Physically Structured Variational Inference for Bayesian Full Waveform Inversion

Xuebin Zhao[1] and Andrew Curtis[1]

[1]University of Edinburgh

May 28, 2024

Abstract

Full waveform inversion (FWI) creates high resolution models of the Earth's subsurface structures from seismic waveform data. Due to the non-linearity and non-uniqueness of FWI problems, finding globally best-fitting model solutions is not necessarily desirable since they fit noise as well as signal in the data. Bayesian FWI calculates a so-called posterior probability distribution function, which describes all possible model solutions and their uncertainties. In this paper, we solve Bayesian FWI using variational inference and propose a new methodology called physically structured variational inference, in which a physics-based structure is imposed on the variational distribution. In a simple example motivated by prior information from past FWI solutions, we include parameter correlations between pairs of spatial locations within a dominant wavelength of each other, and set other correlations to zero. This makes the method far more efficient in terms of both memory requirements and computation, at the cost of some loss of generality in the solution found. We demonstrate the proposed method with a 2D acoustic FWI scenario, and compare the results with those obtained using other methods. This verifies that the method can produce accurate statistical information about the posterior distribution with hugely improved efficiency (in our FWI example, 1 order of magnitude in computation). We further demonstrate that despite the possible reduction in generality of the solution, the posterior uncertainties can be used to solve post-inversion interrogation problems connected to estimating volumes of subsurface reservoirs and of stored CO2, with minimal bias, creating a highly efficient FWI-based decision-making workflow.

Physically Structured Variational Inference for Bayesian Full Waveform Inversion

Xuebin Zhao[1] **and Andrew Curtis**[1]

[1]School of Geosciences, University of Edinburgh, Edinburgh, United Kingdom

Key Points:

- We propose a new variational inference methodology to construct a Bayesian posterior solution with a desired correlation structure.
- The method is far more efficient in terms of both memory requirements and computation, with some loss of generality in the solution.
- We apply the inversion results to two post-inversion problems where the volume of stored CO_2 in a subsurface reservoir is estimated.

Corresponding author: Xuebin Zhao, `xuebin.zhao@ed.ac.uk`

Abstract

Full waveform inversion (FWI) creates high resolution models of the Earth's subsurface structures from seismic waveform data. Due to the non-linearity and non-uniqueness of FWI problems, finding globally best-fitting model solutions is not necessarily desirable since they fit noise as well as signal in the data. Bayesian FWI calculates a so-called posterior probability distribution function, which describes all possible model solutions and their uncertainties. In this paper, we solve Bayesian FWI using variational inference, and propose a new methodology called physically structured variational inference, in which a physics-based structure is imposed on the variational distribution. In a simple example motivated by prior information from past FWI solutions, we include parameter correlations between pairs of spatial locations within a dominant wavelength of each other, and set other correlations to zero. This makes the method far more efficient in terms of both memory requirements and computation, at the cost of some loss of generality in the solution found. We demonstrate the proposed method with a 2D acoustic FWI scenario, and compare the results with those obtained using other methods. This verifies that the method can produce accurate statistical information about the posterior distribution with hugely improved efficiency (in our FWI example, 1 order of magnitude in computation). We further demonstrate that despite the possible reduction in generality of the solution, the posterior uncertainties can be used to solve post-inversion interrogation problems connected to estimating volumes of subsurface reservoirs and of stored CO_2, with minimal bias, creating a highly efficient FWI-based decision-making workflow.

Plain Language Summary

This paper introduces a method to assess uncertainties in seismic images of the subsurface at substantially reduced cost, and to use the information within those uncertainties to answer explicit high-level questions about volumes of subsurface reservoirs and of stored CO2. Computational efficiency is achieved by explicitly imposing known (always observed) trade-offs between parameters that describe local properties of the subsurface. This prevents computing power from being used to re-discover such trade-offs each time an imaging process is performed. In our two-dimensional example in which we image using seismic Full Waveform Inversion, computational cost is reduced by an

order of magnitude and fully nonlinear uncertainties can be characterized both in subsurface structural parameters, and in answers to high-level questions.

1 Introduction

Seismic full waveform inversion (FWI) is a method that generates models of the subsurface seismic velocity structure of the Earth given recorded seismograms. This is achieved using both kinematic (phase) and dynamic (amplitude) information in the waveforms (Tarantola, 1984). FWI has been applied in various fields, for example including regional and global seismology (Fichtner et al., 2009; Tape et al., 2010; French & Romanowicz, 2014; Bozdağ et al., 2016; Fichtner et al., 2018), seismic exploration (Pratt et al., 1998; Virieux & Operto, 2009; Prieux et al., 2013; Warner et al., 2013), medical imaging (Bernard et al., 2017; Guasch et al., 2020; Lucka et al., 2021), and non-destructive detection (He et al., 2021; Patsia et al., 2023).

Traditionally, FWI problems are solved using gradient-based local optimisation methods, where a misfit function between observed and predicted waveform data is minimised iteratively (Plessix, 2006). This process often requires additional regularisation terms, such as smoothing and damping terms, to stabilise the optimisation and improve convergence rates (Zhdanov, 2002; Sen & Roy, 2003; Asnaashari et al., 2013). However, these terms may introduce biases to the final inversion results. In addition, it is challenging to find a good approximation to the true Earth structure that generated the observed waveforms due to the strong non-linearity of the forward function and the non-uniqueness of the inverse problem solution (Boyd & Vandenberghe, 2004).

Recently, FWI has been solved probabilistically using a suite of methods collectively referred to as *Bayesian inference*. In Bayesian FWI, *prior* knowledge about Earth model parameters is updated with new information from the observed waveform data to calculate a *posterior* probability distribution function (pdf), according to Bayes' rule. In principle this distribution incorporates all prior information combined with all information from the data, and expresses the information in terms of constraints on the model parameters. It thus solves the FWI problem by describing all possible model parameter values that fit the dataset to within its uncertainty. The range and probability of different possible models can be used to reduce risk during subsequent decision-making when solving real-world interrogation problems (Poliannikov & Malcolm, 2016; Arnold & Curtis, 2018; Ely et al., 2018; X. Zhao et al., 2022; X. Zhang & Curtis, 2022; Siahkoohi et al., 2022).

Different kinds of Bayesian inference methods have been employed to perform probabilistic FWI. A direct generalisation from deterministic FWI involves approximating the posterior pdf with a Gaussian distribution, centred around an estimated maximum a posteriori (MAP) model obtained using local optimisation methods (Gouveia & Scales, 1998; Bui-Thanh et al., 2013; Zhu et al., 2016; Fang et al., 2018), or through local, low rank pdf approximations using a data assimilation technique (Thurin et al., 2019). If both the likelihood function and prior distribution are assumed to be Gaussians, then this MAP velocity model is equivalent to that obtained using l_2 regularised deterministic FWI (W. Wang et al., 2023). While this kind of methods can produce probabilistic results, the resulting posterior distribution may be affected by the starting point of the inversion, and may not fully capture uncertainty arising from non-linearity of the forward function (Z. Zhao & Sen, 2021).

Fully non-linear Bayesian FWI can be solved using sampling techniques such as Markov chain Monte Carlo (McMC), where random samples are drawn from the posterior distribution. The inversion results are represented by statistics of the sampled models, such as the mean and standard deviation. However, due to the typical high dimensionality (number of parameters to be estimated) of FWI problems, direct sampling methods, including the commonly used Metropolis-Hastings (MH)-McMC (Metropolis et al., 1953; Hastings, 1970; Mosegaard & Tarantola, 1995; Sambridge & Mosegaard, 2002), become impractical. Nevertheless, it is worth noting the existence of studies that employ a target-oriented strategy to reduce the dimensionality of parts of the Earth model of interest, and employ a localised wavefield injection method to calculate wavefields corresponding to each model variation. This reduces the computational complexity of FWI, and allows Metropolis-Hastings McMC to be applied effectively (Ely et al., 2018; Kotsi et al., 2020b; Fu & Innanen, 2022).

Several advanced techniques have been introduced to improve the sampling efficiency of McMC for Bayesian FWI. In reversible-jump McMC (RJ-McMC) (Green, 1995, 2003; Sambridge et al., 2006), a trans-dimensional approach is used to change the parametrisation, including the dimensionality of the model parameter vector. This can significantly improve efficiency by reducing dimensionality to only parameters that are necessary to explain the data and the forward function, and RJ-McMC has been successfully applied to Bayesian FWI (Ray et al., 2016, 2018; Visser et al., 2019; P. Guo et al., 2020). Hamiltonian Monte Carlo (HMC) has also been introduced to improve the sampling efficiency of FWI. In HMC, the sampling process is guided by the gradient of the posterior pdf with respect to the model parameters, and it has been demonstrated that HMC can improve the convergence rate over non-gradient based McMC (Gebraad et al., 2020; Kotsi et al., 2020a; de Lima, Corso, et al., 2023; de Lima, Ferreira, et al., 2023; Zunino et al., 2023; Dhabaria & Singh, 2024). Biswas and Sen (2022) introduced a reversible-jump Hamiltonian Monte Carlo (RJHMC) algorithm for 2D FWI, Z. Zhao and Sen (2021) and Berti et al. (2023) used gradient-based McMC methods to sample the posterior distribution efficiently, and Khoshkholgh et al. (2022) solved FWI using informed-proposal Monte Carlo (Khoshkholgh et al., 2021). Nevertheless, as with other classes of methods, Monte Carlo sampling is known to become computationally intractable for high-dimensional parameter spaces due to the curse of dimensionality (Curtis & Lomax, 2001).

In this study, we focus instead on variational inference, a method that solves Bayesian inversion through optimisation. In variational methods, we define a family of known and tractable distributions, referred to as the variational family. From this family, an optimal member is chosen to approximate the true posterior pdf by minimising the difference between the variational and posterior distributions (Bishop, 2006; Blei et al., 2017; C. Zhang et al., 2018; X. Zhang et al., 2021). Variational inference solves Bayesian problems under an optimisation framework, and the optimisation result is fully probabilistic. In some classes of problems it can therefore be relatively more efficient and scalable to high dimensional problems with large datasets. Variational inference has been applied to different geophysical inverse problems, including travel time tomography (X. Zhang & Curtis, 2020a; X. Zhao et al., 2021; Levy et al., 2022), seismic migration (Siahkoohi et al., 2020; Siahkoohi & Herrmann, 2021; Siahkoohi et al., 2021, 2023), seismic amplitude inversion (Zidan et al., 2022), earthquake hypocentre inversion (Smith et al., 2022), and slip distribution inversion (Sun et al., 2023). However, most of these applications have relatively lower dimensionality and weaker non-linearities compared to FWI.

X. Zhang and Curtis (2020b) introduced a variational method called Stein variational gradient descent (SVGD – Liu & Wang, 2016) to transmission FWI where sources emulating earthquakes are located underneath the velocity structure to be imaged, with receivers on the top surface. SVGD was then applied to 2D reflection FWI with realistic priors (X. Zhang & Curtis, 2021a; Izzatullah et al., 2023), and 3D acoustic FWI using synthetic data (X. Zhang et al., 2023) and field data (Lomas et al., 2023). A stochastic version of SVGD (Gallego & Insua, 2018) was also employed to improve performance for 3D FWI (X. Zhang et al., 2023). X. Zhao and Curtis (2024) introduced boosting variational inference (BVI – F. Guo et al., 2016; Miller et al., 2017) for 2D acoustic FWI, where a mixture of Gaussian distributions is used to approximate the true posterior distribution, resulting in an analytic expression for the posterior distribution. Bates et al. (2022) performed medical ultrasound tomography of the brain using FWI, where a mean field (diagonal) Gaussian distribution is employed as the variational distribution. Alternatively, W. Wang et al. (2023) improved the resolution of inversion results by decomposing the variational objective function into two terms and re-weighting them, however the method tends to underestimate posterior uncertainties. Yin et al. (2024) used conditional normalizing flows to quantify uncertainties in migration-velocity models.

Other than in W. Wang et al. (2023), in the above studies variational methods were applied to improve the efficiency of Bayesian FWI. For 2D FWI, the required number of forward simulations used to estimate means and variances of subsurface parameters was reduced to the order of 100,000 by X. Zhao and Curtis (2024), marking a significant reduction given that the dimensionality of the FWI problem tackled was higher than 10,000. Unfortunately, despite this improvement, the computational cost of solving the forward function in FWI remains prohibitively expensive for many practitioners. Consequently, performing Bayesian FWI in realistic projects using current variational methods is still impractical, even with advanced forward simulation strategies (Treeby & Cox, 2010; Y. Wang et al., 2019; X. Zhao et al., 2020).

In this paper, we propose an efficient and accurate variational methodology for Bayesian FWI by imposing physics-based structure on the variational family. The new method incorporates expected posterior parameter correlations explicitly. We show that this leads to significantly improved accuracy with nearly the same computational cost compared to several existing variational methods, or put another way, reduced cost for the same accuracy.

This rest of this paper is organised as follows. In section 2, we first establish the framework of variational full waveform inversion. Then we introduce the concept of ADVI, and present our new method which we refer to as *physically structured variational in-*

ference (PSVI). In section 3, we demonstrate the proposed method with a 2D synthetic FWI example and compare the inversion results with those obtained using three other variational methods. In section 4, we interpret the inversion results by solving two post-inversion interrogation problems. Finally, we provide a brief discussion of the proposed method and draw conclusions.

2 Methodology

2.1 Variational Full Waveform Inversion (FWI)

FWI uses full waveform data recorded by seismometers to constrain the Earth's interior structure, typically described by a subsurface velocity model. The forward function is defined to predict waveform data that could be recorded at receivers given a subsurface velocity model. This prediction involves solving a wave equation, either in the time or frequency domain, often in two or three dimensions, and potentially adding measurement noise to the data. For simplicity, we assume that the subsurface consists of an acoustic, isotropic, lossless medium with constant density, thereby ignoring exclusively elastic properties including shear waves, attenuation, and anisotropic properties. This simplification allows the scalar acoustic wave equation to be used in forward simulations which reduces computational load. The data-model gradients are calculated using the adjoint state method (Plessix, 2006).

In Bayesian FWI, information about the velocity model is characterized by a *posterior* probability distribution function (pdf) which describes the uncertainties associated with different potential models given the observed data. This can be calculated using Bayes' rule:

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

where $p(\cdot)$ denotes a probability distribution. Symbol $x|y$ indicates conditional dependence between two random variables x and y, and reads as x given y. Term $p(\mathbf{m})$ describes the *prior* information available on the model parameter $\mathbf{m}$, and $p(\mathbf{d}_{obs}|\mathbf{m})$ is the *likelihood*, meaning the probability of the synthetic waveform data $\mathbf{d}_{syn}$ generated by a given model $\mathbf{m}$ through forward simulation matching the observed data $\mathbf{d}_{obs}$. A Gaussian distribution is often used to define the data likelihood function:

$$p(\mathbf{d}_{obs}|\mathbf{m}) \propto \exp\left[-\frac{(\mathbf{d}_{syn}-\mathbf{d}_{obs})^T\Sigma_{\mathbf{d}}^{-1}(\mathbf{d}_{syn}-\mathbf{d}_{obs})}{2}\right] \tag{2}$$

where Σ_d is the covariance matrix of the data error. The denominator $p(\mathbf{d}_{obs})$ in equation 1 is referred to as the *evidence* and is a normalisation constant to ensure that the result of equation 1 is a valid probability distribution.

Bayesian inversion is often solved by Monte Carlo sampling methods. However, the required number of samples increases exponentially with the dimensionality of the inverse problem (the number of unknown model parameters), due to the curse of dimensionality (Curtis & Lomax, 2001). It is very expensive to obtain statistics of posterior pdf's in FWI using Monte Carlo methods, especially when the Earth model $\mathbf{m}$ contains more than 10,000 parameters, as is standard in such problems (Gebraad et al., 2020).

In this paper, we use variational inference to solve Bayesian FWI. In variational methods, a family of distributions (called the variational family) $\mathcal{Q}(\mathbf{m}) = \{q(\mathbf{m})\}$ is defined, from which we select an optimal member to approximate the true (unknown) posterior distribution. The optimal distribution can be found by minimising the difference (distance) between the posterior and variational distributions. Typically, the Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951) is used to measure the distance between two probability distributions, defined as the following expectation term

$$\mathrm{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{obs})] = \mathbb{E}_{q(\mathbf{m})}[\log q(\mathbf{m}) - \log p(\mathbf{m}|\mathbf{d}_{obs})] \tag{3}$$

The KL divergence of two distributions is non-negative, and equals zero only when the two distributions are identical. Substituting equation 1 into 3, we find that minimising the $\mathrm{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{obs})]$ is equivalent to maximising the following *evidence lower bound* of $\log p(\mathbf{d}_{obs})$ ($\mathrm{ELBO}[q(\mathbf{m})]$):

$$\mathrm{ELBO}[q(\mathbf{m})] = \mathbb{E}_{q(\mathbf{m})}[\log p(\mathbf{m}, \mathbf{d}_{obs}) - \log q(\mathbf{m})] \tag{4}$$

In this way, we convert a random sampling problem into a numerical optimisation, while the optimisation result is still a probability distribution that approximates the true posterior pdf.

A key challenge in variational inference is to choose the variational family $\mathcal{Q}(\mathbf{m})$. This determines both the accuracy and efficiency of the variational methods: increasing the complexity (and hence, expressivity) of $\mathcal{Q}(\mathbf{m})$ increases the approximation accuracy as well as the optimisation complexity. Given the expensive nature of forward simulations in FWI, our primary goal is to reduce computational costs (by reducing the number of forward simulations) while maintaining accuracy at an acceptable level. In the following sections we introduce a method called automatic differentiation variational inference (ADVI – Kucukelbir et al., 2017), and propose an alternative effective variational methodology for FWI.

2.2 Automatic Differentiation Variational Inference (ADVI)

ADVI is a well-established variational method that defines a Gaussian variational distribution $q = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, parametrised by a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$ (Kucukelbir et al., 2017). In addition, since a Gaussian distribution is defined over the space of real numbers and since in most geophysical imaging problems model parameters are bounded by physical constraints (e.g., seismic velocity should be a positive number), an invertible transform (a bijection) is applied to convert the Gaussian variational distribution into a bounded space that defines model parameter $\mathbf{m}$. The transformed distribution is then used to approximate the true posterior distribution.

To determine the optimal Gaussian distribution in the unbounded space, we maximise the ELBO$[q(\mathbf{m})]$ in equation 4 with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. This can be solved using a gradient based optimisation method. According to Kucukelbir et al. (2017), the gradient of the ELBO with respect to the covariance matrix $\boldsymbol{\Sigma}$ involves computing $|\boldsymbol{\Sigma}|$, where $|\cdot|$ denotes the determinant of a matrix. Direct calculation of $|\boldsymbol{\Sigma}|$ has a computational complexity of $O(n^3)$, which becomes prohibitively expensive for high dimensional inference problems such as FWI. Therefore, we often use a Cholesky factorisation to parametrise $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^{\mathbf{T}} \tag{5}$$

where $\mathbf{L}$ is a lower triangular matrix. Since $|\mathbf{L}|$ can be calculated easily as the product of its diagonal elements, the determinant $|\boldsymbol{\Sigma}|$ can be obtained by $|\boldsymbol{\Sigma}| = |\mathbf{L}|^2$. Note that the diagonal elements of $\mathbf{L}$ are associated with the variances of model parameters, and should be non-negative to ensure that $\mathbf{L}$ and $\boldsymbol{\Sigma}$ are positive semidefinite. The off-diagonal values of $\mathbf{L}$ contain correlation information between model parameters.

For a n-dimensional problem, we need $n(n+1)/2$ parameters to construct a full matrix $\mathbf{L}$, and consequently a full covariance matrix $\boldsymbol{\Sigma}$. The corresponding method is known as full rank ADVI (Kucukelbir et al., 2017). For example, in Figure 1a, the velocity model comprising 110 $\times$ 250 pixels requires 378,138,750 parameters to describe the full matrix $\mathbf{L}$. This number becomes computationally intractable for large scale 2D and 3D FWI problems.

Alternatively, a mean field approximation is often used to reduce computational complexity, where $\mathbf{L}$ and $\mathbf{\Sigma}$ are parametrised by diagonal matrices. The variational distribution becomes a diagonal Gaussian distribution, which neglects correlation information between different model parameters. In this way, the total number of variables that must be optimised is $2n$ (both $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ contain n independent elements), so is doubled compared to a conventional deterministic inversion. Therefore, the computational overhead is manageable for most problems. Mean field ADVI has been applied to Bayesian FWI in several studies (Bates et al., 2022; W. Wang et al., 2023; X. Zhang et al., 2023), demonstrating that the method is computationally efficient and is able to provide an accurate mean model of the posterior distribution. However, in problems with significant posterior correlations, it tends to strongly underestimate posterior uncertainties since correlation information is neglected *a priori* (X. Zhang et al., 2023).

2.3 Physically Structured Variational Inference (PSVI)

Full rank ADVI and mean field ADVI represent two extreme approaches to construct $\mathbf{L}$: the former aims to optimise all off-diagonal elements of $\mathbf{L}$ to capture the full correlation information of $\mathbf{m}$, whereas the latter sets the off-diagonal elements to zero to reduce computational requirements. In the following, we parametrise $\mathbf{L}$ using a physics-guided structure, which models a subset of its off-diagonal elements.

In most imaging problems, accurate correlation information plays an important role in capturing true structures such as the continuity of properties across neighbouring spatial. Since modelling a full covariance matrix (i.e., full rank ADVI) for high dimensional problems is practically intractable, another approach is to model the most important correlation in vector $\mathbf{m}$, guided by physical properties (prior knowledge) of imaging problems. To illustrate, Figure 1d shows a 2D velocity structure discretized using $nx \times nz$ square grid cells in horizontal and vertical directions, with each cell representing a velocity value at the corresponding spatial location. It is often the case that any grid cell, such as the one marked by a black dot in Figure 1d, is strongly correlated with its surrounding cells (e.g., cells marked by white pluses). The magnitude of correlations between this central cell and other cells decreases as the distance between two locations

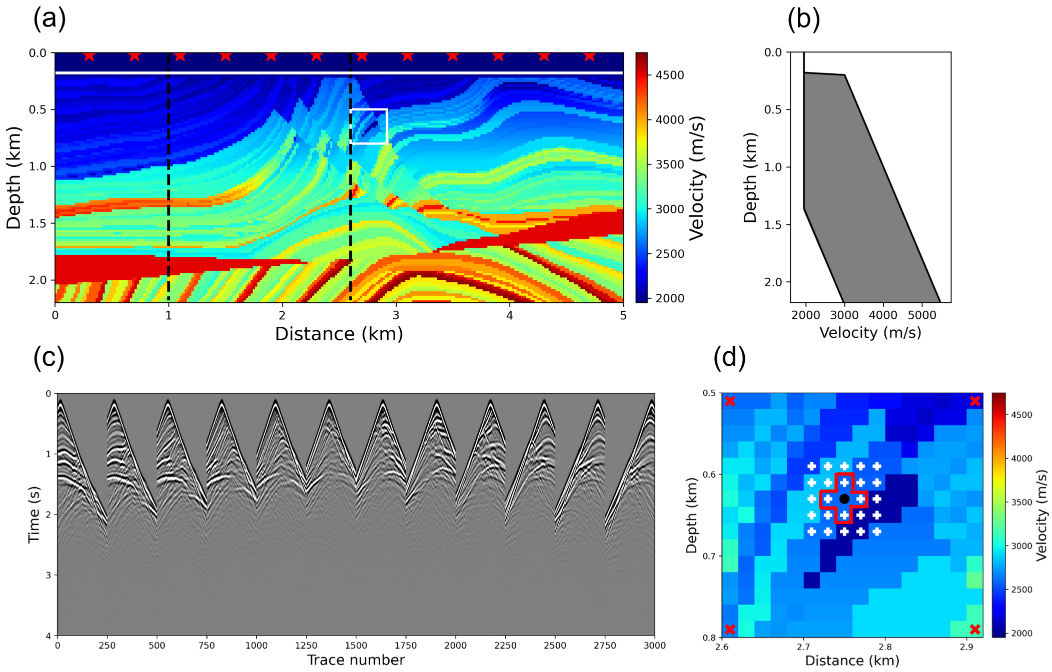

Figure 1. (a) P wave velocity of the Marmousi model used in a 2D acoustic FWI test. Source locations are indicated by red stars and the receiver line is marked by a white line. Dashed black lines display the locations of two vertical profiles used to compare the posterior marginal probability distributions in Figure 4. (b) Upper and lower bounds of the Uniform prior distribution at different depths. (c) Observed dataset which contains twelve common shot gathers. (d) Velocity structure inside the white box in (a), and crosses in cells discussed in the main text.

increases. Cells that are far away from the black dot (e.g., cells denoted by red crosses in Figure 1d) are only weakly correlated with the black-dotted cell, so these correlations can safely be ignored. This feature has been observed in many different imaging problems (Ardizzone et al., 2018; Gebraad et al., 2020; Biswas & Sen, 2022); a clear example displaying such correlations in a velocity profile with depth is shown in Figure 6 of X. Zhang and Curtis (2021b), from the results of surface wave dispersion inversion using two independent nonlinear inversion methods (invertible neural networks and Monte Carlo).

This suggests that it might suffice to model correlations only between parameter values that are spatially close to each other, i.e. which lie within a dominant wavelength, and ignore those that are far away by assuming a particular sparse structure for $\mathbf{L}$. We therefore set off-diagonal elements of $\mathbf{L}$ which represent the main correlations of interest as parameters to be optimised during variational inversion, while imposing all other off-diagonal elements to be zero. Note that we thus impose only a structure on $\mathbf{L}$ rather than placing constraints on the values of its (non-zero) off-diagonal elements: those values are updated freely during inversion.

Suppose that the 2D velocity model displayed in Figure 1d is defined by vector $\mathbf{m}$ in row-major order (i.e., the first nx elements of $\mathbf{m}$ comprise the first row of the 2D image, the second nx elements comprise the second row, and so on). As illustrated in equation 6 below, the first-order off-diagonal elements (blue ones in equation 6 that are directly below the diagonal elements) contain correlation information between two horizontally adjacent grid cells, and off-diagonal elements that are nx rows below the main diagonal elements (red ones in equation 6) describe correlations between two vertically adjacent cells

$$\mathbf{L} = \begin{bmatrix} l_{0,1} & & & & & & \\ {\color{blue}l_{1,1}} & l_{0,2} & & & & & \\ 0 & {\color{blue}l_{1,2}} & l_{0,3} & & & & \\ \ldots & 0 & {\color{blue}l_{1,3}} & \ldots & & & \\ {\color{red}l_{nx,1}} & \ldots & 0 & {\color{blue}\ldots} & l_{0,n-2} & & \\ 0 & {\color{red}\ldots} & \ldots & \ldots & {\color{blue}l_{1,n-2}} & l_{0,n-1} & \\ \ldots & 0 & {\color{red}l_{nx,n-nx}} & \ldots & 0 & {\color{blue}l_{1,n-1}} & l_{0,n} \end{bmatrix} \tag{6}$$

Note that in equation 6, the first subscript i indicates a block of off-diagonal elements that are i rows below the main diagonal (i.e., at an offset of i from the main diagonal), and the second subscript j indicates that $l_{i,j}$ is the jth element of that off-diagonal block. This differs from the commonly used indexing scheme in which the two subscripts imply the row and column number of an element. If we set all remaining elements of $\mathbf{L}$ to zero, then covariance matrix $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^\mathbf{T}$ also has non-zero entities only at two off-diagonal blocks located 1 and nx rows below and above the main diagonal elements (similar to the red and blue elements in equation 6). If such a covariance matrix $\mathbf{\Sigma}$ is used, the variational distribution would also capture a specific spatial correlation structure that only includes parameter correlations between pairs of adjacent cells in both horizontal and vertical directions. Thus, for the grid cell denoted by the black dot in Figure 1d, we would model correlations between this cell and its four adjacent cells inside the red box in Figure 1d: all other correlations are set to zero.

We can impose any desired correlation structure on $\mathbf{\Sigma}$, by setting the corresponding off-diagonal blocks in $\mathbf{L}$ as unknown hyperparameters and optimising them during inversion. The size of the defined correlation template should be relatively small compared to the dimensionality of the problem, so the total number of parameters required to construct $\mathbf{L}$ would also be relatively small compared to that in full rank ADVI. For example, if the white pluses in Figure 1d are used to define a 5×5 correlation kernel then the required number of parameters to construct $\mathbf{\Sigma}$ is smaller than $13n$. Here n is the dimensionality of model vector $\mathbf{m}$, and the number 13 consists of 1 main diagonal block and 12 off-diagonal blocks representing 12 different offsets between cells marked by the white crosses and the central cell in the $5{\times}5$ kernel. Since each off-diagonal block contains fewer parameters than the main diagonal block (i.e., the blue and red elements in equation 6 are fewer than the diagonal elements), the total number of parameters is smaller than $13n$, which is a significant reduction compared to $n(n+1)/2$ parameters used in full rank ADVI.

We implement the aforementioned approach to parametrise the matrix $\mathbf{L}$ and obtain a sparse approximation of the covariance matrix. The inversion results thus effectively and efficiently capture structured correlation information. Since this originated from the inherent physical properties of imaging problems, we name the method as *physically structured variational inference* (PSVI).

To update the variational parameters, we use gradient based optimisation methods. The gradient of the ELBO with respect to the variational parameters can be calculated easily using advanced automatic differentiation libraries such as TensorFlow (Abadi et al., 2016) and PyTorch (Paszke et al., 2019). The expectation term in the EBLO (equation 4) can be estimated by Monte Carlo integration with a small number of samples, which is reasonable because the optimisation is typically carried out over many iterations, allowing the gradients to converge statistically towards the correct solution (Kucukelbir et al., 2017). Given that the computational cost of updating the variational parameters is negligible in comparison to forward modelling in FWI, the proposed method is almost as efficient as mean field ADVI.

3 2D Acoustic FWI Example

In this section, we test the proposed PSVI algorithm in a 2D acoustic FWI example. The true velocity model, shown in Figure 1a, is obtained by truncating the original Marmousi model (Martin et al., 2006) and downsampling it into 110 $\times$ 250 regular grid cells. The grid cell size is 20m in both directions. For simplicity, we maintain a constant density. We simulate 12 sources on the surface with a spacing of 400m (indicated by red stars in Figure 1a). A receiver line containing 250 receivers at an interval of 20m is placed on the seabed at 200m depth (white line in Figure 1a). The observed waveform data are generated by solving the 2D acoustic wave equation using a time-domain finite difference method. The simulation length is 4s with a sample interval of 2ms. The source function is a Ricker wavelet with a dominant frequency of 10 Hz. Figure 1c displays this observed waveform dataset.

We define a Uniform prior distribution for the velocity values in each grid cell. Figure 1b shows the lower and upper bounds of the prior distribution at different depths. We set the velocity in the water layer (down to 200m depth) to its true value during inversion. The likelihood function is a Gaussian distribution (equation 2) with a diagonal covariance matrix Σ_d assuming independence among all data points. We take the maximum amplitude value of each trace and average them. The data noise is assumed to be 1% of the obtained average value. The same finite difference solver is used to calculate the synthetic waveform data $\mathbf{d}_{syn}$, and the gradient of the data misfit (negative log-likelihood function) with respect to the velocity model is computed using the adjoint-state method (Plessix, 2006). For variational inversion, we use Monte Carlo integration

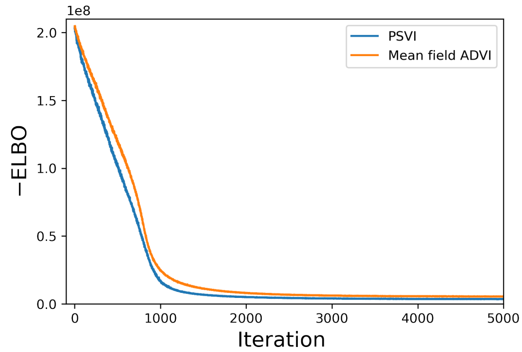

Figure 2. Variation of the negative ELBO with respect to iterations.

to estimate the ELBO in equation 4, and use the automatic differentiation framework provided by PyTorch to build a computational graph, which (automatically) calculates the ELBO and its gradient with respect to the variational parameters (Paszke et al., 2019). Optimization process is carried out using the Adam algorithm (Kingma & Ba, 2014).

We apply mean field ADVI and PSVI to this Bayesian FWI problem. Considering the dimensionality of this problem ($100\times250 = 25{,}000$), full rank ADVI is not performed since constructing a full covariance matrix would be extremely expensive in terms of both memory requirements and computational cost. For mean field ADVI, we use a diagonal Gaussian distribution to approximate the posterior distribution in the unbounded space. For PSVI, a 5×5 correlation kernel is employed to model the main correlations between model parameters, as illustrated by the white pluses in Figure 1d for the central black dotted cell. The choice of this correlation kernel is based on the estimated dominant wavelength of this problem (approximately 200m in shallow subsurface). In both tests, variational parameters ($\boldsymbol{\mu}$ and $\mathbf{L}$) are updated for 5000 iterations, with 2 random samples per iteration used to approximate the ELBO$[q(\mathbf{m})]$ and its gradients with respect to $\boldsymbol{\mu}$ and $\mathbf{L}$. Figure 2 displays the negative ELBOs for these two tests as a function of iterations, indicating that both algorithms achieve a reasonable level of convergence with nearly the same convergence speed, even though PSVI has far more parameters to optimise.

Figures 3a and 3b depict the inversion results. The mean, standard deviation and the relative error (computed by dividing the absolute error between the true and mean models by the standard deviation model) of the posterior distribution are displayed from top to bottom row. The two mean velocity maps exhibit similar features across most locations, generally resembling the true velocity map in Figure 1a. The inversion results struggle to recover some thin layers in the deeper part of the model, potentially due to the relatively low frequency (10 Hz) data used for FWI. Additionally, certain discrepancies are observed between these two maps at specific locations. For example, in the tilt layers annotated by red and black arrows in Figures 3a and 3b, the mean velocity model from mean field ADVI displays discontinuities, while the PSVI results show more continuity, closely resembling the true velocity model. One possible reason for this discrepancy is that accurate correlation information is crucial for recovering the continuity of spatial locations, especially for these thin layers. All correlations between pairs of model parameters are neglected in mean field ADVI, and thus the results may fail to recover the true velocity structures at these locations. By incorporating physically structured correlations between cells within a dominant wavelength, the proposed method improves the inversion accuracy.

Both inversion results show increased uncertainties with greater depth, since the sensitivity of observed seismic data decreases at depth, thus deeper parts of the model are less constrained by the data. The standard deviation values obtained from mean field ADVI are generally smaller than those from PSVI, especially in the shallower subsurface above 1.5km depth. This is because mean field ADVI tends to underestimate posterior uncertainties by neglecting correlations. Similar phenomena have been observed in previous studies (Ely et al., 2018; W. Wang et al., 2023; X. Zhao & Curtis, 2024). Therefore, the relative errors from mean field ADVI are larger compared to those from the proposed method, especially at locations with a depth of 1km and a distance between $0-1.5$km, where the mean model deviates from the true model by more than 3 standard deviations. This discrepancy suggests a low credibility of the inversion results obtained from mean field ADVI. As marked by a white arrow in Figure 3a, lower uncertainty noise is observed, which correspond to layers that are not continuous in the mean velocity map marked by a red arrow. This feature again proves that mean field ADVI provides biased uncertain estimates. By contrast, such uncertainty structures are not observed in Figure 3b, indicating that PSVI has the capability to correct some biases introduced by mean field ADVI.

To validate the inversion results displayed in Figure 3b, we apply two additional variational methods to this problem: *boosting variational inference* (BVI – F. Guo et al.,

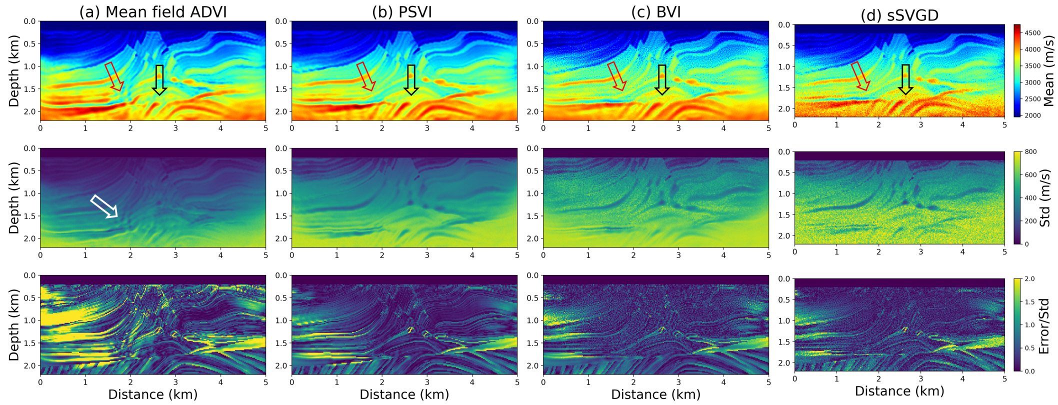

Figure 3. Mean (top row), standard deviation (middle row) and relative error (bottom row) of the posterior distribution obtained using (a) mean field ADVI, (b) PSVI, (c) boosting variational inference (BVI) and (d) stochastic SVGD (sSVGD), respectively. The relative error is the absolute error between the mean and true models divided by the corresponding standard deviation.

2016; Miller et al., 2017) and *stochastic Stein variational gradient descent* (sSVGD – Gallego & Insua, 2018). In BVI, a mixture distribution, in this case a mixture of Gaussians, is used to approximate the posterior distribution considering the fact that a mixture distribution can approximate any target distribution to any level of accuracy. sSVGD is a Monte Carlo based variational method that iteratively pushes a set of random samples towards the posterior distribution by minimising the KL divergence. In addition, a noise term is introduced to these samples at each iteration such that the algorithm converges to the true posterior distribution asymptotically. These two methods have been applied to acoustic FWI problems, and have proved to provide reasonable posterior solutions in two and three dimensional Earth models (X. Zhang et al., 2023; X. Zhao & Curtis, 2024). Figures 3c and 3d depict the inversion results obtained using BVI and sSVGD, respectively. They present very similar features compared to those displayed in Figure 3b: the same continuous structures in the deeper part of the model (denoted by red and black arrows) are observed in the mean velocity maps, and similar higher standard deviation values associated with lower relative errors (distributed within 2 standard deviations) are also present.

To further analyse the accuracy of the inversion results, in Figure 4 we compare the posterior marginal distributions obtained from the four tested methods along two

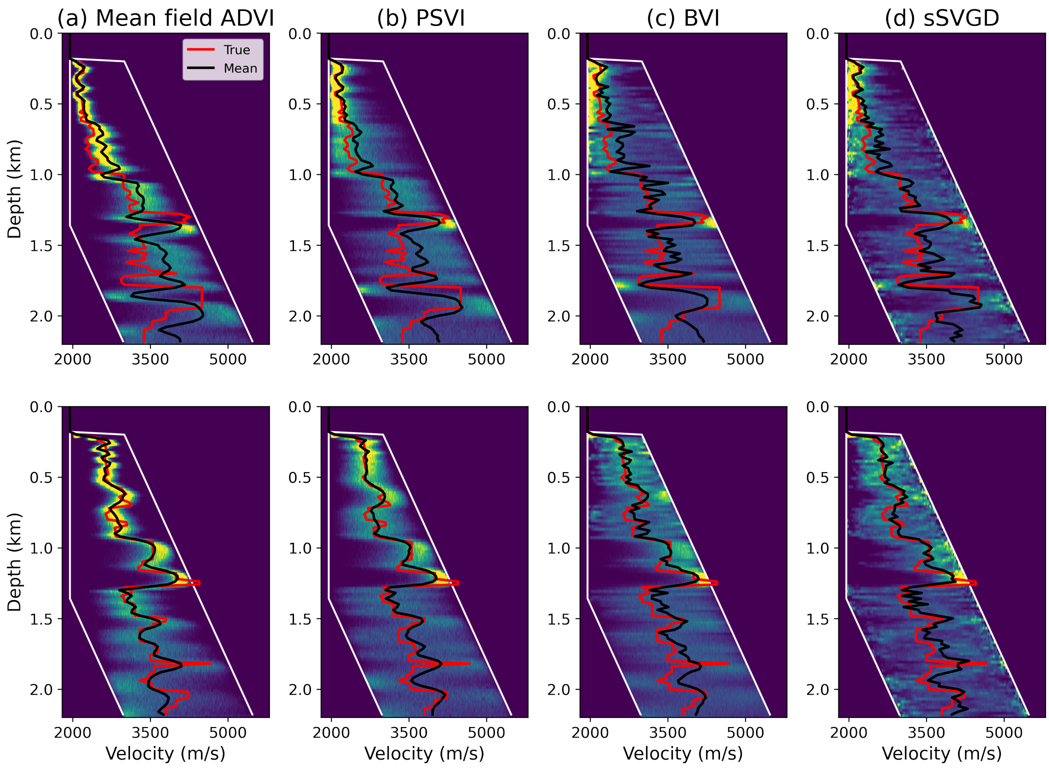

Figure 4. Posterior marginal distributions coloured from dark blue (zero probability) to yellow (maximum value of marginal pdf's in each plot), along two vertical profiles at distances of 1km (top row) and 2.6km (bottom row) obtained using (a) mean field ADVI, (b) PSVI, (c) BVI and (d) sSVGD. The locations of these two profiles are represented by black dashed lines in Figure 1a. In each figure, two white lines show the prior bounds, and black and red lines show the mean and true velocity values.

vertical profiles at horizontal locations of 1km (top row) and 2.6km (bottom row), respectively. The location of these two profiles are displayed by dashed black lines in Figure 1a. The first profile (at a distance of 1km) is strategically placed in regions where the relative errors from mean field ADVI (Figure 3a) are higher, while the second one (at 2.6km) is centrally located within the imaging region. Red lines show the true velocity values and black lines show the mean velocity values obtained using different methods. Overall, the marginal distributions in Figure 4a are narrower compared to those in Figures 4b to 4d, indicating lower posterior uncertainties akin to Figure 3. In the first row of Figure 4 between depths of 0.7km – 1 km and 1.3km – 1.8km, the true velocity values are excluded from the posterior distribution obtained using mean field ADVI, whereas those values correctly reside within the high probability region of the posterior pdfs obtained using the other three methods. These phenomena again prove that mean field ADVI tends to underestimate the posterior uncertainties and introduce biases into the inversion results. By including the main correlation information between adjacent grid cells, PSVI yields better inversion results that are highly consistent with two entirely independent methods. Therefore, we assert that the posterior standard deviations derived from PSVI are likely to be correct.

Given that PSVI is designed to capture correlations between spatially close grid cells, we compare the posterior correlation coefficients between model parameters estimated using different methods. Figure 5 shows the covariance matrices for velocity values within the white box in Figure 1a, obtained using the above four inversion methods. Mean field ADVI uses a transformed diagonal Gaussian distribution to approximate the posterior pdf and disregards correlations between model parameters, thus the posterior covariance matrix predominantly exhibits strong diagonal values corresponding to the variances of model parameters. By incorporating a specific (desired) correlation structure into the variational distribution, the covariance matrix obtained using PSVI displays off-diagonal values representing correlations between different parameters, which are not observed from the results using mean field ADVI. Due to the use of a 5×5 correlation kernel (as represented by the white pluses in Figure 1d), we only include correlation information between a given grid cell and cells within two layers of cells surrounding it. As a result, Figure 5b displays four off-diagonal blocks (two above and two below the diagonal elements). We observe negative correlations between neighbouring cells (in the first off-diagonal block below and above the diagonal values) and positive correlations between every second neighbouring cells (found in the second off-diagonal block).

In Figures 5c and 5d, similar negative off-diagonal correlation blocks are observed in the covariance matrices obtained using BVI and sSVGD. This confirms that in this test we successfully capture the correct correlation information between adjacent cells by using PSVI. While there may be positive correlations with cells two layers apart, these are not visible; this may be because Figures 5c and 5d show a general 'speckle' of non-zero background correlation values that are absent in Figure 5b. In PSVI, we construct a sparse covariance matrix with specific non-zero off-diagonal elements, and set all other values to zero. This neglects correlations between locations that are spatially far away from each other. It should be noted that we do not know whether any of these values in Figures 5c and 5d are correct, since they do not match between the two panels. In the next section, we also prove that these non-zero background correlations play a less significant role in a simulation of a real-world decision-making process. So again we suggest that our implementation of PSVI has modelled the most prominent and consistent features of the correlation structure.

Finally, we analyse the efficiency of the proposed method and compare its cost with other methods. As mentioned in Section 2, the number of hyperparameters that need to be optimised in PSVI is higher than that in mean field ADVI but is significantly lower than that in full rank ADVI. In our test, we find that the computational cost for optimising these variational parameters is much cheaper (almost negligible) compared to the cost used for forward and adjoint simulations in FWI. Therefore, the number of simulations serves as a good metric for the overall cost in this example.

Table 1 summarises the number of simulations used in each tested method. The same simulation settings are used in mean field ADVI and PSVI (10,000 simulations consisting of 5000 iterations with 2 samples per iteration). For BVI, we use a mixture of 24 diagonal Gaussian distributions to approximate the posterior distribution. Each component is updated by 2500 iterations with 2 samples per iteration. Note that the number of simulations used to optimise each component for BVI is smaller than that for ADVI, as full convergence of each component is not necessarily required in BVI (X. Zhao & Curtis, 2024). For sSVGD, we run 5000 iterations with 24 samples, resulting in a total of 120,000 forward evaluations for both BVI and sSVGD. In these two tests, relatively larger

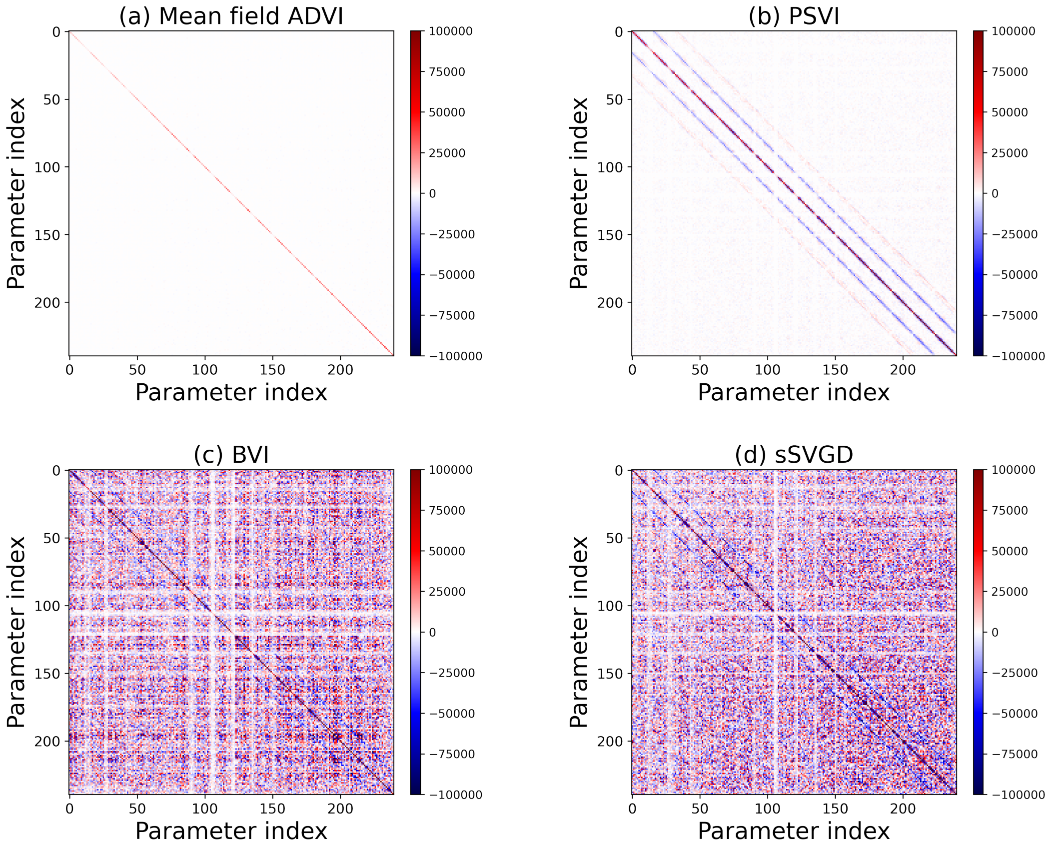

Figure 5. Covariance matrices for velocity values inside the white box in Figure 1a, calculated using the inversion results from (a) mean field ADVI, (b) PSVI, (c) BVI and (d) sSVGD.

Table 1. Number of forward and gradient evaluations for mean field ADVI, PSVI, BVI, and sSVGD. The values represent an indication of the computational cost of each method, as the evaluation of data-model gradients in FWI is by far the most expensive part of each calculation.

Method	Number of Gradient Evaluations
Mean field ADVI	10,000
PSVI	10,000
BVI	120,000
sSVGD	120,000

step sizes are used to speedup the convergence of BVI and sSVGD. However, they still remain one order of magnitude more computationally expensive than mean field ADVI and PSVI. In addition, Figure 2 shows that mean field ADVI and PSVI present roughly the same convergence rate given the same number of forward simulations. This verifies the statement that PSVI is almost as efficient as mean field ADVI. The latter is known to be a particularly inexpensive (yet biased) method for Bayesian inversion from previous studies (X. Zhang & Curtis, 2020a; X. Zhao et al., 2021; Bates et al., 2022; Sun et al., 2023). On the other hand, the PSVI method improves the inversion accuracy and provides similar results compared to two accurate but more computationally demanding methods (BVI and sSVGD). Thus, the proposed method shown to be an efficient algorithm that has provided reliable uncertainty estimates.

4 Interrogating FWI results

The objective of scientific investigations is typically to answer some specific and high-level questions. Examples of these questions in the field of geophysics can be: *How large is a subsurface structure? Is this a good location for carbon capture and storage (CCS)?* Normally these questions are answered in a biased manner without evaluating uncertainties in the results. Interrogation theory provides a systematic way to obtain the least-biased answer to these questions (Arnold & Curtis, 2018). In this section, we solve two interrogation problems using the FWI results obtained above, to evaluate the potential practical value of the correlations estimated by PSVI.

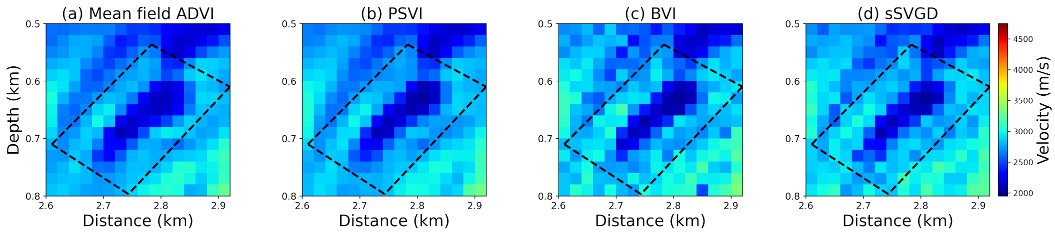

Figure 6. Mean velocity maps inside the white box in Figure 1a (corresponding to the true velocity map displayed in Figure 1d), obtained using (a) mean field ADVI, (b) PSVI, (c) BVI and (d) sSVGD. Black dashed boxes show the region where interrogation is performed.

Interrogation theory shows that the optimal answer a^* to a specific question Q that has a continuous space of possible answers is expressed by the following expectation term:

$$a^* = \mathbb{E}[T(\mathbf{m}|Q)] = \int_{\mathbf{m}} T(\mathbf{m}|Q)p(\mathbf{m}|\mathbf{d}_{obs})\ d\mathbf{m}, \tag{7}$$

where optimality is defined with respect to a squared utility (Arnold & Curtis, 2018). The expectation is taken with respect to the posterior distribution $p(\mathbf{m}|\mathbf{d}_{obs})$ of model parameter $\mathbf{m}$. Term $T(\mathbf{m}|Q)$ is a target function conditioned on the question Q of interest. It is defined to map the high dimensional model parameter $\mathbf{m}$ into a low dimensional target function value t in a target space $\mathbb{T}$, within which the question Q can be answered directly. In such cases the optimal answer in equation 7 is simply the expectation or mean of the posterior target function.

4.1 Interrogation for reservoir size

Figure 6 shows the inverted mean models of the velocity structure within the white box in Figure 1a, obtained through (a) mean field ADVI, (b) PSVI, (c) BVI, and (d) sSVGD. In each figure, we observe a low velocity body at the centre of the model section, outlined by a dashed black box. In this first example, we treat this low velocity zone as a reservoir and use interrogation theory to estimate its size.

Previously, volume-related questions were answered using seismic imaging results obtained from travel time tomographic inversion (X. Zhao et al., 2022) and FWI (X. Zhang & Curtis, 2022; X. Zhao & Curtis, 2024). Following these studies, we define a target function $T(\mathbf{m}|Q)$ as the area of the largest continuous low velocity body, which converts a high dimensional velocity model into a scalar value, representing the estimated reservoir

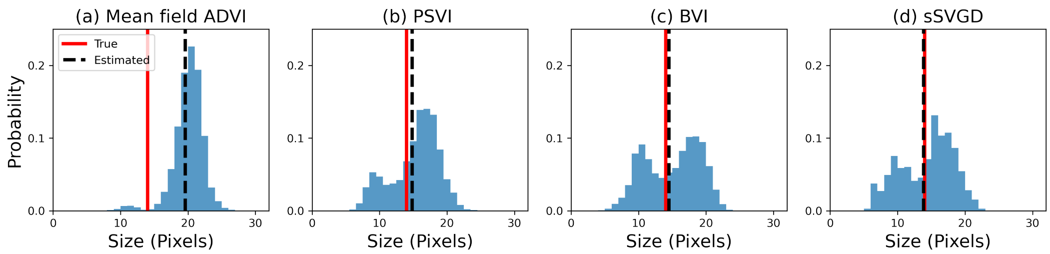

Figure 7. Posterior distributions of the low velocity reservoir size using FWI results obtained from (a) mean field ADVI, (b) PSVI, (c) BVI and (d) sSVGD, respectively. Red lines denote the true reservoir size, and black dashed lines denote the optimal size obtained using interrogation theory.

area from a given posterior sample. Note that this process involves using a velocity threshold to distinguish between low and high velocities. We use the same data-driven method introduced in X. Zhao et al. (2022) to determine the least biased estimate of this threshold value. This involves selecting some cells that are almost definitely inside the low velocity anomaly, others that are almost definitely outside; we then choose the threshold value such that the expected probability of interior cells being below that value equals the expected probability of exterior cells being above that value, according to the posterior pdf. We are then able to calculate the target function for every posterior sample.

Figure 7 displays the posterior distributions of the target function (reservoir size) using the four inversion results obtained previously. In this synthetic test, the true reservoir area is precisely known from Figure 1d and is denoted by red lines in Figure 7. The optimal (least-biased) answer estimated from each inversion method corresponds to the mean value of the respective posterior target function (as per equation 7), and is displayed by a dashed black line in each figure. As discussed in previous sections, mean field ADVI tends to underestimate posterior uncertainties and provides biased inversion results. We see that, the corresponding interrogation results in Figure 7a are also biased: the optimal answer shows a significant error and is far from the true answer, and indeed the true answer is even excluded from the posterior distribution of the estimated volume. By contrast, if we impose physically structured correlation information on model parameter, the optimal answer estimated by PSVI aligns closely with the true answer (Figure 7b). The posterior distribution of the target function also successfully captures bimodal uncertainties, similar to those obtained using BVI and sSVGD.

4.2 Interrogation for CO_2 storage

In the second example, we apply the inversion results to answer a more realistic and practically interesting question. Assume the low velocity reservoir identified above is used in a carbon capture and storage (CCS) project and is injected with CO_2. The injection of CO_2 into a porous rock produces changes in petrophysical parameters of the rock, such as pore fluid phase and water saturation. These changes further result in variations in seismic response of a reservoir, such as seismic velocity. Leveraging the FWI results, we can use these variations to monitor the injected CO_2 in a subsurface CCS project by answering the question: *what is the total volume of CO_2 stored in this reservoir?*

For the characterisation of changes in seismic velocity due to physical parameters related to CO_2, especially CO_2 saturation (S_{co_2}) in the reservoir, we first represent the P wave velocity v_p of a saturated rock using the bulk modulus K_{sat}, shear modulus G_{sat} and density ρ_{sat} of the rock by

$$v_p = \sqrt{\frac{K_{sat} + 4G_{sat}/3}{\rho_{sat}}} \tag{8}$$

The bulk modulus can be calculated using the Gassmann equation (Gassmann, 1951):

$$K_{sat} = K_d + \frac{(1 - \frac{K_d}{K_m})^2}{\frac{\phi}{K_f} + \frac{1-\phi}{K_m} - \frac{K_d}{K_m^2}} \tag{9}$$

where ϕ is the porosity, and K_d, K_m and K_f are the bulk moduli of dry rock, solid matrix and pore fluid. The density of a saturated rock can be calculated as

$$\rho_{sat} = (1 - \phi)\rho_m + \phi\rho_f \tag{10}$$

where ρ_m and ρ_f are the densities of grain matrix and fluid, respectively. The shear modulus G_{sat} is not affected by fluid and only depends on the shear modulus of dry rock G_d

$$G_{sat} = G_d \tag{11}$$

Assuming the reservoir is saturated by two distinct fluids, water and CO_2, the saturation values for water (S_w) and CO_2 (S_{co_2}) are constrained by the relation: $S_w + S_{co_2} = 1$. Then, the bulk modulus and density of fluid can be calculated using the mixing rules

$$\rho_f = S_w\rho_w + S_{co_2}\rho_{co_2} \tag{12}$$

$$K_f = S_w^e K_w + (1 - S_w^e)K_{co_2} \tag{13}$$

Table 2. Rock physics parameters and their associated standard deviations (uncertainties) estimated from the Sleipner field (Dupuy et al., 2017; Ghosh & Ojha, 2020).

Parameter	K_m (GPa)	K_d (GPa)	K_w (GPa)	K_{co_2} (GPa)	G_m (GPa)	G_d (GPa)	ρ_m (kg/m^3)	ρ_w (kg/m^3)	ρ_{co_2} (kg/m^3)	ϕ (%)
Mean value	39.3	2.56	2.31	0.08	44.8	8.1	2664	1030	700	0.3
Uncertainty	1.41	0.08	0.07	0.04	0.81	0.24	3	20	77	0.02

where ρ_w, ρ_{co_2}, K_w and K_{co_2} are the densities and bulk moduli of water and CO_2, and e is an empirical value (Brie et al., 1995). In this example, we use $e = 11$ as suggested by Kim et al. (2013). The injection of CO_2 into a reservoir alters the saturation values S_w and S_{co_2}, changing K_f and ρ_f, and thus also v_p through equations 8 to 13. Therefore, we can estimate S_{co_2} using P wave velocity values obtained from FWI.

To simplify the problem, we assume that some of the aforementioned rock physics parameters follow Gaussian distributions. Their means and standard deviations are estimated from the Sleipner field (Dupuy et al., 2017; Ghosh & Ojha, 2020; Strutz & Curtis, 2024), as listed in Table 2. Given these parameters, we build a direct relationship between P wave velocity v_p and CO_2 saturation S_{co_2}. The results are depicted by the joint probability distribution of v_p and S_{co_2} displayed in Figure 8a. The red curve is the reference $v_p - S_{co_2}$ curve obtained using the mean values from Table 2. In Figure 8a, the posterior distribution of CO_2 saturation for any P-wave velocity value can be obtained. For example, Figures 8b and 8c illustrate two such posterior pdfs corresponding to velocity values of 2045m/s (solid white line in Figure 8a) and 1840m/s (dashed white line). In Figure 8 we observe that seismic velocity is sensitive to small CO_2 saturations (below 0.2) and is insensitive for larger S_{co_2} values (Kim et al., 2013).

In the previous interrogation example, we defined the largest continuous low velocity body as the reservoir of interest for a posterior velocity sample. For each grid cell within the identified reservoir, we substitute its velocity value into Figure 8a to obtain the posterior pdf of CO_2 saturation. Finally, the total (2D) CO_2 volume V_{co_2} stored in the reservoir can be calculated by

$$V_{co_2} = \sum V \phi S_{co_2} \tag{14}$$

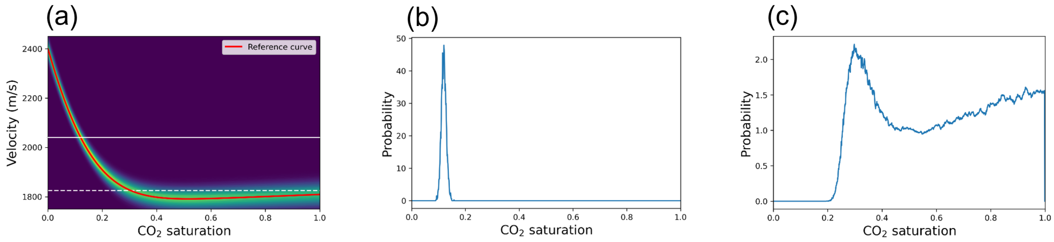

Figure 8. (a) Joint probability distribution of P wave velocity and CO_2 saturation given other parameters listed in Table 2. Red curve shows a one-to-one mapping between v_p and S_{co_2} obtained using the mean values in Table 2, and the colour scale from red through green to dark blue represents the probability distribution of velocity, given any value of CO_2 and the Gaussian distributions defined in Table 2. (b) and (c) display the posterior distributions of CO_2 saturation for velocity values of 2045m/s and 1840m/s, marked by solid and dashed white lines, respectively, in (a).

where V is the (2D) volume (i.e. area) of each grid cell in FWI, and the summation is taken over all grid cells within the reservoir. This defines the target function for this interrogation problem.

Figure 9 displays the posterior distributions of the estimated (2D) CO_2 volume obtained using different inversion methods. Similar to the reservoir size displayed in Figure 7, mean field ADVI provides rather biased interrogation results since it tends to underestimate posterior uncertainties. In contrast, the other three methods provide similar (and possibly correct) posterior distributions with two distinct modes. The three estimated answers are close to the true value, which lies inside the high probability region of the posterior distributions. Figures 7 and 9 prove that PSVI provides accurate uncertainty information that can be used to answer real-world questions correctly. Moreover, the non-zero background correlations ignored by PSVI (displayed in Figures 5c and 5d) are shown to be less important for post-inversion decision-making.

5 Discussion

PSVI can be considered as an intermediate approach between mean field ADVI and full rank ADVI (Kucukelbir et al., 2017). Mean field ADVI neglects all correlations to reduce computations and thus strongly underestimates posterior uncertainties. Full rank ADVI includes full correlation information between model parameters but is computa-

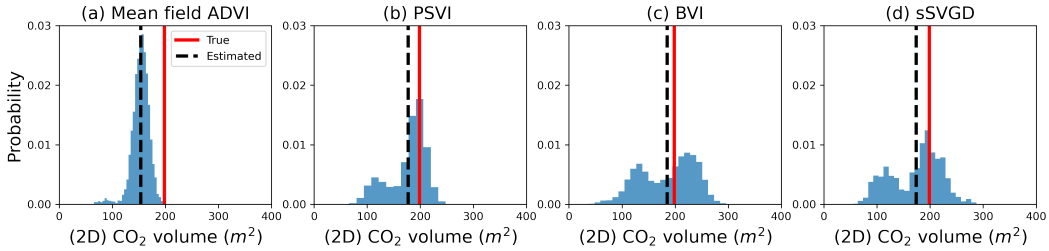

Figure 9. Posterior distributions of the (2D) CO_2 volume stored in the low velocity reservoir, calculated using (a) mean field ADVI, (b) PSVI, (c) BVI and (d) sSVGD. Red lines denote the true CO_2 volume, and black dashed lines denote the least-biased CO_2 volume estimated using interrogation theory.

tionally intractable for high dimensional problems such as 2D or 3D FWI. PSVI, with its ability to capture structured correlations, strikes a balance between efficiency and accuracy. In the context of Bayesian FWI, where problems are often high dimensional and non-linear, PSVI offers improved inversion results while maintaining a computational cost comparable to mean field ADVI. For inverse problems with lower dimensionality such that modelling a full covariance matrix is affordable, full rank ADVI could be a more suitable choice. When dealing with problems with strong multimodality, these Gaussian-based methods are not suitable. It is then advisable to use other variational methods such as normalizing flows (Rezende & Mohamed, 2015), BVI (F. Guo et al., 2016; Miller et al., 2017) or deterministic or stochastic SVGD (Liu & Wang, 2016; Gallego & Insua, 2018). These methods have shown effectiveness in solving multimodal problems, albeit at the cost of a larger number of forward simulations. The No Free Lunch theorem (Wolpert & Macready, 1997) can be paraphrased as: no method is better than any other method when averaged across all problems. There is therefore no possibility to find a 'best' method in general. Nevertheless, individual classes of problems may have more or less efficient algorithms, so having a variety of methods allows for tailored decisions to be based on the nature of the problem to be addressed.

In the 2D FWI example, we use a 5×5 correlation kernel as displayed in Figure 1d. To investigate the impact of the correlation kernel size on inversion results, we conduct an additional test using an 11×11 kernel. The mean, standard deviation and relative error maps of the obtained posterior distribution are displayed in Figure 10a, which reveal nearly identical features, such as the continuous layers discussed previously, when

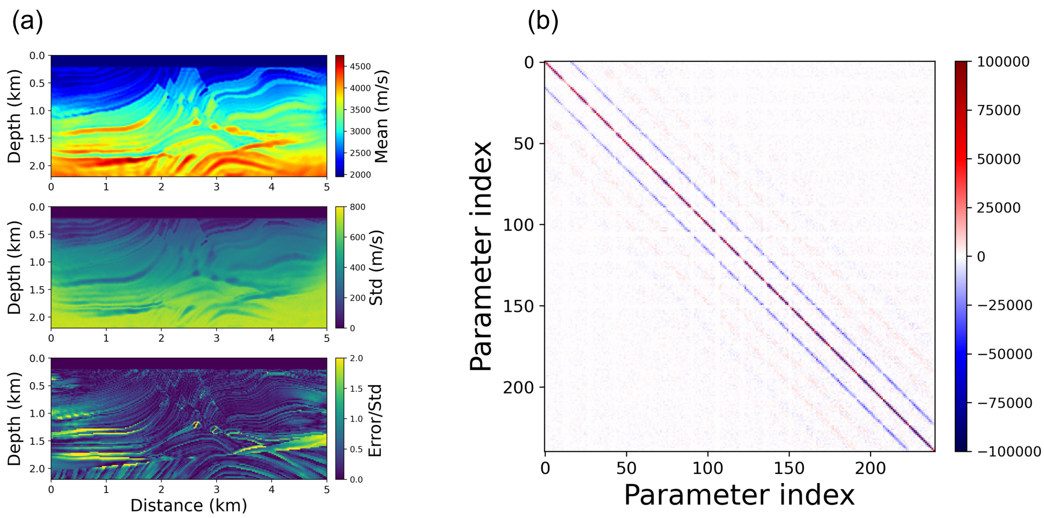

Figure 10. Inversion results obtained from PSVI using an 11×11 correlation kernel. (a) Mean, standard deviation and relative error maps. (b) Covariance matrix inside the white box in Figure 1a.

compared to those obtained using the 5×5 correlation kernel (Figure 3b). Figure 10b displays the posterior covariance matrix, which as expected presents more non-zero off-diagonal covariance blocks than the 5×5 kernel (Figure 5b). The covariance magnitudes decay from the main diagonal block, and become relatively small from the second off-diagonal block. However, modelling these additional covariances requires more parameters to construct the matrix $\mathbf{L}$. In addition, from Figures 5c and 5d, the covariance matrices calculated using BVI and sSVGD exhibit only one prominent off-diagonal block, probably because the non-linearity of FWI makes it challenging to capture a broader correlation structure with embedding prior knowledge of the type of structure sought. Therefore, we conclude that the 5×5 correlation kernel used above is a reasonable choice that trades off both accuracy and efficiency.

In real applications, if other prior knowledge about the subsurface structure is available (e.g., from seismic travel time tomography), we can design specific correlation kernels to capture target-oriented correlation information. Furthermore, the underlying principles of PSVI can be adapted to address temporal problems such as time-lapse (4D) seismic monitoring in which we might expect spatial regularity in the location of injected fluids, or in earthquake forecasting where correlations between seismic events over time might be captured effectively.

PSVI is not merely an extension of mean field ADVI as proposed by Kucukelbir et al. (2017). In fact it can be used to extend a variety of variational methods to enhance their accuracy and efficiency. For example, in BVI the physically structured approach in PSVI can replace diagonal Gaussians in modelling the Gaussian component distributions used in X. Zhao and Curtis (2024). This substitution is likely to improve the accuracy of each component while maintaining similar computational efficiency, potentially leading to a reduction in the required number of components and overall computational cost for BVI.

Similar to BVI, PSVI produces an analytic posterior expression. Therefore, saving and loading inversion results, generating new posterior samples, and sharing the posterior distribution with others post inversion is simple (Scheiter et al., 2022). The proposed method can also be extended to other general Gaussian-based methods such as Gaussian processes (Ray & Myer, 2019; Valentine & Sambridge, 2020a, 2020b; Ray, 2021; Blatter et al., 2021) and mixture density networks (Bishop, 1994; Devilee et al., 1999; Meier et al., 2007; Shahraeeni & Curtis, 2011; Shahraeeni et al., 2012; Earp & Curtis, 2020; Hansen & Finlay, 2022; Bloem et al., 2023), to capture desired correlation structures. Interestingly, special neural network structures are designed for the same purpose, such as the *coupling layer* (Dinh et al., 2015, 2017; Durkan et al., 2019; X. Zhao et al., 2021; X. Zhang & Curtis, 2021b) and the *autoregressive layer* (Kingma et al., 2016; Papamakarios et al., 2017; Huang et al., 2018; De Cao et al., 2019; Levy et al., 2022). However, they often come with a higher number of hyperparameters, making PSVI an attractive and practical choice.

Considering that solving the forward function in 2D FWI is not hugely expensive, we use a relatively smaller step size and more iterations during variational inversion to ensure that the optimisation process has converged stably. Figure 2 illustrates that the negative ELBOs stop decreasing after 2500 - 3000 iterations, indicating that the full 5000 iterations used here might be redundant. For higher dimensional problems such as 3D FWI, we can potentially use larger step sizes with fewer iterations, thereby optimising the balance between computational resources and convergence speed.

The two interrogation examples presented here underscore the significance of estimating accurate uncertainties, even if that demands a substantial increase in computational input. Biased uncertainty information (such as that provided by mean field ADVI) leads to incorrect answers about Earth properties. Therefore, while obtaining an accurate mean velocity model in Bayesian inversion, or just the best-fit model in deterministic inversion, may appear useful, they are far from sufficient for an unbiased and quantitative interpretation of the true Earth. The pursuit of not only precision in mean velocity models but also robust and reliable uncertainty estimates is important for a comprehensive understanding of subsurface structures.

In the first interrogation example, we estimated the size of a subsurface reservoir, where we use relative velocity values and classify them as either low or high based on a velocity threshold value (X. Zhao et al., 2022). In the second example, we take the absolute velocity values and convert them into CO_2 saturation estimates using a non-linear rock physics relationship. If the inversion is performed with higher frequency data, the inverted velocity values would be better constrained and become more accurate. Consequently, the posterior distribution of the estimated CO_2 volume can be improved. In future, 3D Bayesian FWI, together with more advanced reservoir simulation and rock physics inversion techniques, can facilitate more sophisticated and realistic interrogation applications in subsurface carbon capture and storage, or other subsurface projects. This comprehensive approach, enriched with full uncertainty assessments, could significantly contribute to our understanding and improve decision-making in the context of such endeavours.

6 Conclusion

In this work, we propose physically structured variational inference (PSVI) to perform 2D Bayesian full waveform inversion (FWI), in which a physical structure is imposed on the uncertainties in variational distributions based on prior information about imaging problem solutions. In our application, correlations between specific pairs of spatial locations are parametrised and inferred during inversion. Thus, we are able to capture the main correlations with a desired structure in a computationally efficient manner. We apply the proposed method together with three other variational methods: mean field automatic differentiation variational inference (ADVI), boosting variational inference (BVI) and stochastic Stein variational gradient descent (sSVGD), to a synthetic FWI example. This demonstrates that PSVI yields accurate first-order statistical information, including the mean and standard deviation maps as well as the marginal distributions, which are all consistent with those obtained using BVI and sSVGD. It also provides other second-order statistical information, specifically the posterior covariances. In addition, the obtained full uncertainty information is verified through the application of the inversion results to two post-inversion interrogation problems: one estimating a subsurface reservoir size and another estimating CO_2 volume in a carbon capture and storage project. In our examples, PSVI exhibits nearly the same computational efficiency as mean field ADVI while enhancing the inversion accuracy significantly. This opens the possibility that 3D probabilistic FWI with full uncertainty estimation can be performed both efficiently and accurately.

7 Open Research

Software used to perform variational inference can be found at Pyro website (`https://pyro.ai/`, Bingham et al., 2018) and in X. Zhang and Curtis (2023). Software used to perform Automatic Differentiation can be found at PyTorch website (`https://pytorch.org/`, Paszke et al., 2019).

Acknowledgments

The authors thank Edinburgh Imaging Project (EIP - `https://blogs.ed.ac.uk/imaging/`) sponsors (BP and TotalEnergies) for supporting this research.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., … others (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Ardizzone, L., Kruse, J., Wirkert, S., Rahner, D., Pellegrini, E. W., Klessen, R. S., … Köthe, U. (2018). Analyzing inverse problems with invertible neural networks. *arXiv preprint arXiv:1808.04730*.

Arnold, R., & Curtis, A. (2018). Interrogation theory. *Geophysical Journal International*, *214*(3), 1830–1846.

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P., & Virieux, J. (2013). Regularized seismic full waveform inversion with prior model information. *Geophysics*, *78*(2), R25–R36.

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

Bernard, S., Monteiller, V., Komatitsch, D., & Lasaygues, P. (2017). Ultrasonic computed tomography based on full-waveform inversion for bone quantitative imaging. *Physics in Medicine & Biology*, *62*(17), 7011.

Berti, S., Aleardi, M., & Stucchi, E. (2023). A computationally efficient bayesian approach to full-waveform inversion. *Geophysical Prospecting*. doi: https://doi.org/10.1111/1365-2478.13437

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., ... Goodman, N. D. (2018). Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*.

Bishop, C. M. (1994). Mixture density networks. *Aston University*.

Bishop, C. M. (2006). *Pattern recognition and machine learning.* springer.

Biswas, R., & Sen, M. K. (2022). Transdimensional 2d full-waveform inversion and uncertainty estimation. *arXiv preprint arXiv:2201.09334*.

Blatter, D., Ray, A., & Key, K. (2021). Two-dimensional bayesian inversion of magnetotelluric data using trans-dimensional gaussian processes. *Geophysical Journal International*, *226*(1), 548–563.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association*, *112*(518), 859–877.

Bloem, H., Curtis, A., & Tetzlaff, D. (2023). Introducing conceptual geological information into Bayesian tomographic imaging. *Basin Research*. doi: doi:10.1111/bre.12811

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization.* Cambridge university press.

Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., ... Pugmire, D. (2016). Global adjoint tomography: first-generation model. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, *207*(3), 1739–1766.

Brie, A., Pampuri, F., Marsala, A., & Meazza, O. (1995). Shear sonic interpretation in gas-bearing sands. In *Spe annual technical conference and exhibition* (pp. SPE–30595).

Bui-Thanh, T., Ghattas, O., Martin, J., & Stadler, G. (2013). A computational framework for infinite-dimensional bayesian inverse problems part I: The linearized case, with application to global seismic inversion. *SIAM Journal on Scientific Computing*, *35*(6), A2494–A2523.

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

De Cao, N., Titov, I., & Aziz, W. (2019). Block neural autoregressive flow. *arXiv preprint arXiv:1904.04676*.

de Lima, P. D. S., Corso, G., Ferreira, M. S., & de Araújo, J. M. (2023). Acoustic full waveform inversion with hamiltonian monte carlo method. *Physica A: Statistical Mechanics and its Applications*, *617*, 128618.

de Lima, P. D. S., Ferreira, M. S., Corso, G., & de Araújo, J. M. (2023). Bayesian time-lapse full waveform inversion using hamiltonian monte carlo. *arXiv preprint arXiv:2311.02999*.

Devilee, R., Curtis, A., & Roy-Chowdhury, K. (1999). An efficient, probabilistic neural network approach to solving inverse problems: inverting surface wave velocities for eurasian crustal thickness. *Journal of Geophysical Research: Solid Earth*, *104*(B12), 28841–28857.

Dhabaria, N., & Singh, S. C. (2024). Hamiltonian monte carlo based elastic full-waveform inversion of wide-angle seismic data. *Geophysical Journal International*, ggae112.

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

Dinh, L., Sohl-Dickstein, J., & Bengio, S. (2017). Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*.

Dupuy, B., Ghaderi, A., Querendez, E., Mezyk, M., et al. (2017). Constrained avo for co2 storage monitoring at sleipner. *Energy Procedia*, *114*, 3927–3936.

Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. (2019). Neural spline flows. In *Advances in neural information processing systems* (pp. 7509–7520).

Earp, S., & Curtis, A. (2020). Probabilistic neural network-based 2d travel-time tomography. *Neural Computing and Applications*, *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.

Fang, Z., Da Silva, C., Kuske, R., & Herrmann, F. J. (2018). Uncertainty quantification for inverse problems with weak partial-differential-equation constraints. *Geophysics*, *83*(6), R629–R647.

Fichtner, A., Kennett, B. L., Igel, H., & Bunge, H.-P. (2009). Full seismic waveform tomography for upper-mantle structure in the australasian region using adjoint methods. *Geophysical Journal International*, *179*(3), 1703–1725.

Fichtner, A., van Herwaarden, D.-P., Afanasiev, M., Simutė, S., Krischer, L., Çubuk-Sabuncu, Y., … others (2018). The collaborative seismic earth model: generation 1. *Geophysical research letters*, *45*(9), 4007–4016.

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

Fu, X., & Innanen, K. A. (2022). A time-domain multisource bayesian/markov chain monte carlo formulation of time-lapse seismic waveform inversion. *Geophysics*, *87*(4), R349–R361.

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

Gassmann, F. (1951). Elastic waves through a packing of spheres. *Geophysics*, *16*(4), 673–685.

Gebraad, L., Boehm, C., & Fichtner, A. (2020). Bayesian elastic full-waveform inversion using hamiltonian monte carlo. *Journal of Geophysical Research: Solid Earth*, *125*(3), e2019JB018428.

Ghosh, R., & Ojha, M. (2020). Prediction of elastic properties within co2 plume at sleipner field using avs inversion modified for thin-layer reflections guided by uncertainty estimation. *Journal of Geophysical Research: Solid Earth*, *125*(11), e2020JB019782.

Gouveia, W. P., & Scales, J. A. (1998). Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis. *Journal of Geophysical Research: Solid Earth*, *103*(B2), 2759–2779.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, *82*(4), 711–732.

Green, P. J. (2003). Trans-dimensional markov chain monte carlo. *Oxford Statistical Science Series*, 179–198.

Guasch, L., Calderón Agudo, O., Tang, M.-X., Nachev, P., & Warner, M. (2020). Full-waveform inversion imaging of the human brain. *NPJ digital medicine*, *3*(1), 28.

Guo, F., Wang, X., Fan, K., Broderick, T., & Dunson, D. B. (2016). Boosting variational inference. *Advances in Neural Information Processing Systems*.

Guo, P., Visser, G., & Saygin, E. (2020). Bayesian trans-dimensional full waveform inversion: synthetic and field data application. *Geophysical Journal International*, *222*(1), 610–627.

Hansen, T. M., & Finlay, C. C. (2022). Use of machine learning to estimate statistics of the posterior distribution in probabilistic inverse problems—an application to airborne em data. *Journal of Geophysical Research: Solid Earth*, *127*(11), e2022JB024703.

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

He, J., Rao, J., Fleming, J. D., Gharti, H. N., Nguyen, L. T., & Morrison, G. (2021). Numerical ultrasonic full waveform inversion (fwi) for complex structures in coupled 2d solid/fluid media. *Smart Materials and Structures*, *30*(8), 085044.

Huang, C.-W., Krueger, D., Lacoste, A., & Courville, A. (2018). Neural autoregressive flows. *arXiv preprint arXiv:1804.00779*.

Izzatullah, M., Ravasi, M., & Alkhalifah, T. (2023). Physics reliable frugal uncertainty analysis for full waveform inversion. *arXiv preprint arXiv:2305.07921*.

Khoshkholgh, S., Zunino, A., & Mosegaard, K. (2021). Informed proposal monte carlo. *Geophysical Journal International*, *226*(2), 1239–1248.

Khoshkholgh, S., Zunino, A., & Mosegaard, K. (2022). Full-waveform inversion by informed-proposal monte carlo. *Geophysical Journal International*, *230*(3), 1824–1833.

Kim, J., Nam, M. J., & Matsuoka, T. (2013). Estimation of co2 saturation during both co2 drainage and imbibition processes based on both seismic velocity and electrical resistivity measurements. *Geophysical Journal International*, *195*(1), 292–300.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kingma, D. P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I., & Welling, M. (2016). Improved variational inference with inverse autoregressive flow. In *Advances in neural information processing systems* (pp. 4743–4751).

Kotsi, M., Malcolm, A., & Ely, G. (2020a). Time-lapse full-waveform inversion using hamiltonian monte carlo: a proof of concept. In *Seg technical program expanded abstracts 2020* (pp. 845–849). Society of Exploration Geophysicists.

Kotsi, M., Malcolm, A., & Ely, G. (2020b). Uncertainty quantification in time-lapse seismic imaging: A full-waveform approach. *Geophysical Journal International*, *222*(2), 1245–1263.

Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., & Blei, D. M. (2017). Automatic differentiation variational inference. *The Journal of Machine Learning Research*, *18*(1), 430–474.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, *22*(1), 79–86.

Levy, S., Laloy, E., & Linde, N. (2022). Variational bayesian inference with complex geostatistical priors using inverse autoregressive flows. *Computers & Geosciences*, 105263.

Liu, Q., & Wang, D. (2016). Stein variational gradient descent: A general purpose bayesian inference algorithm. In *Advances in neural information processing systems* (pp. 2378–2386).

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 & exhibition* (Vol. 2023, pp. 1–5).

Lucka, F., Pérez-Liva, M., Treeby, B. E., & Cox, B. T. (2021). High resolution 3d ultrasonic breast imaging by time-domain full waveform inversion. *Inverse Problems*, *38*(2), 025008.

Martin, G. S., Wiley, R., & Marfurt, K. J. (2006). Marmousi2: An elastic upgrade for marmousi. *The leading edge*, *25*(2), 156–166.

Meier, U., Curtis, A., & Trampert, J. (2007). Fully nonlinear inversion of fundamental mode surface waves for a global crustal model. *Geophysical Research Letters*, *34*(16).

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, *21*(6), 1087–1092.

Miller, A. C., Foti, N. J., & Adams, R. P. (2017). Variational boosting: Iteratively refining posterior approximations. In *International conference on machine learning* (pp. 2420–2429).

Mosegaard, K., & Tarantola, A. (1995). Monte carlo sampling of solutions to inverse problems. *Journal of Geophysical Research: Solid Earth*, *100*(B7), 12431–12447.

Papamakarios, G., Pavlakou, T., & Murray, I. (2017). Masked autoregressive flow for density estimation. In *Advances in neural information processing systems* (pp. 2338–2347).

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., … others (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*.

Patsia, O., Giannopoulos, A., & Giannakis, I. (2023). GPR full-waveform inversion with deep-learning forward modelling: A case study from non-destructive testing. *IEEE Transactions on Geoscience and Remote Sensing*.

Plessix, R.-E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, *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.

Pratt, R. G., Shin, C., & Hick, G. (1998). Gauss–newton and full newton methods in frequency–space seismic waveform inversion. *Geophysical journal international*, *133*(2), 341–362.

Prieux, V., Brossier, R., Operto, S., & Virieux, J. (2013). Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the valhall field. part 1: Imaging compressional wave speed, density and attenuation. *Geophysical Journal International*, *194*(3), 1640–1664.

Ray, A. (2021). Bayesian inversion using nested trans-dimensional gaussian processes. *Geophysical Journal International*, *226*(1), 302–326.

Ray, A., Kaplan, S., Washbourne, J., & Albertin, U. (2018). Low frequency full waveform seismic inversion within a tree based bayesian framework. *Geophysical Journal International*, *212*(1), 522–542.

Ray, A., & Myer, D. (2019). Bayesian geophysical inversion with trans-dimensional gaussian process machine learning. *Geophysical Journal International*, *217*(3), 1706–1726.

Ray, A., Sekar, A., Hoversten, G. M., & Albertin, U. (2016). Frequency domain full waveform elastic inversion of marine seismic data from the alba field using a bayesian trans-dimensional algorithm. *Geophysical Journal International*, *205*(2), 915–937.

Rezende, D. J., & Mohamed, S. (2015). Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*.

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

Sambridge, M., & Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, *40*(3), 3–1.

Scheiter, M., Valentine, A., & Sambridge, M. (2022). Upscaling and downscaling monte carlo ensembles with generative models. *Geophysical Journal International*, *230*(2), 916–931.

Sen, M. K., & Roy, I. G. (2003). Computation of differential seismograms and iteration adaptive regularization in prestack waveform inversion. *Geophysics*, *68*(6), 2026–2039.

Shahraeeni, M. S., & Curtis, A. (2011). Fast probabilistic nonlinear petrophysical inversion. *Geophysics*, *76*(2), E45–E58.

Shahraeeni, M. S., Curtis, A., & Chao, G. (2012). Fast probabilistic petrophysical mapping of reservoirs from 3d seismic data. *Geophysics*, *77*(3), O1–O19.

Siahkoohi, A., & Herrmann, F. J. (2021). Learning by example: fast reliability-aware seismic imaging with normalizing flows. In *First international meeting for applied geoscience & energy* (pp. 1580–1585).

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., Louboutin, M., Witte, P. A., & Herrmann, F. J. (2021).

Preconditioned training of normalizing flows for variational inference in inverse problems. *arXiv preprint arXiv:2101.03709*.

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.

Siahkoohi, A., Rizzuti, G., Witte, P. A., & Herrmann, F. J. (2020). Faster uncertainty quantification for inverse problems with conditional normalizing flows. *arXiv preprint arXiv:2007.07985*.

Smith, J. D., Ross, Z. E., Azizzadenesheli, K., & Muir, J. B. (2022). Hyposvi: Hypocentre inversion with stein variational inference and physics informed neural networks. *Geophysical Journal International*, *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 CO2 saturations in a subsurface reservoir. *Geophysical Journal International*, *236*(3), 1309–1331.

Sun, L., Wang, L., Xu, G., & Wu, Q. (2023). A new method of variational bayesian slip distribution inversion. *Journal of Geodesy*, *97*(1), 10.

Tape, C., Liu, Q., Maggi, A., & Tromp, J. (2010). Seismic tomography of the southern california crust based on spectral-element and adjoint methods. *Geophysical Journal International*, *180*(1), 433–462.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, *49*(8), 1259–1266.

Thurin, J., Brossier, R., & Métivier, L. (2019). Ensemble-based uncertainty estimation in full waveform inversion. *Geophysical Journal International*, *219*(3), 1613–1635.

Treeby, B. E., & Cox, B. T. (2010). k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields. *Journal of biomedical optics*, *15*(2), 021314–021314.

Valentine, A. P., & Sambridge, M. (2020a). Gaussian process models — I. a framework for probabilistic continuous inverse theory. *Geophysical Journal International*, *220*(3), 1632–1647.

Valentine, A. P., & Sambridge, M. (2020b). Gaussian process models — II. lessons for discrete inversion. *Geophysical Journal International*, *220*(3), 1648–1656.

Virieux, J., & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, *74*(6), WCC1–WCC26.

Visser, G., Guo, P., & Saygin, E. (2019). Bayesian transdimensional seismic full-waveform inversion with a dipping layer parameterization. *Geophysics*, *84*(6), R845–R858.

Wang, W., McMechan, G. A., & Ma, J. (2023). Re-weighted variational full waveform inversions. *Geophysics*, *88*(4), 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.

Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., … others (2013). Anisotropic 3d full-waveform inversion. *Geophysics*, *78*(2), R59–R80.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, *1*(1), 67–82.

Yin, Z., Orozco, R., Louboutin, M., & Herrmann, F. J. (2024). Wise: full-waveform variational inference via subsurface extensions. *Geophysics*, *89*(4), 1–31.

Zhang, C., Bütepage, J., Kjellström, H., & Mandt, S. (2018). Advances in variational inference. *IEEE transactions on pattern analysis and machine intelligence*, *41*(8), 2008–2026.

Zhang, X., & Curtis, A. (2020a). Seismic tomography using variational inference methods. *Journal of Geophysical Research: Solid Earth*, *125*(4), e2019JB018589.

Zhang, X., & Curtis, A. (2020b). Variational full-waveform inversion. *Geophysical Journal International*, *222*(1), 406-411.

Zhang, X., & Curtis, A. (2021a). Bayesian full-waveform inversion with realistic priors. *Geophysics*, *86*(5), 1–20.

Zhang, X., & Curtis, A. (2021b). Bayesian geophysical inversion using invertible neural networks. *Journal of Geophysical Research: Solid Earth*, e2021JB022320.

Zhang, X., & Curtis, A. (2022). Interrogating probabilistic inversion results for subsurface structural information. *Geophysical Journal International*, *229*(2), 750–757.

Zhang, X., & Curtis, A. (2023). VIP – Variational inversion package with example implementations of Bayesian tomographic imaging. *arXiv preprint arXiv:2310.13325*.

Zhang, X., Lomas, A., Zhou, M., Zheng, Y., & Curtis, A. (2023). 3-d bayesian variational full waveform inversion. *Geophysical Journal International*, *234*(1), 546–561.

Zhang, X., Nawaz, A., Zhao, X., & Curtis, A. (2021). An introduction to variational inference in geophysical inverse problems. *Advances in Geophysics*, *62*, 73-140.

Zhao, X., & Curtis, A. (2024). Bayesian inversion, uncertainty analysis and interrogation using boosting variational inference. *Journal of Geophysical Research: Solid Earth*, *129*(1), e2023JB027789.

Zhao, X., Curtis, A., & Zhang, X. (2021). Bayesian seismic tomography using normalizing flows. *Geophysical Journal International*, *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. *Journal of Geophysical Research: Solid Earth*, *127*(4), e2022JB024098.

Zhao, X., Zhou, H., Chen, H., & Wang, Y. (2020). Domain decomposition for large-scale viscoacoustic wave simulation using localized pseudo-spectral method. *IEEE Transactions on Geoscience and Remote Sensing*, 1-14.

Zhao, Z., & Sen, M. K. (2021). A gradient-based markov chain monte carlo method for full-waveform inversion and uncertainty analysis. *Geophysics*, *86*(1), R15–R30.

Zhdanov, M. S. (2002). *Geophysical inverse theory and regularization problems* (Vol. 36). Elsevier.

Zhu, H., Li, S., Fomel, S., Stadler, G., & Ghattas, O. (2016). A bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration. *Geophysics*, *81*(5), R307–R323.

Zidan, A., Li, Y., & Cheng, A. (2022). Regularized seismic amplitude inversion via variational inference. *Geophysical Prospecting*, *70*(9), 1507–1527.

Zunino, A., Gebraad, L., Ghirotto, A., & Fichtner, A. (2023, 10). HMCLab: a framework for solving diverse geophysical inverse problems using the hamiltonian monte carlo method. *Geophysical Journal International*, *235*(3), 2979-2991.

Physically Structured Variational Inference for Bayesian Full Waveform Inversion

Xuebin Zhao[1] and Andrew Curtis[1]

[1]School of Geosciences, University of Edinburgh, Edinburgh, United Kingdom

Key Points:

- We propose a new variational inference methodology to construct a Bayesian posterior solution with a desired correlation structure.
- The method is far more efficient in terms of both memory requirements and computation, with some loss of generality in the solution.
- We apply the inversion results to two post-inversion problems where the volume of stored CO_2 in a subsurface reservoir is estimated.

Corresponding author: Xuebin Zhao, xuebin.zhao@ed.ac.uk

Abstract

Full waveform inversion (FWI) creates high resolution models of the Earth's subsurface structures from seismic waveform data. Due to the non-linearity and non-uniqueness of FWI problems, finding globally best-fitting model solutions is not necessarily desirable since they fit noise as well as signal in the data. Bayesian FWI calculates a so-called posterior probability distribution function, which describes all possible model solutions and their uncertainties. In this paper, we solve Bayesian FWI using variational inference, and propose a new methodology called physically structured variational inference, in which a physics-based structure is imposed on the variational distribution. In a simple example motivated by prior information from past FWI solutions, we include parameter correlations between pairs of spatial locations within a dominant wavelength of each other, and set other correlations to zero. This makes the method far more efficient in terms of both memory requirements and computation, at the cost of some loss of generality in the solution found. We demonstrate the proposed method with a 2D acoustic FWI scenario, and compare the results with those obtained using other methods. This verifies that the method can produce accurate statistical information about the posterior distribution with hugely improved efficiency (in our FWI example, 1 order of magnitude in computation). We further demonstrate that despite the possible reduction in generality of the solution, the posterior uncertainties can be used to solve post-inversion interrogation problems connected to estimating volumes of subsurface reservoirs and of stored CO_2, with minimal bias, creating a highly efficient FWI-based decision-making workflow.

Plain Language Summary

This paper introduces a method to assess uncertainties in seismic images of the subsurface at substantially reduced cost, and to use the information within those uncertainties to answer explicit high-level questions about volumes of subsurface reservoirs and of stored CO2. Computational efficiency is achieved by explicitly imposing known (always observed) trade-offs between parameters that describe local properties of the subsurface. This prevents computing power from being used to re-discover such trade-offs each time an imaging process is performed. In our two-dimensional example in which we image using seismic Full Waveform Inversion, computational cost is reduced by an order of magnitude and fully nonlinear uncertainties can be characterized both in subsurface structural parameters, and in answers to high-level questions.

1 Introduction

Seismic full waveform inversion (FWI) is a method that generates models of the subsurface seismic velocity structure of the Earth given recorded seismograms. This is achieved using both kinematic (phase) and dynamic (amplitude) information in the waveforms (Tarantola, 1984). FWI has been applied in various fields, for example including regional and global seismology (Fichtner et al., 2009; Tape et al., 2010; French & Romanowicz, 2014; Bozdağ et al., 2016; Fichtner et al., 2018), seismic exploration (Pratt et al., 1998; Virieux & Operto, 2009; Prieux et al., 2013; Warner et al., 2013), medical imaging (Bernard et al., 2017; Guasch et al., 2020; Lucka et al., 2021), and non-destructive detection (He et al., 2021; Patsia et al., 2023).

Traditionally, FWI problems are solved using gradient-based local optimisation methods, where a misfit function between observed and predicted waveform data is minimised iteratively (Plessix, 2006). This process often requires additional regularisation terms, such as smoothing and damping terms, to stabilise the optimisation and improve convergence rates (Zhdanov, 2002; Sen & Roy, 2003; Asnaashari et al., 2013). However, these terms may introduce biases to the final inversion results. In addition, it is challenging to find a good approximation to the true Earth structure that generated the observed waveforms due to the strong non-linearity of the forward function and the non-uniqueness of the inverse problem solution (Boyd & Vandenberghe, 2004).

Recently, FWI has been solved probabilistically using a suite of methods collectively referred to as *Bayesian inference*. In Bayesian FWI, *prior* knowledge about Earth model parameters is updated with new information from the observed waveform data to calculate a *posterior* probability distribution function (pdf), according to Bayes' rule. In principle this distribution incorporates all prior information combined with all information from the data, and expresses the information in terms of constraints on the model parameters. It thus solves the FWI problem by describing all possible model parameter values that fit the dataset to within its uncertainty. The range and probability of different possible models can be used to reduce risk during subsequent decision-making when solving real-world interrogation problems (Poliannikov & Malcolm, 2016; Arnold & Curtis, 2018; Ely et al., 2018; X. Zhao et al., 2022; X. Zhang & Curtis, 2022; Siahkoohi et al., 2022).

Different kinds of Bayesian inference methods have been employed to perform probabilistic FWI. A direct generalisation from deterministic FWI involves approximating the posterior pdf with a Gaussian distribution, centred around an estimated maximum a posteriori (MAP) model obtained using local optimisation methods (Gouveia & Scales, 1998; Bui-Thanh et al., 2013; Zhu et al., 2016; Fang et al., 2018), or through local, low rank pdf approximations using a data assimilation technique (Thurin et al., 2019). If both the likelihood function and prior distribution are assumed to be Gaussians, then this MAP velocity model is equivalent to that obtained using l_2 regularised deterministic FWI (W. Wang et al., 2023). While this kind of methods can produce probabilistic results, the resulting posterior distribution may be affected by the starting point of the inversion, and may not fully capture uncertainty arising from non-linearity of the forward function (Z. Zhao & Sen, 2021).

Fully non-linear Bayesian FWI can be solved using sampling techniques such as Markov chain Monte Carlo (McMC), where random samples are drawn from the posterior distribution. The inversion results are represented by statistics of the sampled models, such as the mean and standard deviation. However, due to the typical high dimensionality (number of parameters to be estimated) of FWI problems, direct sampling methods, including the commonly used Metropolis-Hastings (MH)-McMC (Metropolis et al., 1953; Hastings, 1970; Mosegaard & Tarantola, 1995; Sambridge & Mosegaard, 2002), become impractical. Nevertheless, it is worth noting the existence of studies that employ a target-oriented strategy to reduce the dimensionality of parts of the Earth model of interest, and employ a localised wavefield injection method to calculate wavefields corresponding to each model variation. This reduces the computational complexity of FWI, and allows Metropolis-Hastings McMC to be applied effectively (Ely et al., 2018; Kotsi et al., 2020b; Fu & Innanen, 2022).

Several advanced techniques have been introduced to improve the sampling efficiency of McMC for Bayesian FWI. In reversible-jump McMC (RJ-McMC) (Green, 1995, 2003; Sambridge et al., 2006), a trans-dimensional approach is used to change the parametrisation, including the dimensionality of the model parameter vector. This can significantly improve efficiency by reducing dimensionality to only parameters that are necessary to explain the data and the forward function, and RJ-McMC has been successfully applied to Bayesian FWI (Ray et al., 2016, 2018; Visser et al., 2019; P. Guo et al., 2020). Hamiltonian Monte Carlo (HMC) has also been introduced to improve the sampling efficiency of FWI. In HMC, the sampling process is guided by the gradient of the posterior pdf with respect to the model parameters, and it has been demonstrated that HMC can improve the convergence rate over non-gradient based McMC (Gebraad et al., 2020; Kotsi et al., 2020a; de Lima, Corso, et al., 2023; de Lima, Ferreira, et al., 2023; Zunino et al., 2023; Dhabaria & Singh, 2024). Biswas and Sen (2022) introduced a reversible-jump Hamiltonian Monte Carlo (RJHMC) algorithm for 2D FWI, Z. Zhao and Sen (2021) and Berti et al. (2023) used gradient-based McMC methods to sample the posterior distribution efficiently, and Khoshkholgh et al. (2022) solved FWI using informed-proposal Monte Carlo (Khoshkholgh et al., 2021). Nevertheless, as with other classes of methods, Monte Carlo sampling is known to become computationally intractable for high-dimensional parameter spaces due to the curse of dimensionality (Curtis & Lomax, 2001).

In this study, we focus instead on variational inference, a method that solves Bayesian inversion through optimisation. In variational methods, we define a family of known and tractable distributions, referred to as the variational family. From this family, an optimal member is chosen to approximate the true posterior pdf by minimising the difference between the variational and posterior distributions (Bishop, 2006; Blei et al., 2017; C. Zhang et al., 2018; X. Zhang et al., 2021). Variational inference solves Bayesian problems under an optimisation framework, and the optimisation result is fully probabilistic. In some classes of problems it can therefore be relatively more efficient and scalable to high dimensional problems with large datasets. Variational inference has been applied to different geophysical inverse problems, including travel time tomography (X. Zhang & Curtis, 2020a; X. Zhao et al., 2021; Levy et al., 2022), seismic migration (Siahkoohi et al., 2020; Siahkoohi & Herrmann, 2021; Siahkoohi et al., 2021, 2023), seismic amplitude inversion (Zidan et al., 2022), earthquake hypocentre inversion (Smith et al., 2022), and slip distribution inversion (Sun et al., 2023). However, most of these applications have relatively lower dimensionality and weaker non-linearities compared to FWI.

X. Zhang and Curtis (2020b) introduced a variational method called Stein variational gradient descent (SVGD – Liu & Wang, 2016) to transmission FWI where sources emulating earthquakes are located underneath the velocity structure to be imaged, with receivers on the top surface. SVGD was then applied to 2D reflection FWI with realistic priors (X. Zhang & Curtis, 2021a; Izzatullah et al., 2023), and 3D acoustic FWI using synthetic data (X. Zhang et al., 2023) and field data (Lomas et al., 2023). A stochastic version of SVGD (Gallego & Insua, 2018) was also employed to improve performance for 3D FWI (X. Zhang et al., 2023). X. Zhao and Curtis (2024) introduced boosting variational inference (BVI – F. Guo et al., 2016; Miller et al., 2017) for 2D acoustic FWI, where a mixture of Gaussian distributions is used to approximate the true posterior distribution, resulting in an analytic expression for the posterior distribution. Bates et al. (2022) performed medical ultrasound tomography of the brain using FWI, where a mean field (diagonal) Gaussian distribution is employed as the variational distribution. Alternatively, W. Wang et al. (2023) improved the resolution of inversion results by decomposing the variational objective function into two terms and re-weighting them, however the method tends to underestimate posterior uncertainties. Yin et al. (2024) used conditional normalizing flows to quantify uncertainties in migration-velocity models.

Other than in W. Wang et al. (2023), in the above studies variational methods were applied to improve the efficiency of Bayesian FWI. For 2D FWI, the required number of forward simulations used to estimate means and variances of subsurface parameters was reduced to the order of 100,000 by X. Zhao and Curtis (2024), marking a significant reduction given that the dimensionality of the FWI problem tackled was higher than 10,000. Unfortunately, despite this improvement, the computational cost of solving the forward function in FWI remains prohibitively expensive for many practitioners. Consequently, performing Bayesian FWI in realistic projects using current variational methods is still impractical, even with advanced forward simulation strategies (Treeby & Cox, 2010; Y. Wang et al., 2019; X. Zhao et al., 2020).

In this paper, we propose an efficient and accurate variational methodology for Bayesian FWI by imposing physics-based structure on the variational family. The new method incorporates expected posterior parameter correlations explicitly. We show that this leads to significantly improved accuracy with nearly the same computational cost compared to several existing variational methods, or put another way, reduced cost for the same accuracy.

This rest of this paper is organised as follows. In section 2, we first establish the framework of variational full waveform inversion. Then we introduce the concept of ADVI, and present our new method which we refer to as *physically structured variational in-*

ference (PSVI). In section 3, we demonstrate the proposed method with a 2D synthetic FWI example and compare the inversion results with those obtained using three other variational methods. In section 4, we interpret the inversion results by solving two post-inversion interrogation problems. Finally, we provide a brief discussion of the proposed method and draw conclusions.

2 Methodology

2.1 Variational Full Waveform Inversion (FWI)

FWI uses full waveform data recorded by seismometers to constrain the Earth's interior structure, typically described by a subsurface velocity model. The forward function is defined to predict waveform data that could be recorded at receivers given a subsurface velocity model. This prediction involves solving a wave equation, either in the time or frequency domain, often in two or three dimensions, and potentially adding measurement noise to the data. For simplicity, we assume that the subsurface consists of an acoustic, isotropic, lossless medium with constant density, thereby ignoring exclusively elastic properties including shear waves, attenuation, and anisotropic properties. This simplification allows the scalar acoustic wave equation to be used in forward simulations which reduces computational load. The data-model gradients are calculated using the adjoint state method (Plessix, 2006).

In Bayesian FWI, information about the velocity model is characterized by a *posterior* probability distribution function (pdf) which describes the uncertainties associated with different potential models given the observed data. This can be calculated using Bayes' rule:

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

where $p(\cdot)$ denotes a probability distribution. Symbol $x|y$ indicates conditional dependence between two random variables x and y, and reads as x given y. Term $p(\mathbf{m})$ describes the *prior* information available on the model parameter $\mathbf{m}$, and $p(\mathbf{d}_{obs}|\mathbf{m})$ is the *likelihood*, meaning the probability of the synthetic waveform data $\mathbf{d}_{syn}$ generated by a given model $\mathbf{m}$ through forward simulation matching the observed data $\mathbf{d}_{obs}$. A Gaussian distribution is often used to define the data likelihood function:

$$p(\mathbf{d}_{obs}|\mathbf{m}) \propto \exp\left[-\frac{(\mathbf{d}_{syn}-\mathbf{d}_{obs})^T\Sigma_{\mathbf{d}}^{-1}(\mathbf{d}_{syn}-\mathbf{d}_{obs})}{2}\right] \tag{2}$$

where Σ_d is the covariance matrix of the data error. The denominator $p(\mathbf{d}_{obs})$ in equation 1 is referred to as the *evidence* and is a normalisation constant to ensure that the result of equation 1 is a valid probability distribution.

Bayesian inversion is often solved by Monte Carlo sampling methods. However, the required number of samples increases exponentially with the dimensionality of the inverse problem (the number of unknown model parameters), due to the curse of dimensionality (Curtis & Lomax, 2001). It is very expensive to obtain statistics of posterior pdf's in FWI using Monte Carlo methods, especially when the Earth model $\mathbf{m}$ contains more than 10,000 parameters, as is standard in such problems (Gebraad et al., 2020).

In this paper, we use variational inference to solve Bayesian FWI. In variational methods, a family of distributions (called the variational family) $\mathcal{Q}(\mathbf{m}) = \{q(\mathbf{m})\}$ is defined, from which we select an optimal member to approximate the true (unknown) posterior distribution. The optimal distribution can be found by minimising the difference (distance) between the posterior and variational distributions. Typically, the Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951) is used to measure the distance between two probability distributions, defined as the following expectation term

$$\mathrm{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{obs})] = \mathbb{E}_{q(\mathbf{m})}[\log q(\mathbf{m}) - \log p(\mathbf{m}|\mathbf{d}_{obs})] \tag{3}$$

The KL divergence of two distributions is non-negative, and equals zero only when the two distributions are identical. Substituting equation 1 into 3, we find that minimising the $\mathrm{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{obs})]$ is equivalent to maximising the following *evidence lower bound* of $\log p(\mathbf{d}_{obs})$ ($\mathrm{ELBO}[q(\mathbf{m})]$):

$$\mathrm{ELBO}[q(\mathbf{m})] = \mathbb{E}_{q(\mathbf{m})}[\log p(\mathbf{m}, \mathbf{d}_{obs}) - \log q(\mathbf{m})] \tag{4}$$

In this way, we convert a random sampling problem into a numerical optimisation, while the optimisation result is still a probability distribution that approximates the true posterior pdf.

A key challenge in variational inference is to choose the variational family $\mathcal{Q}(\mathbf{m})$. This determines both the accuracy and efficiency of the variational methods: increasing the complexity (and hence, expressivity) of $\mathcal{Q}(\mathbf{m})$ increases the approximation accuracy as well as the optimisation complexity. Given the expensive nature of forward simulations in FWI, our primary goal is to reduce computational costs (by reducing the number of forward simulations) while maintaining accuracy at an acceptable level. In the following sections we introduce a method called automatic differentiation variational inference (ADVI – Kucukelbir et al., 2017), and propose an alternative effective variational methodology for FWI.

2.2 Automatic Differentiation Variational Inference (ADVI)

ADVI is a well-established variational method that defines a Gaussian variational distribution $q = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, parametrised by a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$ (Kucukelbir et al., 2017). In addition, since a Gaussian distribution is defined over the space of real numbers and since in most geophysical imaging problems model parameters are bounded by physical constraints (e.g., seismic velocity should be a positive number), an invertible transform (a bijection) is applied to convert the Gaussian variational distribution into a bounded space that defines model parameter $\mathbf{m}$. The transformed distribution is then used to approximate the true posterior distribution.

To determine the optimal Gaussian distribution in the unbounded space, we maximise the ELBO$[q(\mathbf{m})]$ in equation 4 with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. This can be solved using a gradient based optimisation method. According to Kucukelbir et al. (2017), the gradient of the ELBO with respect to the covariance matrix $\boldsymbol{\Sigma}$ involves computing $|\boldsymbol{\Sigma}|$, where $|\cdot|$ denotes the determinant of a matrix. Direct calculation of $|\boldsymbol{\Sigma}|$ has a computational complexity of $O(n^3)$, which becomes prohibitively expensive for high dimensional inference problems such as FWI. Therefore, we often use a Cholesky factorisation to parametrise $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^{\mathbf{T}} \tag{5}$$

where $\mathbf{L}$ is a lower triangular matrix. Since $|\mathbf{L}|$ can be calculated easily as the product of its diagonal elements, the determinant $|\boldsymbol{\Sigma}|$ can be obtained by $|\boldsymbol{\Sigma}| = |\mathbf{L}|^2$. Note that the diagonal elements of $\mathbf{L}$ are associated with the variances of model parameters, and should be non-negative to ensure that $\mathbf{L}$ and $\boldsymbol{\Sigma}$ are positive semidefinite. The off-diagonal values of $\mathbf{L}$ contain correlation information between model parameters.

For a n-dimensional problem, we need $n(n+1)/2$ parameters to construct a full matrix $\mathbf{L}$, and consequently a full covariance matrix $\boldsymbol{\Sigma}$. The corresponding method is known as full rank ADVI (Kucukelbir et al., 2017). For example, in Figure 1a, the velocity model comprising 110 $\times$ 250 pixels requires 378,138,750 parameters to describe the full matrix $\mathbf{L}$. This number becomes computationally intractable for large scale 2D and 3D FWI problems.

Alternatively, a mean field approximation is often used to reduce computational complexity, where $\mathbf{L}$ and $\boldsymbol{\Sigma}$ are parametrised by diagonal matrices. The variational distribution becomes a diagonal Gaussian distribution, which neglects correlation information between different model parameters. In this way, the total number of variables that must be optimised is $2n$ (both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ contain n independent elements), so is doubled compared to a conventional deterministic inversion. Therefore, the computational overhead is manageable for most problems. Mean field ADVI has been applied to Bayesian FWI in several studies (Bates et al., 2022; W. Wang et al., 2023; X. Zhang et al., 2023), demonstrating that the method is computationally efficient and is able to provide an accurate mean model of the posterior distribution. However, in problems with significant posterior correlations, it tends to strongly underestimate posterior uncertainties since correlation information is neglected *a priori* (X. Zhang et al., 2023).

2.3 Physically Structured Variational Inference (PSVI)

Full rank ADVI and mean field ADVI represent two extreme approaches to construct $\mathbf{L}$: the former aims to optimise all off-diagonal elements of $\mathbf{L}$ to capture the full correlation information of $\mathbf{m}$, whereas the latter sets the off-diagonal elements to zero to reduce computational requirements. In the following, we parametrise $\mathbf{L}$ using a physics-guided structure, which models a subset of its off-diagonal elements.

In most imaging problems, accurate correlation information plays an important role in capturing true structures such as the continuity of properties across neighbouring spatial. Since modelling a full covariance matrix (i.e., full rank ADVI) for high dimensional problems is practically intractable, another approach is to model the most important correlation in vector $\mathbf{m}$, guided by physical properties (prior knowledge) of imaging problems. To illustrate, Figure 1d shows a 2D velocity structure discretized using $nx \times nz$ square grid cells in horizontal and vertical directions, with each cell representing a velocity value at the corresponding spatial location. It is often the case that any grid cell, such as the one marked by a black dot in Figure 1d, is strongly correlated with its surrounding cells (e.g., cells marked by white pluses). The magnitude of correlations between this central cell and other cells decreases as the distance between two locations

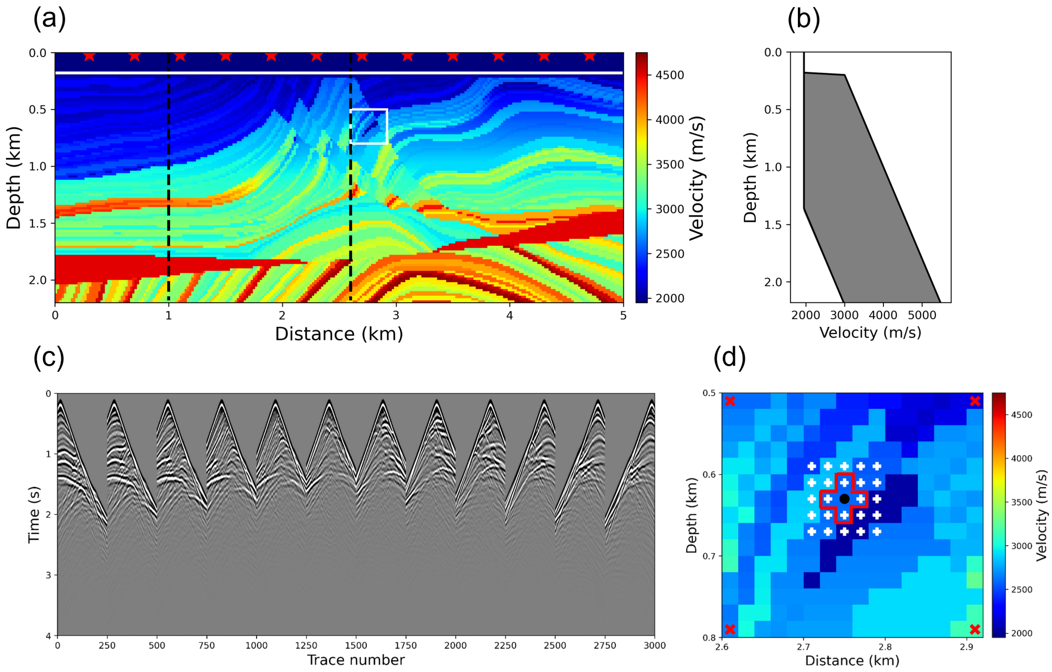

Figure 1. (a) P wave velocity of the Marmousi model used in a 2D acoustic FWI test. Source locations are indicated by red stars and the receiver line is marked by a white line. Dashed black lines display the locations of two vertical profiles used to compare the posterior marginal probability distributions in Figure 4. (b) Upper and lower bounds of the Uniform prior distribution at different depths. (c) Observed dataset which contains twelve common shot gathers. (d) Velocity structure inside the white box in (a), and crosses in cells discussed in the main text.

increases. Cells that are far away from the black dot (e.g., cells denoted by red crosses in Figure 1d) are only weakly correlated with the black-dotted cell, so these correlations can safely be ignored. This feature has been observed in many different imaging problems (Ardizzone et al., 2018; Gebraad et al., 2020; Biswas & Sen, 2022); a clear example displaying such correlations in a velocity profile with depth is shown in Figure 6 of X. Zhang and Curtis (2021b), from the results of surface wave dispersion inversion using two independent nonlinear inversion methods (invertible neural networks and Monte Carlo).

This suggests that it might suffice to model correlations only between parameter values that are spatially close to each other, i.e. which lie within a dominant wavelength, and ignore those that are far away by assuming a particular sparse structure for $\mathbf{L}$. We therefore set off-diagonal elements of $\mathbf{L}$ which represent the main correlations of interest as parameters to be optimised during variational inversion, while imposing all other off-diagonal elements to be zero. Note that we thus impose only a structure on $\mathbf{L}$ rather than placing constraints on the values of its (non-zero) off-diagonal elements: those values are updated freely during inversion.

Suppose that the 2D velocity model displayed in Figure 1d is defined by vector $\mathbf{m}$ in row-major order (i.e., the first nx elements of $\mathbf{m}$ comprise the first row of the 2D image, the second nx elements comprise the second row, and so on). As illustrated in equation 6 below, the first-order off-diagonal elements (blue ones in equation 6 that are directly below the diagonal elements) contain correlation information between two horizontally adjacent grid cells, and off-diagonal elements that are nx rows below the main diagonal elements (red ones in equation 6) describe correlations between two vertically adjacent cells

$$\mathbf{L} = \begin{bmatrix} l_{0,1} & & & & & & \\ \color{blue}{l_{1,1}} & l_{0,2} & & & & & \\ 0 & \color{blue}{l_{1,2}} & l_{0,3} & & & & \\ \ldots & 0 & \color{blue}{l_{1,3}} & \ldots & & & \\ \color{red}{l_{nx,1}} & \ldots & 0 & \color{blue}{\ldots} & l_{0,n-2} & & \\ 0 & \color{red}{\ldots} & \ldots & \ldots & \color{blue}{l_{1,n-2}} & l_{0,n-1} & \\ \ldots & 0 & \color{red}{l_{nx,n-nx}} & \ldots & 0 & \color{blue}{l_{1,n-1}} & l_{0,n} \end{bmatrix} \tag{6}$$

Note that in equation 6, the first subscript i indicates a block of off-diagonal elements that are i rows below the main diagonal (i.e., at an offset of i from the main diagonal), and the second subscript j indicates that $l_{i,j}$ is the jth element of that off-diagonal block. This differs from the commonly used indexing scheme in which the two subscripts imply the row and column number of an element. If we set all remaining elements of $\mathbf{L}$ to zero, then covariance matrix $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$ also has non-zero entities only at two off-diagonal blocks located 1 and nx rows below and above the main diagonal elements (similar to the red and blue elements in equation 6). If such a covariance matrix $\mathbf{\Sigma}$ is used, the variational distribution would also capture a specific spatial correlation structure that only includes parameter correlations between pairs of adjacent cells in both horizontal and vertical directions. Thus, for the grid cell denoted by the black dot in Figure 1d, we would model correlations between this cell and its four adjacent cells inside the red box in Figure 1d: all other correlations are set to zero.

We can impose any desired correlation structure on $\mathbf{\Sigma}$, by setting the corresponding off-diagonal blocks in $\mathbf{L}$ as unknown hyperparameters and optimising them during inversion. The size of the defined correlation template should be relatively small compared to the dimensionality of the problem, so the total number of parameters required to construct $\mathbf{L}$ would also be relatively small compared to that in full rank ADVI. For example, if the white pluses in Figure 1d are used to define a 5×5 correlation kernel then the required number of parameters to construct $\mathbf{\Sigma}$ is smaller than $13n$. Here n is the dimensionality of model vector $\mathbf{m}$, and the number 13 consists of 1 main diagonal block and 12 off-diagonal blocks representing 12 different offsets between cells marked by the white crosses and the central cell in the 5×5 kernel. Since each off-diagonal block contains fewer parameters than the main diagonal block (i.e., the blue and red elements in equation 6 are fewer than the diagonal elements), the total number of parameters is smaller than $13n$, which is a significant reduction compared to $n(n+1)/2$ parameters used in full rank ADVI.

We implement the aforementioned approach to parametrise the matrix $\mathbf{L}$ and obtain a sparse approximation of the covariance matrix. The inversion results thus effectively and efficiently capture structured correlation information. Since this originated from the inherent physical properties of imaging problems, we name the method as *physically structured variational inference* (PSVI).

To update the variational parameters, we use gradient based optimisation methods. The gradient of the ELBO with respect to the variational parameters can be calculated easily using advanced automatic differentiation libraries such as TensorFlow (Abadi et al., 2016) and PyTorch (Paszke et al., 2019). The expectation term in the EBLO (equation 4) can be estimated by Monte Carlo integration with a small number of samples, which is reasonable because the optimisation is typically carried out over many iterations, allowing the gradients to converge statistically towards the correct solution (Kucukelbir et al., 2017). Given that the computational cost of updating the variational parameters is negligible in comparison to forward modelling in FWI, the proposed method is almost as efficient as mean field ADVI.

3 2D Acoustic FWI Example

In this section, we test the proposed PSVI algorithm in a 2D acoustic FWI example. The true velocity model, shown in Figure 1a, is obtained by truncating the original Marmousi model (Martin et al., 2006) and downsampling it into 110 $\times$ 250 regular grid cells. The grid cell size is 20m in both directions. For simplicity, we maintain a constant density. We simulate 12 sources on the surface with a spacing of 400m (indicated by red stars in Figure 1a). A receiver line containing 250 receivers at an interval of 20m is placed on the seabed at 200m depth (white line in Figure 1a). The observed waveform data are generated by solving the 2D acoustic wave equation using a time-domain finite difference method. The simulation length is 4s with a sample interval of 2ms. The source function is a Ricker wavelet with a dominant frequency of 10 Hz. Figure 1c displays this observed waveform dataset.

We define a Uniform prior distribution for the velocity values in each grid cell. Figure 1b shows the lower and upper bounds of the prior distribution at different depths. We set the velocity in the water layer (down to 200m depth) to its true value during inversion. The likelihood function is a Gaussian distribution (equation 2) with a diagonal covariance matrix Σ_d assuming independence among all data points. We take the maximum amplitude value of each trace and average them. The data noise is assumed to be 1% of the obtained average value. The same finite difference solver is used to calculate the synthetic waveform data $\mathbf{d}_{syn}$, and the gradient of the data misfit (negative log-likelihood function) with respect to the velocity model is computed using the adjoint-state method (Plessix, 2006). For variational inversion, we use Monte Carlo integration

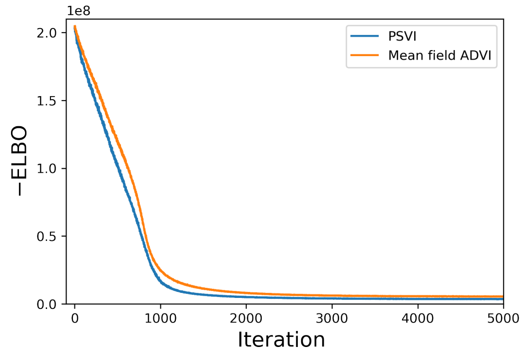

Figure 2. Variation of the negative ELBO with respect to iterations.

to estimate the ELBO in equation 4, and use the automatic differentiation framework provided by PyTorch to build a computational graph, which (automatically) calculates the ELBO and its gradient with respect to the variational parameters (Paszke et al., 2019). Optimization process is carried out using the Adam algorithm (Kingma & Ba, 2014).

We apply mean field ADVI and PSVI to this Bayesian FWI problem. Considering the dimensionality of this problem ($100\times250 = 25{,}000$), full rank ADVI is not performed since constructing a full covariance matrix would be extremely expensive in terms of both memory requirements and computational cost. For mean field ADVI, we use a diagonal Gaussian distribution to approximate the posterior distribution in the unbounded space. For PSVI, a 5×5 correlation kernel is employed to model the main correlations between model parameters, as illustrated by the white pluses in Figure 1d for the central black dotted cell. The choice of this correlation kernel is based on the estimated dominant wavelength of this problem (approximately 200m in shallow subsurface). In both tests, variational parameters ($\boldsymbol{\mu}$ and $\mathbf{L}$) are updated for 5000 iterations, with 2 random samples per iteration used to approximate the ELBO$[q(\mathbf{m})]$ and its gradients with respect to $\boldsymbol{\mu}$ and $\mathbf{L}$. Figure 2 displays the negative ELBOs for these two tests as a function of iterations, indicating that both algorithms achieve a reasonable level of convergence with nearly the same convergence speed, even though PSVI has far more parameters to optimise.

Figures 3a and 3b depict the inversion results. The mean, standard deviation and the relative error (computed by dividing the absolute error between the true and mean models by the standard deviation model) of the posterior distribution are displayed from top to bottom row. The two mean velocity maps exhibit similar features across most locations, generally resembling the true velocity map in Figure 1a. The inversion results struggle to recover some thin layers in the deeper part of the model, potentially due to the relatively low frequency (10 Hz) data used for FWI. Additionally, certain discrepancies are observed between these two maps at specific locations. For example, in the tilt layers annotated by red and black arrows in Figures 3a and 3b, the mean velocity model from mean field ADVI displays discontinuities, while the PSVI results show more continuity, closely resembling the true velocity model. One possible reason for this discrepancy is that accurate correlation information is crucial for recovering the continuity of spatial locations, especially for these thin layers. All correlations between pairs of model parameters are neglected in mean field ADVI, and thus the results may fail to recover the true velocity structures at these locations. By incorporating physically structured correlations between cells within a dominant wavelength, the proposed method improves the inversion accuracy.

Both inversion results show increased uncertainties with greater depth, since the sensitivity of observed seismic data decreases at depth, thus deeper parts of the model are less constrained by the data. The standard deviation values obtained from mean field ADVI are generally smaller than those from PSVI, especially in the shallower subsurface above 1.5km depth. This is because mean field ADVI tends to underestimate posterior uncertainties by neglecting correlations. Similar phenomena have been observed in previous studies (Ely et al., 2018; W. Wang et al., 2023; X. Zhao & Curtis, 2024). Therefore, the relative errors from mean field ADVI are larger compared to those from the proposed method, especially at locations with a depth of 1km and a distance between $0-1.5$km, where the mean model deviates from the true model by more than 3 standard deviations. This discrepancy suggests a low credibility of the inversion results obtained from mean field ADVI. As marked by a white arrow in Figure 3a, lower uncertainty noise is observed, which correspond to layers that are not continuous in the mean velocity map marked by a red arrow. This feature again proves that mean field ADVI provides biased uncertain estimates. By contrast, such uncertainty structures are not observed in Figure 3b, indicating that PSVI has the capability to correct some biases introduced by mean field ADVI.

To validate the inversion results displayed in Figure 3b, we apply two additional variational methods to this problem: *boosting variational inference* (BVI – F. Guo et al.,

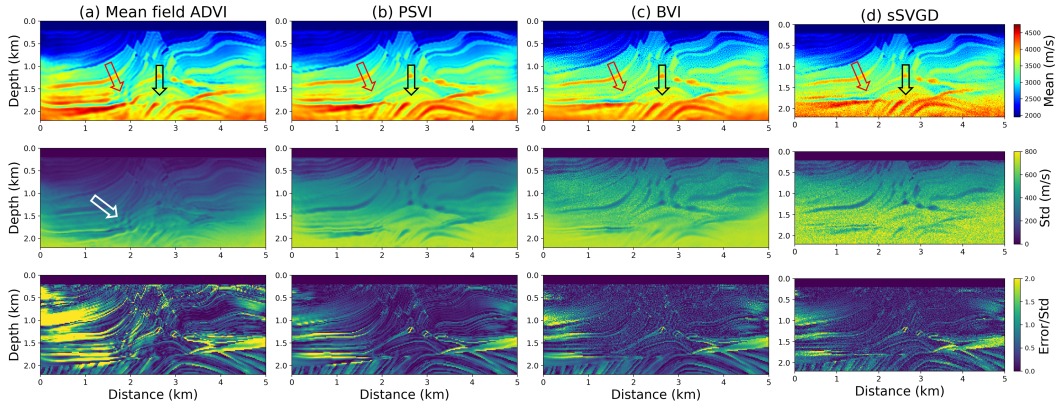

Figure 3. Mean (top row), standard deviation (middle row) and relative error (bottom row) of the posterior distribution obtained using (a) mean field ADVI, (b) PSVI, (c) boosting variational inference (BVI) and (d) stochastic SVGD (sSVGD), respectively. The relative error is the absolute error between the mean and true models divided by the corresponding standard deviation.

2016; Miller et al., 2017) and *stochastic Stein variational gradient descent* (sSVGD – Gallego & Insua, 2018). In BVI, a mixture distribution, in this case a mixture of Gaussians, is used to approximate the posterior distribution considering the fact that a mixture distribution can approximate any target distribution to any level of accuracy. sSVGD is a Monte Carlo based variational method that iteratively pushes a set of random samples towards the posterior distribution by minimising the KL divergence. In addition, a noise term is introduced to these samples at each iteration such that the algorithm converges to the true posterior distribution asymptotically. These two methods have been applied to acoustic FWI problems, and have proved to provide reasonable posterior solutions in two and three dimensional Earth models (X. Zhang et al., 2023; X. Zhao & Curtis, 2024). Figures 3c and 3d depict the inversion results obtained using BVI and sSVGD, respectively. They present very similar features compared to those displayed in Figure 3b: the same continuous structures in the deeper part of the model (denoted by red and black arrows) are observed in the mean velocity maps, and similar higher standard deviation values associated with lower relative errors (distributed within 2 standard deviations) are also present.

To further analyse the accuracy of the inversion results, in Figure 4 we compare the posterior marginal distributions obtained from the four tested methods along two

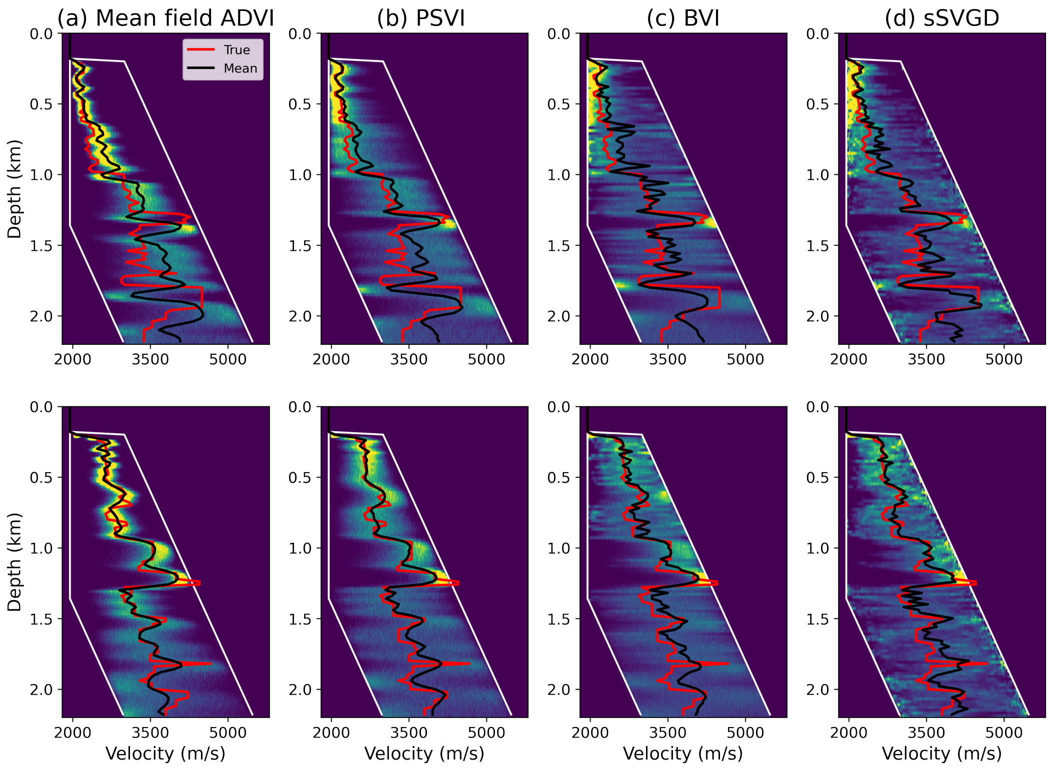

Figure 4. Posterior marginal distributions coloured from dark blue (zero probability) to yellow (maximum value of marginal pdf's in each plot), along two vertical profiles at distances of 1km (top row) and 2.6km (bottom row) obtained using (a) mean field ADVI, (b) PSVI, (c) BVI and (d) sSVGD. The locations of these two profiles are represented by black dashed lines in Figure 1a. In each figure, two white lines show the prior bounds, and black and red lines show the mean and true velocity values.

vertical profiles at horizontal locations of 1km (top row) and 2.6km (bottom row), respectively. The location of these two profiles are displayed by dashed black lines in Figure 1a. The first profile (at a distance of 1km) is strategically placed in regions where the relative errors from mean field ADVI (Figure 3a) are higher, while the second one (at 2.6km) is centrally located within the imaging region. Red lines show the true velocity values and black lines show the mean velocity values obtained using different methods. Overall, the marginal distributions in Figure 4a are narrower compared to those in Figures 4b to 4d, indicating lower posterior uncertainties akin to Figure 3. In the first row of Figure 4 between depths of 0.7km – 1 km and 1.3km – 1.8km, the true velocity values are excluded from the posterior distribution obtained using mean field ADVI, whereas those values correctly reside within the high probability region of the posterior pdfs obtained using the other three methods. These phenomena again prove that mean field ADVI tends to underestimate the posterior uncertainties and introduce biases into the inversion results. By including the main correlation information between adjacent grid cells, PSVI yields better inversion results that are highly consistent with two entirely independent methods. Therefore, we assert that the posterior standard deviations derived from PSVI are likely to be correct.

Given that PSVI is designed to capture correlations between spatially close grid cells, we compare the posterior correlation coefficients between model parameters estimated using different methods. Figure 5 shows the covariance matrices for velocity values within the white box in Figure 1a, obtained using the above four inversion methods. Mean field ADVI uses a transformed diagonal Gaussian distribution to approximate the posterior pdf and disregards correlations between model parameters, thus the posterior covariance matrix predominantly exhibits strong diagonal values corresponding to the variances of model parameters. By incorporating a specific (desired) correlation structure into the variational distribution, the covariance matrix obtained using PSVI displays off-diagonal values representing correlations between different parameters, which are not observed from the results using mean field ADVI. Due to the use of a 5×5 correlation kernel (as represented by the white pluses in Figure 1d), we only include correlation information between a given grid cell and cells within two layers of cells surrounding it. As a result, Figure 5b displays four off-diagonal blocks (two above and two below the diagonal elements). We observe negative correlations between neighbouring cells (in the first off-diagonal block below and above the diagonal values) and positive correlations between every second neighbouring cells (found in the second off-diagonal block).

In Figures 5c and 5d, similar negative off-diagonal correlation blocks are observed in the covariance matrices obtained using BVI and sSVGD. This confirms that in this test we successfully capture the correct correlation information between adjacent cells by using PSVI. While there may be positive correlations with cells two layers apart, these are not visible; this may be because Figures 5c and 5d show a general ‘speckle’ of non-zero background correlation values that are absent in Figure 5b. In PSVI, we construct a sparse covariance matrix with specific non-zero off-diagonal elements, and set all other values to zero. This neglects correlations between locations that are spatially far away from each other. It should be noted that we do not know whether any of these values in Figures 5c and 5d are correct, since they do not match between the two panels. In the next section, we also prove that these non-zero background correlations play a less significant role in a simulation of a real-world decision-making process. So again we suggest that our implementation of PSVI has modelled the most prominent and consistent features of the correlation structure.

Finally, we analyse the efficiency of the proposed method and compare its cost with other methods. As mentioned in Section 2, the number of hyperparameters that need to be optimised in PSVI is higher than that in mean field ADVI but is significantly lower than that in full rank ADVI. In our test, we find that the computational cost for optimising these variational parameters is much cheaper (almost negligible) compared to the cost used for forward and adjoint simulations in FWI. Therefore, the number of simulations serves as a good metric for the overall cost in this example.

Table 1 summarises the number of simulations used in each tested method. The same simulation settings are used in mean field ADVI and PSVI (10,000 simulations consisting of 5000 iterations with 2 samples per iteration). For BVI, we use a mixture of 24 diagonal Gaussian distributions to approximate the posterior distribution. Each component is updated by 2500 iterations with 2 samples per iteration. Note that the number of simulations used to optimise each component for BVI is smaller than that for ADVI, as full convergence of each component is not necessarily required in BVI (X. Zhao & Curtis, 2024). For sSVGD, we run 5000 iterations with 24 samples, resulting in a total of 120,000 forward evaluations for both BVI and sSVGD. In these two tests, relatively larger

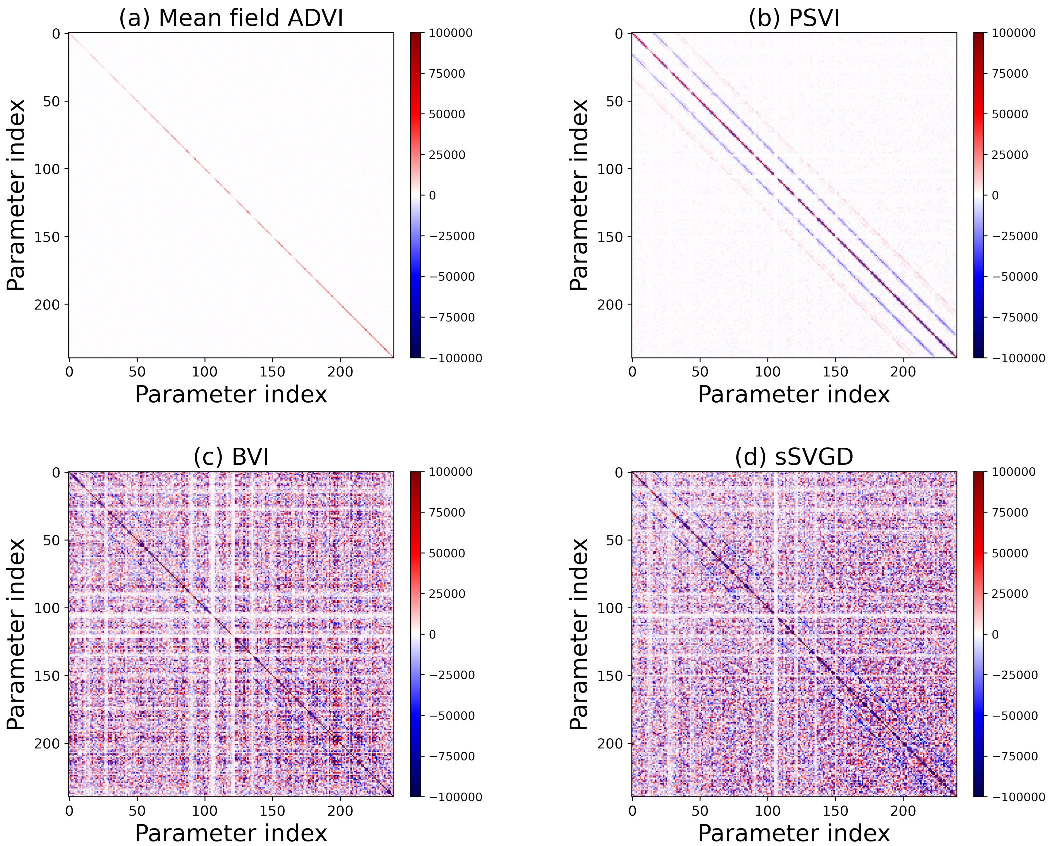

Figure 5. Covariance matrices for velocity values inside the white box in Figure 1a, calculated using the inversion results from (a) mean field ADVI, (b) PSVI, (c) BVI and (d) sSVGD.

Table 1. Number of forward and gradient evaluations for mean field ADVI, PSVI, BVI, and sSVGD. The values represent an indication of the computational cost of each method, as the evaluation of data-model gradients in FWI is by far the most expensive part of each calculation.

Method	Number of Gradient Evaluations
Mean field ADVI	10,000
PSVI	10,000
BVI	120,000
sSVGD	120,000

step sizes are used to speedup the convergence of BVI and sSVGD. However, they still remain one order of magnitude more computationally expensive than mean field ADVI and PSVI. In addition, Figure 2 shows that mean field ADVI and PSVI present roughly the same convergence rate given the same number of forward simulations. This verifies the statement that PSVI is almost as efficient as mean field ADVI. The latter is known to be a particularly inexpensive (yet biased) method for Bayesian inversion from previous studies (X. Zhang & Curtis, 2020a; X. Zhao et al., 2021; Bates et al., 2022; Sun et al., 2023). On the other hand, the PSVI method improves the inversion accuracy and provides similar results compared to two accurate but more computationally demanding methods (BVI and sSVGD). Thus, the proposed method shown to be an efficient algorithm that has provided reliable uncertainty estimates.

4 Interrogating FWI results

The objective of scientific investigations is typically to answer some specific and high-level questions. Examples of these questions in the field of geophysics can be: *How large is a subsurface structure? Is this a good location for carbon capture and storage (CCS)?* Normally these questions are answered in a biased manner without evaluating uncertainties in the results. Interrogation theory provides a systematic way to obtain the least-biased answer to these questions (Arnold & Curtis, 2018). In this section, we solve two interrogation problems using the FWI results obtained above, to evaluate the potential practical value of the correlations estimated by PSVI.

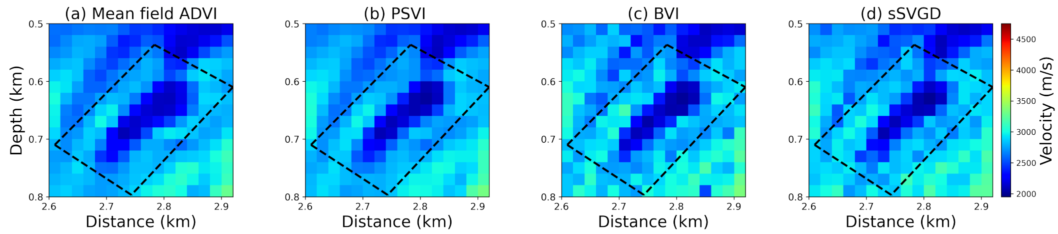

Figure 6. Mean velocity maps inside the white box in Figure 1a (corresponding to the true velocity map displayed in Figure 1d), obtained using (a) mean field ADVI, (b) PSVI, (c) BVI and (d) sSVGD. Black dashed boxes show the region where interrogation is performed.

Interrogation theory shows that the optimal answer a^* to a specific question Q that has a continuous space of possible answers is expressed by the following expectation term:

$$a^* = \mathbb{E}[T(\mathbf{m}|Q)] = \int_{\mathbf{m}} T(\mathbf{m}|Q)p(\mathbf{m}|\mathbf{d}_{obs})\ d\mathbf{m}, \tag{7}$$

where optimality is defined with respect to a squared utility (Arnold & Curtis, 2018). The expectation is taken with respect to the posterior distribution $p(\mathbf{m}|\mathbf{d}_{obs})$ of model parameter $\mathbf{m}$. Term $T(\mathbf{m}|Q)$ is a target function conditioned on the question Q of interest. It is defined to map the high dimensional model parameter $\mathbf{m}$ into a low dimensional target function value t in a target space $\mathbb{T}$, within which the question Q can be answered directly. In such cases the optimal answer in equation 7 is simply the expectation or mean of the posterior target function.

4.1 Interrogation for reservoir size

Figure 6 shows the inverted mean models of the velocity structure within the white box in Figure 1a, obtained through (a) mean field ADVI, (b) PSVI, (c) BVI, and (d) sSVGD. In each figure, we observe a low velocity body at the centre of the model section, outlined by a dashed black box. In this first example, we treat this low velocity zone as a reservoir and use interrogation theory to estimate its size.

Previously, volume-related questions were answered using seismic imaging results obtained from travel time tomographic inversion (X. Zhao et al., 2022) and FWI (X. Zhang & Curtis, 2022; X. Zhao & Curtis, 2024). Following these studies, we define a target function $T(\mathbf{m}|Q)$ as the area of the largest continuous low velocity body, which converts a high dimensional velocity model into a scalar value, representing the estimated reservoir

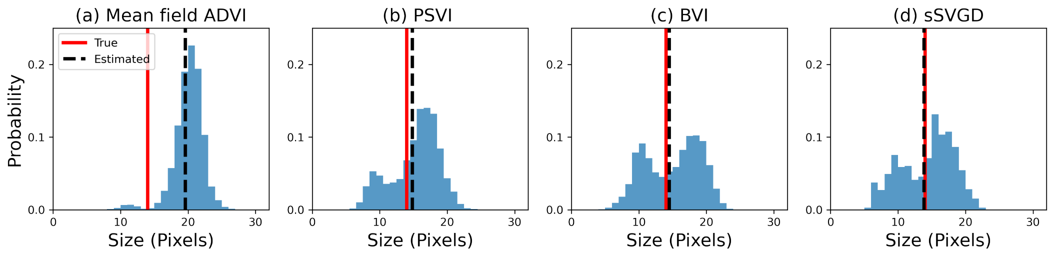

Figure 7. Posterior distributions of the low velocity reservoir size using FWI results obtained from (a) mean field ADVI, (b) PSVI, (c) BVI and (d) sSVGD, respectively. Red lines denote the true reservoir size, and black dashed lines denote the optimal size obtained using interrogation theory.

area from a given posterior sample. Note that this process involves using a velocity threshold to distinguish between low and high velocities. We use the same data-driven method introduced in X. Zhao et al. (2022) to determine the least biased estimate of this threshold value. This involves selecting some cells that are almost definitely inside the low velocity anomaly, others that are almost definitely outside; we then choose the threshold value such that the expected probability of interior cells being below that value equals the expected probability of exterior cells being above that value, according to the posterior pdf. We are then able to calculate the target function for every posterior sample.

Figure 7 displays the posterior distributions of the target function (reservoir size) using the four inversion results obtained previously. In this synthetic test, the true reservoir area is precisely known from Figure 1d and is denoted by red lines in Figure 7. The optimal (least-biased) answer estimated from each inversion method corresponds to the mean value of the respective posterior target function (as per equation 7), and is displayed by a dashed black line in each figure. As discussed in previous sections, mean field ADVI tends to underestimate posterior uncertainties and provides biased inversion results. We see that, the corresponding interrogation results in Figure 7a are also biased: the optimal answer shows a significant error and is far from the true answer, and indeed the true answer is even excluded from the posterior distribution of the estimated volume. By contrast, if we impose physically structured correlation information on model parameter, the optimal answer estimated by PSVI aligns closely with the true answer (Figure 7b). The posterior distribution of the target function also successfully captures bimodal uncertainties, similar to those obtained using BVI and sSVGD.

4.2 Interrogation for CO_2 storage

In the second example, we apply the inversion results to answer a more realistic and practically interesting question. Assume the low velocity reservoir identified above is used in a carbon capture and storage (CCS) project and is injected with CO_2. The injection of CO_2 into a porous rock produces changes in petrophysical parameters of the rock, such as pore fluid phase and water saturation. These changes further result in variations in seismic response of a reservoir, such as seismic velocity. Leveraging the FWI results, we can use these variations to monitor the injected CO_2 in a subsurface CCS project by answering the question: *what is the total volume of CO_2 stored in this reservoir?*

For the characterisation of changes in seismic velocity due to physical parameters related to CO_2, especially CO_2 saturation (S_{co_2}) in the reservoir, we first represent the P wave velocity v_p of a saturated rock using the bulk modulus K_{sat}, shear modulus G_{sat} and density ρ_{sat} of the rock by

$$v_p = \sqrt{\frac{K_{sat} + 4G_{sat}/3}{\rho_{sat}}} \tag{8}$$

The bulk modulus can be calculated using the Gassmann equation (Gassmann, 1951):

$$K_{sat} = K_d + \frac{(1 - \frac{K_d}{K_m})^2}{\frac{\phi}{K_f} + \frac{1-\phi}{K_m} - \frac{K_d}{K_m^2}} \tag{9}$$

where ϕ is the porosity, and K_d, K_m and K_f are the bulk moduli of dry rock, solid matrix and pore fluid. The density of a saturated rock can be calculated as

$$\rho_{sat} = (1 - \phi)\rho_m + \phi\rho_f \tag{10}$$

where ρ_m and ρ_f are the densities of grain matrix and fluid, respectively. The shear modulus G_{sat} is not affected by fluid and only depends on the shear modulus of dry rock G_d

$$G_{sat} = G_d \tag{11}$$

Assuming the reservoir is saturated by two distinct fluids, water and CO_2, the saturation values for water (S_w) and CO_2 (S_{co_2}) are constrained by the relation: $S_w + S_{co_2} = 1$. Then, the bulk modulus and density of fluid can be calculated using the mixing rules

$$\rho_f = S_w\rho_w + S_{co_2}\rho_{co_2} \tag{12}$$

$$K_f = S_w^e K_w + (1 - S_w^e)K_{co_2} \tag{13}$$

Table 2. Rock physics parameters and their associated standard deviations (uncertainties) estimated from the Sleipner field (Dupuy et al., 2017; Ghosh & Ojha, 2020).

Parameter	K_m (GPa)	K_d (GPa)	K_w (GPa)	K_{co_2} (GPa)	G_m (GPa)	G_d (GPa)	ρ_m (kg/m^3)	ρ_w (kg/m^3)	ρ_{co_2} (kg/m^3)	ϕ (%)
Mean value	39.3	2.56	2.31	0.08	44.8	8.1	2664	1030	700	0.3
Uncertainty	1.41	0.08	0.07	0.04	0.81	0.24	3	20	77	0.02

where ρ_w, ρ_{co_2}, K_w and K_{co_2} are the densities and bulk moduli of water and CO_2, and e is an empirical value (Brie et al., 1995). In this example, we use $e = 11$ as suggested by Kim et al. (2013). The injection of CO_2 into a reservoir alters the saturation values S_w and S_{co_2}, changing K_f and ρ_f, and thus also v_p through equations 8 to 13. Therefore, we can estimate S_{co_2} using P wave velocity values obtained from FWI.

To simplify the problem, we assume that some of the aforementioned rock physics parameters follow Gaussian distributions. Their means and standard deviations are estimated from the Sleipner field (Dupuy et al., 2017; Ghosh & Ojha, 2020; Strutz & Curtis, 2024), as listed in Table 2. Given these parameters, we build a direct relationship between P wave velocity v_p and CO_2 saturation S_{co_2}. The results are depicted by the joint probability distribution of v_p and S_{co_2} displayed in Figure 8a. The red curve is the reference $v_p - S_{co_2}$ curve obtained using the mean values from Table 2. In Figure 8a, the posterior distribution of CO_2 saturation for any P-wave velocity value can be obtained. For example, Figures 8b and 8c illustrate two such posterior pdfs corresponding to velocity values of 2045m/s (solid white line in Figure 8a) and 1840m/s (dashed white line). In Figure 8 we observe that seismic velocity is sensitive to small CO_2 saturations (below 0.2) and is insensitive for larger S_{co_2} values (Kim et al., 2013).

In the previous interrogation example, we defined the largest continuous low velocity body as the reservoir of interest for a posterior velocity sample. For each grid cell within the identified reservoir, we substitute its velocity value into Figure 8a to obtain the posterior pdf of CO_2 saturation. Finally, the total (2D) CO_2 volume V_{co_2} stored in the reservoir can be calculated by

$$V_{co_2} = \sum V \phi S_{co_2} \tag{14}$$

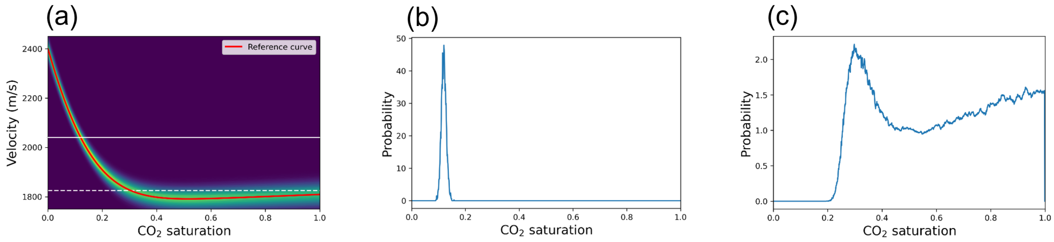

Figure 8. (a) Joint probability distribution of P wave velocity and CO_2 saturation given other parameters listed in Table 2. Red curve shows a one-to-one mapping between v_p and S_{co_2} obtained using the mean values in Table 2, and the colour scale from red through green to dark blue represents the probability distribution of velocity, given any value of CO_2 and the Gaussian distributions defined in Table 2. (b) and (c) display the posterior distributions of CO_2 saturation for velocity values of 2045m/s and 1840m/s, marked by solid and dashed white lines, respectively, in (a).

where V is the (2D) volume (i.e. area) of each grid cell in FWI, and the summation is taken over all grid cells within the reservoir. This defines the target function for this interrogation problem.

Figure 9 displays the posterior distributions of the estimated (2D) CO_2 volume obtained using different inversion methods. Similar to the reservoir size displayed in Figure 7, mean field ADVI provides rather biased interrogation results since it tends to underestimate posterior uncertainties. In contrast, the other three methods provide similar (and possibly correct) posterior distributions with two distinct modes. The three estimated answers are close to the true value, which lies inside the high probability region of the posterior distributions. Figures 7 and 9 prove that PSVI provides accurate uncertainty information that can be used to answer real-world questions correctly. Moreover, the non-zero background correlations ignored by PSVI (displayed in Figures 5c and 5d) are shown to be less important for post-inversion decision-making.

5 Discussion

PSVI can be considered as an intermediate approach between mean field ADVI and full rank ADVI (Kucukelbir et al., 2017). Mean field ADVI neglects all correlations to reduce computations and thus strongly underestimates posterior uncertainties. Full rank ADVI includes full correlation information between model parameters but is computa-

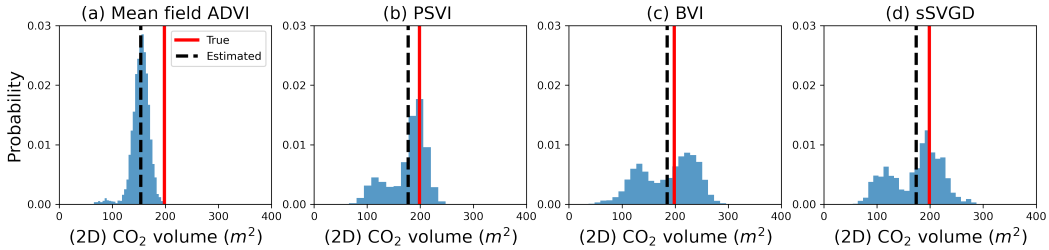

Figure 9. Posterior distributions of the (2D) CO_2 volume stored in the low velocity reservoir, calculated using (a) mean field ADVI, (b) PSVI, (c) BVI and (d) sSVGD. Red lines denote the true CO_2 volume, and black dashed lines denote the least-biased CO_2 volume estimated using interrogation theory.

tionally intractable for high dimensional problems such as 2D or 3D FWI. PSVI, with its ability to capture structured correlations, strikes a balance between efficiency and accuracy. In the context of Bayesian FWI, where problems are often high dimensional and non-linear, PSVI offers improved inversion results while maintaining a computational cost comparable to mean field ADVI. For inverse problems with lower dimensionality such that modelling a full covariance matrix is affordable, full rank ADVI could be a more suitable choice. When dealing with problems with strong multimodality, these Gaussian-based methods are not suitable. It is then advisable to use other variational methods such as normalizing flows (Rezende & Mohamed, 2015), BVI (F. Guo et al., 2016; Miller et al., 2017) or deterministic or stochastic SVGD (Liu & Wang, 2016; Gallego & Insua, 2018). These methods have shown effectiveness in solving multimodal problems, albeit at the cost of a larger number of forward simulations. The No Free Lunch theorem (Wolpert & Macready, 1997) can be paraphrased as: no method is better than any other method when averaged across all problems. There is therefore no possibility to find a 'best' method in general. Nevertheless, individual classes of problems may have more or less efficient algorithms, so having a variety of methods allows for tailored decisions to be based on the nature of the problem to be addressed.

In the 2D FWI example, we use a 5×5 correlation kernel as displayed in Figure 1d. To investigate the impact of the correlation kernel size on inversion results, we conduct an additional test using an 11×11 kernel. The mean, standard deviation and relative error maps of the obtained posterior distribution are displayed in Figure 10a, which reveal nearly identical features, such as the continuous layers discussed previously, when

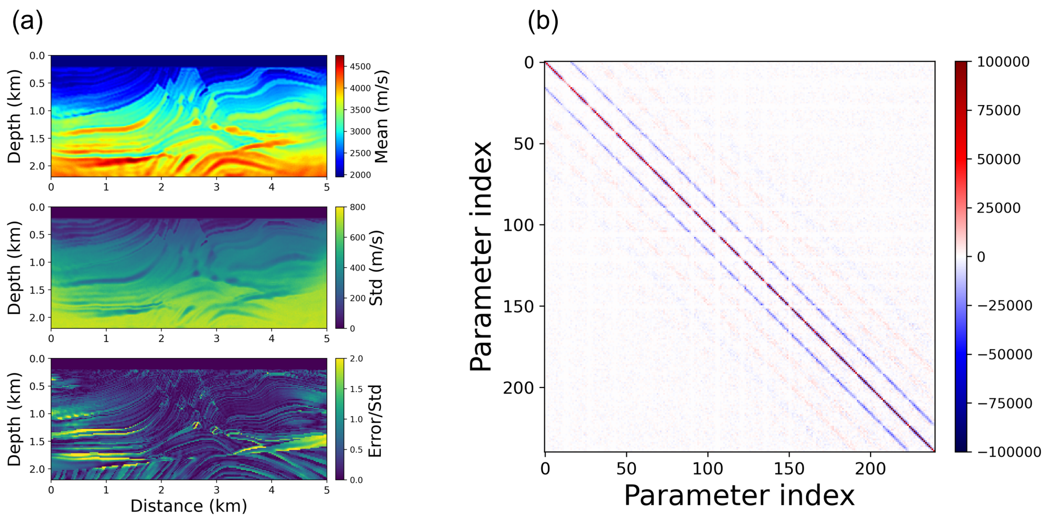

Figure 10. Inversion results obtained from PSVI using an 11×11 correlation kernel. (a) Mean, standard deviation and relative error maps. (b) Covariance matrix inside the white box in Figure 1a.

compared to those obtained using the 5×5 correlation kernel (Figure 3b). Figure 10b displays the posterior covariance matrix, which as expected presents more non-zero off-diagonal covariance blocks than the 5×5 kernel (Figure 5b). The covariance magnitudes decay from the main diagonal block, and become relatively small from the second off-diagonal block. However, modelling these additional covariances requires more parameters to construct the matrix **L**. In addition, from Figures 5c and 5d, the covariance matrices calculated using BVI and sSVGD exhibit only one prominent off-diagonal block, probably because the non-linearity of FWI makes it challenging to capture a broader correlation structure with embedding prior knowledge of the type of structure sought. Therefore, we conclude that the 5×5 correlation kernel used above is a reasonable choice that trades off both accuracy and efficiency.

In real applications, if other prior knowledge about the subsurface structure is available (e.g., from seismic travel time tomography), we can design specific correlation kernels to capture target-oriented correlation information. Furthermore, the underlying principles of PSVI can be adapted to address temporal problems such as time-lapse (4D) seismic monitoring in which we might expect spatial regularity in the location of injected fluids, or in earthquake forecasting where correlations between seismic events over time might be captured effectively.

PSVI is not merely an extension of mean field ADVI as proposed by Kucukelbir et al. (2017). In fact it can be used to extend a variety of variational methods to enhance their accuracy and efficiency. For example, in BVI the physically structured approach in PSVI can replace diagonal Gaussians in modelling the Gaussian component distributions used in X. Zhao and Curtis (2024). This substitution is likely to improve the accuracy of each component while maintaining similar computational efficiency, potentially leading to a reduction in the required number of components and overall computational cost for BVI.

Similar to BVI, PSVI produces an analytic posterior expression. Therefore, saving and loading inversion results, generating new posterior samples, and sharing the posterior distribution with others post inversion is simple (Scheiter et al., 2022). The proposed method can also be extended to other general Gaussian-based methods such as Gaussian processes (Ray & Myer, 2019; Valentine & Sambridge, 2020a, 2020b; Ray, 2021; Blatter et al., 2021) and mixture density networks (Bishop, 1994; Devilee et al., 1999; Meier et al., 2007; Shahraeeni & Curtis, 2011; Shahraeeni et al., 2012; Earp & Curtis, 2020; Hansen & Finlay, 2022; Bloem et al., 2023), to capture desired correlation structures. Interestingly, special neural network structures are designed for the same purpose, such as the *coupling layer* (Dinh et al., 2015, 2017; Durkan et al., 2019; X. Zhao et al., 2021; X. Zhang & Curtis, 2021b) and the *autoregressive layer* (Kingma et al., 2016; Papamakarios et al., 2017; Huang et al., 2018; De Cao et al., 2019; Levy et al., 2022). However, they often come with a higher number of hyperparameters, making PSVI an attractive and practical choice.

Considering that solving the forward function in 2D FWI is not hugely expensive, we use a relatively smaller step size and more iterations during variational inversion to ensure that the optimisation process has converged stably. Figure 2 illustrates that the negative ELBOs stop decreasing after 2500 - 3000 iterations, indicating that the full 5000 iterations used here might be redundant. For higher dimensional problems such as 3D FWI, we can potentially use larger step sizes with fewer iterations, thereby optimising the balance between computational resources and convergence speed.

The two interrogation examples presented here underscore the significance of estimating accurate uncertainties, even if that demands a substantial increase in computational input. Biased uncertainty information (such as that provided by mean field ADVI) leads to incorrect answers about Earth properties. Therefore, while obtaining an accurate mean velocity model in Bayesian inversion, or just the best-fit model in deterministic inversion, may appear useful, they are far from sufficient for an unbiased and quantitative interpretation of the true Earth. The pursuit of not only precision in mean velocity models but also robust and reliable uncertainty estimates is important for a comprehensive understanding of subsurface structures.

In the first interrogation example, we estimated the size of a subsurface reservoir, where we use relative velocity values and classify them as either low or high based on a velocity threshold value (X. Zhao et al., 2022). In the second example, we take the absolute velocity values and convert them into CO_2 saturation estimates using a non-linear rock physics relationship. If the inversion is performed with higher frequency data, the inverted velocity values would be better constrained and become more accurate. Consequently, the posterior distribution of the estimated CO_2 volume can be improved. In future, 3D Bayesian FWI, together with more advanced reservoir simulation and rock physics inversion techniques, can facilitate more sophisticated and realistic interrogation applications in subsurface carbon capture and storage, or other subsurface projects. This comprehensive approach, enriched with full uncertainty assessments, could significantly contribute to our understanding and improve decision-making in the context of such endeavours.

6 Conclusion

In this work, we propose physically structured variational inference (PSVI) to perform 2D Bayesian full waveform inversion (FWI), in which a physical structure is imposed on the uncertainties in variational distributions based on prior information about imaging problem solutions. In our application, correlations between specific pairs of spatial locations are parametrised and inferred during inversion. Thus, we are able to capture the main correlations with a desired structure in a computationally efficient manner. We apply the proposed method together with three other variational methods: mean field automatic differentiation variational inference (ADVI), boosting variational inference (BVI) and stochastic Stein variational gradient descent (sSVGD), to a synthetic FWI example. This demonstrates that PSVI yields accurate first-order statistical information, including the mean and standard deviation maps as well as the marginal distributions, which are all consistent with those obtained using BVI and sSVGD. It also provides other second-order statistical information, specifically the posterior covariances. In addition, the obtained full uncertainty information is verified through the application of the inversion results to two post-inversion interrogation problems: one estimating a subsurface reservoir size and another estimating CO_2 volume in a carbon capture and storage project. In our examples, PSVI exhibits nearly the same computational efficiency as mean field ADVI while enhancing the inversion accuracy significantly. This opens the possibility that 3D probabilistic FWI with full uncertainty estimation can be performed both efficiently and accurately.

7 Open Research

Software used to perform variational inference can be found at Pyro website (`https://pyro.ai/`, Bingham et al., 2018) and in X. Zhang and Curtis (2023). Software used to perform Automatic Differentiation can be found at PyTorch website (`https://pytorch.org/`, Paszke et al., 2019).

Acknowledgments

The authors thank Edinburgh Imaging Project (EIP - `https://blogs.ed.ac.uk/imaging/`) sponsors (BP and TotalEnergies) for supporting this research.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., … others (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Ardizzone, L., Kruse, J., Wirkert, S., Rahner, D., Pellegrini, E. W., Klessen, R. S., … Köthe, U. (2018). Analyzing inverse problems with invertible neural networks. *arXiv preprint arXiv:1808.04730*.

Arnold, R., & Curtis, A. (2018). Interrogation theory. *Geophysical Journal International*, *214*(3), 1830–1846.

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P., & Virieux, J. (2013). Regularized seismic full waveform inversion with prior model information. *Geophysics*, *78*(2), R25–R36.

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

Bernard, S., Monteiller, V., Komatitsch, D., & Lasaygues, P. (2017). Ultrasonic computed tomography based on full-waveform inversion for bone quantitative imaging. *Physics in Medicine & Biology*, *62*(17), 7011.

Berti, S., Aleardi, M., & Stucchi, E. (2023). A computationally efficient bayesian approach to full-waveform inversion. *Geophysical Prospecting*. doi: https://doi.org/10.1111/1365-2478.13437

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., ... Goodman, N. D. (2018). Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*.

Bishop, C. M. (1994). Mixture density networks. *Aston University*.

Bishop, C. M. (2006). *Pattern recognition and machine learning.* springer.

Biswas, R., & Sen, M. K. (2022). Transdimensional 2d full-waveform inversion and uncertainty estimation. *arXiv preprint arXiv:2201.09334*.

Blatter, D., Ray, A., & Key, K. (2021). Two-dimensional bayesian inversion of magnetotelluric data using trans-dimensional gaussian processes. *Geophysical Journal International*, *226*(1), 548–563.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association*, *112*(518), 859–877.

Bloem, H., Curtis, A., & Tetzlaff, D. (2023). Introducing conceptual geological information into Bayesian tomographic imaging. *Basin Research*. doi: doi:10.1111/bre.12811

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization.* Cambridge university press.

Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., ... Pugmire, D. (2016). Global adjoint tomography: first-generation model. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, *207*(3), 1739–1766.

Brie, A., Pampuri, F., Marsala, A., & Meazza, O. (1995). Shear sonic interpretation in gas-bearing sands. In *Spe annual technical conference and exhibition* (pp. SPE–30595).

Bui-Thanh, T., Ghattas, O., Martin, J., & Stadler, G. (2013). A computational framework for infinite-dimensional bayesian inverse problems part I: The linearized case, with application to global seismic inversion. *SIAM Journal on Scientific Computing*, *35*(6), A2494–A2523.

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

De Cao, N., Titov, I., & Aziz, W. (2019). Block neural autoregressive flow. *arXiv preprint arXiv:1904.04676*.

de Lima, P. D. S., Corso, G., Ferreira, M. S., & de Araújo, J. M. (2023). Acoustic full waveform inversion with hamiltonian monte carlo method. *Physica A: Statistical Mechanics and its Applications*, *617*, 128618.

de Lima, P. D. S., Ferreira, M. S., Corso, G., & de Araújo, J. M. (2023). Bayesian time-lapse full waveform inversion using hamiltonian monte carlo. *arXiv preprint arXiv:2311.02999*.

Devilee, R., Curtis, A., & Roy-Chowdhury, K. (1999). An efficient, probabilistic neural network approach to solving inverse problems: inverting surface wave velocities for eurasian crustal thickness. *Journal of Geophysical Research: Solid Earth*, *104*(B12), 28841–28857.

Dhabaria, N., & Singh, S. C. (2024). Hamiltonian monte carlo based elastic full-waveform inversion of wide-angle seismic data. *Geophysical Journal International*, ggae112.

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

Dinh, L., Sohl-Dickstein, J., & Bengio, S. (2017). Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*.

Dupuy, B., Ghaderi, A., Querendez, E., Mezyk, M., et al. (2017). Constrained avo for co2 storage monitoring at sleipner. *Energy Procedia*, *114*, 3927–3936.

Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. (2019). Neural spline flows. In *Advances in neural information processing systems* (pp. 7509–7520).

Earp, S., & Curtis, A. (2020). Probabilistic neural network-based 2d travel-time tomography. *Neural Computing and Applications*, *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.

Fang, Z., Da Silva, C., Kuske, R., & Herrmann, F. J. (2018). Uncertainty quantification for inverse problems with weak partial-differential-equation constraints. *Geophysics*, *83*(6), R629–R647.

Fichtner, A., Kennett, B. L., Igel, H., & Bunge, H.-P. (2009). Full seismic waveform tomography for upper-mantle structure in the australasian region using adjoint methods. *Geophysical Journal International*, *179*(3), 1703–1725.

Fichtner, A., van Herwaarden, D.-P., Afanasiev, M., Simutė, S., Krischer, L., Çubuk-Sabuncu, Y., … others (2018). The collaborative seismic earth model: generation 1. *Geophysical research letters*, *45*(9), 4007–4016.

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

Fu, X., & Innanen, K. A. (2022). A time-domain multisource bayesian/markov chain monte carlo formulation of time-lapse seismic waveform inversion. *Geophysics*, *87*(4), R349–R361.

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

Gassmann, F. (1951). Elastic waves through a packing of spheres. *Geophysics*, *16*(4), 673–685.

Gebraad, L., Boehm, C., & Fichtner, A. (2020). Bayesian elastic full-waveform inversion using hamiltonian monte carlo. *Journal of Geophysical Research: Solid Earth*, *125*(3), e2019JB018428.

Ghosh, R., & Ojha, M. (2020). Prediction of elastic properties within co2 plume at sleipner field using avs inversion modified for thin-layer reflections guided by uncertainty estimation. *Journal of Geophysical Research: Solid Earth*, *125*(11), e2020JB019782.

Gouveia, W. P., & Scales, J. A. (1998). Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis. *Journal of Geophysical Research: Solid Earth*, *103*(B2), 2759–2779.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, *82*(4), 711–732.

Green, P. J. (2003). Trans-dimensional markov chain monte carlo. *Oxford Statistical Science Series*, 179–198.

Guasch, L., Calderón Agudo, O., Tang, M.-X., Nachev, P., & Warner, M. (2020). Full-waveform inversion imaging of the human brain. *NPJ digital medicine*, *3*(1), 28.

Guo, F., Wang, X., Fan, K., Broderick, T., & Dunson, D. B. (2016). Boosting variational inference. *Advances in Neural Information Processing Systems*.

Guo, P., Visser, G., & Saygin, E. (2020). Bayesian trans-dimensional full waveform inversion: synthetic and field data application. *Geophysical Journal International*, *222*(1), 610–627.

Hansen, T. M., & Finlay, C. C. (2022). Use of machine learning to estimate statistics of the posterior distribution in probabilistic inverse problems—an application to airborne em data. *Journal of Geophysical Research: Solid Earth*, *127*(11), e2022JB024703.

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

He, J., Rao, J., Fleming, J. D., Gharti, H. N., Nguyen, L. T., & Morrison, G. (2021). Numerical ultrasonic full waveform inversion (fwi) for complex structures in coupled 2d solid/fluid media. *Smart Materials and Structures*, *30*(8), 085044.

Huang, C.-W., Krueger, D., Lacoste, A., & Courville, A. (2018). Neural autoregressive flows. *arXiv preprint arXiv:1804.00779*.

Izzatullah, M., Ravasi, M., & Alkhalifah, T. (2023). Physics reliable frugal uncertainty analysis for full waveform inversion. *arXiv preprint arXiv:2305.07921*.

Khoshkholgh, S., Zunino, A., & Mosegaard, K. (2021). Informed proposal monte carlo. *Geophysical Journal International*, *226*(2), 1239–1248.

Khoshkholgh, S., Zunino, A., & Mosegaard, K. (2022). Full-waveform inversion by informed-proposal monte carlo. *Geophysical Journal International*, *230*(3), 1824–1833.

Kim, J., Nam, M. J., & Matsuoka, T. (2013). Estimation of co2 saturation during both co2 drainage and imbibition processes based on both seismic velocity and electrical resistivity measurements. *Geophysical Journal International*, *195*(1), 292–300.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kingma, D. P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I., & Welling, M. (2016). Improved variational inference with inverse autoregressive flow. In *Advances in neural information processing systems* (pp. 4743–4751).

Kotsi, M., Malcolm, A., & Ely, G. (2020a). Time-lapse full-waveform inversion using hamiltonian monte carlo: a proof of concept. In *Seg technical program expanded abstracts 2020* (pp. 845–849). Society of Exploration Geophysicists.

Kotsi, M., Malcolm, A., & Ely, G. (2020b). Uncertainty quantification in time-lapse seismic imaging: A full-waveform approach. *Geophysical Journal International*, *222*(2), 1245–1263.

Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., & Blei, D. M. (2017). Automatic differentiation variational inference. *The Journal of Machine Learning Research*, *18*(1), 430–474.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, *22*(1), 79–86.

Levy, S., Laloy, E., & Linde, N. (2022). Variational bayesian inference with complex geostatistical priors using inverse autoregressive flows. *Computers & Geosciences*, 105263.

Liu, Q., & Wang, D. (2016). Stein variational gradient descent: A general purpose bayesian inference algorithm. In *Advances in neural information processing systems* (pp. 2378–2386).

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 & exhibition* (Vol. 2023, pp. 1–5).

Lucka, F., Pérez-Liva, M., Treeby, B. E., & Cox, B. T. (2021). High resolution 3d ultrasonic breast imaging by time-domain full waveform inversion. *Inverse Problems*, *38*(2), 025008.

Martin, G. S., Wiley, R., & Marfurt, K. J. (2006). Marmousi2: An elastic upgrade for marmousi. *The leading edge*, *25*(2), 156–166.

Meier, U., Curtis, A., & Trampert, J. (2007). Fully nonlinear inversion of fundamental mode surface waves for a global crustal model. *Geophysical Research Letters*, *34*(16).

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, *21*(6), 1087–1092.

Miller, A. C., Foti, N. J., & Adams, R. P. (2017). Variational boosting: Iteratively refining posterior approximations. In *International conference on machine learning* (pp. 2420–2429).

Mosegaard, K., & Tarantola, A. (1995). Monte carlo sampling of solutions to inverse problems. *Journal of Geophysical Research: Solid Earth*, *100*(B7), 12431–12447.

Papamakarios, G., Pavlakou, T., & Murray, I. (2017). Masked autoregressive flow for density estimation. In *Advances in neural information processing systems* (pp. 2338–2347).

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... others (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*.

Patsia, O., Giannopoulos, A., & Giannakis, I. (2023). GPR full-waveform inversion with deep-learning forward modelling: A case study from non-destructive testing. *IEEE Transactions on Geoscience and Remote Sensing*.

Plessix, R.-E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, *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.

Pratt, R. G., Shin, C., & Hick, G. (1998). Gauss–newton and full newton methods in frequency–space seismic waveform inversion. *Geophysical journal international*, *133*(2), 341–362.

Prieux, V., Brossier, R., Operto, S., & Virieux, J. (2013). Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the valhall field. part 1: Imaging compressional wave speed, density and attenuation. *Geophysical Journal International*, *194*(3), 1640–1664.

Ray, A. (2021). Bayesian inversion using nested trans-dimensional gaussian processes. *Geophysical Journal International*, *226*(1), 302–326.

Ray, A., Kaplan, S., Washbourne, J., & Albertin, U. (2018). Low frequency full waveform seismic inversion within a tree based bayesian framework. *Geophysical Journal International*, *212*(1), 522–542.

Ray, A., & Myer, D. (2019). Bayesian geophysical inversion with trans-dimensional gaussian process machine learning. *Geophysical Journal International*, *217*(3), 1706–1726.

Ray, A., Sekar, A., Hoversten, G. M., & Albertin, U. (2016). Frequency domain full waveform elastic inversion of marine seismic data from the alba field using a bayesian trans-dimensional algorithm. *Geophysical Journal International*, *205*(2), 915–937.

Rezende, D. J., & Mohamed, S. (2015). Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*.

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

Sambridge, M., & Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, *40*(3), 3–1.

Scheiter, M., Valentine, A., & Sambridge, M. (2022). Upscaling and downscaling monte carlo ensembles with generative models. *Geophysical Journal International*, *230*(2), 916–931.

Sen, M. K., & Roy, I. G. (2003). Computation of differential seismograms and iteration adaptive regularization in prestack waveform inversion. *Geophysics*, *68*(6), 2026–2039.

Shahraeeni, M. S., & Curtis, A. (2011). Fast probabilistic nonlinear petrophysical inversion. *Geophysics*, *76*(2), E45–E58.

Shahraeeni, M. S., Curtis, A., & Chao, G. (2012). Fast probabilistic petrophysical mapping of reservoirs from 3d seismic data. *Geophysics*, *77*(3), O1–O19.

Siahkoohi, A., & Herrmann, F. J. (2021). Learning by example: fast reliability-aware seismic imaging with normalizing flows. In *First international meeting for applied geoscience & energy* (pp. 1580–1585).

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., Louboutin, M., Witte, P. A., & Herrmann, F. J. (2021).

Preconditioned training of normalizing flows for variational inference in inverse problems. *arXiv preprint arXiv:2101.03709*.

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.

Siahkoohi, A., Rizzuti, G., Witte, P. A., & Herrmann, F. J. (2020). Faster uncertainty quantification for inverse problems with conditional normalizing flows. *arXiv preprint arXiv:2007.07985*.

Smith, J. D., Ross, Z. E., Azizzadenesheli, K., & Muir, J. B. (2022). Hyposvi: Hypocentre inversion with stein variational inference and physics informed neural networks. *Geophysical Journal International*, *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 CO2 saturations in a subsurface reservoir. *Geophysical Journal International*, *236*(3), 1309–1331.

Sun, L., Wang, L., Xu, G., & Wu, Q. (2023). A new method of variational bayesian slip distribution inversion. *Journal of Geodesy*, *97*(1), 10.

Tape, C., Liu, Q., Maggi, A., & Tromp, J. (2010). Seismic tomography of the southern california crust based on spectral-element and adjoint methods. *Geophysical Journal International*, *180*(1), 433–462.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, *49*(8), 1259–1266.

Thurin, J., Brossier, R., & Métivier, L. (2019). Ensemble-based uncertainty estimation in full waveform inversion. *Geophysical Journal International*, *219*(3), 1613–1635.

Treeby, B. E., & Cox, B. T. (2010). k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields. *Journal of biomedical optics*, *15*(2), 021314–021314.

Valentine, A. P., & Sambridge, M. (2020a). Gaussian process models — I. a framework for probabilistic continuous inverse theory. *Geophysical Journal International*, *220*(3), 1632–1647.

Valentine, A. P., & Sambridge, M. (2020b). Gaussian process models — II. lessons for discrete inversion. *Geophysical Journal International*, *220*(3), 1648–1656.

Virieux, J., & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, *74*(6), WCC1–WCC26.

Visser, G., Guo, P., & Saygin, E. (2019). Bayesian transdimensional seismic full-waveform inversion with a dipping layer parameterization. *Geophysics*, *84*(6), R845–R858.

Wang, W., McMechan, G. A., & Ma, J. (2023). Re-weighted variational full waveform inversions. *Geophysics*, *88*(4), 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.

Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., ... others (2013). Anisotropic 3d full-waveform inversion. *Geophysics*, *78*(2), R59–R80.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, *1*(1), 67–82.

Yin, Z., Orozco, R., Louboutin, M., & Herrmann, F. J. (2024). Wise: full-waveform variational inference via subsurface extensions. *Geophysics*, *89*(4), 1–31.

Zhang, C., Bütepage, J., Kjellström, H., & Mandt, S. (2018). Advances in variational inference. *IEEE transactions on pattern analysis and machine intelligence*, *41*(8), 2008–2026.

Zhang, X., & Curtis, A. (2020a). Seismic tomography using variational inference methods. *Journal of Geophysical Research: Solid Earth*, *125*(4), e2019JB018589.

Zhang, X., & Curtis, A. (2020b). Variational full-waveform inversion. *Geophysical Journal International*, *222*(1), 406-411.

Zhang, X., & Curtis, A. (2021a). Bayesian full-waveform inversion with realistic priors. *Geophysics*, *86*(5), 1–20.

Zhang, X., & Curtis, A. (2021b). Bayesian geophysical inversion using invertible neural networks. *Journal of Geophysical Research: Solid Earth*, e2021JB022320.

Zhang, X., & Curtis, A. (2022). Interrogating probabilistic inversion results for subsurface structural information. *Geophysical Journal International*, *229*(2), 750–757.

Zhang, X., & Curtis, A. (2023). VIP – Variational inversion package with example implementations of Bayesian tomographic imaging. *arXiv preprint arXiv:2310.13325*.

Zhang, X., Lomas, A., Zhou, M., Zheng, Y., & Curtis, A. (2023). 3-d bayesian variational full waveform inversion. *Geophysical Journal International*, *234*(1), 546–561.

Zhang, X., Nawaz, A., Zhao, X., & Curtis, A. (2021). An introduction to variational inference in geophysical inverse problems. *Advances in Geophysics*, *62*, 73-140.

Zhao, X., & Curtis, A. (2024). Bayesian inversion, uncertainty analysis and interrogation using boosting variational inference. *Journal of Geophysical Research: Solid Earth*, *129*(1), e2023JB027789.

Zhao, X., Curtis, A., & Zhang, X. (2021). Bayesian seismic tomography using normalizing flows. *Geophysical Journal International*, *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. *Journal of Geophysical Research: Solid Earth*, *127*(4), e2022JB024098.

Zhao, X., Zhou, H., Chen, H., & Wang, Y. (2020). Domain decomposition for large-scale viscoacoustic wave simulation using localized pseudo-spectral method. *IEEE Transactions on Geoscience and Remote Sensing*, 1-14.

Zhao, Z., & Sen, M. K. (2021). A gradient-based markov chain monte carlo method for full-waveform inversion and uncertainty analysis. *Geophysics*, *86*(1), R15–R30.

Zhdanov, M. S. (2002). *Geophysical inverse theory and regularization problems* (Vol. 36). Elsevier.

Zhu, H., Li, S., Fomel, S., Stadler, G., & Ghattas, O. (2016). A bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration. *Geophysics*, *81*(5), R307–R323.

Zidan, A., Li, Y., & Cheng, A. (2022). Regularized seismic amplitude inversion via variational inference. *Geophysical Prospecting*, *70*(9), 1507–1527.

Zunino, A., Gebraad, L., Ghirotto, A., & Fichtner, A. (2023, 10). HMCLab: a framework for solving diverse geophysical inverse problems using the hamiltonian monte carlo method. *Geophysical Journal International*, *235*(3), 2979-2991.