Geophysical Journal International

Geophys. J. Int. (2026) **245,** 1–22
Advance Access publication 2026 April 10
Research Paper

https://doi.org/10.1093/gji/ggag138

GeoPVI: a Python package for geoscientific inversion using parametric variational inference

Xuebin Zhao and Andrew Curtis

School of GeoSciences, University of Edinburgh, Edinburgh EH9 3FE, UK. E-mail: xuebin.zhao@ed.ac.uk

Accepted 2026 April 9. Received 2026 March 17; in original form 2025 October 20

SUMMARY

In many fields of geoscience, researchers study the Earth's properties by solving inverse or inference problems. Probabilistic approaches have gained increased attention over the past decade because they address the nonlinearity and non-uniqueness properties of many naturally-inspired inverse problems and allow uncertainties in the solutions to be estimated. However, implementing such methods is computationally expensive and requires expertise in inverse and inference theory, high performance computing and the geoscientific theory to be inverted. This makes the methods inaccessible to many geoscientists. In this paper, we first review the theoretical background of a particular suite of probabilistic algorithms referred to as *parametric variational inference* (PVI), and introduce GeoPVI, an open-source Python package designed to facilitate the implementation of these methods. With GeoPVI, users can model uncertainties in their geophysical parameter estimates efficiently given their expertise in inverse theory. It differs from sampling-based, non-parametric variational methods in that the probabilistic solution—the posterior or post-inversion probability distribution function that describes uncertainty in the model parameters of interest—is parametrized by explicit mathematical expressions. These expressions allow for the efficient storage and transfer, and for the evaluation of the posterior probability density for any set of parameter values. We demonstrate how to use the package to solve a set of problems, including tomographic imaging using traveltime data, full waveform inversion, surface wave dispersion inversion and vertical electrical sounding. We provide built-in forward functions to simulate first arrival traveltimes and full acoustic waveform data (in two spatial dimensions), and external forward functions can be incorporated into the package easily. We also demonstrate how to change prior information efficiently post-inversion, using the method of variational prior replacement. Contributions from the community are welcome, to make the package more broadly applicable.

Key words: Bayesian inference; Inverse theory; Probability distributions; Tomography; Computational Seismology.

1 INTRODUCTION

In many fields of geoscience, researchers study the Earth's properties by solving inverse or inference problems. Almost all such problems have solutions that are non-unique, so probabilities are used to quantify the solution components that are more or less likely to be true. Unfortunately calculating these probabilities can be computationally intensive, or even prohibitive. In this paper we therefore present a code package, with accompanying relevant theoretical background, user manual and worked examples, that make finding probabilistic solutions to inverse or inference problems more computationally efficient, and to make a relatively new class of *variational inference* methods easier to adopt.

It is relatively common that geoscientists only have access to data collected on or around the Earth's surface, yet wish to estimate properties of the Earth's interior that are not directly observable. Various inverse or inference problems then need to be solved, in which unknown model parameter values that describe the Earth's interior are estimated based on the observed data. Our examples therefore focus on methods to image the interior of the Earth using a variety of methods, but they can readily be adapted to a wide range of other problems.

To tackle such an inverse problem, a forward relationship (or forward function) is defined to predict data that would have been recorded if any particular set of model parameters represented the true Earth. For most geoscientific problems, a direct inverse of this relationship either does not exist or cannot be calculated. Consequently, advanced inference algorithms must be deployed to estimate possible parameter values.

Over past decades, a number of software packages have been developed to solve geophysical inverse problems (T.M. Hansen *et al.* 2013; R. Cockett *et al.* 2015; J.A. Vrugt 2016; M. Afanasiev *et al.* 2017; C. Rücker *et al.* 2017; N.L. Foks & B.J. Minsley 2020; B.J. Minsley *et al.* 2021; A.P. Valentine & M. Sambridge 2022; F. Magrini *et al.* 2022, 2025; A. Zunino *et al.* 2023; X. Zhang & A. Curtis 2024). Most of these employ deterministic methods in which a set of parameter values is calculated that produce simulated data which best fit the

observed data. Typically, these values are found using a procedure that involves iterative linearization of the forward function then minimization of a measure of misfit between observed and predicted data. This is repeated, until the calculated misfit value falls below a pre-defined threshold. However in nonlinear problems, finding a solution that provides good data fits is not sufficient to guarantee that the parameter values are close to those representing the true state of nature. Geophysical inverse problems always have non-unique solutions, usually meaning that many model parameter values fit the observed data to within the range of possible data errors (S. Boyd & L. Vandenberghe 2004; A. Tarantola 2005; A.P. Valentine & M. Sambridge 2023). It is therefore important to find all such parameter values, and account for those uncertainties in post inversion tasks (O.V. Poliannikov & A.E. Malcolm 2016; R. Arnold & A. Curtis 2018; G. Ely *et al.* 2018; A. Siahkoohi *et al.* 2022; X. Zhang & A. Curtis 2022; X. Zhao *et al.* 2022).

In this paper, we explore Bayesian inference, which solves inverse problems under a probabilistic framework. Bayesian inference encapsulates all available information in probability distributions, which are combined according to Bayes' rule (T. Bayes 1763; A. Tarantola 2005):

$$p(\mathbf{m}|\mathbf{d}_{\text{obs}}) = \frac{p(\mathbf{d}_{\text{obs}}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d}_{\text{obs}})}. \tag{1}$$

Here, $p(\mathbf{m})$ is the *prior* probability distribution function (PDF) which describes information about model parameter $\mathbf{m}$ available before inversion (prior to observing the data), and $p(\mathbf{d}_{\text{obs}}|\mathbf{m})$ is the *likelihood* function which quantifies the probability of observing data vector $\mathbf{d}_{\text{obs}}$ given any parameter value vector $\mathbf{m}$. The likelihood is used to measure how well $\mathbf{d}_{\text{obs}}$ matches synthetic data generated by a particular model $\mathbf{m}$. Term $p(\mathbf{d}_{\text{obs}}) = \int_{\mathbf{m}} p(\mathbf{d}_{\text{obs}}|\mathbf{m})p(\mathbf{m})d\mathbf{m}$ is a normalization constant known as the *evidence*. Bayesian inference therefore combines data and prior knowledge in a probabilistic manner, and the resultant *posterior* PDF $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$ describes all possible model solutions that fit both the observed data and prior knowledge, and their corresponding probabilities.

Generally, the posterior PDF cannot be calculated directly due to the evidence term which involves the evaluation of a high dimensional integral that is computationally infeasible. Consequently, sampling-based methods are often employed to solve Bayesian problems, in which a set of model samples is generated within the parameter space, such that the probability density of these samples is proportional to the posterior PDF in the limit as the number of samples tends to infinity. Markov chain Monte Carlo (McMC) is widely used for this purpose: it constructs chains of samples distributed according to a target distribution (in this case the posterior PDF). After sampling, the collected samples can be used to calculate statistics of the posterior distribution (K. Mosegaard & A. Tarantola 1995; K. Mosegaard & M. Sambridge 2002; M. Sambridge & K. Mosegaard 2002). These methods have proved to be extremely useful because in principle they can be used to approximate the solution to almost any inference problem with almost any set of prior information. However because of their generality, all such sampling methods tend to be computationally costly when prior information is weak in higher dimensional inference problems (i.e. the number of unknowns in model vector $\mathbf{m}$ increases), due to the curse of dimensionality (A. Curtis & A. Lomax 2001). There is therefore a growing interest in exploring alternative methods that may be more efficient for particular classes of problems encountered in geosciences.

Variational inference presents an efficient alternative to McMC and has gained traction in geophysics in recent years. The efficiency of variational methods stems from the philosophy of solving Bayesian inverse problems within an optimization framework, by seeking one optimal approximation to the true posterior pdf from a predefined family of distributions. Variational methods therefore convert a random sampling problem into a numerical optimization, yet still produce fully probabilistic results. While the No Free Lunch Theorem states that no algorithm performs better than any other when averaged over all problems (D.H. Wolpert & W.G. Macready 1997), for particular classes of problems, variational methods can be more efficient than random sampling. They can also scale better to higher dimensional problems and their convergence is easier to detect (C.M. Bishop 2006; D.M. Blei *et al.* 2017; X. Zhang *et al.* 2021).

Since first introduced to geophysics in A. Nawaz & A. Curtis (2018, 2019) and A. Nawaz *et al.* (2020), a variety of variational methods have been applied to geophysical inverse problems. For example, X. Zhang & A. Curtis (2020) and X. Zhao *et al.* (2021) performed variational traveltime tomographic inversions, demonstrating the efficiency of the method with real data examples. A. Siahkoohi *et al.* (2021, 2023) applied *amortized* variational inference to seismic migration, in which a variational model is pre-trained to provide an approximate posterior solution conditioned on observed data, followed by fine-tuning to improve the accuracy. S. Levy *et al.* (2022) and M. Liu *et al.* (2024) performed variational inversion within a *latent space* parametrized by deep generative models, thereby allowing geologically plausible prior information to be incorporated into the inversion. Additional geophysical applications include earthquake hypocentre inversion (J.D. Smith *et al.* 2022), slip distribution inversion (L. Sun *et al.* 2023), full waveform inversion (X. Zhang & A. Curtis 2021; A. Lomas *et al.* 2023; W. Wang *et al.* 2023; X. Zhang *et al.* 2023; M. Izzatullah *et al.* 2024; Z. Yin *et al.* 2025; X. Zhao & A. Curtis 2025b, 2026; Chen *et al.* 2025). and experimental design (D. Strutz & A. Curtis 2024).

Existing variational methods can generally be categorized into two groups based on the way that they represent an approximation to the posterior distribution. In the first group, the variational distribution is represented implicitly through a set of samples (particles). These samples are optimized such that their density best approximates the posterior distribution. This approach includes Stein variational gradient descent (SVGD–Q. Liu & D. Wang 2016) and its variants (V. Gallego & D.R. Insua 2018; P. Chen & O. Ghattas 2020; W. Gong *et al.* 2020; F. D'Angelo & V. Fortuin 2021). The second group employs a parametric representation, by building an explicit mathematical expression (an equation) that approximates the form of the posterior PDF (D. Rezende & S. Mohamed 2015; F. Guo *et al.* 2016; A. Kucukelbir *et al.* 2017; J. Sjölund 2023; X. Zhao & A. Curtis 2024b). We refer to the latter approach as *parametric variational inference* (PVI) – the focus of this paper.

Several parametric variational methods have been explored in previous literature. A. Kucukelbir *et al.* (2017) introduced automatic differentiation variational inference (ADVI) where a transformed Gaussian distribution is used as the variational distribution. D. Rezende & S. Mohamed (2015) developed a method using a sequence of invertible transforms called normalizing flows, which converts a simple and known probability distribution into a potentially far more complex approximation of the posterior PDF. F. Guo *et al.* (2016) and A.C. Miller *et al.* (2017) proposed boosting variational inference (BVI), where a mixture of simpler distributions (e.g. Gaussians) is used to better approximate the posterior PDF. A. Nawaz & A. Curtis (2018, 2019) and A. Nawaz *et al.* (2020) showed how to construct bespoke, highly efficient variational methods for problems with specific classes of parameter–data relationships. X. Zhao & A. Curtis (2024b) proposed physically structured variational inference (PSVI) for full waveform inversion, crafting a variational distribution with a tailored correlation structure.

Several advantages make parametric variational inference methods particularly beneficial and appealing for practical applications. First, the parametric (explicit) nature allows for efficient evaluation of the posterior probability density of any model sample post inversion. This also allows prior information to be updated post inversion using a variational prior replacement method (M. Walker & A. Curtis 2014b; X. Zhao & A. Curtis 2024c). Secondly, these methods enable easy generation of new posterior samples, without the need for additional forward simulations or evaluations of Bayes' rule. The same is not straightforward using non-parametric methods such as Markov chain Monte Carlo (K. Mosegaard & A. Tarantola 1995; M. Sambridge & K. Mosegaard 2002) or SVGD (Q. Liu & D. Wang 2016): in both cases drawing new samples requires that the algorithm is re-started which typically requires many new forward evaluations. Thirdly, the parametric nature of the variational posterior distribution allows some statistics and inferred properties to be calculated analytically, which is not possible using sample-based methods without assuming an explicit form for the sample density. And fourthly, it is efficient to save, share and load inversion results (M. Scheiter *et al.* 2022): only the control parameters of the parametric expression are required, instead of a batch of posterior samples which by necessity can each be large (e.g. in 3-D FWI problems). On the other hand, the suite of parametric methods requires the prior probability density values of model samples to be evaluated explicitly, which could prevent the use of many implicit prior distributions in which random realizations can be generated easily yet evaluating their probability density values is not straightforward (e.g. I. Goodfellow *et al.* 2014; G. Mariethoz & J. Caers 2014; D. Tetzlaff 2023).

Currently, there is a lack of a comprehensive introduction (both theoretical and practical) and tutorial specifically on parametric variational inference within the field of geophysics. Some software packages that implement variational methods in geophysical contexts exist—for example, SVGD is implemented in HMCLab (A. Zunino *et al.* 2023); InvertibleNetworks package (R. Orozco *et al.* 2023) is developed to implement invertible neural networks and can be used to perform amortized variational inference (Z. Yin *et al.* 2025); X. Zhang & A. Curtis (2024) developed VIP, a variational inversion package to perform Bayesian tomographic imaging problems, in which SVGD, sSVGD and ADVI are implemented. These packages focus on individual parametric (ADVI in X. Zhang & A. Curtis 2024) or non-parametric methods, so do not allow a variety of parametric methods with different levels of complexity to be applied to particular geophysical problems easily, and to be compared and studied systematically by general geophysicists. In addition, the implementation of parametric methods currently requires a high level of expertise in coding even for experts in inverse theory, discouraging their widespread use within the community.

To address these short-comings, this paper systematically discusses and compares several of the current state-of-art parametric variational inference methods, to educate geophysicists and to make a group of such methods significantly easier to use. It introduces a Python package to ease the methods' implementation in particular problems. And it includes a set of worked examples in seismology and electrical resistivity inversion. Our code package is named **GeoPVI** for **Geo** physical inversion using **P**arametric **V**ariational **I**nference methods. While GeoPVI allows easy implementation of several parametric variational inversion algorithms, users need to know basic inverse theory (e.g. A. Tarantola 2005) and the theoretical background behind these methods (provided below) to use the package.

In the following sections we first introduce the concept of parametric variational inference, and then outline the theoretical background of several specific algorithms that have been implemented in the GeoPVI package. In Section 3, we describe the package, after which we present several application examples for benchmark testing and easy adaptation to other problems by users, and to enable future algorithms to be tested against those included here. Finally we discuss potential uses and conclude.

2 THEORETICAL BACKGROUND

2.1 Parametric variational inference

In variational inference, we define a family of known probability distributions $\mathcal{Q}(\mathbf{m}) = \{q(\mathbf{m})\}$ (known as the variational family). From this family, we select one optimal member $q^*(\mathbf{m})$ which best approximates the true posterior distribution $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$. This is typically achieved by minimizing the Kullback–Leibler (KL) divergence (a measure of difference between two probability distributions—S. Kullback & R.A. Leibler 1951) between $q(\mathbf{m})$ and $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$:

$$\text{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{\text{obs}})] = \mathbb{E}_{q(\mathbf{m})}[\log q(\mathbf{m}) - \log p(\mathbf{m}|\mathbf{d}_{\text{obs}})], \tag{2}$$

where $\mathbb{E}_{q(\mathbf{m})}[f(\mathbf{m})] = \int_{\mathbf{m}} q(\mathbf{m}) f(\mathbf{m}) d\mathbf{m}$ is the expectation of function $f(\mathbf{m})$ with respect to $q(\mathbf{m})$. By substituting Bayes rule (eq. 1) into eq. (2), it can be shown that minimizing the KL divergence in eq. (2) is equivalent to maximizing the *evidence lower bound* (ELBO)

on $\log p(\mathbf{d}_{\text{obs}})$ (D.M. Blei *et al.* 2017; C. Zhang *et al.* 2018):

$$\text{ELBO}[q(\mathbf{m})] = \mathbb{E}_{q(\mathbf{m})}[\log p(\mathbf{m}, \mathbf{d}_{\text{obs}}) - \log q(\mathbf{m})]. \tag{3}$$

In practice, the latter is a more attractive optimization criterion than eq. (2) because it does not require the value of the logarithmic evidence term $\log p(\mathbf{d}_{\text{obs}})$ which is often computationally intractable, and which is implicit in the posterior pdf in eq. (2). The optimization result is a probability distribution $q^*(\mathbf{m})$ that satisfies

$$\begin{aligned} q^*(\mathbf{m}) &= \underset{q \in Q}{\text{argmax}}\, \text{ELBO}[q(\mathbf{m})] \\ &\approx p(\mathbf{m}|\mathbf{d}_{\text{obs}}) \end{aligned}, \tag{4}$$

which is the best approximation to $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$ within $\mathcal{Q}(\mathbf{m})$.

In parametric variational inference, the variational distribution is represented by parametric and explicit expressions. For example, a Gaussian variational distribution denoted by $q(\mathbf{m}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is controlled by parameters $\boldsymbol{\mu}$ (mean vector) and $\boldsymbol{\Sigma}$ (covariance matrix) and can be written as an explicit mathematical equation. As described in Section 1, several advantages are offered by parametric variational methods, making them attractive in certain geophysical problems (e.g. M. Scheiter *et al.* 2022; X. Zhao & A. Curtis 2024c).

In parametric variational methods, the $\text{ELBO}[q(\mathbf{m})]$ in eq. (3) is maximized iteratively with respect to control parameters Θ which determine the parametric expression. This is accomplished using a gradient-based approach:

$$\Theta^{i+1} = \Theta^i + \epsilon\, \nabla_\Theta \text{ELBO}[q(\mathbf{m})], \tag{5}$$

where ϵ is the step size (learning rate). Term $\nabla_\Theta \text{ELBO}[q(\mathbf{m})]$ is the gradient of the $\text{ELBO}[q(\mathbf{m})]$ with respect to Θ. This gradient requires an expectation term to be evaluated, which contains derivatives of the terms in eq. (3). The expectation can be estimated using Monte Carlo integration, in which only a low number of samples (potentially even a single sample) is necessary in each iteration: given that optimization is performed over many iterations, such reduced-accuracy gradient estimates allow statistical convergence towards a reasonable solution (A. Kucukelbir *et al.* 2017). In addition, a reparametrization trick is often used to ensure that the Monte Carlo estimate of the expectation is differentiable with respect to Θ, and to reduce the variance of the estimated gradient at each iteration (D.P. Kingma & M. Welling 2014).

2.2 ADVI and PSVI

The simplest parametric variational method implemented in the GeoPVI package is mean field automatic differentiation variational inference (ADVI—A. Kucukelbir *et al.* 2017), which has been used extensively in geophysical inverse problems (X. Zhang and A. Curtis 2020; X. Zhao *et al.* 2021; O. Bates *et al.* 2022 ; A. Siahkoohi *et al.* 2023; L. Sun *et al.* 2023; W. Wang *et al.* 2023; X. Zhang *et al.* 2023). In ADVI, we first define a parametric variational family with an explicit expression, such as a Gaussian variational family. An invertible transform is applied to convert the (Gaussian) random variables (which have infinite support, from minus infinity to plus infinity) into geophysical model parameters defined within a bounded space (e.g. seismic velocities should be positive so are bounded by zero). The following function is often used for this purpose:

$$m_i = T(\eta_i) = a_i + \frac{b_i - a_i}{1 + \exp(-\eta_i)}, \tag{6}$$

where m_i and η_i represent the i-th parameter in the constrained (geophysical) and unbounded spaces respectively, and a_i and b_i are the lower and upper bounds on m_i. The transformed distribution is then used to approximate the true posterior distribution. The inverse transform of eq. (6)

$$\eta_i = T^{-1}(m_i) = \log(m_i - a_i) - \log(b_i - m_i) \tag{7}$$

converts model parameters from a bounded space back into the space of Real numbers. Note that a_i and b_i are fixed with pre-defined values, so this transform does not contain parameters to be varied (optimized during inversion).

Consider a Gaussian variational family: we need to optimize two sets of control parameters: a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$, containing n and $n(n+1)/2$ independent parameters (where n is the dimensionality of the problem), respectively. They can be optimized using a gradient-based optimization method. Three strategies to parametrize the covariance matrix $\boldsymbol{\Sigma}$ have been implemented in GeoPVI. Mean field ADVI assumes a diagonal posterior covariance matrix, effectively disregarding all correlations between model parameter pairs (known as a mean field approximation—C.M. Bishop 2006). This enhances the algorithmic efficiency by simplifying the variational optimization when maximizing the ELBO in eq. (3). The number of control parameters to be optimized is then double the number in a deterministic inversion for the maximum-likelihood parameter values alone. While this method is efficient and can provide accurate mean estimates, several studies have shown that it tends to strongly underestimate the posterior uncertainties and bias the inversion results, especially in problems with strong or complex posterior correlation structures (X. Zhang *et al.* 2023).

A. Kucukelbir *et al.* (2017) introduced full rank ADVI, which constructs a full posterior covariance matrix $\boldsymbol{\Sigma}$. Since evaluating the gradient of the $\text{ELBO}[q(\mathbf{m})]$ with respect to $\boldsymbol{\Sigma}$ requires its determinant $|\boldsymbol{\Sigma}|$ to be calculated which is computationally expensive,

the covariance matrix $\mathbf{\Sigma}$ is often reparametrized by a Cholesky factorization $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T$. Here, $\mathbf{L}$ is a lower triangular matrix with non-zero diagonal elements (ensuring that $\mathbf{\Sigma}$ is positive semidefinite). Its determinant can be calculated efficiently as the product of its diagonal elements. While full rank ADVI improves the performance of mean field ADVI, it also increases computational demands and memory usage substantially. Particularly in high dimensional (many parameter) inverse problems, this method can become expensive or computationally intractable, because an n-dimensional parameter vector requires $n(n+1)/2$ parameters to construct the lower triangular matrix $\mathbf{L}$ which defines $\mathbf{\Sigma}$.

In response to the limitations in both mean field ADVI and full rank ADVI, X. Zhao & A. Curtis (2024b) introduced physically structured variational inference (PSVI). This approach seeks an intermediate balance between the former two methods, by imposing a physics-guided structure on $\mathbf{L}$ (thus also on $\mathbf{\Sigma}$). Specifically, only parameter correlations that are expected to be dominant based on physical principles, are represented in $\mathbf{L}$; all other correlation information is ignored *a priori*. For example, in many spatial imaging inverse problems, model correlations are observed to be strongest between locations that are spatially proximal, because model parametrizations usually specify information at a scale below the resolution of observed data. The magnitude of correlations decays rapidly with increasing distance between locations, so correlations between distant parameters can more reasonably be ignored during inversions. Given that the diagonal elements of $\mathbf{L}$ primarily represent the variances of model parameters, and that the off-diagonal elements mainly capture the correlations between parameter pairs, we impose the following type of sparse parametrization on $\mathbf{L}$:

$$\mathbf{L} = \begin{bmatrix} l_{0,1} & & & & & & \\ l_{1,1} & l_{0,2} & & & & & \\ 0 & l_{1,2} & l_{0,3} & & & & \\ \dots & 0 & l_{1,3} & \dots & & & \\ l_{i,1} & \dots & 0 & \dots & l_{0,n-2} & & \\ 0 & \dots & \dots & \dots & l_{1,n-2} & l_{0,n-1} & \\ \dots & 0 & l_{i,n-i} & \dots & 0 & l_{1,n-1} & l_{0,n} \end{bmatrix}, \tag{8}$$

where sparsely distributed non-zero off-diagonal elements are used to encapsulate dominant spatial correlations—in this case spatially proximal parameters. These elements are treated as tunable parameters that are optimized freely during inversion. All other values are set to 0, assuming spatial independence for the corresponding parameter pairs and thus imposing a sparse structure on $\mathbf{L}$.

In the implementation of PSVI within the GeoPVI software package, any pre-defined sparse structure can be imposed on matrix $\mathbf{L}$ to encode desired correlations on the variational distribution, based on physical (prior) knowledge of the specific inverse problem of interest. By doing so, the required number of parameters in $\mathbf{L}$ remains more manageable for large problems, and the variational distribution is still able to capture the most significant expected correlations. Note that while PSVI provides an intermediate method between mean field and full rank ADVI for Gaussian distributions, in fact it can be used to extend a variety of variational methods to enhance their accuracy and efficiency. For example, while this is not yet implemented in GeoPVI, and PSVI method could be used to model higher order correlations which typically occur in non-Gaussian variational distributions.

2.3 Normalizing flows

Normalizing flows offer a flexible and expressive mechanism for modelling complex probability distributions, by passing a simple and known initial (base) distribution (e.g. a standard Gaussian or a uniform distribution) through a sequence of invertible and differentiable transforms which change its shape, known as the flows (L. Dinh *et al.* 2015; D. Rezende & S. Mohamed 2015; I. Kobyzev *et al.* 2020; G. Papamakarios *et al.* 2021).

Let $\mathbf{m}_0 \in \mathbb{R}^n$ be an n-dimensional random variable distributed according to a known probability distribution $q_0(\mathbf{m})$. The main concept behind normalizing flows is to apply a function $f_\theta : \mathbb{R}^n \to \mathbb{R}^n$ to $\mathbf{m}_0$, where θ represents a set of hyperparameters that parametrize the function f_θ. Based on the change of variable formula

$$\begin{cases} \mathbf{m}_1 = f_\theta(\mathbf{m}_0) \\ \log q_1(\mathbf{m}_1) = \log q_0(\mathbf{m}_0) + \log \left| \det \dfrac{\partial f_\theta^{-1}}{\partial \mathbf{m}_1} \right| = \log q_0(\mathbf{m}_0) - \log \left| \det \dfrac{\partial f_\theta}{\partial \mathbf{m}_0} \right| \end{cases}, \tag{9}$$

where $|\det(\cdot)|$ calculates the absolute value of the determinant of a matrix. Term $\dfrac{\partial f_\theta}{\partial \mathbf{m}_0}$ is the Jacobian matrix of f_θ and $\left| \det \dfrac{\partial f_\theta}{\partial \mathbf{m}_0} \right|$ quantifies the change in parameter space volume around $\mathbf{m}_0$ induced by f_θ. In normalizing flows, $q_0(\mathbf{m}_0)$ is defined to be the initial (base) distribution, while f_θ is called a flow and should be differentiable and invertible to ensure that $\left| \det \dfrac{\partial f_\theta}{\partial \mathbf{m}_0} \right|$ exists and is non-zero. The transformed distribution $q_1(\mathbf{m}_1)$ is called the target distribution. Typically, a series of multiple flows (transforms) is used to construct a flows-based model through the chain rule

$$\begin{cases} \mathbf{m}_K = \left(f_{\theta_K} \circ f_{\theta_{K-1}} \cdots \circ f_{\theta_2} \circ f_{\theta_1} \right)(\mathbf{m}_0) \\ \log q_K(\mathbf{m}_K) = \log q_0(\mathbf{m}_0) + \sum_{i=1}^K \log \left| \det \dfrac{\partial f_{\theta_i}^{-1}}{\partial \mathbf{m}_i} \right| = \log q_0(\mathbf{m}_0) - \sum_{i=1}^K \log \left| \det \dfrac{\partial f_{\theta_i}}{\partial \mathbf{m}_{i-1}} \right| \end{cases}, \tag{10}$$

where symbol $\circ$ stands for the composition of functions. In this way, the base distribution 'flows' through the trajectory of these transforms, and transforms step-wise towards a target distribution with any desired level of complexity.

In variational inference, our objective is to approximate the posterior distribution using the transformed distribution $q_K(\mathbf{m}_K)$. To achieve this, we maximize the ELBO$[q_K(\mathbf{m}_K)]$ in eq. (3) with respect to $q_K(\mathbf{m}_K)$ by optimizing the flows parameters $\Theta = \{\theta_1, \theta_2, ..., \theta_K\}$ using a gradient-based optimization method. Since both the base distribution and the flows (transforms) are known and have parametric expressions, the target distribution (in this case the approximate posterior PDF) is parametric.

Generally, normalizing flows should be: (1) invertible; (2) expressive enough to allow the target distribution to adequately approximate the posterior distribution; (3) computationally efficient both in terms of performing the forward transform and in calculating the determinant of the Jacobian matrix. Many effective flows have been proposed to perform flexible and nonlinear transforms. Reviews of these methods can be found in I. Kobyzev *et al.* (2020), G. Papamakarios *et al.* (2021) and X. Zhao *et al.* (2021). In GeoPVI, we implement several flows that have been tested in geophysical problems (X. Zhao *et al.* 2021; S. Levy *et al.* 2022; C. Sun *et al.* 2024), including planar and radial flows (D. Rezende & S. Mohamed 2015), various coupling flows with different nonlinear functions (L. Dinh *et al.* 2015, 2017), autoregressive flows (D.P. Kingma *et al.* 2016; G. Papamakarios *et al.* 2017; D.P. Kingma & P. Dhariwal 2018) and neural spline flows (C. Durkan *et al.* 2019).

2.4 Boosting variational inference

In boosting variational inference (BVI), the variational distribution is defined as a mixture of n probability distributions:

$$q^n(\mathbf{m}) = \sum_{i=1}^{n} w_i g_i(\mathbf{m}), \tag{11}$$

where $g_i(\mathbf{m})$ represents a component distribution, and w_i is the corresponding weight coefficient for $g_i(\mathbf{m})$ which controls the magnitude of the contribution of this component, satisfying $0 \leq w_i \leq 1$ and $\sum_{i=1}^{n} w_i = 1$. Theoretically, a mixture distribution can approximate any probability distribution up to any level of accuracy, even if a simple component distribution is used (C.M. Bishop 1994; U. Meier *et al.* 2007; M.S. Shahraeeni *et al.* 2012; S. Earp & A. Curtis 2020). If the component distribution is parametric, the mixture distribution is also parametric.

In BVI we optimize the set $\{w_i, g_i(\mathbf{m}); i = 1, 2, ..., n\}$ by maximizing the ELBO$[q^n(\mathbf{m})]$. Due to the non-convex nature of this maximization with respect to the variational parameters, there is a risk of converging to local maxima where one component dominates the mixture with a large weight coefficient (close to 1). To mitigate this, F. Guo *et al.* (2016) and A.C. Miller *et al.* (2017) recommended calculating each component iteratively using a greedy algorithm, in which only one component distribution is introduced and optimized at each BVI step. This optimized component is then incorporated into the mixture model to progressively refine the approximation accuracy of the mixture distribution. In other words, each new component is optimized to reduce the discrepancy between the existing mixture distribution and the posterior distribution. Mathematically, the process begins with a single component $q^1(\mathbf{m}) = g_1(\mathbf{m})$ and $w_1 = 1$. Subsequently, each new component $g_t(\mathbf{m})$ and its weight $w_t \in [0, 1]$ are optimized iteratively over t, resulting in a new mixture distribution:

$$q^t(\mathbf{m}) = (1 - w_t)q^{t-1}(\mathbf{m}) + w_t g_t(\mathbf{m}). \tag{12}$$

Following F. Guo *et al.* (2016), we optimize each component distribution by first calculating $g_t(\mathbf{m})$, and then determining the weight coefficient w_t given $g_t(\mathbf{m})$. By taking the first-order Taylor expansion of the ELBO$[q^t(\mathbf{m})]$ around $q^{t-1}(\mathbf{m})$ (F. Guo *et al.* 2016; F. Locatello *et al.* 2018), it can be shown that the optimal $g_t^*(\mathbf{m})$ satisfies

$$g_t^*(\mathbf{m}) = \underset{g_t(\mathbf{m})}{\operatorname{argmax}} \left\{ \mathbb{E}_{g_t(\mathbf{m})}[\log p(\mathbf{m}, \mathbf{d}_{\text{obs}})] - \mathbb{E}_{g_t(\mathbf{m})}[\log q^{t-1}(\mathbf{m})] - \lambda \mathbb{E}_{g_t(\mathbf{m})}[\log g_t(\mathbf{m})] \right\}, \tag{13}$$

$$g_t^*(\mathbf{m}) = \underset{g_t(\mathbf{m})}{\operatorname{argmax}} \left\{ \mathbb{E}_{g_t(\mathbf{m})}[\log p(\mathbf{m}, \mathbf{d}_{obs})] - \mathbb{E}_{g_t(\mathbf{m})}[\log q^{t-1}(\mathbf{m})] - \lambda \mathbb{E}_{g_t(\mathbf{m})}[\log g_t(\mathbf{m})] \right\},$$

where $\left(\mathbb{E}_{g_t(\mathbf{m})}[\log p(\mathbf{m}, \mathbf{d}_{\text{obs}})] - \mathbb{E}_{g_t(\mathbf{m})}[\log q^{t-1}(\mathbf{m})]\right)$ quantifies the expected residual discrepancy between the current mixture distribution $q^{t-1}(\mathbf{m})$ and the unnormalized posterior PDF $p(\mathbf{m}, \mathbf{d}_{\text{obs}})$. By maximizing this objective function, we introduce a new component $g_t(\mathbf{m})$ which reduces the difference between $q^{t-1}(\mathbf{m})$ and $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$. This ensures that BVI improves the accuracy of the variational distribution as each new component is added.

The entropy $(-\mathbb{E}_{g_t(\mathbf{m})}[\log g_t(\mathbf{m})])$ in eq. (13) is a regularization term. Since entropy measures the uncertainty within a probability distribution, maximizing the entropy helps to ensure that $g_t(\mathbf{m})$ does not converge (collapse) into a narrow distribution or a degenerate point solution, where the distribution would have a significant probability mass concentrated only at a single point (F. Guo *et al.* 2016). Parameter λ controls the contribution of this regularization term; a larger value typically results in a broader, more dispersed component distribution $g_t(\mathbf{m})$. Note that whatever the value of λ, the introduction of each component should improve the fit to the posterior PDF, and the value of λ can be reduced with increasing iteration number to improve the final result.

Weight coefficient $w_t \in [0, 1]$ for $g_t(\mathbf{m})$ is obtained by maximizing the ELBO$[q^t(\mathbf{m})]$ with respect to w_t using gradient-based optimization methods (F. Guo *et al.* 2016; F. Locatello *et al.* 2018; R. Giaquinto & A. Banerjee 2020). Such a process necessitates additional

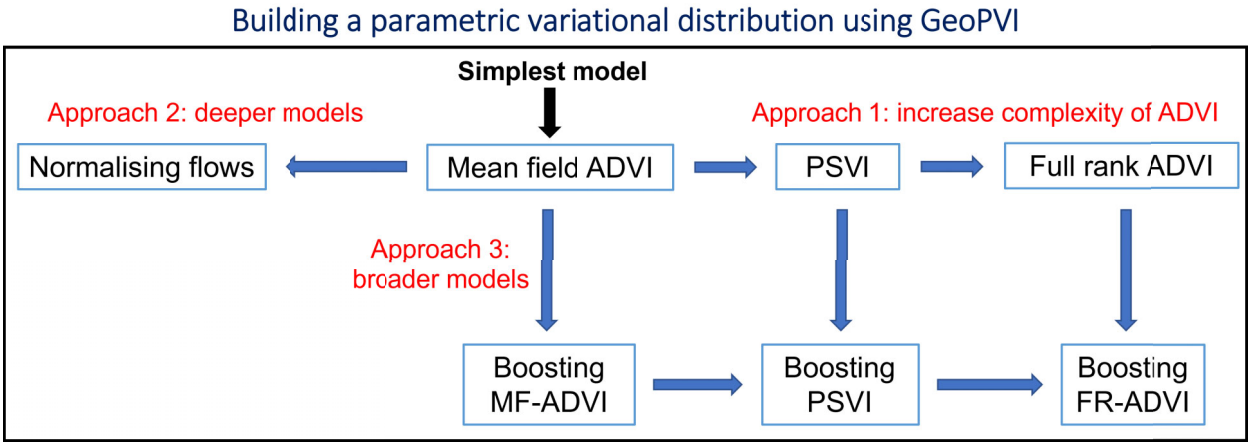

Figure 1. Structure of the GeoPVI package for building a parametric variational distribution. Arrows between different methods indicate increasing complexity of the variational PDF. Starting from the simplest case of mean field ADVI, three different approaches are employed to construct an improved variational PDF: either increase complexity of the ADVI covariance matrix (right arrows), implement a flows-based model (left arrow) or apply boosting-based methods (downward arrows).

forward simulations. However, X. Zhao & A. Curtis (2024a) showed that for specific geophysical problems such as traveltime tomography and full waveform inversion, an empirical formula for w_t can yield good posterior solutions efficiently. While this approach may not converge to the theoretically optimal weight coefficients, it offers a simpler, more computationally tractable method of updating weights without the need for further forward evaluations. Any discrepancies introduced by the sub-optimality of this approach can be corrected by adding more components to the mixture distribution.

3 GEOPVI PACKAGE OVERVIEW

GeoPVI is a Python-based code package designed for variational Bayesian inversion using a suite of tools that support various parametric variational methods. Current features include mean field ADVI and full rank ADVI, PSVI, normalizing flows and BVI. In addition, the package provides an efficient tool to replace prior information in Bayesian posterior solutions.

In eq. (3), the calculation of the variational objective function requires both $\log p(\mathbf{m}, \mathbf{d}_{\text{obs}})$ and $\log q(\mathbf{m})$ to be calculated, which can be constructed separately in GeoPVI. The former is the sum of the logarithmic prior $\log p(\mathbf{m})$ and likelihood $\log p(\mathbf{d}_{\text{obs}}|\mathbf{m})$ values. GeoPVI supports normal and uniform prior probability distributions by default, or allows users to define custom prior distributions by implementing a function that evaluates $\log p(\mathbf{m})$ and its gradient $\nabla_{\mathbf{m}} \log p(\mathbf{m})$ for any model sample $\mathbf{m}$. Alternatively, if the codes that calculate $\log p(\mathbf{m})$ are written in PyTorch, the gradient $\nabla_{\mathbf{m}} \log p(\mathbf{m})$ can be calculated analytically through the automatic differentiation engine provided by PyTorch, so users do not need to calculate the gradient manually. To calculate the likelihood, forward simulation of model parameter $\mathbf{m}$ is required. GeoPVI includes two commonly used forward functions tailored for seismic imaging problems: the fast marching method (FMM—N. Rawlinson & M. Sambridge 2005) for traveltime tomography, and a 2-D acoustic full waveform simulator (Y. Wang *et al.* 2019; X. Zhao *et al.* 2020) for full waveform inversion. Additionally, users can integrate external forward functions into GeoPVI by providing appropriate mechanisms for both forward simulation and gradient evaluation. Examples that integrate external forward functions for surface wave dispersion inversion and electrical resistivity inversion are provided within the package and presented in the next section.

Fig. 1 illustrates the structure of the GeoPVI package to construct a parametric variational distribution and to calculate its probability density value $\log q(\mathbf{m})$. Starting from most basic method, mean field ADVI, we introduce three approaches to improve its performance. The first approach is to increase the complexity of the Gaussian covariance matrix, from diagonal to the more sophisticated forms used in PSVI and full rank ADVI. These methods optimize posterior correlations: PSVI focuses on dominant parameter correlations that are known *a priori*, while full rank ADVI includes all correlations. In the second approach, we construct a 'deeper' learning model using normalizing flows. In this case we consider a series of parametric transforms (flows), rather than one single parametric distribution as employed in the first approach. The third approach builds a 'broader' model using BVI. Here, broader refers to the number of component functions that are added, while deeper refers to the number of functions that are composed in series, to obtain the result. In principle, we can use all methods in the first row in Fig. 1 to construct component distributions, then apply the boosting concept of BVI to build a mixture distribution. Currently, GeoPVI features boosting mean field ADVI, PSVI and full rank ADVI, as illustrated in the bottom row in Fig. 1.

GeoPVI uses the automatic differentiation engine provided by PyTorch to build a computational graph, which calculates the ELBO and its gradient with respect to the variational parameters automatically (A. Paszke *et al.* 2019). Specifically, the calculation of $\log q(\mathbf{m})$

for all implemented methods in GeoPVI is written in PyTorch. Therefore, its gradient is already registered through automatic differentiation and does not require explicit derivation, so users do not necessarily need to derive analytic gradient expressions nor write the corresponding code manually, both of which might be complicated to implement. If the calculation of $\log p(\mathbf{m}, \mathbf{d}_{\text{obs}})$ is also written in PyTorch, then the whole inversion process is implemented through automatic differentiation and there is no requirement to evaluate gradients explicitly. If $\log p(\mathbf{m}, \mathbf{d}_{\text{obs}})$ is calculated using non-PyTorch codes (for example many commonly used forward simulation functions in geophysics were programmed decades ago using non-Python programming languages), users are allowed to register both $\log p(\mathbf{m}, \mathbf{d}_{\text{obs}})$ and $\nabla_{\mathbf{m}} \log p(\mathbf{m}, \mathbf{d}_{\text{obs}})$ into the automatic differentiation framework. Then the full inversion (optimization) process can still be implemented through automatic differentiation. Such examples are provided in the GeoPVI package. For parallelization, basic calculations within PyTorch are inherently parallelized. Given that the most computationally expensive component in Bayesian geophysical inversion is the forward model and gradient evaluation, GeoPVI use libraries such as Dask and Multiprocessing to facilitate the parallelization of forward simulations at the sample level, to improve efficiency. The package can therefore be deployed on either a desktop computer or modern high performance computation facilities.

A user manual is provided within the GeoPVI package for in-depth guidance. In addition, several Jupyter notebooks are available, showcasing geophysical inversion examples to demonstrate the package's basic applications. Finally, GeoPVI is still being updated and expanded: contributions from the community are highly encouraged and would be welcomed.

4 APPLICATIONS

In the following we provide examples that demonstrate how GeoPVI can be used to solve nonlinear inverse problems and estimate uncertainties in the inversion results. These are included as worked examples in the GeoPVI package.

4.1 Traveltime tomography

We first apply GeoPVI to a 2-D traveltime tomography example, in which subsurface velocity structure is estimated using measured first arrival traveltimes of seismic waves travelling between source and receiver locations. In this scenario, the forward function simulates traveltimes of waves propagating through a given velocity model field. Typically, this function is characterized by the Eikonal equation

$$(\nabla t(\mathbf{x}; \mathbf{x}_s))^2 = c^{-2}(\mathbf{x}), \tag{14}$$

where $c(\mathbf{x})$ is the medium velocity and $t(\mathbf{x}; \mathbf{x}_s)$ is the arrival time of waves from a fixed source location $\mathbf{x}_s$ to any point $\mathbf{x}$. The forward problem in eq. (14) is usually solved using the fast marching method (N. Rawlinson & M. Sambridge 2004, 2005). We use regularly gridded models, so the gradient of the synthetic traveltime data with respect to the velocity model parameters is computed by tracing the fastest ray and then calculating the corresponding ray path length within each grid cell. This forward model is integrated within the GeoPVI package.

We invert for the group velocity map of Love waves around the British Isles. The British Isles have been studied extensively using different seismic imaging methods including earthquake body wave tomography (S.J. Arrowsmith *et al.* 2005; R. Luckett & B. Baptie 2015), ambient noise surface wave tomography (H. Nicolson *et al.* 2012, 2014) and a combination of both (E. Asencio *et al.* 2003). Recently, probabilistic inversion methods have also been applied to this region (E. Galetti *et al.* 2017; X. Zhao *et al.* 2022, 2026). In this test, we use seismic ambient noise data recorded by 61 seismometers around the British Isles (orange triangles in Fig. 2). These data are cross-correlated to estimate inter-receiver first arrival traveltimes, which are used as the observed data set in this study—a detailed description of the data processing is available in E. Galetti *et al.* (2017). We consider Love wave traveltimes at 10 s periods.

For inversion, the target area in Fig. 2 is discretized into a regular grid of 37×40 cells with a spacing of 0.33° in both longitude and latitude directions. A uniform prior distribution bounded between 1.5 to 4.8 km s^{-1} is defined for the Love wave group velocity in each cell: the mean of this uniform PDF is obtained by measuring the average velocity across all valid ray paths, and the upper and lower bounds are chosen to contain all possible velocities observed on the dispersion curves (E. Galetti *et al.* 2017). The likelihood function is formulated as a diagonal Gaussian distribution measuring the misfit between modelled and observed traveltimes and its uncertainty is set to be consistent with the results of E. Galetti *et al.* (2017).

Since the dimensionality of this problem (1480) is not prohibitively high, it is possible to implement all variational methods in the GeoPVI package, as well as Metropolis-Hastings Markov chain Monte Carlo (MH-McMC—W.K. Hastings 1970). We also apply Hamiltonian Monte Carlo (HMC—M.D. Hoffman & A. Gelman 2014; A. Zunino *et al.* 2023) to this problem as another independent method. Therefore, this example serves as a good benchmark example with real data to compare and analyse different inversion methods and to test the code package, and also allows comparison with previous results.

Fig. 3 shows the mean and standard deviation maps of the posterior PDF's obtained using the above suite of methods. Inside the receiver arrays where data have higher sensitivity, the average group velocity maps obtained using different methods are highly consistent and align closely with previous studies and the known geology of the British Isles (H. Nicolson *et al.* 2012, 2014; E. Galetti *et al.* 2017; X. Zhao *et al.* 2026). To further demonstrate that each method finds reasonable posterior samples (model realizations) and thus uncertainty structures, in Appendix A we display several posterior samples obtained from different methods.

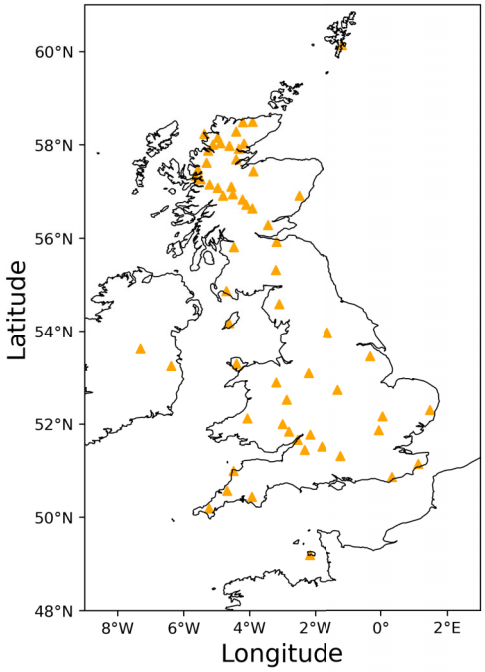

Figure 2. Study area around the British Isles used in the traveltime tomography example. The location of 61 seismometers are represented by triangles. Inter-receiver traveltimes are estimated by cross-correlating recorded ambient noise data using seismic interferometry (H. Nicolson *et al.* 2012, 2014; E. Galetti *et al.* 2017), and are used as the observed data in this test.

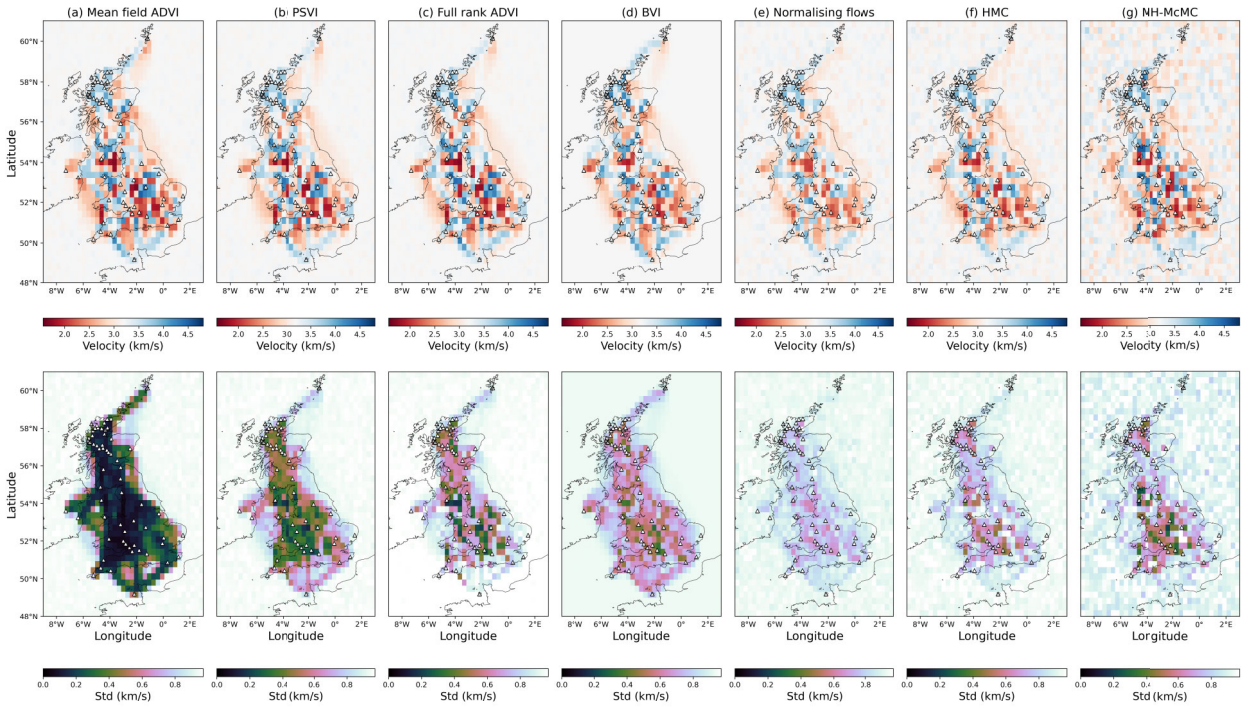

Figure 3. Mean (top row) and standard deviation (bottom row) maps of 10 s-period Love wave tomography results of the British Isles using different inversion methods indicated in the top titles. White triangles show the locations of receivers used in this example.

The standard deviation maps displayed from Figs 3(b) to (g) exhibit similar features, which differ from that in Fig. 3(a) obtained using mean field ADVI. This is because mean field ADVI ignores posterior correlations, resulting in a substantial underestimation of uncertainties; this is reflected in low posterior standard deviation values and leads to biased inversion results (X. Zhang *et al.* 2023). Different levels of posterior uncertainties displayed in each figure in the bottom row reveal different theoretical assumptions and inversion accuracy from the corresponding methods. In particular, standard deviation maps from HMC and normalizing flows are consistent. Given that we obtain similar results from two completely different methods, and that HMC is a Monte Carlo method which

Table 1. A comparison of computational cost in terms of number of forward evaluations required for different methods to provide the Love wave tomography results across the British Isles shown in Fig. 3. In the second column, N_{iter} represents number of iterations used in variational and Monte Carlo methods; N_{MC} stands for the number of samples to estimate the expectation term in eq. (3) in variational methods, and the number of Markov chains in HMC and MH-McMC. A special case is BVI, in which we use a mixture of 20 Gaussian components to approximate the true posterior distribution, resulting in a total number of 200 000 forward simulations. Total CPU time in the fourth column is estimated using an AMD EPYC 9654P processor, and is estimated based on per CPU core execution (i.e. the actual elapsed time multiplies the number of CPU cores used). Since the Markov chains used in MH-McMC have not fully converged (from Fig. 3g), the actual computational requirement for this method is likely to be even higher than that reported here.

Method	$N_{\text{iter}} \times N_{\text{MC}}$	Forward evaluations	CPU Time (hr)
Mean field ADVI	10 000 × 1	10 000	6.1
PSVI	10 000 × 1	10 000	6.3
Full rank ADVI	10 000 × 1	10 000	6.9
Normalizing flows	5000 × 20	100 000	78.3
BVI	5000 × 2 (20 components)	200 000	124.7
HMC	100 000 × 4	400 000	235.1
MH-McMC	1500 000 × 10	15 000 000	6600.0

is supposed to sample the true posterior distribution once the chains have converged, we conclude that the results from HMC represent a good reference solution to this Bayesian inversion problem. Therefore, by comparing standard deviation maps from different methods with that from HMC, we can analyse the performance of the corresponding methods.

For example, mean field ADVI, PSVI and full rank ADVI all employ a transformed Gaussian pdf, but they assume different inter-parameter correlation structures *a priori*, and thus produce different results. As expected, mean field ADVI and full rank ADVI provide the worst and best results among the three methods, respectively, with PSVI exhibiting intermediate performance. Nevertheless, they all provide lower posterior uncertainties compared to HMC, implying that the Gaussian assumption might not be appropriate for this real tomographic problem and so may underestimate posterior uncertainties as demonstrated in previous studies (X. Zhang *et al.* 2023). However, they are computationally efficient (see Table 1).

Normalizing flows and BVI are implemented in GeoPVI to address more complex inverse problems with potentially multimodal posterior distributions. In Fig. 3, the corresponding standard deviation maps show higher standard deviations compared to those from the three Gaussian-based methods. At certain spatial locations, posterior standard deviation values from BVI are lower than those from HMC, demonstrating that BVI may provide less accurate solution to this problem. This might be caused by a limited number of Gaussian components used to approximate the true posterior distribution in BVI. In this example, normalizing flows provide the best posterior solution among the tested variational methods. The standard deviation values from MH-McMC are smaller than those from HMC and normalizing flows. In addition, both mean and standard deviation maps from MH-McMC present more heterogeneous imaging noise compared to other results, notably in the offshore areas where the data offer little or no constraint so posterior uncertainties should be approximately laterally constant (since prior uncertainties are constant). This indicates that the chains have not fully converged even after 1.5 million iterations (Table 1) due to the high dimensionality (1480) of this problem, demonstrating the high computational cost of MH-McMC.

Table 1 compares the computational cost in this example. For parametric variational methods, inverse problems are solved by maximizing the ELBO in eq. (3) iteratively. Within each iteration, a small number of samples is used to approximate the ELBO and its gradient using Monte Carlo integration (A. Kucukelbir *et al.* 2017). Forward simulation of these samples is parallelized during inversion, each of which is performed using one CPU core. A special case is BVI, in which we use a mixture of 20 Gaussian components to approximate the true posterior PDF, each of which is optimized iteratively by maximizing the term in brackets in eq. (13). For MH-McMC, we run 10 independent Markov chains, and draw 1.5 million samples per chain. For each chain, the first 1 million samples are discarded as burn-in, and after the burn-in period every 100th sample is retained to approximate an ensemble of posterior samples. This results in a total number of 15 million forward simulations. For HMC, we run four chains with 100 000 iterations each, and discard the first 40 000 steps as burn-in period. For the two Monte Carlo methods, the sampling process involving multiple chains is parallelized, each chain using one CPU core.

In this example, different methods provide similar inversion results, but with quite different computational costs. Table 1 shows that all variational methods, including SVGD and stochastic SVGD (sSVGD) reported in X. Zhang & A. Curtis (2024), appear to be more efficient than the basic MH-McMC algorithm employed here. In MH-McMC, we used an adaptive Gaussian proposal distribution whose covariance matrix is calculated from previous samples in each Markov chain. Therefore, a large number of iterations is used to estimate a good proposal distribution and to reach convergence. Note that this MH-McMC algorithm is used only as a baseline to be consistent with several of our previous papers on this and related imaging problems. The number of iterations used in MH-McMC (and HMC) might also be biased due to the subjective detection of the convergence of Markov chains.

Variational methods and HMC require the calculation of the data-model gradient. In traveltime tomography, the gradient is calculated by tracing the fastest ray through the computed traveltime field. This process requires an effort equivalent to approximately 0.08 of a forward simulation (X. Zhang & A. Curtis 2024). Due to its relative insignificance, this computation is not accounted for in Table 1.

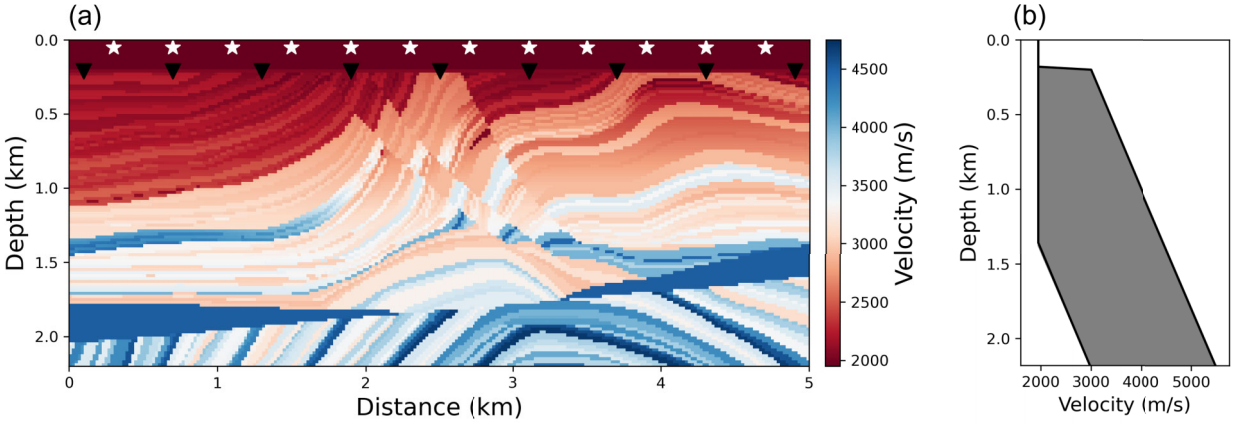

Figure 4. (a) True velocity model used in the 2-D full waveform inversion example. White stars represent locations of 12 shots, and black triangles sparsely denote the location of a receiver line (250 receivers are placed along this line from left to right). (b) Range of the uniform probability distribution used as prior information about the velocity value at each depth.

However, in other problems where gradient computation is expensive such as typically in full waveform inversion, these additional computations cannot be ignored.

4.2 2-D acoustic full waveform inversion

In this example, we perform Bayesian 2-D acoustic full waveform inversion (FWI) using the GeoPVI package. FWI uses full waveform information (both phase and amplitude) of the recorded seismograms to constrain subsurface properties, and offers higher resolution subsurface images compared to traveltime tomography. In acoustic FWI, the forward problem involves solving the constant-density acoustic wave equation:

$$\frac{1}{c^2(\mathbf{x})}\frac{\partial^2 p(\mathbf{x},t)}{\partial t^2} = \nabla^2 p(\mathbf{x},t) + s(\mathbf{x},t), \tag{15}$$

where t is time, $\mathbf{x}$ denotes a spatial location, $p(\mathbf{x},t)$ is the pressure field, symbol ∇^2 is the Laplacian operator and function $s(\mathbf{x},t)$ is a source term. Forward calculations are often carried out using a finite difference method (C. Bunks *et al.* 1995; D. Pasalic & R. McGarry 2010), finite element method (K.J. Marfurt 1984; M.J. Grote *et al.* 2006), pseudo analytical method (J.T. Etgen & S. Brandsberg-Dahl 2009) or pseudospectral method (B. Fornberg 1987, 1998).

Fig. 4 shows the true velocity model used as the target in this example, which is a part of the Marmousi model (G.S. Martin *et al.* 2006). The model is discretized using a regular grid of 110 $\times$ 250 cells with a spacing of 20 m in each direction. During inversion, velocities in the top ten rows of the grid are fixed at their true values. Twelve sources (white stars in Fig. 4a) are placed on the surface with a spacing of 400 m, and 250 receivers (sparsely represented by black triangles in Fig. 4a) are distributed evenly at 200 m depth. We use a finite difference method to generate synthetic waveform data of 4 s long using a Ricker wavelet with a dominant frequency of 10 Hz, and add uncorrelated Gaussian noise with zero mean and a standard deviation of 0.1 to each sample of the waveform. The resulting waveforms are used as the observed data to be inverted. The likelihood function is defined to be an uncorrelated Gaussian distribution with a standard deviation of 0.1 (i.e. we assume that noise level is known in the observed data). The prior distribution is the uniform distribution displayed in Fig. 4(b).

This FWI problem poses significant challenge due to its high dimensionality (25 000 parameters). Constructing and optimizing a full covariance matrix in full rank ADVI is impossible in terms of both memory requirement and computational cost. Similarly, implementing normalizing flows is costly as it requires large neural networks to be trained. Therefore, we restrict testing to mean field ADVI, PSVI and BVI in this example.

Fig. 5 shows the mean and standard deviation maps of the posterior PDF's obtained using these three methods. The mean models produce similar results, all resembling the true velocity model shown in Fig. 4(a). The standard deviation maps present different magnitudes of the posterior uncertainties, similarly to those observed in the traveltime tomography example in Fig. 3, due to the different theoretical assumptions implicit in the different methods. Mean field ADVI ignores all correlations between different parameters, thus provides underestimated uncertainty results; PSVI considers important (dominant) correlations and provides higher standard deviation values; BVI tends to provide a reasonably accurate posterior distribution for similar FWI problems (X. Zhao & A. Curtis 2024a) but requires a relatively large computational budget (see below). Nevertheless, the results from PSVI and BVI are similar, providing more confidence in the solution.

Table 2 lists the computational cost for these methods. In FWI, the gradient of the misfit function between observed and simulated waveform data with respect to a velocity model is typically computed by the adjoint state method (R.E. Plessix 2006). This requires

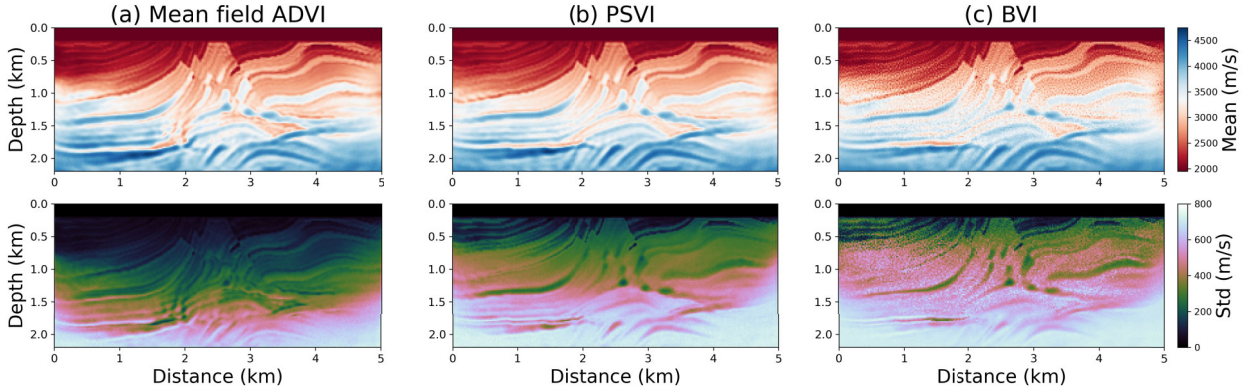

Figure 5. Mean (top row) and standard deviation (bottom row) maps of the posterior distribution obtained using (a) mean field ADVI, (b) PSVI and (c) boosting variational inference (BVI), respectively.

Table 2. A comparison of computational cost for different methods performed in the 2-D FWI example. In the second column, N_{iter} represents the number of iterations and N_{MC} stands for the number of samples used to perform the Monte Carlo approximation in variational methods. In this test, the data-model gradient is estimated using the adjoint state method, which requires an additional forward simulation (included in the tabulated values). CPU time in the fourth column is estimated using an Intel Xeon Max 9480 processor, and is estimated based on single CPU core execution as in Table 1.

Method	$N_{\text{iter}} \times N_{\text{MC}}$	Equivalent forward simulations	CPU Time (hr)
Mean field ADVI	5000 × 2	20 000	338
PSVI	5000 × 2	20 000	340
BVI	2500 × 2 (24 components)	240 000	4129

one additional forward simulation (for time-reversed modelling). Therefore, each iteration requires two simulations per model sample, resulting in the number of the equivalent forward simulations in Table 2. This table shows the advantage of PSVI: it delivers a reasonably accurate Bayesian solution with significantly improved efficiency (in this example one order of magnitude reduction in computation) compared to a more complex method (BVI). The cost is thus almost reduced to that of the commonly used (and significantly less accurate) mean field ADVI.

Finally, we briefly illustrate how to change prior information within the parametric variational framework using the method of variational prior replacement (M. Walker & A. Curtis 2014b; X. Zhao & A. Curtis 2024c). The parametric methods used here, combined with those in the two cited papers (especially the latter which introduces a vriational version of the prior replacement method), are particularly important for FWI problems in which forward simulations can be extremely costly. Similarly to the example in the latter paper above, say we want to replace the uniform prior PDF $p_u(\mathbf{m})$ used in the above FWI example by a smoothed prior distribution $p_s(\mathbf{m})$, to inject smoothness into the inverted models. If the posterior PDF corresponding to the uniform prior is obtained using parametric variational methods, we simply introduce a second variational distribution $q_s(\mathbf{m})$ to approximate the posterior PDF corresponding to the smoothed prior:

$$\begin{aligned} p_s(\mathbf{m}|\mathbf{d}_{\text{obs}}) &= \frac{p(\mathbf{d}_{\text{obs}}|\mathbf{m})p_u(\mathbf{m})}{p_u(\mathbf{d}_{\text{obs}})}\,\frac{p_s(\mathbf{m})}{p_u(\mathbf{m})}\,\frac{p_u(\mathbf{d}_{\text{obs}})}{p_s(\mathbf{d}_{\text{obs}})}, \\ &\approx k\, q_u(\mathbf{m})\,\frac{p_s(\mathbf{m})}{p_u(\mathbf{m})}, \\ &\approx q_s(\mathbf{m}). \end{aligned} \tag{16}$$

The first line is Bayes' rule applied to the smoothed prior pdf $p_s(\mathbf{m})$ and data $\mathbf{d}_{\text{obs}}$, written in terms of the Bayesian posterior obtained with the uniform prior (first term on the right). The second line is obtained by introducing $q_u(\mathbf{m}) \approx p_u(\mathbf{m}|\mathbf{d}_{\text{obs}})$ (i.e. the variational approximation to the solution with a uniform prior) and defining a constant $k = p_u(\mathbf{d}_{\text{obs}})/p_s(\mathbf{d}_{\text{obs}})$. In the third line, we introduce and optimize another variational distribution $q_s(\mathbf{m})$ to approximate the new posterior distribution corresponding to the smoothed prior PDF, by minimizing the KL divergence between two PDF's denoted in the second and third lines. Since the old variational distribution $q_u(\mathbf{m})$ is parametric, its density value can be evaluated directly without any new forward simulation. Therefore, the new variational PDF $q_s(\mathbf{m})$ can be obtained from $q_u(\mathbf{m})$ with almost zero computational effort compared to solving a variational Bayesian inversion problem from scratch. While wave propagation is a highly nonlinear physics problem and is sensitive to small scatterers, X. Zhao & A. Curtis (2024c) demonstrated that in spatially 2-D FWI problems, results from variational prior replacement are reasonably accurate compared to those from independent Bayesian inversion. Its performance becomes worse in spatially 3-D problems (X. Zhao & A. Curtis 2025a), yet can be improved by a refinement step with a relatively small amount of additional computation.

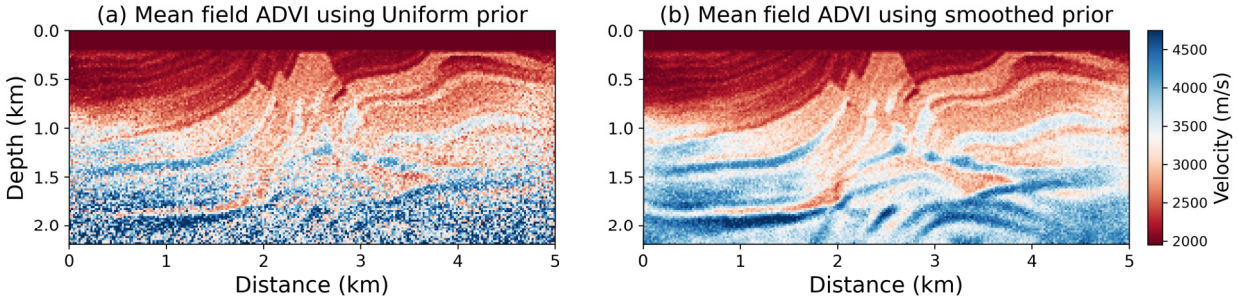

Figure 6. One posterior sample obtained from the mean field ADVI solution found using (a) a uniform prior distribution and (b) a smoothed prior distribution. The latter is obtained from the former at low cost using the method of variational prior replacement.

We apply the method to the inversion results from mean field ADVI displayed in Fig. 5(a). Fig. 6 shows random samples drawn from the two posterior PDF's, respectively. The sample in Fig. 6(b) is smoother, exhibiting far less spatially abrupt heterogeneity after replacing the prior information. This efficient update of the prior information cannot be achieved if the posterior PDF corresponding to the uniform prior is not constructed using a parametric variational method.

4.3 Surface wave dispersion inversion

The third example is an inversion of 1-D shear wave velocity structure from observations of surface wave group or phase velocity data—a common problem in seismology referred to as surface wave dispersion inversion. When propagating through a vertically heterogeneous medium like the solid Earth, seismic surface waves show dispersive properties (K. Aki & P.G. Richards 2002). This means that different frequencies of waves travel at different speeds because they are sensitive to different depth ranges—and thus different elastic properties—of the Earth's interior. This phenomenon is typically represented by a dispersion curve or relation, which describes the speed of the wave as a function of frequency. We can invert for shear wave velocity profiles as a function of depth beneath the Earth's surface using dispersion curves extracted from recorded seismograms (e.g. using the methods in Section 4.1 for a range of frequencies), since surface wave dispersion data are primarily sensitive to shear wave velocity variations.

In dispersion inversion, the forward problem involves modelling surface wave dispersion curves corresponding to a given layered shear wave velocity model, which can be accomplished using the SURF96 program (R.B. Herrmann 2013). Since the dimensionality of a typical dispersion inversion problem is relatively low (commonly of the order of 10), we use a finite difference method to estimate the gradient of modelled dispersion curves with respect to shear wave velocities: we perturb each shear velocity parameter in turn by a small amount, then calculate the change in simulated surface wave dispersion.

We consider a synthetic example in which the subsurface is parametrized using 10 layers. The top nine layers each has a thickness of 0.1 km, and the final layer extends to infinite depth to represent a half-space. Fig. 7(a) displays the true shear wave velocity profile as a function of depth. We model Rayleigh wave phase velocities across 14 periods (frequencies), and add uncorrelated Gaussian random noise with a magnitude of 0.02 km s^{-1}. These data form the observed Rayleigh wave phase velocity dispersion curve, as shown by a black line in Fig. 7(b). For each layer, we define a uniform prior distribution for the shear wave velocity; the lower and upper bounds are represented by dashed blue lines in Fig. 7(a), which encompass the true velocity model. The likelihood function is set to be a diagonal Gaussian distribution with a standard deviation of 0.004 km s^{-1} for each data point (black errorbars in Fig. 7b) to represent a real situation in which we underestimate the noise level in observed data.

We apply full rank ADVI and normalizing flows to this example, since these two methods are not implemented in the 2-D FWI example due to the high dimensionality of that problem. To validate the inversion results, we conduct a Monte Carlo inversion which involves three chains, each generating 150 000 samples, including a burn-in period of 50 000 samples. We discard the burn-in samples, and select every 20th of the remaining 100 000 samples. These retained samples form the final set of McMC samples used to calculate the posterior statistics. Note that we run Monte Carlo with a large number of iterations to ensure that the chains have converged and that the obtained results can be served as a reference solution to this Bayesian problem. This number is reasonable compared to other Monte Carlo dispersion inversion studies (E. Galetti *et al.* 2017; E. Crowder *et al.* 2021; F. Magrini *et al.* 2025).

Fig. 8 displays the posterior marginal distributions of the inversion results for different layers. The results from McMC and normalizing flows are highly consistent, while the results from full rank ADVI exhibit notable differences. Full rank ADVI seems to underestimate posterior uncertainties, as evidenced by the narrower marginal posterior PDF's compared to those from the other methods. This suggests that the true posterior distribution has heavier tails than a Gaussian PDF. In this example, the fourth layer represents a low velocity anomaly situated between two relatively higher velocity layers. Both McMC and normalizing flows recover this feature; full rank ADVI skews the corresponding posterior marginal distribution towards low values but fails to demonstrate a clear decrease in mean velocity with depth. To further demonstrate the accuracy of the inversion results from normalising flows, we draw 80 posterior samples and calculate the corresponding synthetic dispersion curves, as displayed by red lines in Fig. 7(b). All synthetic curves

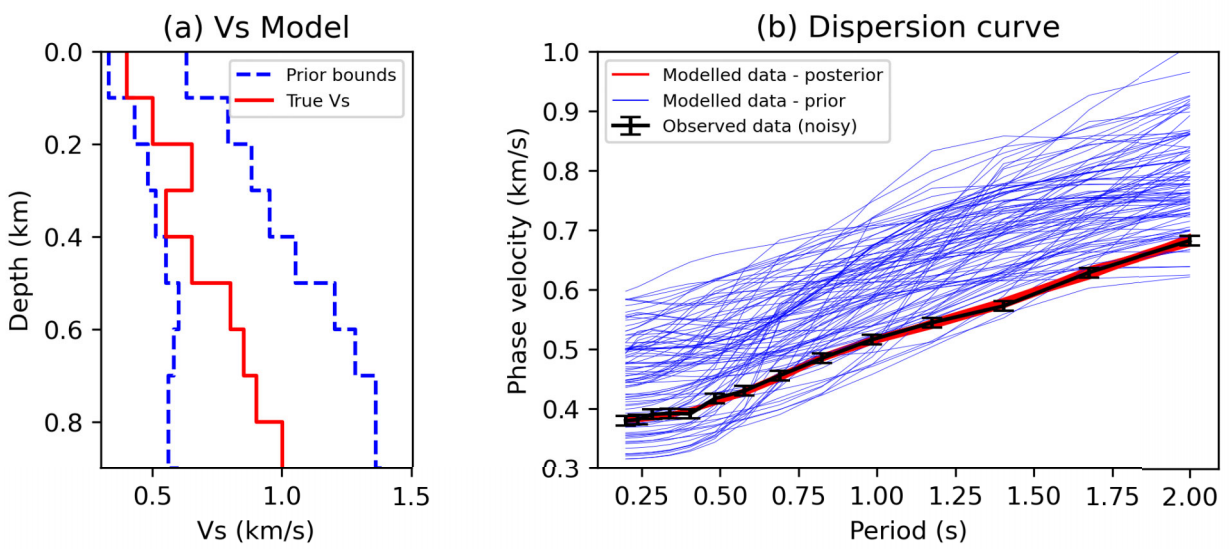

Figure 7. (a) True shear wave velocity profile (red curve) and the bounds on the uniform prior distribution (dashed blue curves) in each layer. (b) Observed Rayleigh wave dispersion curve (black) and 80 modelled dispersion curves predicted by 80 prior samples (blue) and 80 posterior samples from the posterior distribution estimated using normalizing flows (red). Black errorbars show data uncertainties used to define the likelihood function in this example.

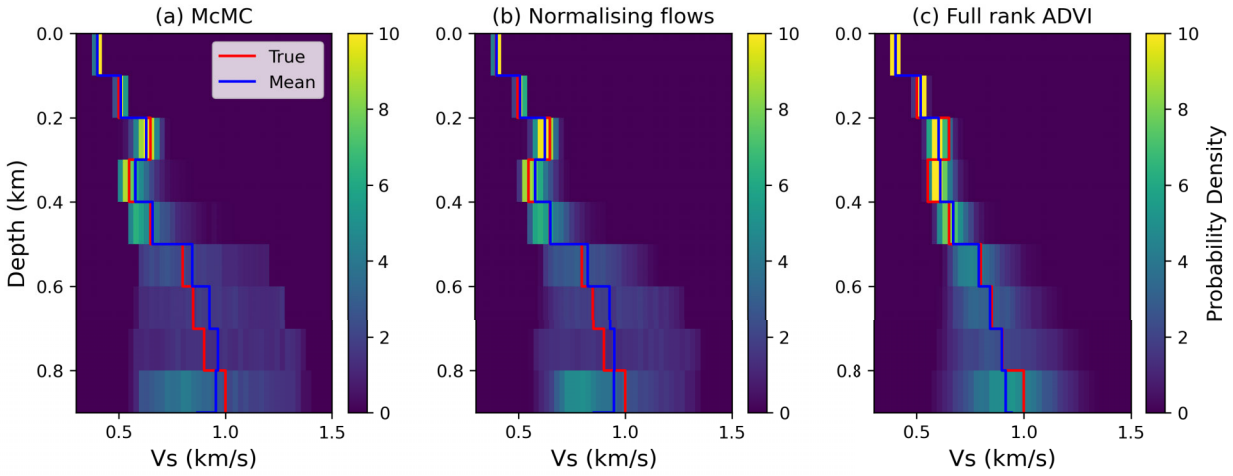

Figure 8. Posterior marginal distributions of shear velocity at each depth from the surface wave dispersion inversion example, obtained using (a) Markov chain Monte Carlo (McMC), (b) normalizing flows, and (c) full rank ADVI. Red lines show the true shear velocity and blue lines show the posterior mean velocity profile.

align well with the observed data (black line). Blue lines show 80 synthetic dispersion curves simulated by 80 prior samples. Data fit is improved significantly after the inversion. Fig. 9 shows the posterior correlation coefficients between parameters estimated from the three methods. Again, the results are highly consistent, particularly between the results shown in Figs 9(a) and (b), which are almost identical. Since we can treat the results from Monte Carlo as the true solution to a Bayesian inverse problem if that algorithm has reached a reasonable state of convergence, and since we obtain nearly the same posterior statistics using McMC and normalizing flows, it is reasonable to assume that the results from normalizing flows are accurate.

Table 3 compares the computational cost of these three methods. Note that the data-model gradient is estimated using a finite difference method, requiring 10 additional forward simulations per iteration to perturb velocity values in the 10 different layers. The equivalent forward simulation number from variational methods is therefore scaled by a factor of 11 for a direct comparison with those from MH-McMC. Despite this scaling, in this example variational methods can still provide almost identical (statistical properties of the) inversion results compared to those from our implementation of MH-McMC which uses more samples. For higher dimensional inverse problems with more efficient gradient computation methods (e.g. those presented in previous sections), any efficiency gain provided by variational methods becomes more significant.

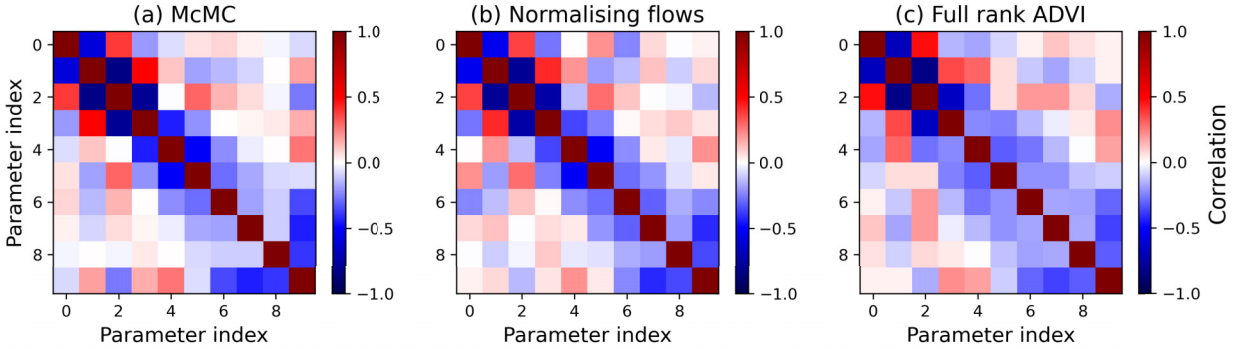

Figure 9. Posterior correlation coefficients between shear wave velocities of different layers in the surface wave dispersion inversion example, obtained using (a) Markov chain Monte Carlo (McMC), (b) normalizing flows and (c) full rank ADVI. In each panel, parameter indices on both axes denote layer indices from top to bottom, and the value in each pixel represents the posterior correlation between the two indexed layers.

Table 3. A comparison of computational cost for different methods used in the dispersion inversion example. In the second column, N_{iter} represents the number of iterations used in variational and McMC methods; N_{MC} stands for the number of samples used to perform the Monte Carlo approximation in variational methods, and the number of Markov chains in McMC. In this test, the data-model gradient is estimated using finite differences, which requires 10 additional forward simulations to perturb velocity values in 10 different layers. Therefore, each iteration requires 11 simulations per model sample, in which one simulation is used to estimate the corresponding data and another ten to obtain derivatives. CPU time is estimated using an AMD EPYC 9654P processor, and is estimated based on single CPU core execution.

Method	$N_{\text{iter}} \times N_{\text{MC}}$	Equivalent forward simulations	CPU time (s)
Full rank ADVI	5000×2	110 000	27
Normalizing Flows	$10\,000 \times 2$	220 000	97
MH-McMC	$150\,000 \times 3$	450 000	158

4.4 Vertical electrical sounding

The last example is a vertical electrical sounding (VES) inversion, a method used to estimate 1-D layered subsurface electrical resistivity structure using measurements of voltage from an electrical field induced by distant grounded electrodes. VES is primarily employed for exploration of subsurface water resources, and can be useful in many environmental and engineering areas. In VES, apparent resistivity measurements are typically obtained by deploying an array of current electrodes or potential electrodes into the ground, at a set of probe spacings. This creates an electrical field which allows potential differences to be measured, and apparent electrical resistivity to be calculated. By varying the electrode spacings, the observations are sensitive to subsurface resistivity over different depth ranges.

We consider a four-layered subsurface model. The thickness and resistivity values for each layer are displayed by a blue line in Fig. 10(a). The third layer represents a near surface water layer with a relatively low resistivity and the fourth layer extends to infinite depth, representing a half-space. We set up eight electrode spacings to measure apparent resistivity values, with each measurement subjected to 2 per cent of uncorrelated Gaussian random noise, as presented by a blue line in Fig. 10(e). The result is used as the observed data set in this test. For inversion, we treat both the thickness of the first three layers and the resistivity values of all layers as unknown, resulting in seven model parameters to be inverted. A uniform prior distribution is defined for each parameter, and the lower and upper bounds are represented by the range of each horizontal axis displayed in Figs 10(b) and (c). Note that in Figs 10(a) and (e), we use logarithmic axis scales to better display the results, and the actual prior parameter values are defined without applying an exponential transform. The forward simulation is performed through the pyGIMLi library (C. Rücker *et al.* 2017).

We employ normalizing flows for this simple example, and display the inversion results in Fig. 10. A dashed red line in Fig. 10(a) shows the inverted mean resistivity profile over depth, which resembles the true model especially within the top two layers. Figs 10(b) and (c) show the posterior marginal PDF's for layer thickness and resistivity values, respectively. Notably, the true values fall within the high probability regions of the posterior marginal distributions, except for the third layer (the water layer) for which biases towards larger values are observed in both thickness and resistivity parameters. This is reasonable in reality since a low resistivity layer that is thin and a higher resistivity layer that is thicker may fit the apparent resistivity data almost equally well. This is verified by Fig. 10(d), in which the joint posterior probability distribution of the third layer's thickness and resistivity values are displayed. Positive correlations are observed between these two parameters, confirming the above proposition. Fig. 10(e) compares observed and synthetic data predicted using the inverted mean model (dashed red line in Fig. 10a). High consistency between these two curves proves the accuracy of the inversion results. The posterior marginal distribution for the resistivity value of the fourth layer remains broadly similarly to the uniform prior distribution, which likely stems from the limited range of the current electrode spacings applied during the experiment, thus capturing little information about this deep layer. All of these features indicate that the inversion results capture reasonable posterior uncertainty information.

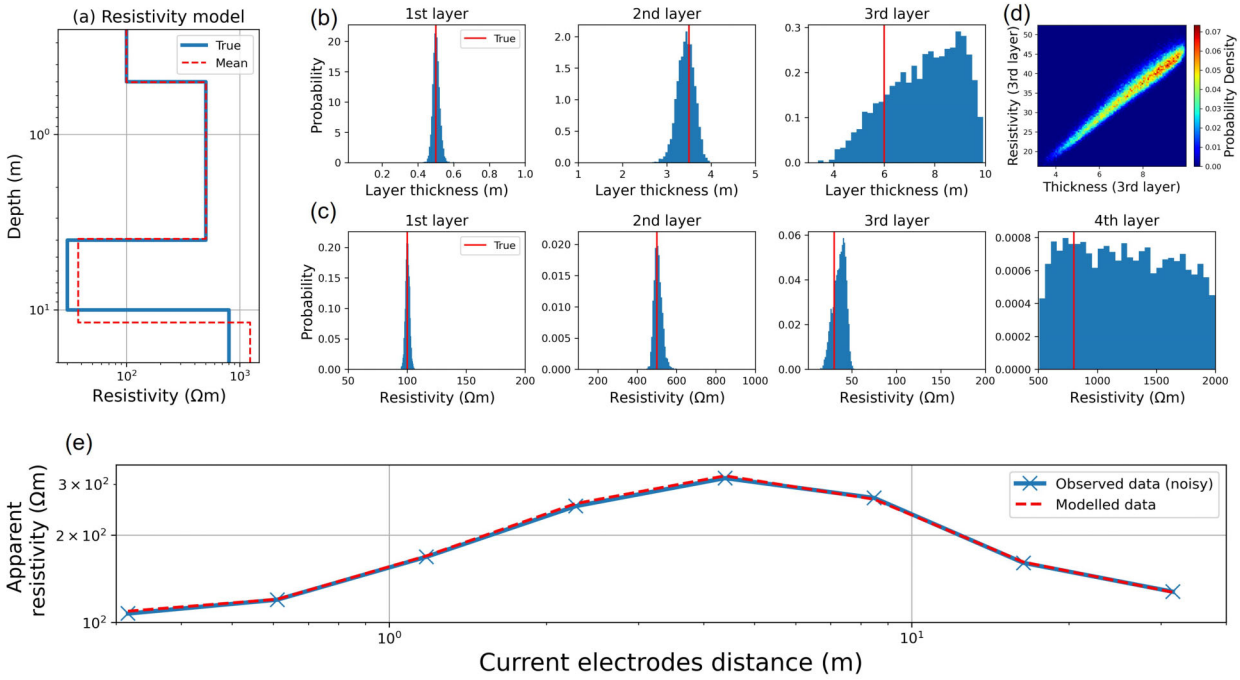

Figure 10. (a) True (blue line) and inverted mean (dashed red line) resistivity profiles over depth. (b) Posterior marginal PDF's for layer thickness of the top three layers. (c) Posterior marginal PDF's for resistivity value of all four layers. (d) Joint posterior distribution of layer thickness and resistivity values of the third layer. (e) Observed and modelled apparent resistivity values (the latter predicted using the inverted mean model displayed in panel (a)) as a function of distance between two current electrodes. Uncertainties on the true data are 2 per cent (too small to represent clearly in the figure).

5 DISCUSSION

The GeoPVI package incorporates several parametric variational inference methods, each with distinct algorithmic complexities to approximate the posterior PDF. In our experience, the relative performance of all of the methods illustrated in Fig. 3 is typical. Mean field ADVI is usually computationally cheapest to reach convergence, but the method employs a mean field approximation, which often creates uncertainty estimates that are too low. PSVI offers improved accuracy by incorporating the *a priori* most dominant correlations between model parameters. If the dimensionality of an inverse problem is not too high such that modelling a full covariance matrix is feasible, then full rank ADVI would normally be the best choice out of these three methods. However, all of these methods rely on a (transformed) Gaussian distribution and can not represent posterior distributions accurately in complex problems characterized by multimodal posterior distributions. Normalizing flows and BVI can be deployed under these circumstances, although they require greater computational resources (Tables 1 to 3). Therefore, as with any class of methods, one always has to choose an algorithm that balances efficiency and accuracy for any given inverse problem. This can be proven, and is summarized by the No Free Lunch theorem which states that no method is better than any other method when averaged across all possible problems (D.H. Wolpert & W.G. Macready 1997). However, for any particular subclass of problems, there do exist better and worse algorithms. Therefore, in the GeoPVI package we provide a suite of methods with different efficiencies and accuracies, which allows all methods to be tested easily and efficiently, and/or tailored decisions to be made about which method to use based on the nature of the problem at hand.

In GeoPVI, the choice of many hyperparameters affects its performance, some of which have a larger effect than others. Different algorithms have different assumptions, performance and computational requirements, so users should choose a particular algorithm based on the complexity of the problem to be solved and their available computational resources. For BVI, the number of components affects its performance and efficiency. Given that each new component is added sequentially in a greedy manner, X. Zhao & A. Curtis (2024a) suggested a practical way to determine the component number, in which statistical properties of the approximate posterior distribution are checked after each component is optimized and the algorithm is stopped if no obvious changes in those statistical properties two successive BVI iterations is observed. For normalizing flows, the flows structure is an important factor. Generally, a deeper flows model performs better than a wider one (I. Kobyzev *et al.* 2020; G. Papamakarios *et al.* 2021). The number of flows (transforms) determines the accuracy and efficiency of the method, and should be chosen based on the particular problem and computational resources. Based on our tests, transforms with high nonlinearity (e.g. C. Durkan *et al.* 2019) often perform better than simpler ones (e.g. L. Dinh *et al.* 2015, 2017). For PSVI, the size and structure of the correlation kernel imposed to the posterior solution are key factors, which can be designed based on the sensitivity of data to the underlying physics. Nevertheless from various problems that we tested, the quality of the PSVI solution does not change much while enlarging the size of the correlation kernel after a certain size (e.g. X. Zhao & A. Curtis 2024b), meaning that it is not wise to define the kernel to be as large as possible.

Other hyperparameters related to variational optimization also affect the performance of GeoPVI; they generally have smaller effects compared to the above ones, yet their selection is also more subjective. Here we provide some criteria that we used when applying GeoPVI (however note that they are still subjective). First, we normally determine the total number of samples used for inversion based on the overall computational budget. Then, the number of iterations and number of samples per iteration used for variational optimization can be decided by computational hardware (e.g. number of CPU and GPU) available. Specifically, to save wall-clock (elapsed) time of variational inversion, we suggest selecting the number of samples per iteration such that (almost) full computational capacity can be used at the same time; the number of iterations can then be determined accordingly. As in many modern deep learning algorithms, a learning rate affects the efficiency and accuracy. Nevertheless, the use of adaptive and stochastic optimization algorithms reduces the sensitivity of the results to the learning rate (D.P. Kingma & J. Ba 2014).

Non-parametric variational methods, such as SVGD (Q. Liu & D. Wang 2016) and sSVGD (V. Gallego & D.R. Insua 2018), provide a set of posterior samples without explicit parametric expressions for the modelled variational distribution. The samples produced by these methods might more easily capture higher order statistics describing trade-offs in the posterior PDF than the methods analysed herein (X. Zhang *et al.* 2023). When using parametric variational methods, we would need to construct a variational model designed to achieve this purpose, for example by designing a (predefined) posterior correlation structure in PSVI.

For nonlinear inverse problems, especially those of high dimensionality, accurate prior information is crucial if we wish to find a useful and correct posterior solution. Currently in the GeoPVI package we implement uniform and normal prior distributions. Additional options for incorporating smoothing prior information are available in the package. We provide a template that enables users to define their own prior PDF's easily, such that more sophisticated geological prior information can be injected into variational inversion (M. Walker & A. Curtis 2014a; L. Mosser *et al.* 2020; S. Levy *et al.* 2022; H. Bloem *et al.* 2024; M. Liu *et al.* 2024; G.A. Meles *et al.* 2025). In GeoPVI, examples are also provided to perform variational prior replacement to update (replace) prior information efficiently post-Bayesian inversion. Note that to ensure that variational optimization can be performed properly, the probability density value (and its gradient) of the prior distribution used in GeoPVI needs to be evaluated explicitly.

Several software packages have been developed to address Bayesian geophysical inverse problems, each tailored to unique aspects and challenges. HMCLab (A. Zunino *et al.* 2023) employs Hamiltonian Monte Carlo (R.M. Neal *et al.* 2011; A. Fichtner *et al.* 2019) to solve a diverse array of inversion problems; BayesBay (F. Magrini *et al.* 2025) is a Python package designed for generalized transdimensional and hierarchical Bayesian inference (M. Sambridge *et al.* 2006; T. Bodin & M. Sambridge 2009; T. Bodin *et al.* 2012); CoFI (J. He *et al.* 2024) provides a common framework that bridges generic inference algorithms with specific geoscientific applications; InvertibleNetworks package (R. Orozco *et al.* 2023) is developed to implement invertible neural networks and can be used to perform amortized variational inference (Z. Yin *et al.* 2025). X. Zhang & A. Curtis (2024) developed VIP, a variational inversion code package for Bayesian tomographic imaging problems using SVGD, sSVGD and ADVI. In this work, we introduce GeoPVI as a systematic framework to implement various parametric variational inference methods, and variational prior replacement, efficiently. All of these packages make it easier for researchers in the geophysical community to apply advanced nonlinear inversion methods to their own inference problems, promoting enhanced understanding and innovative solutions in geophysical research.

Note that one limitation for GeoPVI (and potentially for almost all of the packages mentioned in the Introduction) is that the prior probability density values of posterior samples need to be evaluated explicitly. Therefore, in this paper and the current package, relatively simple prior PDF's are defined and used. It is not straightforward to employ some prior distributions from which it is easy to generate samples yet difficult to evaluate their probability density values, such as those defined by geostatistics (G. Mariethoz & J. Caers 2014), geological process models (D. Tetzlaff 2023) and deep generative neural networks (D.P. Kingma & M. Welling 2014; I. Goodfellow *et al.* 2014; J. Ho *et al.* 2020). One possible solution is to use neural density estimators to learn approximate prior PDF's produced by each of these methods (e.g. C.M. Bishop 1994; G. Papamakarios 2019; G. Papamakarios *et al.* 2021). However, these methods are generally difficult to scale to high dimensional problems such as FWI. Future work could investigate possible solutions for this problem.

6 CONCLUSION

We introduce GeoPVI, a Python package designed for implementing Bayesian probabilistic inverse problems using a range of parametric variational inference methods. These methods use parametric and explicit expressions to construct approximate posterior probability distributions. GeoPVI includes several parametric variational methods that have been tested in geophysics, including mean field ADVI, PSVI, full rank ADVI, normalizing flows and BVI. We apply all of the above methods to a Love wave traveltime tomography example across the British Isles, which illustrates how inversion accuracy and efficiency can be driven by the different theoretical assumptions in each method. This example serves as a guide for how users might select the most suitable algorithms based on their particular problem to be solved. In a 2-D acoustic full waveform inversion example, we show that GeoPVI is able to solve high dimensional and highly nonlinear inverse problems. We further demonstrate that the parametric nature of the inversion results can be used to update prior information post-inversion with almost no additional computation. Application of the package to a surface wave dispersion inversion example proves that variational methods can provide almost identical results to those from Monte Carlo methods.

A simple vertical electrical sounding experiment confirms the framework's ability to accurately capture posterior uncertainty information in problems with entirely different (non-seismological) physics. We hope that the GeoPVI package will have a wide range of applications in geosciences in the future.

ACKNOWLEDGMENTS

We thank the Edinburgh Imaging Project (EIP—https://blogs.ed.ac.uk/imaging/) sponsors (BP and TotalEnergies) for supporting this research. For the purpose of open access, we have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

DATA AVAILABILITY

A tagged version of the GeoPVI package at the time of publication of this paper can be accessed on GitHub. https://github.com/XuebinZhaoZXB/GeoPVI/releases/tag/v2.0), which contains the full package, a user manual and all working examples presented in this paper. Traveltime data used in the Love wave tomography example are obtained from E. Galetti *et al.* (2017). Raw ambient noise seismic data can be accessed at the British Geological Survey (https://eida.bgs.ac.uk/).

REFERENCES

Afanasiev, M., Boehm, C., van Driel, M., Krischer, L., May, D., Rietmann, M. & Fichtner, A., 2017. Salvus: a flexible high-performance and open-source package for waveform modelling and inversion from laboratory to global scales. *EGU General Assembly Conference Abstracts,* Vienna, Austria, p. 9456.

Aki, K. & Richards, P.G., 2002. Quantitative Seismology, University Science Books.

Arnold, R. & Curtis, A., 2018. Interrogation theory, *Geophys. J. Int.,* **214**(3), 1830–1846.

Arrowsmith, S.J., Kendall, M., White, N., VanDecar, J.C. & Booth, D.C., 2005. Seismic imaging of a hot upwelling beneath the British Isles, *Geology,* **33**(5), 345–348.

Asencio, E., Knapp, J.H., Owens, T.J. & Helffrich, G., 2003. Mapping fine-scale heterogeneities within the continental mantle lithosphere beneath scotland: Combining active-and passive-source seismology, *Geology,* **31**(6), 477–480.

Bates, O., Guasch, L., Strong, G., Robins, T.C., Calderon-Agudo, O., Cueto, C., Cudeiro, J. & Tang, M., 2022. A probabilistic approach to tomography and adjoint state methods, with an application to full waveform inversion in medical ultrasound, *Inverse Probl.,* **38**(4), 045008.

Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances., *Phil. Trans. R. Soc. London,* **1763,** 370–418.

Bishop, C.M., 1994. Mixture Density Networks, *Aston University*.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning, Springer.

Blei, D.M., Kucukelbir, A. & McAuliffe, J.D., 2017. Variational inference: a review for statisticians, *J. Am. Stat. Assoc.,* **112**(518), 859–877.

Bloem, H., Curtis, A. & Tetzlaff, D., 2024. Introducing conceptual geological information into Bayesian tomographic imaging, *Basin Res.,* **36**(1), e12811.

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., 2012. Transdimensional tomography with unknown data noise, *Geophys. J. Int.,* **189**(3), 1536–1556.

Boyd, S. & Vandenberghe, L., 2004. Convex optimization, Cambridge Univ. Press.

Bunks, C., Saleck, F.M., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics,* **60**(5), 1457–1473.

Chen, P. & Ghattas, O., 2020. Projected Stein variational gradient descent, *Adv. Neur. Inform. Process. Syst.,* **33,** 1947–1958.

Chen, H., Cheng, W., Wang, L., Zhou, H. & Yu, B., 2025. Variational inference full-waveform inversion using geostatistical prior particles, *Geophysicists,* Society of Exploration Geophysicists.

Cockett, R., Kang, S., Heagy, L.J., Pidlisecky, A. & Oldenburg, D.W., 2015. SimPEG: An open source framework for simulation and gradient based parameter estimation in geophysical applications, *Comput. Geosci.,* **85,** 142–154.

Crowder, E., Rawlinson, N., Cornwell, D., Sammarco, C., Galetti, E. & Curtis, A., 2021. New insights into north sea deep crustal structure and extension from transdimensional ambient noise tomography, *Geophys. J. Int.,* **224**(2), 1197–1210.

Curtis, A. & Lomax, A., 2001. Prior information, sampling distributions, and the curse of dimensionality, *Geophysics,* **66**(2), 372–378.

D'Angelo, F. & Fortuin, V., 2021. Annealed Stein variational gradient descent, *Third Symposium on Advances in Approximate Bayesian Inference.*

Dinh, L., Krueger, D. & Bengio, Y., 2015. NICE: Non-linear independent components estimation, preprint (arXiv:1410.8516).

Dinh, L., Sohl-Dickstein, J. & Bengio, S., 2017. Density estimation using real NVP, *International Conference on Learning Representations.*

Durkan, C., Bekasov, A., Murray, I. & Papamakarios, G., 2019. Neural spline flows, Advances in Neural Information Processing Systems, **32,** 7509–7520, eds, Wallach, H., Larochelle, H., Beygelzimer, A., Fox, E. & Garnett, R.

Earp, S. & Curtis, A., 2020. Probabilistic neural network-based 2D travel-time tomography, *Neur. Comput. Appl.,* **32**(22), 17077–17095.

Ely, G., Malcolm, A. & Poliannikov, O.V., 2018. Assessing uncertainties in velocity models and images with a fast nonlinear uncertainty quantification method, *Geophysics,* **83**(2), R63–R75.

Etgen, J.T. & Brandsberg-Dahl, S., 2009. The pseudo-analytical method: Application of pseudo-Laplacians to acoustic and acoustic anisotropic wave propagation, *SEG Technical Program Expanded Abstracts 2009,* Society of Exploration Geophysicists, pp. 2552–2556.

Fichtner, A., Zunino, A. & Gebraad, L., 2019. Hamiltonian Monte Carlo solution of tomographic inverse problems, *Geophys. J. Int.,* **216**(2), 1344–1363.

Foks, N.L. & Minsley, B.J., 2020. GeoBIPy - Geophysical Bayesian Inference in Python,.

Fornberg, B., 1987. The pseudospectral method: Comparisons with finite differences for the elastic wave equation, *Geophysics,* **52**(4), 483–501.

Fornberg, B., 1998. A practical guide to pseudospectral methods, no. 1, Cambridge Univ. Press.

Galetti, E., Curtis, A., Baptie, B., Jenkins, D. & Nicolson, H., 2017. Transdimensional Love-wave tomography of the British Isles and shear-velocity structure of the East Irish Sea Basin from ambient-noise interferometry, *Geophys. J. Int.,* **208**(1), 36–58.

Gallego, V. & Insua, D.R., 2018. Stochastic gradient MCMC with repulsive forces, preprint (arXiv:1812.00071).

Giaquinto, R. & Banerjee, A., 2020. Gradient boosted normalizing flows, *Adv. Neural Inform. Process. Syst.,* **33,** 22104–22117.
Gong, W., Li, Y. & Hernández-Lobato, J.M., 2020. Sliced kernelized Stein discrepancy, preprint (arXiv:2006.16531).
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y., 2014. Generative adversarial nets. Advances in neural information processing systems **27,** 2672 –268, eds, Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. & Weinberger, KQ., Curran Associates, Inc.
Grote, M.J., Schneebeli, A. & Schötzau, D., 2006. Discontinuous galerkin finite element method for the wave equation, *SIAM J. Numer. Anal.,* **44**(6), 2408–2431.
Guo, F., Wang, X., Fan, K., Broderick, T. & Dunson, D.B., 2016. Boosting variational inference, preprint (arXiv:1611.05559)
Hansen, T.M., Cordua, K.S., Looms, M.C. & Mosegaard, K., 2013. SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 2 - Application to crosshole GPR tomography, *Comput. Geosci.,* **52,** 481–492.
Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika,* **57**(1), 97–109.
He, J., Hauser, J., Sambridge, M., Magrini, F., Valentine, A. & Marignier, A., 2024. CoFI-Linking geoscience inference problems with tools for their solution, *European Geosciences Union General Assembly 2024 (EGU24),* 14483. Tech. Rep., Copernicus Meeting, https://inlab.au/cofi
Herrmann, R.B., 2013. Computer programs in seismology: An evolving tool for instruction and research, *Seismol. Res. Lett.,* **84**(6), 1081–1088.
Ho, J., Jain, A. & Abbeel, P., 2020. Denoising diffusion probabilistic models, *Adv. Neur. Inform. Process. Syst.,* **33,** 6840–6851.
Hoffman, M.D. & Gelman, A., 2014. The no-u-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo., *J. Mach. Learn. Res.,* **15**(1), 1593–1623.
Izzatullah, M., Alali, A., Ravasi, M. & Alkhalifah, T., 2024. Physics-reliable frugal local uncertainty analysis for full waveform inversion, *Geophys. Prospect.,* **72**(7), 2718–2738.
Kingma, D.P. & Ba, J., 2014. ADAM: A method for stochastic optimization, preprint (arXiv:1412.6980).
Kingma, D.P. & Dhariwal, P., 2018. Glow: Generative flow with invertible 1x1 convolutions, Advances in Neural Information Processing Systems, **31,** 10215 –10224, eds, Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N. & Garnett, R., Curran Associates, Inc.
Kingma, D.P. & Welling, M., 2014. Auto-encoding variational Bayes, preprint (arXiv:1312.6114).
Kingma, D.P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I. & Welling, M., 2016. Improved variational inference with inverse autoregressive flow. *Advances in neural information processing systems,* **29,** pp.4743–4751, eds, Lee, S., Sugiyama, M., Luxburg, U., Guyon, I. & Garnett, R.
Kobyzev, I., Prince, S.J. & Brubaker, M.A., 2020. Normalizing flows: an introduction and review of current methods, *IEEE Trans. Pattern Anal. Mach. Intell.,* **43**(11), 3964–3979.
Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A. & Blei, D.M., 2017. Automatic differentiation variational inference, *J. Mach. Learn. Res.,* **18**(1), 430–474.
Kullback, S. & Leibler, R.A., 1951. On information and sufficiency, *Ann. Math. Stat.,* **22**(1), 79–86.
Levy, S., Laloy, E. & Linde, N., 2022. Variational Bayesian inference with complex geostatistical priors using inverse autoregressive flows, *Comput. Geosci.,* 105263.
Liu, M., Grana, D. & Mukerji, T., 2024. Geostatistical inversion for subsurface characterization using stein variational gradient descent with autoencoder neural network: an application to geologic carbon sequestration, *J. geophys. Res.: Solid Earth,* **129**(7), e2024JB029073.
Liu, Q. & Wang, D., 2016. Stein variational gradient descent: A general purpose Bayesian inference algorithm, Advances in neural information processing systems, **29,** pp.2378–2386, eds, Lee, D., Sugiyama, M., Luxburg, U., Guyon, I. & Garnett, R., Curran Associates, Inc.
Locatello, F., Dresdner, G., Khanna, R., Valera, I. & Rätsch, G., 2018. Boosting black box variational inference, *Advances in Neural Information Processing Systems,* **31**.
Lomas, A., Luo, S., Irakarama, M., Johnston, R., Vyas, M. & Shen, X., 2023. 3D probabilistic full waveform inversion: Application to Gulf of Mexico field data, in *84th EAGE Annual Conference and Exhibition,* European Association of Geoscientists and Engineers, Vienna, Austria, pp. 1–5.
Luckett, R. & Baptie, B., 2015. Local earthquake tomography of Scotland, *Geophys. J. Int.,* **200**(3), 1538–1554.
Magrini, F., Lauro, S., Kästle, E. & Boschi, L., 2022. Surface-wave tomography using Seislib: a Python package for multiscale seismic imaging, *Geophys. J. Int.,* **231**(2), 1011–1030.
Magrini, F., He, J. & Sambridge, M., 2025. BayesBay: a versatile Bayesian inversion framework written in Python, *Seismol. Res. Lett.,* **96**(3), 2052–2064.
Marfurt, K.J., 1984. Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations, *Geophysics,* **49**(5), 533–549.
Mariethoz, G. & Caers, J., 2014. Multiple-point Geostatistics: Stochastic Modeling with Training Images, John Wiley and Sons.
Martin, G.S., Wiley, R. & Marfurt, K.J., 2006. Marmousi2: An elastic upgrade for Marmousi, *Leading Edge,* **25**(2), 156–166.
Meier, U., Curtis, A. & Trampert, J., 2007. Global crustal thickness from neural network inversion of surface wave data, *Geophys. J. Int.,* **169**(2), 706–722.
Meles, G.A., Marelli, S. & Linde, N., 2025. Bayesian full waveform inversion with sequential surrogate model refinement, *Geophys. J. Int.,* **243**(2), ggaf349.
Miller, A.C., Foti, N.J. & Adams, R.P., 2017. Variational boosting: Iteratively refining posterior approximations. *International Conference on Machine Learning,* PMLR, pp. 2420–2429.
Minsley, B.J., Foks, N.L. & Bedrosian, P.A., 2021. Quantifying model structural uncertainty using airborne electromagnetic data, *Geophys. J. Int.,* **224**(1), 590–607.
Mosegaard, K. & Sambridge, M., 2002. Monte Carlo analysis of inverse problems, *Inverse probl.,* **18**(3), R29.
Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.: Solid Earth,* **100**(B7), 12431–12447.
Mosser, L., Dubrule, O. & Blunt, M.J., 2020. Stochastic seismic waveform inversion using generative adversarial networks as a geological prior, *Math. Geosci.,* **52**(1), 53–79.
Nawaz, A. & Curtis, A., 2018. Variational Bayesian inversion (VBI) of quasi-localized seismic attributes for the spatial distribution of geological facies, *Geophys. J. Int.,* **214**(2), 845–875.
Nawaz, A. & Curtis, A., 2019. Rapid discriminative variational Bayesian inversion of geophysical data for the spatial distribution of geological properties, *J. geophys. Res.: Solid Earth,* **124**(6), 5867–5887.
Nawaz, A., Curtis, A., Shahraeeni, M.S. & Gerea, C., 2020. Variational Bayesian inversion of seismic attributes jointly for geological facies and petrophysical rock properties, *Geophysics,* **85**(4), 1–78.
Neal, R.M. *et al.*, 2011. MCMC using Hamiltonian dynamics, *Handbook of Markov Chain Monte Carlo,* Vol. **2**(11), pp. 2, eds Brooks, S., Gelman, A., Jones, G. & Meng, X.-L., CRC Press.
Nicolson, H., Curtis, A., Baptie, B. & Galetti, E., 2012. Seismic interferometry and ambient noise tomography in the British Isles, *Proc. Geologists' Assoc.,* **123**(1), 74–86.
Nicolson, H., Curtis, A. & Baptie, B., 2014. Rayleigh wave tomography of the British Isles from ambient seismic noise, *Geophys. J. Int.,* **198**(2), 637–655.

Orozco, R., Witte, P., Louboutin, M., Siahkoohi, A., Rizzuti, G., Peters, B. & Herrmann, F.J., 2023. Invertiblenetworks. jl: A julia package for scalable normalizing flows, preprint (arXiv:2312.13480).
Papamakarios, G., 2019. Neural density estimation and likelihood-free inference, preprint (arXiv:1910.13233).
Papamakarios, G., Pavlakou, T. & Murray, I., 2017. Masked autoregressive flow for density estimation, *Advances in neural information processing systems,* **30**.
Papamakarios, G., Nalisnick, E., Rezende, D.J., Mohamed, S. & Lakshminarayanan, B., 2021. Normalizing flows for probabilistic modeling and inference, *J. Mach. Learn. Res.,* **22**(57), 1–64.
Pasalic, D. & McGarry, R., 2010. Convolutional perfectly matched layer for isotropic and anisotropic acoustic wave equations. *SEG International Exposition and Annual Meeting,* SEG, pp. SEG–2010.
Paszke, A. *et al.*, 2019. Pytorch: An imperative style, high-performance deep learning library, *Advances in neural information processing systems,* **32**.
Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.,* **167**(2), 495–503.
Poliannikov, O.V. & Malcolm, A.E., 2016. The effect of velocity uncertainty on migrated reflectors: Improvements from relative-depth imaging, *Geophysics,* **81**(1), S21–S29.
Rawlinson, N. & Sambridge, M., 2004. Wave front evolution in strongly heterogeneous layered media using the fast marching method, *Geophys. J. Int.,* **156**(3), 631–647.
Rawlinson, N. & Sambridge, M., 2005. The fast marching method: an effective tool for tomographic imaging and tracking multiple phases in complex layered media, *Explor. Geophys.,* **36**(4), 341–350.
Rezende, D. & Mohamed, S., 2015. Variational inference with normalizing flows. *International conference on machine learning,* PMLR, pp. 1530–1538.
Rücker, C., Günther, T. & Wagner, F.M., 2017. pyGIMLi: An open-source library for modelling and inversion in geophysics, *Comput. Geosci.,* **109,** 106–123.
Sambridge, M. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.,* **40**(3), 3–1.
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.
Scheiter, M., Valentine, A. & Sambridge, M., 2022. Upscaling and downscaling Monte Carlo ensembles with generative models, *Geophys. J. Int.,* **230**(2), 916–931.
Shahraeeni, M.S., Curtis, A. & Chao, G., 2012. Fast probabilistic petrophysical mapping of reservoirs from 3D seismic data, *Geophysics,* **77**(4), O1–O19.
Siahkoohi, A., Rizzuti, G., Louboutin, M., Witte, P.A. & Herrmann, F.J., 2021. Preconditioned training of normalizing flows for variational inference in inverse problems, *Third Symposium on Advances in Approximate Bayesian Inference,* preprint (arXiv:2101.03709).
Siahkoohi, A., Rizzuti, G. & Herrmann, F.J., 2022. Deep Bayesian inference for seismic imaging with tasks, *Geophysics,* **87**(5), S281–S302.
Siahkoohi, A., Rizzuti, G., Orozco, R. & Herrmann, F.J., 2023. Reliable amortized variational inference with physics-based latent distribution correction, *Geophysics,* **88**(3), 1–137.
Sjölund, J., 2023. A tutorial on parametric variational inference, preprint (arXiv:2301.01236).
Smith, J.D., Ross, Z.E., Azizzadenesheli, K. & Muir, J.B., 2022. HypoSVI: Hypocentre inversion with Stein variational inference and physics informed neural networks, *Geophys. J. Int.,* **228**(1), 698–710.
Strutz, D. & Curtis, A., 2024. Variational Bayesian experimental design for geophysical applications: seismic source location, amplitude versus offset inversion, and estimating CO_2 saturations in a subsurface reservoir, *Geophys. J. Int.,* **236**(3), 1309–1331.
Sun, C., Malcolm, A., Kumar, R. & Mao, W., 2024. Enabling uncertainty quantification in a standard full-waveform inversion method using normalizing flows, *Geophysics,* **89**(5), R493–R507.
Sun, L., Wang, L., Xu, G. & Wu, Q., 2023. A new method of variational Bayesian slip distribution inversion, *J. Geod.,* **97**(1), 10.
Tarantola, A., 2005. Inverse problem theory and methods for model parameter estimation, vol. **89,** SIAM.
Tetzlaff, D., 2023. Stratigraphic forward modeling software package for research and education, preprint (arXiv:2302.05272).
Valentine, A.P. & Sambridge, M., 2022. pyprop8: A lightweight code to simulate seismic observables in a layered half-space, *J. Open Source Softw.,* **7**(76), 4217.
Valentine, A.P. & Sambridge, M., 2023. Emerging directions in geophysical inversion, in *Applications of Data Assimilation and Inverse Problems in the Earth Sciences,* Vol. **5,** pp. 9, eds Ismail-Zadeh, A., Castelli, F., Jones, D. & Sanchez„ S., Cambridge Univ. Press.
Vrugt, J.A., 2016. Markov chain Monte Carlo simulation using the DREAm software package: theory, concepts, and MATLAB implementation, *Environ. Modell. Softw.,* **75,** 273–316.
Walker, M. & Curtis, A., 2014a. Expert elicitation of geological spatial statistics using genetic algorithms, *Geophys. J. Int,* **198,** 342–356.
Walker, M. & Curtis, A., 2014b. Varying prior information in Bayesian inversion, *Inverse Probl.,* **30**(6), 065002.
Wang, W., McMechan, G.A. & Ma, J., 2023. Re-weighted variational full waveform inversions, *Geophysics,* **88**(1), 1–61.
Wang, Y., Zhou, H., Zhao, X., Zhang, Q., Zhao, P., Yu, X. & Chen, Y., 2019. Cu Q-RTM: A CUDA-based code package for stable and efficient Q-compensated reverse time migration, *Geophysics,* **84**(1), F1–F15.
Wolpert, D.H. & Macready, W.G., 1997. No free lunch theorems for optimization, *IEEE Trans. Evolut. Comput.,* **1**(1), 67–82.
Yin, Z., Orozco, R. & Herrmann, F.J., 2025. WISER: multimodal variational inference for full-waveform inversion without dimensionality reduction, *Geophysics,* **90**(2), A1–A7.
Zhang, C., Bütepage, J., Kjellström, H. & Mandt, S., 2018. Advances in variational inference, *IEEE Trans. Pattern Anal. Mach. Intell.,* **41**(8), 2008–2026.
Zhang, X. & Curtis, A., 2020. Seismic tomography using variational inference methods, *J. geophys. Res.: Solid Earth,* **125**(4), e2019JB018589.
Zhang, X. & Curtis, A., 2021. Bayesian full-waveform inversion with realistic priors, *Geophysics,* **86**(5), A45–A49.
Zhang, X. & Curtis, A., 2022. Interrogating probabilistic inversion results for subsurface structural information, *Geophys. J. Int.,* **229**(2), 750–757.
Zhang, X. & Curtis, A., 2024. VIP–Variational inversion package with example implementations of Bayesian tomographic imaging, *Seismica,* **3**(1).
Zhang, X., Nawaz, A., Zhao, X. & Curtis, A., 2021. An introduction to variational inference in geophysical inverse problems, *Adv. Geophys.,* **62,** 73–140.
Zhang, X., Lomas, A., Zhou, M., Zheng, Y. & Curtis, A., 2023. 3D Bayesian variational full waveform inversion, *Geophys. J. Int.,* **234**(1), 546–561.
Zhao, X. & Curtis, A., 2024a. Bayesian inversion, uncertainty analysis and interrogation using boosting variational inference, *J. geophys. Res.: Solid Earth,* **129**(1), e2023JB027789.
Zhao, X. & Curtis, A., 2024b. Physically structured variational inference for Bayesian full waveform inversion, *J. geophys. Res.: Solid Earth,* **129**(11), e2024JB029557.
Zhao, X. & Curtis, A., 2024c. Variational prior replacement in Bayesian inference and inversion, *Geophys. J. Int.,* **239**(2), 1236–1256.

Zhao, X. & Curtis, A., 2025a. Efficient Bayesian full-waveform inversion and analysis of prior hypotheses in three dimensions, *Geophysics,* **90**(6), R373–R388.
Zhao, X. & Curtis, A., 2025b. On the design of ultra-sparse seismic surveys for monitoring subsurface CO_2 storage sites using full waveform inversion, *Int. J. Greenhouse Gas Control,* **146,** 104433.
Zhao, X. & Curtis, A., 2026. Linearised versus nonlinear estimates of uncertainty in full waveform inversion, preprint (arXiv:2603.11711).
Zhao, X., Zhou, H., Chen, H. & Wang, Y., 2021. Domain decomposition for large-scale viscoacoustic wave simulation using localized pseudo-spectral method, *IEEE Trans. Geosci. Remote Sens.,* **59,** 2666–2679.
Zhao, X., Curtis, A. & Zhang, X., 2021. Bayesian seismic tomography using normalizing flows, *Geophys. J. Int.,* **228**(1), 213–239.
Zhao, X., Curtis, A. & Zhang, X., 2022. Interrogating subsurface structures using probabilistic tomography: an example assessing the volume of Irish Sea basins, *J. geophys. Res.: Solid Earth,* **127**(4), e2022JB024098.
Zhao, X., Irvin, L., Galetti, E. & Curtis, A., 2026. Direct-3D variational Bayesian surface wave inversion and its application to ambient noise tomography beneath Great Britain, *Geophys. J. Int.,* **245**(2), ggag079.
Zunino, A., Gebraad, L., Ghirotto, A. & Fichtner, A., 2023. HMCLab: a framework for solving diverse geophysical inverse problems using the Hamiltonian Monte Carlo method, *Geophys. J. Int.,* **235**(3), 2979–2991.

APPENDIX A: POSTERIOR SAMPLES FROM TRAVELTIME TOMOGRAPHY EXAMPLE

In this appendix, we present five posterior samples (model realizations) obtained from different inversion methods applied in the traveltime tomography example in Section 4.1, as shown in Fig. A1. In this figure we interpolate each pixelated posterior sample such that main velocity features can be observed clearly. Generally at large scale, the main features of posterior samples from different methods on the main island are consistent and are similar to the mean velocity maps displayed in Fig. 3 in the main text (offshore where there are almost no data constraints, velocity heterogeneity is consistent with uniformly random variations, as specified in the prior pdf). This demonstrates that each inversion method finds reasonable posterior samples and thus uncertainty structure.

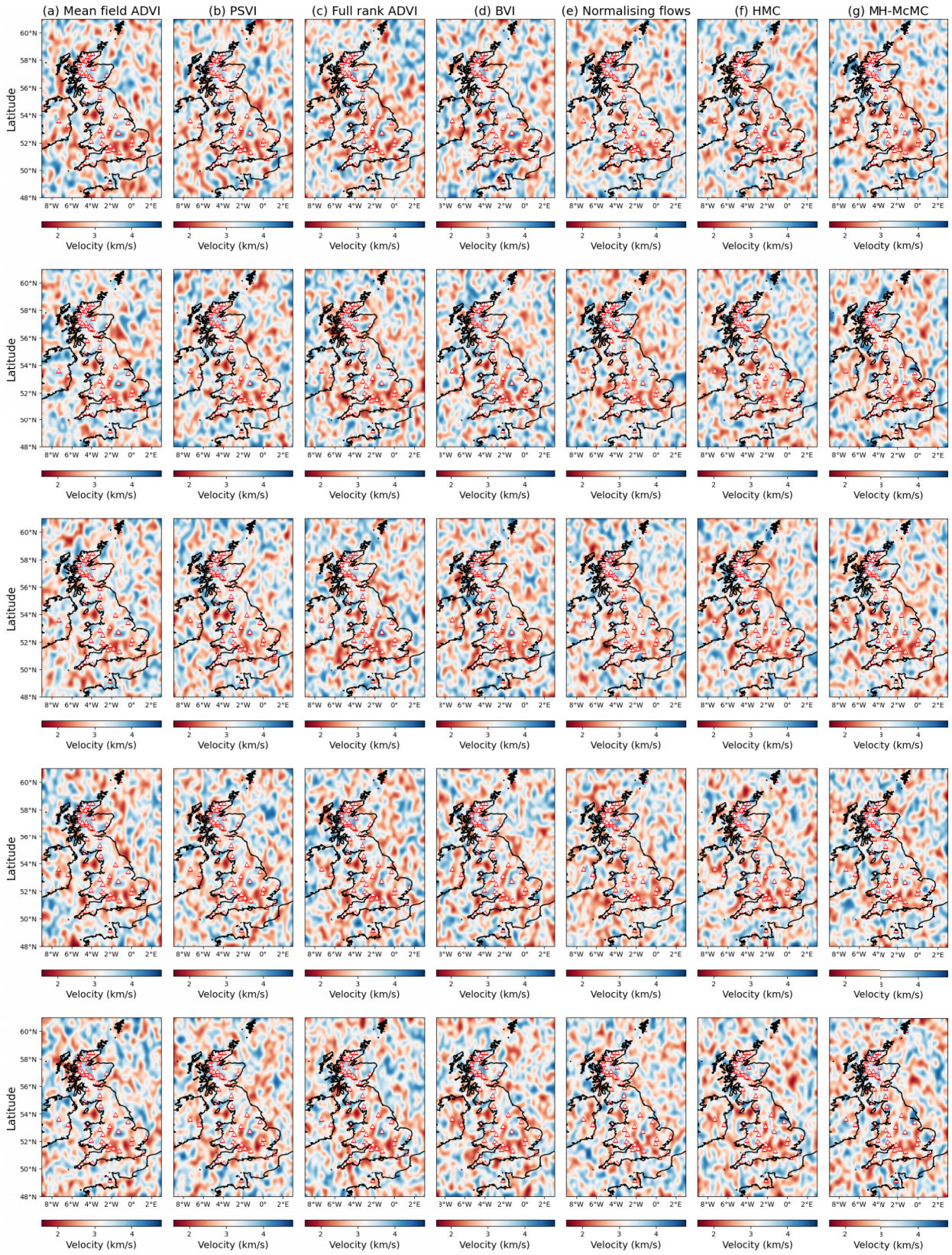

Figure A1. Five posterior samples of 10s-period Love wave traveltime tomography results of the British Isles using different inversion methods indicated in the top titles. From top to bottom, each row shows one sample. White triangles show the locations of receivers used in this example.