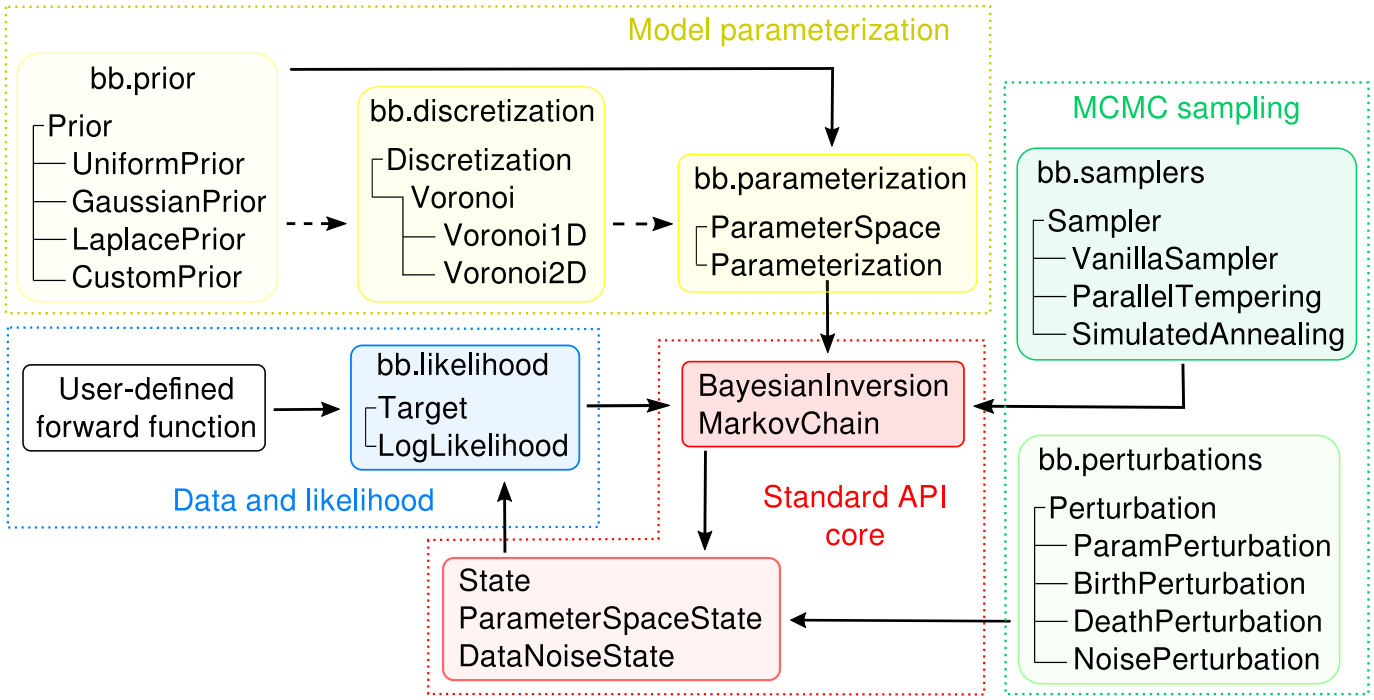


Figure 1. Workflow diagram showing all modules and classes of the BayesBay standard API. A typical BayesBay workflow involves defining class instances of *Parameterization* (encapsulating the prior) and *LogLikelihood*. These are used by *BayesianInversion* to sample the posterior. At each Markov chain iteration, a new model is proposed via a perturbative approach and stored in a BayesBay *State*. The likelihood is then evaluated through one or more user-defined forward functions, and the proposed model is accepted with a probability given by equation (2). The color version of this figure is available only in the electronic edition.

Parameterizing the inverse problem

The process begins by defining the prior probability distribution for the free parameters of the inverse problem via the BayesBay *prior* module. For instance, in a gravity inversion, density would be treated as a free parameter (e.g., Izquierdo *et al.*, 2023); shear-wave velocity and radial anisotropy could be set as unknowns in the inversion of surface-wave dispersion data (e.g., Taira and Yoshizawa, 2020); and so on. The dimensionality of such free parameters can either be treated as a constant or as unknown, with the latter corresponding to a transdimensional parameterization.

In several geoscientific inverse problems, a somewhat special type of parameter that may be treated as unknown is the discretization underlying the free—and often physically meaningful—parameters described earlier. For example, in the aforementioned shear-velocity inversion, the Earth could be discretized using a transdimensional 1D Voronoi tessellation, where each Voronoi cell represents a subsurface layer characterized by a specific V_S value (e.g., Piana Agostinetti and Malinverno, 2010; Bodin, Sambridge, Tkalčić, *et al.*, 2012). The BayesBay *discretization* module facilitates these applications by linking the number of basis functions defining the discretization (e.g., the number of Voronoi cells) to the dimensionality of the associated free parameters (e.g., the V_S value associated with each cell).

The thus-defined prior distributions should be used in BayesBay to initialize a *Parameterization* class instance, which enables the calculation of the prior ratio in equation (2) at each Markov chain iteration.

Defining the likelihood

To facilitate joint inversions involving an arbitrary number of data types, BayesBay provides the *Target* and *LogLikelihood* classes. Whereas the *Target* class stores the observed data $\mathbf{d}$ in a specialized Python object, the *LogLikelihood* class evaluates the model likelihood at each Markov chain iteration. Users can initialize *LogLikelihood* with one or more *Target* instances, each associated with a distinct forward function, or alternatively, by providing a function that directly returns the log likelihood for a given model. In the former case, the *Target* class allows users to either specify the data covariance matrix prior to the inversion or treat it as unknown (further details in the Appendix), and the likelihood is then calculated via equation (3).

Sampling the posterior

The class instances of *Parameterization* and *LogLikelihood* allow sampling the posterior through the *BayesianInversion* class. Posterior sampling in BayesBay can be performed using an arbitrary number of Markov chains, which can be run in parallel on both standard computers and computer clusters. Users need to decide on the total number of Markov chain iterations, the number of burn-in iterations, and the interval between model saves after the burn-in period.

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/96/3/2052/7099576/srl-2024275.1.pdf
by Australian National University user

TABLE 1
Overview of Tutorials Available in the BayesBay Documentation, Categorized by Sampling Problem Class (in Bold)

Type	Features	Parameters	Discretization
QuickStart tutorial	KN	1	None
Regression			
Curve fitting	KN	4	None
	UN	5	
Partition modeling	UN	9	Voronoi 1D
	UN, TD	6–40	
Clustering			
Gaussian mixture modeling	UN	9	None
	UN, TD	5–23	
Imaging			
Dispersion-curve inversion	Rayleigh: UN, TD	10–32	Voronoi 1D
	Rayleigh and Love: UN, TD	11–33	
Surface-wave tomography	UN, TD	122–1502	Voronoi 2D

KN, known noise; TD, transdimensional parameterization; UN, unknown noise.

As previously mentioned, each Markov chain iteration involves the proposal of a new model by randomly perturbing the current one and computing the acceptance probability (equation 2). In BayesBay, this process is managed by the classes in the *perturbations* module (see Fig. 1), which handle the dimensionality and the numerical values of model parameters, as well as the data noise. Notably, the *Perturbation* class also supports the implementation of arbitrary types of perturbations, leaving the calculation of the acceptance probability to the user in such cases.

Finally, relevant to inverse problems characterized by complex posterior distributions, BayesBay facilitates the implementation of sophisticated sampling criteria via the *Sampler* class. A BayesBay *Sampler* allows, for example, altering the standard deviation of the Gaussians used to perturb the current model after or prior to each Markov chain iteration. It provides control over the statistical temperature of the chains, enabling likelihood-tempering approaches such as parallel tempering (Ray *et al.*, 2013; Sambridge, 2014), for which native support is provided. In addition, a BayesBay *Sampler* enables advanced behaviors such as inter-chain communication, a process that temporarily interrupts parallel computation to exchange information between different chains during the inversion.

Example Applications

Having discussed the most relevant technical aspects of BayesBay, we now demonstrate its use by tackling three Bayesian sampling tasks. The first two applications involve inverse problems commonly found in geophysics, and the third addresses a partition modeling problem. All examples are approached using transdimensional parameterizations and by treating the data noise as unknown.

Example 1: seismic tomography

As a first example, we use BayesBay to estimate lateral variations in Rayleigh-wave velocity across Australia at the surface-wave period of 5 s. We use the data set assembled by Magrini *et al.* (2023) based on recordings of seismic ambient noise, which consists of n = 15,661 measurements of average inter-station phase-velocity v (Fig. 2a); these define the data vector $\mathbf{d} = [s_1, s_2, \ldots, s_n]^T$, where $s_i = v_i^{-1}$ denotes the average inter-station slowness.

We parameterize the data covariance matrix $\mathbf{C}_d = \sigma^2\mathbf{I}$ through the noise standard deviation σ, which is assigned a uniform prior distribution within the range 0–0.01 s/km. We discretize the Earth's surface using a transdimensional Voronoi tessellation, with each Voronoi cell corresponding to a phase-velocity value defined by a uniform prior between 2 and 4 km/s. The model parameters to be inferred are therefore $\mathbf{m} = [\sigma, k, \mathbf{c}, \mathbf{v}]^T$, where $\mathbf{c}$ and $\mathbf{v}$ are the k-vectors whose entries represent the Voronoi nuclei coordinates and their velocities, respectively.

Similar to Magrini *et al.* (2023), we assume that surface-wave propagation between two points on the Earth's surface occurs along the great-circle path connecting them. Under this assumption, in discrete form, the forward equation for the average slowness associated with the ith station pair reads

$$s_i = \frac{1}{L_i}\sum_{j=1}^{k}\frac{l_j}{v_j}, \tag{5}$$

where L_i denotes the great-circle distance, l_j/L_i is the fraction of the great-circle path crossing the jth Voronoi cell, and v_j is the cell's phase velocity.

In practice, obtaining forward predictions $\mathbf{d}_{\text{pred}}$ presents two main challenges (e.g., Sambridge and Guđmundsson, 1998; Debayle and Sambridge, 2004). First, there are no analytical expressions for the intersections of a great-circle path with Voronoi cell boundaries, complicating the calculation of l_j in equation (5). Second, such intersections must be recomputed whenever the discretization is perturbed, which becomes increasingly expensive as the number of station pairs increases.

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/96/3/2052/7099576/srl-2024275.1.pdf
by Australian National University user

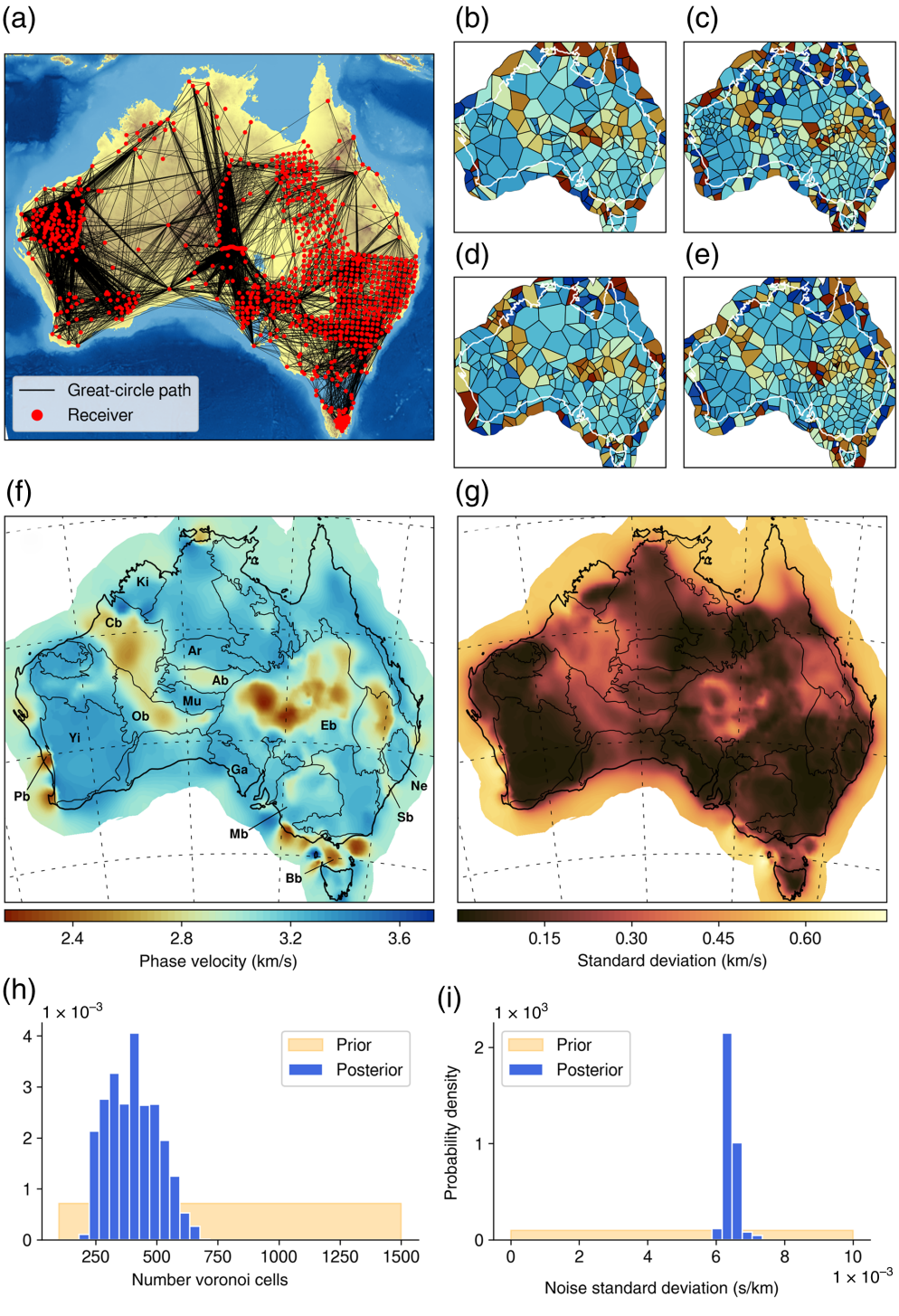


Figure 2. Results of the transdimensional surface-wave tomography. Receivers and ray-path coverage used to map phase-velocity variations across Australia are shown in panel (a). Panels (b–e) display four randomly selected posterior samples. Inferred maps of average phase velocity and standard deviation are shown in panels (f) and (g), respectively. Panels (h,i) present histograms of the number of Voronoi cells in each saved model and the data noise standard deviation σ. Key to marked features: Amadeus Basin (Ab), Arunta Province (Ar), Bass Basin (Bb), Canning Basin (Cb), Eromanga Basin (Eb), Gawler Craton (Ga), Kimberley Block (Ki), Murray Basin (Mb), Musgrave Province (Mu), New England Orogen (Ne), Officer Basin (Ob), Perth Basin (Pb), Sydney Basin (Sb), Yilgarn Craton (Yi). The color version of this figure is available only in the electronic edition.

To address these issues, we define an orthonormal grid finely covering Australia with 11,916 pixels of 0.3° × 0.3°. This allows us to compute, prior to the inversion, the 15,661 × 11,916 matrix **A** whose *ij*th entry is the fraction of the *i*th great-circle path crossing the *j*th grid cell (Magrini *et al.*, 2022). In turn, the resulting matrix enables data predictions $\mathbf{d}_{\text{pred}} = \mathbf{Ax}$, where the *j*th entry of **x** corresponds to the slowness of the Voronoi nucleus hosting the *j*th grid cell, which can be efficiently determined at each Markov chain iteration through a kd-tree algorithm (Bentley, 1975).

Following the approach described earlier, we sample the posterior distribution via 64 Markov chains. We run each chain for 450,000 iterations and save one model every 300 starting from the 150,000th iteration. This results in 64,000 samples distributed according to the posterior, four of which are shown in Figure 2b–e.

We use the collected samples to calculate statistics such as the average phase velocity and standard deviation across Australia (Fig. 2f,g). The inferred uncertainty reflects lateral variations in ray-path coverage, with low standard deviations corresponding to areas of dense coverage and vice versa. Conversely, the inferred phase-velocity map reflects known geologic features of Australia (e.g., Saygin and Kennett, 2010; Salmon *et al.*, 2013; Kennett *et al.*, 2018), exhibiting relatively low velocities in the sedimentary basins and higher velocities in the cratonic blocks (Fig. 2f). Our phase-velocity map also aligns with those in Bodin, Sambridge, Rawlinson, and Arroucau

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/96/3/2052/7099576/srl-2024275.1.pdf
by Australian National University user

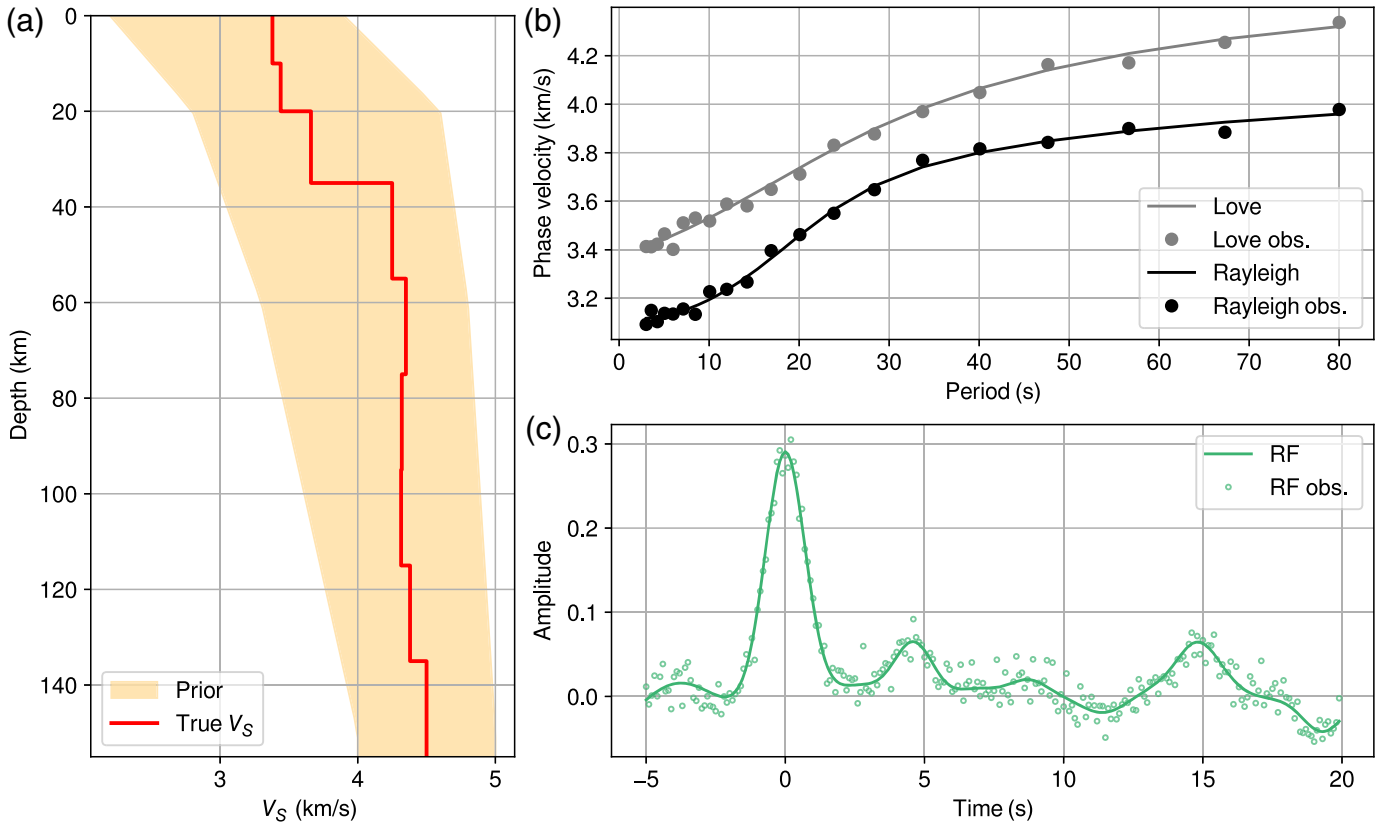


Figure 3. Noisy receiver function (RF) and surface-wave dispersion data (dots in panels b and c), along with their noiseless counterparts (continuous lines), as obtained from the subsurface model in panel (a). Note the depth-dependent prior for V_S (orange shading) used in the Bayesian inversion. The color version of this figure is available only in the electronic edition.

(2012) and Magrini *et al.* (2023), where differences can be attributed to the different data set used in the former and the different method used in the latter. Overall, these results indicate the robustness of our Bayesian sampling approach.

Example 2: joint inversion of receiver functions and surface-wave dispersion

In this section, we demonstrate how BayesBay can jointly invert multiple data sets by tackling a synthetic inverse problem. Specifically, we aim to recover the shear-wave velocity of a 1D stratified Earth model consisting of eight layers overlying a homogeneous half-space (Fig. 3a). Similar to Bodin, Sambridge, Tkalčić, *et al.* (2012), we assume V_P and density ρ are known and given by the relationships $V_P = 1.77V_S$ and $\rho = 0.32V_P + 0.77$ (Berteussen, 1977).

We use the thus-defined subsurface model to generate a synthetic data set, consisting of two surface-wave dispersion curves (one for Rayleigh and one for Love) and a receiver function. Uncorrelated random noise is added to these synthetic observations to define the data vector $\mathbf{d} = [\mathbf{d}_R,\mathbf{d}_L,\mathbf{d}_{RF}]^T$, where R, L, and RF stand for Rayleigh, Love, and receiver function, respectively (Fig. 3b,c). (Noise correlation is not considered here for simplicity, although it is supported by the BayesBay API; see the Appendix.)

Similar to the previous section, we parameterize the data covariance matrices of $\mathbf{d}_R$, $\mathbf{d}_L$, and $\mathbf{d}_{RF}$ using the standard deviations σ_R, σ_L, and σ_{RF}, each of which is assigned a uniform prior distribution. We discretize the subsurface through a transdimensional Voronoi tessellation, resulting in the model $\mathbf{m} = [\sigma_R,\sigma_L,\sigma_{RF},k,\mathbf{c},\mathbf{v}]^T$, where this time the ith entry of the k-vector $\mathbf{v}$ corresponds to the shear velocity of the ith Voronoi cell.

To demonstrate BayesBay capability to accommodate sophisticated parameterizations, we use a depth-dependent prior for V_S (Fig. 3a), reflecting a typical geophysical scenario where detailed subsurface information is available prior to the inversion. Compared with prior distributions that are uniform in depth, this approach reduces the parameter space and can accelerate the decrease in misfit during the burn-in period, potentially speeding up MCMC convergence. This concept is illustrated in Figure 4.

We run 24 Markov chains in parallel for 600,000 iterations, saving every 150th model after an initial burn-in of 150,000 iterations. Visual inspection of the results in Figure 5 indicates that the true model has been satisfactorily recovered. Consistent with Bodin, Sambridge, Tkalčić, *et al.* (2012), the posterior median successfully retrieves both the shear-wave velocity of each layer in the true subsurface model (Fig. 5g) and the major discontinuities (Fig. 5h). On the other hand, the synthetic data set used in

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/96/3/2052/7099576/srl-2024275.1.pdf
by Australian National University user

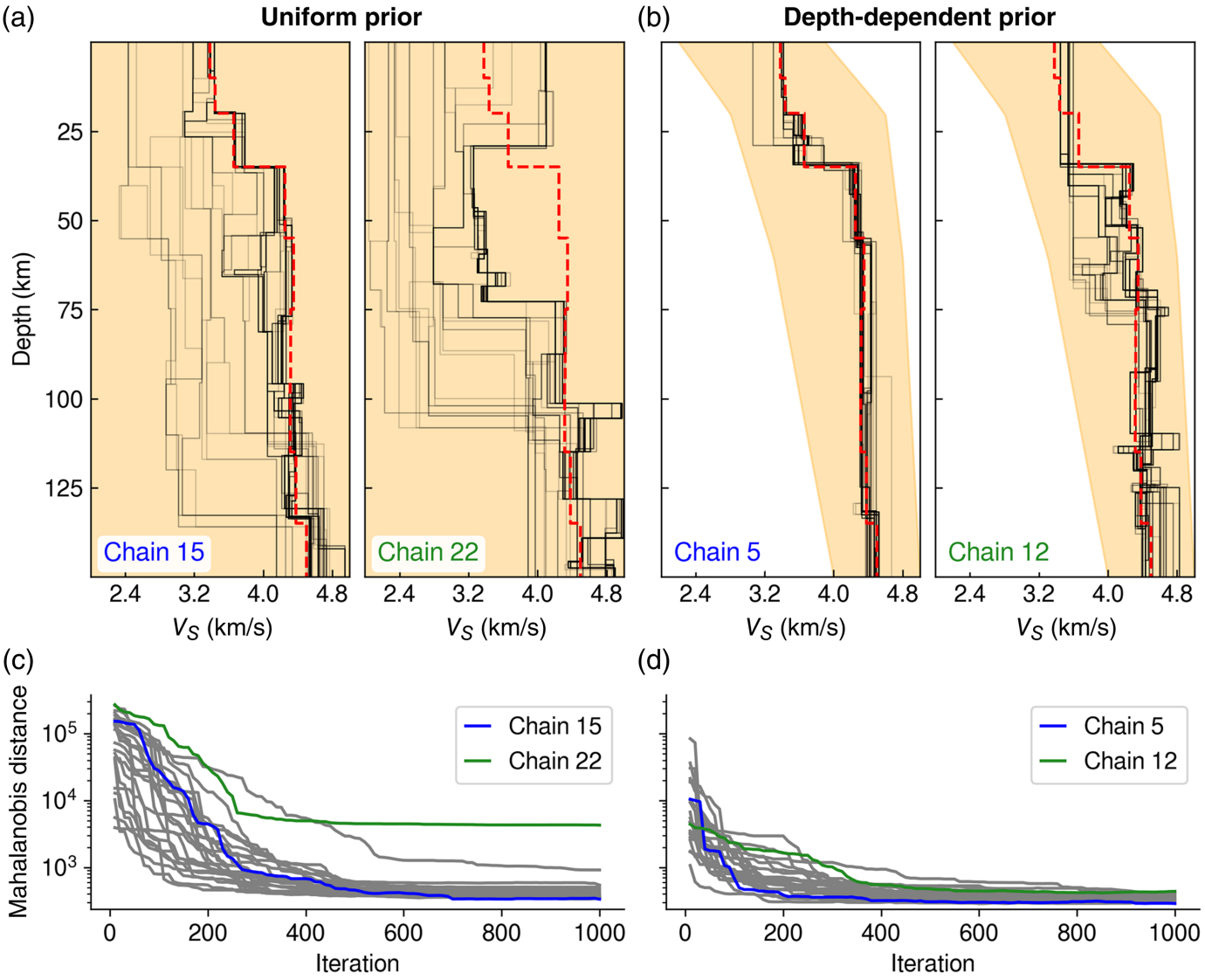


Figure 4. First 1000 iterations of the burn-in period of a trans-dimensional Bayesian inversion for V_S. The results are presented for both (a) a uniform prior defined between 2 and 5 km/s and (b) a depth-dependent prior, indicated by the orange shading. (c, d) Mahalanobis distance Φ is shown as a function of iteration for 24 Markov chains, with colored lines highlighting two selected chains. Panels (a,b) show the shear-velocity models sampled by these highlighted chains, along with the ground truth (red-dashed lines). Knowledge of true data noise is assumed to ensure comparability of Φ across all chains. Note the larger values of Φ for some chains when using a prior uniform in depth. The color version of this figure is available only in the electronic edition.

this experiment does not seem to adequately constrain the three discontinuities in the 75–115 km depth range, likely because of the small velocity gradients in this region. This is reflected in the mode of the inferred number of layers, which is two layers fewer than in the true model (Fig. 5l). As discussed in the Introduction, this observation aligns with the parsimonious nature of RJ-MCMC (e.g., Malinverno, 2002; Sambridge *et al.*, 2006).

Finally, we find the inferred data noise standard deviations to be generally close to the true values ($\sigma_{\mathrm{R}}^{\mathrm{true}} = \sigma_{\mathrm{L}}^{\mathrm{true}} = 0.020$ s/km and $\sigma_{\mathrm{RF}}^{\mathrm{true}} = 0.015$). With uncertainty reported at the two-sigma level, we obtain a posterior mean of 0.027 ± 0.010 s/km for σ_{R} and σ_{L} (Fig. 5i,j) and 0.015 ± 0.002 for σ_{RF} (Fig. 5k), which is particularly well recovered.

Example 3: partition modeling

The final example application of BayesBay involves a regression problem often referred to as partition modeling (e.g., Denison *et al.*, 2002; Stephenson *et al.*, 2004; Gallagher *et al.*, 2011). Specifically, we consider the piecewise polynomial function

$$f(x) = \begin{cases} 0 & \text{if } -10 \le x \le -7 \\ 1.5x + 5 & \text{if } -7 < x \le -3 \\ x^2 & \text{if } -3 < x \le 3 \\ -1.5x + 5 & \text{if } 3 < x \le 7 \\ 0 & \text{if } 7 < x \le 10 \end{cases}. \tag{6}$$

Our goal is to infer $f(x)$, without assuming a specific number of segments, from noisy observations $\mathbf{d} = [d_1, d_2, \ldots, d_{200}]^T$, where $d_i \sim \mathcal{N}(f(x_i), 1.5)$ (Fig. 6b).

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/96/3/2052/7099576/srl-2024275.1.pdf
by Australian National University user

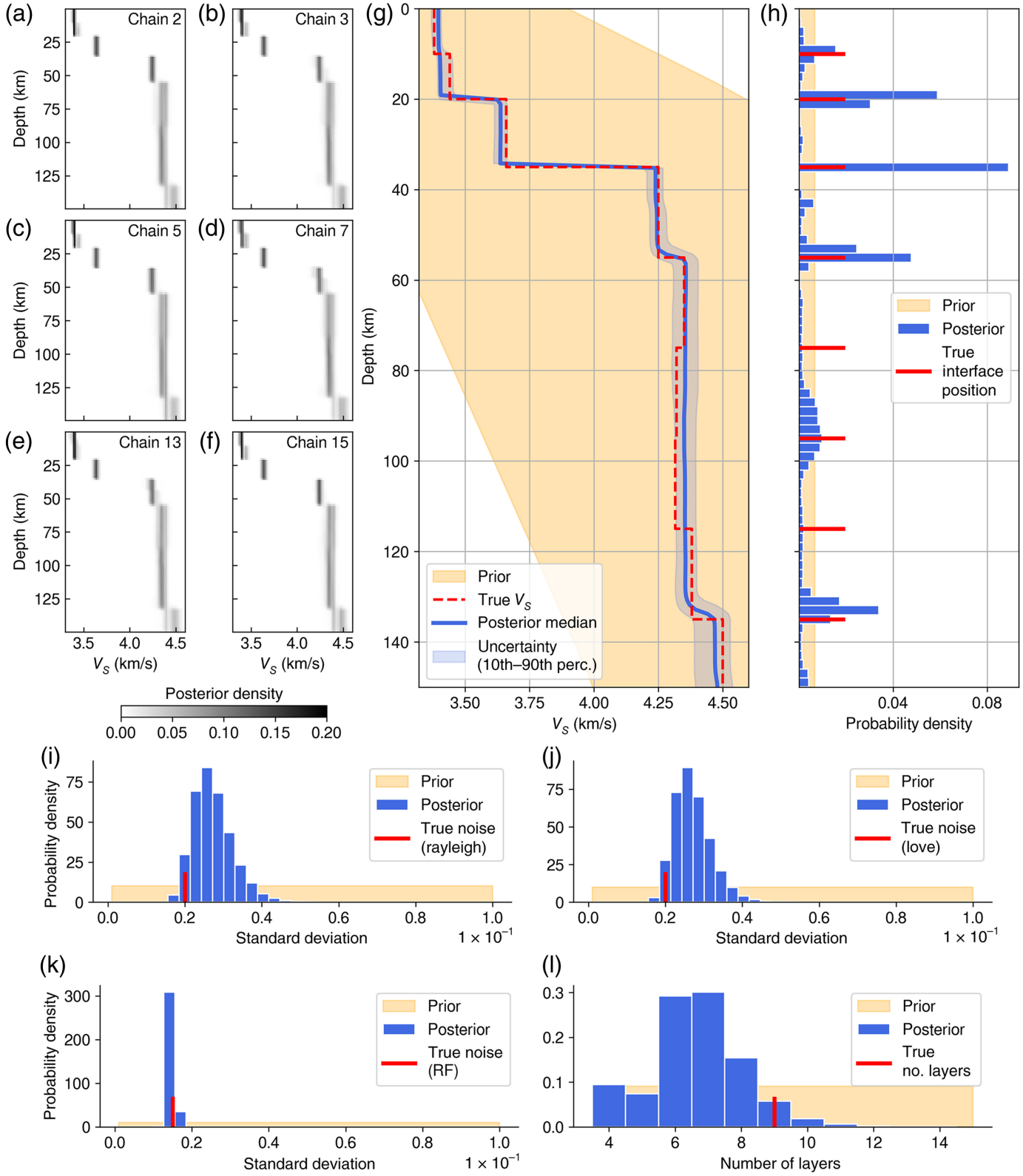
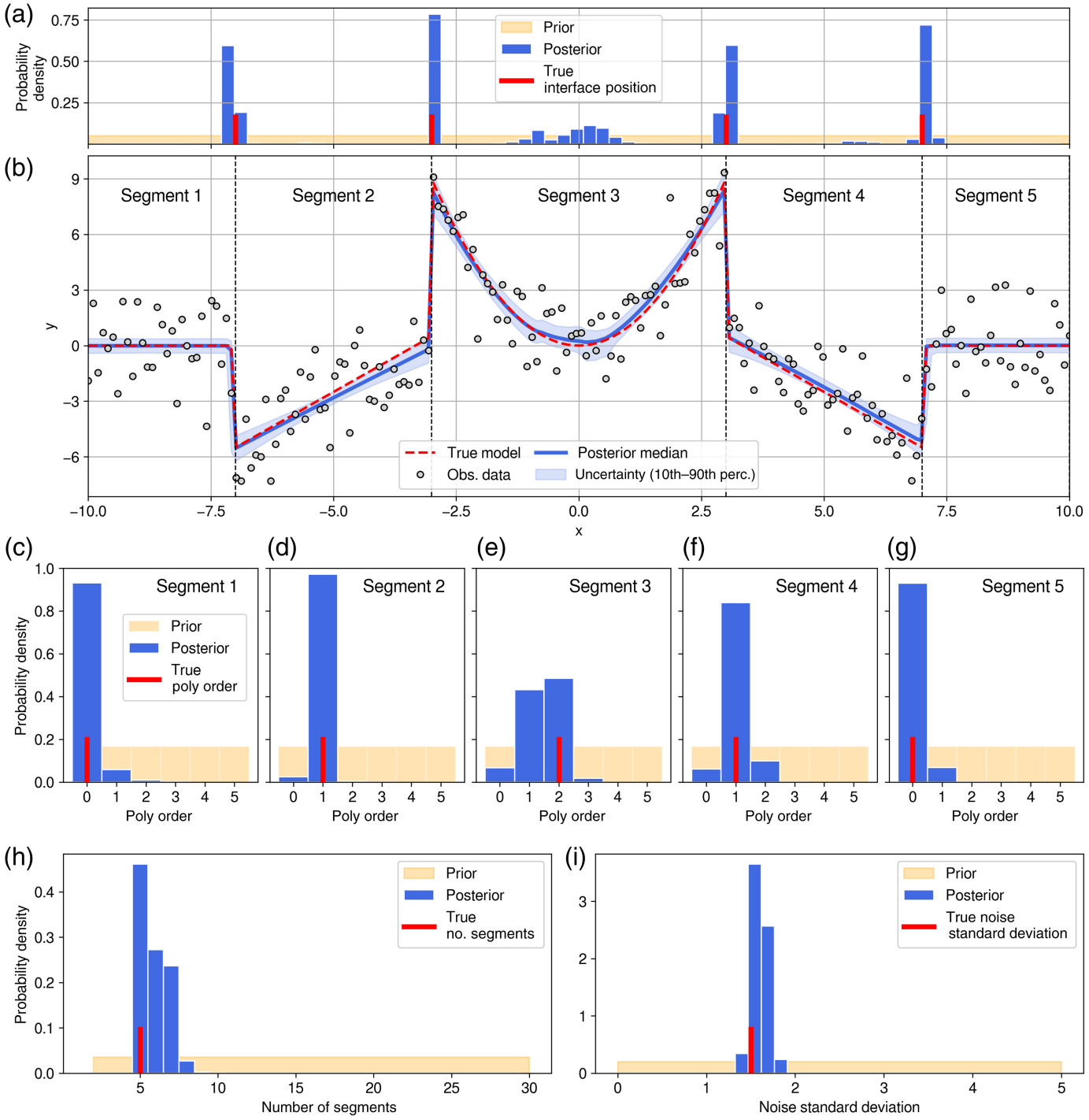


Figure 5. Joint inversion results based on surface-wave dispersion and receiver-function data. Panels (a–f) show inferred posterior distributions from six randomly selected Markov chains. The posterior median (blue line) of all 72,000 collected samples, along with the 10th–90th percentile range (blue shading), is presented in panel (g). Panel (h) displays the inferred probability of occurrence of a discontinuity as a function of depth, and histograms of inferred data noise and number of layers are shown in panels (i–l). The color version of this figure is available only in the electronic edition.

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/96/3/2052/7099576/srl-2024275.1.pdf
by Australian National University user



This seemingly simple problem is, in fact, relatively complex when we consider that each segment of $f(x)$ consists of a polynomial $P(x) = \sum_{j=0}^{n} a_j x^j$ of variable degree n and coefficients a_j. Addressing this situation necessitates a parameterization with two "nested" levels of transdimensionality: the first to model the number and positions of segment interfaces; the second to accommodate the variable degrees of the polynomials in each segment while determining their coefficients (Fig. 7).

The strategy described earlier results in the parameters $\mathbf{m} = [\sigma, k, \mathbf{c}, (n_1, \mathbf{a}_1), (n_2, \mathbf{a}_2), \ldots, (n_k, \mathbf{a}_k)]^T$, where σ denotes the noise standard deviation, the k-vector of Voronoi nuclei $\mathbf{c}$

Figure 6. Results of the Bayesian regression used to recover the piecewise polynomial function (equation 6). Panel (b) displays the true function (red-dashed line) and the observed data (gray dots) along with the inferred posterior median (blue line) and the 10th–90th percentile range (light-blue shading). The posterior distributions of segment interfaces, the number of segments, and data noise are shown in panels (a), (h), and (i), respectively. Histograms of inferred polynomial degrees in each segment are reported in panels (c–g). The color version of this figure is available only in the electronic edition.

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/96/3/2052/7099576/srl-2024275.1.pdf
by Australian National University user

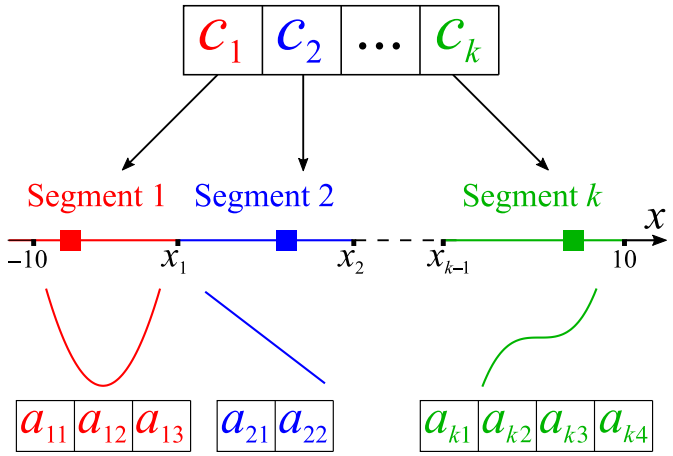


Figure 7. Parameterization used to infer the piecewise polynomial (equation 6). The Voronoi nuclei $c_1, c_2, \ldots, c_k$ identify k segments on the x axis. Within the ith segment, the coefficients $\mathbf{a}_i$ define a polynomial. Both the number of nuclei and the polynomial degrees are treated as unknown. The color version of this figure is available only in the electronic edition.

models the segments, and the $(n_i + 1)$-vector $\mathbf{a}_i$ defines the polynomial in the ith segment. We use a uniform prior for each entry of $\mathbf{a}_i$, ranging between −10 and 10.

We sample the posterior distribution using 10 Markov chains that are run in parallel for 800,000 iterations, saving every 100th model starting from the 400,000th iteration. The results, summarized in Figure 6, point to the robustness of our approach. First, the positions of the interfaces are retrieved convincingly well (Fig. 6a), and the posterior median of the predicted data closely approximates $f(x)$ (Fig. 6b). Second, the mode of the inferred distribution of the number of segments coincides with the ground truth, and the average inferred data noise, 1.603 ± 0.170 (two-sigma), recovers well the true value of 1.5 (Fig. 6h,i). Finally, we find that the mode of the histograms of inferred polynomial degrees corresponds to the ground truth in all segments (Fig. 6c–g).

Conclusions

We have presented BayesBay from both a technical perspective and by applying it to three distinct inverse problems, chosen to demonstrate the broad range of functionalities possible. All problems were addressed using transdimensional parameterizations and treating the data noise as unknown.

The first application involved a 2D tomographic problem, tackled in a manner that is arguably most widespread in the Bayesian geophysical-inversion literature (e.g., Bodin and Sambridge, 2009; Crowder *et al.*, 2019): constraining the medium's physical properties through a single set of observations and a uniform prior independent of location.

In the second application, we used a more sophisticated parameterization to address a 1D geophysical inverse problem. Specifically, we used a depth-dependent prior probability, reflecting a typical scenario for which some kind of detailed Earth knowledge is available prior to the inversion (e.g., Yuan and Bodin, 2018). Compared with simpler—and nonetheless more widespread—approaches, discretization-informed prior probabilities reduce the parameter space and may help accelerate the decrease in misfit during the burn-in period. BayesBay streamlines the setup of such priors, making it easy for users to incorporate their knowledge of the inverse problem.

Another strength of BayesBay, leveraged in the second example, is its capability to seamlessly combine different data sets, each possibly associated with a different forward operator and unknown noise properties. This feature is particularly relevant considering the impressive wealth of geophysical data that now enables joint inversions across a range of applications.

The final application involved a toy problem that required a parameterization with two different, nested levels of transdimensionality. Although we selected this example to further demonstrate BayesBay capability to address complex inverse problems, our approach is also relevant to practical partition modeling problems in the geosciences (e.g., Gallagher *et al.*, 2011).

In all examples, RJ-MCMC sampling through BayesBay allowed us to recover known solutions, demonstrating its robustness and versatility in a wide range of inverse problems.

Data and Resources

BayesBay is open-source and hosted on GitHub at https://github.com/fmagrini/bayes-bay, with documentation available at https://bayes-bay.readthedocs.io/en/latest. In Example I, forward predictions of average interstation slowness were obtained using the *SeisLib* Python library (Magrini *et al.*, 2022). In Example II, surface-wave dispersion curves and receiver functions were calculated using the *disba* (https://pypi.org/project/disba) and *espresso* (https://pypi.org/project/espresso) Python libraries, respectively. The Common Framework for Inference (CoFI) project is available at https://inlab.au/cofi. All websites were last accessed in October 2024.

Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest recorded.

Acknowledgments

The authors are grateful to Alison Malcolm for constructive comments on an earlier version of this article. The authors thank the Commonwealth Scientific Industrial Research Organization Future Science Platform for Deep Earth Imaging for support. The authors acknowledge funding from the Australian Research Council through a Discovery Project (Grant Number DP200100053) and the AuScope Opportunity Fund.

References

Afonso, J. C., M. Moorkamp, and J. Fullea (2016). Imaging the lithosphere and upper mantle: Where we are at and where we are going, in *Integrated Imaging of the Earth: Theory and Applications*, Geophysical Monograph Series, Vol. 218, American Geophysical Union and John Wiley & Sons, Washington, D.C., 191–218.

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/96/3/2052/7099576/srl-2024275.1.pdf
by Australian National University user

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching, *Commun. ACM* **18,** no. 9, 509–517.

Berteussen, K.-A. (1977). Moho depth determinations based on spectral-ratio analysis of NORSAR long-period P waves, *Phys. Earth Planet. In.* **15,** no. 1, 13–27.

Bodin, T., and M. Sambridge (2009). Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.* **178,** no. 3, 1411–1436.

Bodin, T., M. Sambridge, N. Rawlinson, and P. Arroucau (2012). Transdimensional tomography with unknown data noise, *Geophys. J. Int.* **189,** no. 3, 1536–1556.

Bodin, T., M. Sambridge, H. Tkalčić, P. Arroucau, K. Gallagher, and N. Rawlinson (2012). Transdimensional inversion of receiver functions and surface wave dispersion, *J. Geophys. Res.* **117,** no. B2, doi: 10.1029/2011JB008560.

Cowles, M. K., and B. P. Carlin (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review, *J. Am. Stat. Assoc.* **91,** no. 434, 883–904.

Crowder, E., N. Rawlinson, S. Pilia, D. Cornwell, and A. Reading (2019). Transdimensional ambient noise tomography of Bass Strait, southeast Australia, reveals the sedimentary basin and deep crustal structure beneath a failed continental rift, *Geophys. J. Int.* **217,** no. 2, 970–987.

Debayle, E., and M. Sambridge (2004). Inversion of massive surface wave data sets: Model construction and resolution assessment, *J. Geophys. Res.* **109,** no. B2, doi: 10.1029/2003JB002652.

Denison, D. G., C. C. Holmes, B. K. Mallick, and A. F. Smith (2002). *Bayesian Methods for Nonlinear Classification and Regression*, Vol. 386, John Wiley & Sons, Hoboken, New Jersey.

Dettmer, J., R. Benavente, P. R. Cummins, and M. Sambridge (2014). Trans-dimensional finite-fault inversion, *Geophys. J. Int.* **199,** no. 2, 735–751.

Dosso, S. E., J. Dettmer, G. Steininger, and C. W. Holland (2014). Efficient trans-dimensional Bayesian inversion for geoacoustic profile estimation, *Inverse Prob.* **30,** no. 11, 114018.

Dreiling, J., and F. Tilmann (2019). BayHunter-McMC transdimensional Bayesian inversion of receiver functions and surface wave dispersion, doi: 10.5880/GFZ.2.4.2019.001.

Foreman-Mackey, D., D. W. Hogg, D. Lang, and J. Goodman (2013). emcee: The MCMC hammer, *Publ. Astron. Soc. Pac.* 125**,** no. 925, 306.

Gallagher, K., T. Bodin, M. Sambridge, D. Weiss, M. Kylander, and D. Large (2011). Inference of abrupt changes in noisy geochemical records using transdimensional changepoint models, *Earth Planet. Sci. Lett.* **311,** nos. 1/2, 182–194.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika* **82,** no. 4, 711–732.

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

Green, P. J., and D. I. Hastie (2009). Reversible jump MCMC, *Genetics* **155,** no. 3, 1391–1403.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57,** no. 1, 97–109.

Hawkins, R., and M. Sambridge (2015). Geophysical imaging using trans-dimensional trees, *Geophys. J. Int.* **203,** no. 2, 972–1000.

Hawkins, R., T. Bodin, M. Sambridge, G. Choblet, and L. Husson (2019). Trans-dimensional surface reconstruction with different classes of parameterization, *Geochem. Geophys. Geosys.* **20,** no. 1, 505–529.

Hawkins, R., R. C. Brodie, and M. Sambridge (2018). Trans-dimensional Bayesian inversion of airborne electromagnetic data for 2D conductivity profiles, *Explor. Geophys.* **49,** no. 2, 134–147.

Izquierdo, K., V. Lekić, and L. G. Montési (2020). A Bayesian approach to infer interior mass anomalies from the gravity data of celestial bodies, *Geophys. J. Int.* **220,** no. 3, 1687–1699.

Izquierdo, K., V. Lekić, and L. G. Montési (2023). An object-oriented Bayesian gravity inversion scheme for inferring density anomalies in planetary interiors, *Earth Space Sci.* **10,** no. 7, e2023EA002853, doi: 10.1029/2023EA002853.

Kennett, B., R. Chopping, and R. Blewett (2018). *The Australian Continent: A Geophysical Synthesis*, The Australian National University in Canberra, Australian Capital Territory, Australia.

Magrini, F., E. Kästle, S. Pilia, N. Rawlinson, and L. De Siena (2023). A new shear-velocity model of continental Australia based on multiscale surface-wave tomography, *J. Geophys. Res.* **128,** no. 7, e2023JB026688, doi: 10.1029/2023EA002853.

Magrini, F., S. Lauro, E. Kästle, and L. Boschi (2022). Surface-wave tomography using SeisLib: a Python package for multiscale seismic imaging, *Geophys. J. Int.* **231,** no. 2, 1011–1030.

Malinverno, A. (2002). Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.* **151,** no. 3, 675–688.

Malinverno, A., and V. A. Briggs (2004). Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes, *Geophysics* **69,** no. 4, 1005–1016.

Menke, W. (2018). *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, London, United Kingdom.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21,** no. 6, 1087–1092.

Mustać, M., and H. Tkalčić (2016). Point source moment tensor inversion through a Bayesian hierarchical model, *Geophys. J. Int.* **204,** no. 1, 311–323.

Niskanen, M., and T. Lähivaara (2023). COMPOSTI: A Python-based program for seismic trans-dimensional inversion, *SoftwareX* **21,** 101298.

Parker, R. L. (1994). *Geophysical Inverse Theory*, Vol. 1, Princeton University Press, Princeton, New Jersey.

Piana Agostinetti, N., and A. Malinverno (2010). Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.* **181,** no. 2, 858–872.

Piana Agostinetti, N., G. Giacomuzzi, and A. Malinverno (2015). Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.* **201,** no. 3, 1598–1617.

Pugh, D., and R. White (2018). MTfit: A Bayesian approach to seismic moment tensor inversion, *Seismol. Res. Lett.* **89,** no. 4, 1507–1513.

Rawlinson, N., and M. Sambridge (2003). Seismic traveltime tomography of the crust and lithosphere, *Adv. Geophys.* **46,** 81–199.

Rawlinson, N., A. Fichtner, M. Sambridge, and M. K. Young (2014). Seismic tomography and the assessment of uncertainty, *Adv. Geophys.* **55,** 1–76.

Ray, A., D. L. Alumbaugh, G. M. Hoversten, and K. Key (2013). Robust and accelerated Bayesian inversion of marine controlled-source electromagnetic data using parallel tempering, *Geophysics* **78,** no. 6, E271–E280.

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/96/3/2052/7099576/srl-2024275.1.pdf
by Australian National University user

Salmon, M., B. Kennett, and E. Saygin (2013). Australian seismological reference model (AuSREM): Crustal component, *Geophys. J. Int.* **192,** no. 1, 190–206.

Salvatier, J., T. V. Wiecki, and C. Fonnesbeck (2016). Probabilistic programming in Python using PyMC3, *Peer J. Comput. Sci.* **2,** e55, doi: 10.7717/peerj-cs.55.

Sambridge, M. (2014). A parallel tempering algorithm for probabilistic sampling and multimodal optimization, *Geophys. J. Int.* **196,** no. 1, 357–374.

Sambridge, M., and Ó. Guđmundsson (1998). Tomographic systems of equations with irregular cells, *J. Geophys. Res.* **103,** no. B1, 773–781.

Sambridge, M., T. Bodin, K. Gallagher, and H. Tkalčić (2013). Transdimensional inference in the geosciences, *Phil. Trans. Roy. Soc. A* **371,** no. 1984, doi: 10.1098/rsta.2011.0547.

Sambridge, M., K. Gallagher, A. Jackson, and P. Rickwood (2006). Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.* **167,** no. 2, 528–542.

Saygin, E., and B. L. Kennett (2010). Ambient seismic noise tomography of Australian continent, *Tectonophysics* **481,** nos. 1/4, 116–125.

Steininger, G., J. Dettmer, S. E. Dosso, and C. W. Holland (2013). Trans-dimensional joint inversion of seabed scattering and reflection data, *J. Acoust. Soc. Am.* **133,** no. 3, 1347–1357.

Stephenson, J., K. Gallagher, and C. Holmes (2004). *Beyond Kriging: Dealing with Discontinuous Spatial Data Fields Using Adaptive Prior Information and Bayesian Partition Modelling*, Vol. 239, Geological Society, London, United Kingdom, Special Publications, 195–209.

Taira, T., and K. Yoshizawa (2020). Upper-mantle discontinuities beneath Australia from transdimensional Bayesian inversions using multimode surface waves and receiver functions, *Geophys. J. Int.* **223,** no. 3, 2085–2100.

Vozoff, K., and D. Jupp (1975). Joint inversion of geophysical data, *Geophys. J. Int.* **42,** no. 3, 977–991.

Yuan, H., and T. Bodin (2018). A probabilistic shear wave velocity model of the crust in the central West Australian craton constrained by transdimensional inversion of ambient noise dispersion, *Tectonics* **37,** no. 7, 1994–2012.

Appendix

In this appendix, we derive the formulas that BayesBay implements to calculate, at each Markov chain iteration, the acceptance probability α in equation (2). Throughout, we make no assumptions about the number of different data sets used in the inference process; that is, we consider the data vector:

$$\mathbf{d} = [\mathbf{d}_1,\mathbf{d}_2,\ldots,\mathbf{d}_s]^T, \tag{A1}$$

where $\mathbf{d}_i$ and $\mathbf{d}_j$ are independent of one another. This results in the block-diagonal data covariance matrix:

$$\mathbf{C}_d = \begin{pmatrix} \mathbf{C}_{d_1} & 0 & \cdots & 0 \\ 0 & \mathbf{C}_{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{C}_{d_s} \end{pmatrix}, \tag{A2}$$

with $\mathbf{C}_{d_i}$ denoting the covariance matrix of $\mathbf{d}_i$.

Moreover, we consider a set of free parameters $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m$, each being a vector of k dimensions with real-valued entries. To ensure our derivation is in the most general terms, we assume that the jth entry of $\mathbf{v}_i$ corresponds to the jth basis function in a hypothetical discretization $\mathbf{c}$.

With these definitions in place, the prior distribution of the model $\mathbf{m} = [k,\mathbf{c},\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m]^T$ reads

$$p(\mathbf{m}) = p(k)p(\mathbf{c}|k)p(\mathbf{v}_1|\mathbf{c})p(\mathbf{v}_2|\mathbf{c})\ldots p(\mathbf{v}_m|\mathbf{c}), \tag{A3}$$

where $p(\mathbf{v}_i|\mathbf{c})$ (rather than $p(\mathbf{v}_i|k)$, as in Bodin, Sambridge, Tkalčić, *et al.*, 2012) emphasizes that the prior probability of $\mathbf{v}_i$ may depend on the discretization. This is the case in the shear-velocity inversion presented in Example II, which involves a depth-dependent prior probability.

The following sections provide analytic expressions for α resulting from different kinds of perturbations to $\mathbf{m}$.

Perturbation of a "physical" parameter

The first type of perturbation we consider involves a change in the jth entry of $\mathbf{v}_i$. ($\mathbf{v}_i$ might not always represent a physically meaningful quantity in practical applications, and here we refer to it as a "physical parameter" only to distinguish it from the other components of $\mathbf{m}$.) In BayesBay, this is implemented by sampling from a Gaussian probability distribution centered on the current value of $\mathbf{v}_i$, that is, $v'_{ij} \sim \mathcal{N}(v_{ij},\theta_i^2)$, where the standard deviation θ_i is a hyperparameter chosen by the user prior to the inversion.

In this case, the prior ratio $p(\mathbf{m}')/p(\mathbf{m}) = p(\mathbf{v}'|\mathbf{c})/p(\mathbf{v}|\mathbf{c}) = p(v'_{ij})/p(v_{ij})$, and because $q(v'_{ij}|v_{ij}) = q(v_{ij}|v'_{ij})$, the proposal ratio $q(\mathbf{m}|\mathbf{m}')/q(\mathbf{m}'|\mathbf{m}) = 1$. Furthermore, because the perturbation does not involve a change in dimensions, $|\mathbf{J}| = 1$. It follows that

$$\alpha(\mathbf{m}'|\mathbf{m}) = \Lambda(\mathbf{m},\mathbf{m}')\frac{p(v'_{ij})}{p(v_{ij})}, \tag{A4}$$

where

$$\Lambda(\mathbf{m},\mathbf{m}') = \exp\left(-\frac{\Phi(\mathbf{m}') - \Phi(\mathbf{m})}{2}\right) \tag{A5}$$

stems from the likelihood ratio. When $p(v_{ij})$ is uniform, $p(v'_{ij})/p(v_{ij}) = 1$, and equation (A4) simplifies to equation (C17) in Bodin, Sambridge, Tkalčić, *et al.* (2012).

Perturbation of the discretization

When a perturbation affects the discretization underlying $\mathbf{v}_i$, the proposals $q(\mathbf{c}|\mathbf{c}')$ and $q(\mathbf{c}'|\mathbf{c})$ can only be evaluated with detailed information about $\mathbf{c}$. Here, we consider an r-dimensional Voronoi tessellation defined by the k-vector $\mathbf{c}$ of Voronoi nuclei, whose jth element is the coordinate vector $\mathbf{c}_j$. (Alternative discretizations can be implemented in BayesBay through the *Discretization* class.) This vector is displaced via a random deviate from a multivariate normal distribution, resulting in $\mathbf{c}'_j \sim \mathcal{N}(\mathbf{c}_j,\mathrm{diag}(\theta_1^2,\theta_2^2,\ldots,\theta_r^2))$, where θ_l denotes the standard

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/96/3/2052/7099576/srl-2024275.1.pdf
by Australian National University user

deviation of the Gaussian used to perturb the lth spatial dimension. Similar to the previous section, the symmetry of the proposal distributions and the fact that $\mathbf{m}$ and $\mathbf{m}'$ have the same dimensionality yield $q(\mathbf{m}|\mathbf{m}')/q(\mathbf{m}'|\mathbf{m}) = 1$ and $|\mathbf{J}| = 1$, respectively.

To evaluate the prior ratio associated with changes in the discretization, we rely on the "grid trick" (Bodin, Sambridge, Tkalčić, *et al.*, 2012; Steininger *et al.*, 2013). The grid trick assumes a discrete set of N possible positions for Voronoi nuclei, each with equal probability of being selected, such that $p(\mathbf{c}|k) = \binom{N}{k}^{-1}$; it follows that $p(\mathbf{c}'|k)/p(\mathbf{c}|k) = 1$. Moreover, because we are considering free parameters $\mathbf{v}_i$ whose prior may be dependent on the spatial position in the discretization, and the jth Voronoi nuclei has been perturbed, $\prod_i \frac{p(\mathbf{v}_i|\mathbf{c}')}{p(\mathbf{v}_i|\mathbf{c})} = \prod_i \frac{p(v_{ij}|\mathbf{c}'_j)}{p(v_{ij}|\mathbf{c}_j)}$. Therefore,

$$\alpha(\mathbf{m}'|\mathbf{m}) = \Lambda(\mathbf{m},\mathbf{m}')\prod_i \frac{p(v_{ij}|\mathbf{c}'_j)}{p(v_{ij}|\mathbf{c}_j)}. \tag{A6}$$

Again, when $\mathbf{v}_i$ is independent of the discretization, that is, $p(\mathbf{v}_i|\mathbf{c}) = p(\mathbf{v}_i|k)$, the previous expression simplifies to equation (C17) in Bodin, Sambridge, Tkalčić, *et al.* (2012).

Birth and death perturbations

A birth perturbation involves the addition of a new dimension to $\mathbf{c}$ and $\mathbf{v}_i$, resulting in $k' = k + 1$. When a Voronoi tessellation underlies $\mathbf{v}_i$, this is implemented in BayesBay by first adding a randomly chosen Voronoi nucleus to $\mathbf{c}$, thereby obtaining $\mathbf{c}' = [\mathbf{c}_1,\mathbf{c}_2,\ldots,\mathbf{c}_k,\mathbf{c}_{k+1}]^T$. For simplicity, let us assume that the nearest neighbor of the newly added nucleus $\mathbf{c}_{k+1}$ is $\mathbf{c}_k$. New values $v_{i(k+1)}$ are then randomly sampled either from the prior or by perturbation of v_{ik}. (We refer to the latter approach as "birth from neighbor.") In both cases, it can be shown that $|\mathbf{J}| = 1$ (e.g., Bodin, Sambridge, Tkalčić, *et al.*, 2012; Hawkins and Sambridge, 2015).

Using again the grid trick, the proposals $q(\mathbf{c}'|\mathbf{m}) = 1/(N-k)$, $q(\mathbf{c}|\mathbf{m}') = 1/(k+1)$, and $q(\mathbf{v}_i|\mathbf{m}') = 1$ (e.g., Bodin, Sambridge, Tkalčić, *et al.*, 2012). Therefore,

$$\frac{q(\mathbf{m}|\mathbf{m}')}{q(\mathbf{m}'|\mathbf{m})} = \frac{q(\mathbf{c}|\mathbf{m}')}{q(\mathbf{c}'|\mathbf{m})}\prod_i \frac{q(\mathbf{v}_i|\mathbf{m}')}{q(\mathbf{v}'_i|\mathbf{m})} = \frac{N-k}{k+1}\prod_i \frac{1}{q(\mathbf{v}'_i|\mathbf{m})}, \tag{A7}$$

where we assumed that the probability of proposing a birth perturbation is the same as that of proposing a death perturbation. Moreover, upon the assumption that k is uniformly distributed, that is, $p(k) = 1/(k_{\max} - k_{\min})$, evaluating the prior ratio is straightforward. Given that $\prod_i \frac{p(\mathbf{v}'_i|\mathbf{c}')}{p(\mathbf{v}_i|\mathbf{c})} = \prod_i p(v'_{i(k+1)}|\mathbf{c}_{k+1})$ and $\frac{p(\mathbf{c}'|k+1)}{p(\mathbf{c}|k)} = \frac{k+1}{N-k}$, we find

$$\frac{p(\mathbf{m}')}{p(\mathbf{m})} = \frac{k+1}{N-k}\prod_i p(v'_{i(k+1)}|\mathbf{c}_{k+1}). \tag{A8}$$

It follows that for a birth perturbation,

$$\alpha(\mathbf{m}'|\mathbf{m})_{\text{birth}} = \Lambda(\mathbf{m},\mathbf{m}')\prod_i \frac{p(v'_{i(k+1)}|\mathbf{c}_{k+1})}{q(\mathbf{v}'_i|\mathbf{m})}. \tag{A9}$$

Further simplifications require a choice for the proposal $q(\mathbf{v}'_i|\mathbf{m})$. In the case of a birth from prior, $q(\mathbf{v}'_i|\mathbf{m}) = p(v'_{i(k+1)}|\mathbf{c}_{k+1})$, and the right side of equation (A9) reduces to $\Lambda(\mathbf{m},\mathbf{m}')$. A birth-from-neighbor approach instead implies that $q(\mathbf{v}'_i|\mathbf{m}) = \frac{1}{\theta_i\sqrt{2\pi}}\exp\{-\frac{(v'_{i(k+1)}-v_{ik})^2}{2\theta_i^2}\}$, leading to a slightly more complex expression.

Finally, a death perturbation involves randomly removing the jth element of $\mathbf{c}$ and $\mathbf{v}_i$. In this case, it is easy to show that the acceptance probability

$$\alpha(\mathbf{m}'|\mathbf{m})_{\text{death}} = \Lambda(\mathbf{m},\mathbf{m}')\prod_i \frac{q(\mathbf{v}'_i|\mathbf{m})}{p(v'_{i(k+1)}|\mathbf{c}_{k+1})}. \tag{A10}$$

Perturbation of the data noise

In this section, we consider the case where the data noise of one or more components of $\mathbf{d}$ is to be inferred in the inversion process. Specifically, we examine the situation where the first b components of $\mathbf{d}$ have unknown noise properties, corresponding to the hyperparameters $\mathbf{h} = [\mathbf{h}_1,\mathbf{h}_2,\ldots,\mathbf{h}_b]$, where the ith component of $\mathbf{h}$ determines $\mathbf{C}_{d_i}$ in equation (A2).

A data-noise perturbation involves altering the data covariance matrix $\mathbf{C}_d$ through a perturbation of $\mathbf{h}$. Similar to the previous sections, this perturbation is implemented in BayesBay by sampling from a Gaussian distribution such that $h'_{ij} \sim \mathcal{N}(h_{ij},\theta_i^2)$, where the subscript j denotes the jth entry of $\mathbf{h}_i$. Assuming that all components of $\mathbf{h}$ are perturbed and recalling that the proposal distributions for the considered perturbations are symmetric, we find

$$\alpha(\mathbf{m}'|\mathbf{m}) = \Lambda(\mathbf{m},\mathbf{m}')\sqrt{\prod_{i=1}^{b}\frac{|\mathbf{C}_{d_i}|}{|\mathbf{C}'_{d_i}|}}, \tag{A11}$$

where we used the fact that $|\mathbf{C}_d| = \prod_i |\mathbf{C}_{d_i}|$.

Similar to Bodin, Sambridge, Tkalčić, *et al.* (2012), BayesBay implements the noise hyperparameters $\mathbf{h}_i = [\sigma_i,\rho_i]^T$, where $\sigma_i > 0$ denotes the noise standard deviation and $0 \le \rho_i < 1$ the correlation between two adjacent samples in the data vector $\mathbf{d}_i$. We assume the correlation function is given by an exponential decay ρ_i^c, where c defines the correlation between data points that are c samples apart. Conveniently, this parameterization leads to analytical expressions, used in our software, for the inverse of the covariance matrix and its determinant (see equations D2 and D3 in Bodin, Sambridge, Tkalčić, *et al.*, 2012).

Manuscript received 30 June 2024
Published online 27 January 2025

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/96/3/2052/7099576/srl-2024275.1.pdf
by Australian National University user