GEOPHYSICS, VOL. 90, NO. 6 (NOVEMBER-DECEMBER 2025); P. R373–R388, 13 FIGS.
10.1190/GEO2024-0774.1

DOI:10.1190/geo2024-0774.1

Efficient Bayesian full-waveform inversion and analysis of prior hypotheses in three dimensions

Xuebin Zhao[1] and Andrew Curtis[1]

ABSTRACT

Three-dimensional seismic full-waveform inversion (3D FWI) is a highly nonlinear and computationally demanding inverse problem that constructs 3D subsurface seismic velocity structures using seismic waveform data. To characterize nonuniqueness in the solutions, we demonstrate Bayesian 3D FWI using an efficient method called physically structured variational inference and apply it to 3D acoustic Bayesian FWI. The results provide reasonable posterior uncertainty estimates, at a computational cost that is only an order of magnitude greater than that of standard, deterministic FWI. Furthermore, we deploy variational prior replacement to calculate Bayesian solutions corresponding to different classes of prior information at low additional cost. The results obtained using prior information that models should be smooth show loop-like high uncertainty structures that are consistent with the fully nonlinear inversion results presented previously. These structures disappear when smoothing is not imposed, so we conclude that they may be caused by smoothness constraints in tomographic problems. We further analyze a variety of prior hypotheses by constructing Bayesian L-curves, which reveal the sensitivity of the inversion process to different prior assumptions. To our knowledge, this is the first study that allows such prior hypotheses to be compared in probabilistic 3D FWI at a feasible computational cost. This work shows that fully probabilistic 3D FWI can be performed and used to test different prior hypotheses, at a cost that may be practical, at least for small problems.

INTRODUCTION

Seismic full-waveform inversion (FWI) uses phase and amplitude information from observed waveform data to estimate compatible subsurface seismic velocity maps (Virieux and Operto, 2009). FWI is often implemented using deterministic methods, such as gradient-based optimization, in which a measure of the misfit between the observed and simulated seismograms is minimized iteratively. Given the highly nonlinear nature of the FWI problem, achieving a good initial model is crucial to avoid convergence to a locally optimal solution, and due to the ill-posed nature of most FWI problems, additional regularization is also required to stabilize every inversion (Zhdanov, 2002; Asnaashari et al., 2013; Aghamiry et al., 2018). Although this approach provides a single point estimate of a solution, it unfortunately cannot readily be extended to estimate accurate uncertainty in the solutions and, therefore, cannot be used to assess risks in post imaging decision-making processes robustly (Arnold and Curtis, 2018; Ely et al., 2018; Siahkoohi et al., 2022; Zhang and Curtis, 2022; Zhao et al., 2022).

Bayesian inference solves inverse problems within a probabilistic framework, in which a family of all potential solutions and their uncertainties are described by the so-called *posterior* probability distribution function (PDF). The posterior PDF is calculated using Bayes' rule to update *prior* knowledge about model parameters with new information from observed data. Monte Carlo sampling methods are commonly used to sample the Bayesian posterior PDF to quantify uncertainties. However, applying conventional Monte Carlo methods to Bayesian FWI is computationally expensive due to the many model parameters involved, which incurs the curse of dimensionality (Curtis and Lomax, 2001; Sambridge and

Manuscript received by the Editor 6 November 2024; revised manuscript received 17 June 2025; published ahead of production 17 July 2025; published online 18 September 2025.
[1]University of Edinburgh, School of GeoSciences, Edinburgh, United Kingdom. E-mail: xuebin.zhao@ed.ac.uk (corresponding author); andrew.curtis@ed.ac.uk.

Mosegaard, 2002; Scales, 2005). Efficient sampling techniques have been developed in attempts to address this issue by either reducing the dimensionality (the number of unknown parameters) of the sampling problem (Ray et al., 2016; Tsai et al., 2023; Hu et al., 2024; Mulder and Kuvshinov, 2024), using data-model gradient information to enhance sampling efficiency (Gebraad et al., 2020; Zhao and Sen, 2021; Berti et al., 2023; Zunino et al., 2023), or a combination of both (Biswas and Sen, 2022). Nevertheless, direct posterior sampling methods such as these remain a computational challenge.

Monte Carlo methods are expensive, in part, because no structure is imposed a priori on the sampling distribution; sampling algorithms must therefore be sufficiently flexible to adapt to and explore any PDF topography to any level of detail. Variational inference offers an alternative to Monte Carlo methods for solving Bayesian problems. The method searches for an estimate of the posterior PDF that has a particular structure imposed a priori. This usually comes in the form of a family of PDFs, which is expected to contain an acceptable approximation to the posterior PDF. The best estimate of the posterior PDF within that family can then be found by optimization rather than random sampling. Because the range of possible PDF estimates is infinitely smaller than the range of general PDFs explored by Monte Carlo methods, variational methods can be more efficient and easier to scale to high-dimensional problems. Recently, variational methods have been applied to 2D Bayesian FWI and have been shown to reduce its computational cost while still providing acceptable estimates of the posterior distribution (Zhang and Curtis, 2021; Bates et al., 2022; Wang et al., 2023; Izzatullah et al., 2024; Sun et al., 2024; Xie et al., 2024; Yin et al., 2024; Zhao and Curtis, 2025).

This work concerns 3D FWI, which is typically included among the most computationally challenging problems in subsurface science. Nevertheless, Zhang et al. (2023) apply variational methods to a synthetic acoustic 3D Bayesian FWI problem, demonstrating the feasibility of finding solutions. Lomas et al. (2023) and Walker et al. (2024) then apply one of the methods to a field data set collected with an airgun source and ocean bottom nodes over a salt body in the Gulf of Mexico. Although there was no independent test of the solution quality, they estimated that uncertainties were higher around the boundaries of the salt body, as is generally expected in true seismic imaging solutions (Galetti et al., 2015). Hoffmann et al. (2024) also perform 3D FWI and local uncertainty estimation using a source subsampling strategy and ensemble Kalman filter (Evensen, 1994). In most of these 3D FWI studies, a spatial smoothing operator is used *a priori* to improve the convergence rate of the inversion. However, the obtained posterior distribution may then underestimate uncertainties, which in part exclude the true solution (Zhang et al., 2023). Moreover, the type and strength of the regularization applied are effectively subjective hypotheses, and they are difficult to choose in practice because almost any form of regularity in the true Earth varies spatially. Thus, in principle, each such hypothesis should be analyzed and tested. Considering the huge computational cost of 3D FWI, analyzing different hypotheses by comparing the Bayesian inversion results that follow from each one is usually impractical.

We demonstrate an efficient Bayesian 3D FWI and an approach to analyze different prior hypotheses. The novel contributions of this work can be summarized as follows:

First, we show Bayesian 3D FWI with substantially improved efficiency compared with previous methods. The improvement is attributed to the use of a recently introduced variational method: physically structured variational inference (PSVI) (Zhao and Curtis, 2024b), which has been proven to provide relatively accurate inversion results, including uncertainty estimates at greatly reduced computational cost in 2D FWI problems. Our experiment shows that with PSVI, fully nonlinear 3D Bayesian FWI can be solved using only an order of magnitude more computation than traditional, deterministic FWI. To further improve the accuracy of the 3D FWI results reported in previous studies, we use a multiscale inversion strategy (Bunks et al., 1995) by performing Bayesian FWI across three frequency bands, using relatively low, intermediate, and high-frequency data in an attempt to reduce cycle skipping.

Second, to analyze, test, and potentially select between different prior hypotheses, we apply a variational prior replacement (VPR) methodology (Zhao and Curtis, 2024c) to obtain inversion results corresponding to different prior PDFs at a negligible additional computational cost. We compare VPR results with those obtained from independent Bayesian inversion and find that VPR cannot provide perfect estimates of posterior uncertainties due to the extreme complexity of this 3D FWI problem. We therefore introduce a method to fine-tune the VPR results with a small amount of extra computation, which significantly improves the results.

Finally, we use the posterior PDFs obtained using different prior distributions to construct a so-called *Bayesian L-curve*, from which we analyze different prior assumptions and select one optimal choice among different hypotheses.

In the next section, we summarize the methods deployed in this study. Next, we apply PSVI to 3D FWI and compare the inversion results with those from other variational and Monte Carlo methods, from which we identify two potential problems that affect the inversion accuracy. We then show how these issues can be addressed and present improved inversion results. In addition, we compare inversion results obtained using different prior hypotheses by building a *Bayesian L-curve*. Finally, we discuss the implications of this work and draw conclusions.

METHODOLOGY

Variational Bayesian inversion

Bayesian inference calculates the *posterior* PDF of model parameters $\mathbf{m}$ given the observed data $\mathbf{d}_{\text{obs}}$ using Bayes' rule

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

where $p(\mathbf{m})$ represents *prior* information about $\mathbf{m}$, and $p(\mathbf{d}_{\text{obs}}|\mathbf{m})$ is the *likelihood* of observing $\mathbf{d}_{\text{obs}}$ given any value of $\mathbf{m}$. The term $p(\mathbf{d}_{\text{obs}})$ is a normalization constant called the *evidence*.

In this work, we focus on variational inference, in which an optimal variational distribution $q(\mathbf{m})$ is selected from a family of predefined probability distributions to best approximate the true but unknown posterior distribution $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$. This is often accomplished by minimizing the Kullback-Leibler (KL) divergence, a measure of difference, between the posterior and variational distributions (Kullback and Leibler, 1951), or equivalently by maximizing the evidence lower bound (ELBO) of $\log p(\mathbf{d}_{\text{obs}})$ defined as (Blei et al., 2017)

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

DOI:10.1190/geo2024-0774.1

where $\mathbb{E}_{q(\mathbf{m})}[g(\mathbf{m})] = \int_{\mathbf{m}} q(\mathbf{m}) g(\mathbf{m}) d\mathbf{m}$, which calculates the expectation of a function $g(\mathbf{m})$ with respect to the probability distribution $q(\mathbf{m})$. Finding the optimal $q(\mathbf{m})$ by maximizing equation 2 is an optimization problem with a fully probabilistic result.

Physically structured variational inference

PSVI is an efficient variational method that, in one implementation, uses a transformed Gaussian distribution with a specific covariance structure to best fit the posterior distribution (Zhao and Curtis, 2024b). A Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is characterized by a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$. To ensure positive semi-definiteness, the covariance matrix is often re-parameterized using a Cholesky factorization $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$, where $\mathbf{L}$ is a lower triangular matrix. A variational distribution $q(\mathbf{m})$ can then be obtained by applying an invertible transform f to the Gaussian distribution

$$\log\, q(\mathbf{m}) = \log\, \mathcal{N}(\boldsymbol{\mu}, \mathbf{L}\mathbf{L}^{\mathbf{T}}) - \log|\det[\partial_{\mathbf{m}} f^{-1}(\mathbf{m})]|. \quad (3)$$

The term $\det[\cdot]$ calculates the absolute value of the determinant of the Jacobian matrix $\partial_{\mathbf{m}} f^{-1}(\mathbf{m})$, which accounts for the volume change effected by f (Rezende and Mohamed, 2015). This transform converts a Gaussian random variable defined in an unbounded space into the space of physical model parameters $\mathbf{m}$, which is often bounded (Kucukelbir et al., 2017).

Modeling a full covariance matrix requires $n(n+1)/2$ hyperparameters to construct $\mathbf{L}$ (n being the number of model parameters or the dimensionality of an inverse problem), which is infeasible in 3D FWI problems in which n is typically of an order greater than 10^5. However, modeling only the diagonal elements (the commonly applied, so-called "mean-field approximation") ignores all correlations between the model parameters, resulting in a significant underestimation of the uncertainties in the posterior distribution (Kucukelbir et al., 2017; Zhang et al., 2023). To avoid these two extremes, PSVI focuses on capturing the most important (dominant) posterior correlations in the model vector $\mathbf{m}$, guided by prior knowledge of the physics controlling typical imaging inverse problems. Specifically, our implementation of PSVI includes posterior correlations between pairs of locations that are in spatial proximity to each other, typically within one dominant wavelength in FWI problems, because wavefield reflection and refraction respond to physical properties of the subsurface, averaged spatially over approximately 1 wavelength. All other parameter correlations are ignored. This structure of posterior correlations has been observed to emerge naturally in a number of previous nonlinear studies, validating this approach (Zhao and Curtis, 2024b).

We implement PSVI by defining specific off-diagonal elements of $\mathbf{L}$ as the real number parameters to be optimized during inversion; all other elements are set to zero, thus assuming the spatial independence of the corresponding parameter pairs (Zhao and Curtis, 2024b). This results in a sparse structure for $\mathbf{L}$, and thus, also the covariance matrix $\boldsymbol{\Sigma}$, requiring only a manageable number of hyperparameters to be optimized during the inversion while still capturing what are typically found to be the most significant correlations.

We acknowledge that this approach does ignore some inter-parameter correlations, which are not strictly zero. However, now that we have managed to perform fully nonlinear Bayesian FWI in our previous work (e.g., Zhao and Curtis, 2024b), we observe clearly that the dominant posterior correlations almost always occur locally, between cells that are up to a wavelength or so apart horizontally and vertically. This is because, within a wavelength, there is little ability to resolve much more than the average velocity values. Therefore, sub-wavelength velocities can be varied such that they preserve the correct mean value without compromising data fit. Correlations with parameters in cells that are further afield break down very rapidly because those parameters are also most strongly correlated with the values in cells in their local neighborhoods. In other words, although it is true that there are far-field correlations, they are demonstrably (and intuitively) far less important than those in local neighborhoods. In 3D FWI problems, we are obliged to make some approximations to reduce memory requirements when constructing a covariance matrix. To illustrate, in the following section we present a 3D FWI example with a grid size of $101 \times 101 \times 63$. If we consider even a single-precision floating-point format (each element has a size of four bytes), storing a full covariance matrix requires about one TB of memory, which is infeasible for most cases in reality.

Considering the computational cost of 3D Bayesian FWI, a top priority is to improve its efficiency. PSVI does so at the cost of some loss of generality in the solution found due to the Gaussian foundations for the posterior distribution. Zhao and Curtis (2024b) demonstrate that the method, nevertheless, produces reasonable statistical information about the full, nonlinear posterior distribution. We therefore accept the loss of generality and use the method in this work.

Variational prior replacement

The prior replacement was developed for situations wherein we wish to calculate various Bayesian solutions to an inverse problem using different classes of prior information given the same observed data set (Walker and Curtis, 2014). The variational version of the method used here, called variational prior replacement (VPR), was developed to improve its computational efficiency (Zhao and Curtis, 2024c). Suppose we have two different prior PDFs, $p_{\text{old}}(\mathbf{m})$ and $p_{\text{new}}(\mathbf{m})$, (hereafter, subscripts old and new will be used to denote the order in which the PDFs are estimated or used). According to Bayes' rule (equation 1), the two posterior probability distributions can be calculated by

$$p_{\text{old}}(\mathbf{m}|\mathbf{d}_{\text{obs}}) = \frac{p(\mathbf{d}_{\text{obs}}|\mathbf{m}) p_{\text{old}}(\mathbf{m})}{p_{\text{old}}(\mathbf{d}_{\text{obs}})} \approx q_{\text{old}}(\mathbf{m}), \quad (4)$$

and

$$\begin{aligned} p_{\text{new}}(\mathbf{m}|\mathbf{d}_{\text{obs}}) &= \frac{p(\mathbf{d}_{\text{obs}}|\mathbf{m}) p_{\text{new}}(\mathbf{m})}{p_{\text{new}}(\mathbf{d}_{\text{obs}})}, \\ &= \frac{p(\mathbf{d}_{\text{obs}}|\mathbf{m}) p_{\text{old}}(\mathbf{m})}{p_{\text{old}}(\mathbf{d}_{\text{obs}})} \frac{p_{\text{new}}(\mathbf{m})}{p_{\text{old}}(\mathbf{m})} \frac{p_{\text{old}}(\mathbf{d}_{\text{obs}})}{p_{\text{new}}(\mathbf{d}_{\text{obs}})}, \\ &= k p_{\text{old}}(\mathbf{m}|\mathbf{d}_{\text{obs}}) \frac{p_{\text{new}}(\mathbf{m})}{p_{\text{old}}(\mathbf{m})}, \\ &\approx k q_{\text{old}}(\mathbf{m}) \frac{p_{\text{new}}(\mathbf{m})}{p_{\text{old}}(\mathbf{m})}, \\ &\approx q_{\text{new}}(\mathbf{m}). \end{aligned} \quad (5)$$

In equation 4, the (old) Bayesian problem given data $\mathbf{d}_{\text{obs}}$ and prior PDF $p_{\text{old}}(\mathbf{m})$ is solved using variational inference, and the solution $p_{\text{old}}(\mathbf{m}|\mathbf{d}_{\text{obs}})$ is approximated by the variational distribution

DOI:10.1190/geo2024-0774.1

$q_{\text{old}}(\mathbf{m})$. The third and fourth lines in equation 5 are obtained by substituting equation 4 into the second line. The ratio $p_{\text{new}}(\mathbf{m})/p_{\text{old}}(\mathbf{m})$ updates the solution that includes the old prior to a solution that contains the new one, and we set a constant $k = p_{\text{old}}(\mathbf{d}_{\text{obs}})/p_{\text{new}}(\mathbf{d}_{\text{obs}})$. This implies that the new posterior distribution can be calculated from the old one by updating prior information post inversion, as shown by Walker and Curtis (2014): the inversion corresponding to the new prior distribution does not need to be solved from scratch.

Evaluating k requires the two evidence terms to be calculated, which is computationally intractable. Zhao and Curtis (2024c) introduce a second variational distribution $q_{\text{new}}(\mathbf{m})$ to approximate $p_{\text{new}}(\mathbf{m}|\mathbf{d}_{\text{obs}})$ (the fifth line in equation 5) by minimizing the KL divergence between the new posterior and $q_{\text{new}}(\mathbf{m})$, without requiring the value of k to be known explicitly. This step is called VPR.

A variety of variational inference methods can be used to calculate $q_{\text{old}}(\mathbf{m})$ and $q_{\text{new}}(\mathbf{m})$ (Kucukelbir et al., 2017; Zhang and Curtis, 2020; Zhao et al., 2021). In this paper, we use PSVI because it is efficient, and importantly, because its probability value is easy to evaluate for any model parameter values post inversion. Using PSVI to find $q_{\text{old}}(\mathbf{m})$ thus allows us to efficiently perform VPR to find $q_{\text{new}}(\mathbf{m})$ in the fifth line of equation 5, which approximates the solution for any of the alternative prior distributions.

3D BAYESIAN FWI EXAMPLE

We consider a synthetic 3D acoustic FWI problem. As shown in Figure 1a, the true velocity model used in this test is part of the 3D overthrust model (Aminzadeh et al., 1996), which contains $63 \times 101 \times 101$ grid cells in the Z (depth), Y, and X directions, respectively, with a cell size of 50 m in each direction. Figure 1c–1e shows three 2D slices of the true velocity structure at locations Y = 2.5 km, X = 2.5 km, and Z = 1.25 km (horizontal slice), respectively. We deploy 81 impulsive sources (the red stars) and 10,201 receivers (the white dots) at the surface with spacings of 500 m and 50 m, respectively. Both observed and synthetic data are generated by solving the 3D acoustic wave equation using a time-domain pseudo-analytical method (Etgen and Brandsberg-Dahl, 2009) with an Ormsby wavelet (Ryan, 1994).

We define a uniform prior distribution for velocity values at different depths, with the upper and lower bounds shown in Figure 1b. We use this uniform prior PDF and a diagonal Gaussian likelihood function (implying uncorrelated data noise, which is not a requirement of the method) to perform variational Bayesian inversion.

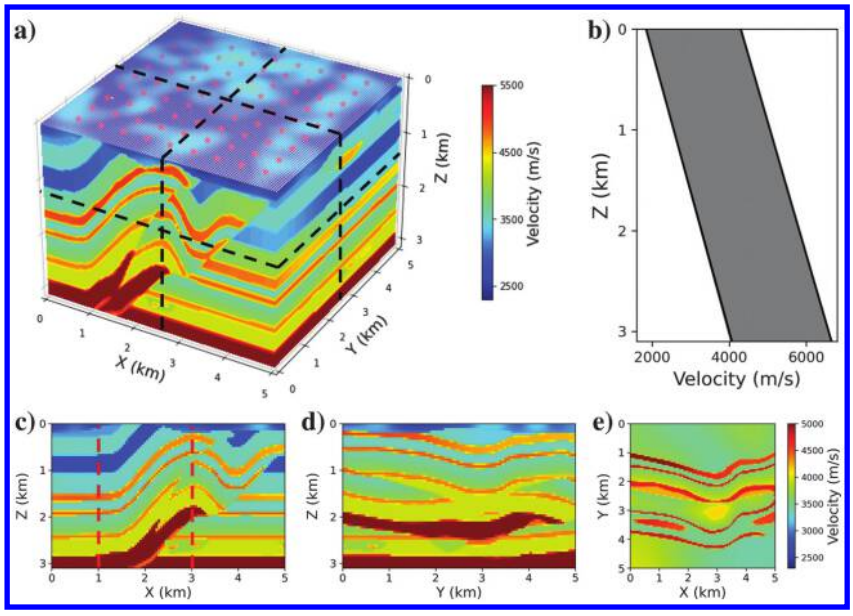

Figure 1. (a) The true velocity model. The source and receiver locations are indicated by the red stars and white dots at the surface (the latter are so close together that they may be observed as a pale haze). (b) The upper and lower bounds of a uniform prior distribution over seismic velocity at each depth. (c), (d), and (e) show two vertical and one horizontal 2D slices of the true velocity model at Y = 2.5 km, X = 2.5 km, and Z = 1.25 km, respectively. These three sections are marked by the dashed black lines in (a). The dashed red lines in (c) display the locations of two vertical profiles along which posterior marginals are compared in the main text.

Comparison with previous results

We apply PSVI to the preceding 3D FWI problem and compare the inversion results with those obtained from three other methods presented in Zhang et al. (2023): mean field automatic differentiation variational inference (mean field ADVI) (Kucukelbir et al., 2017), Stein variational gradient descent (SVGD) (Liu and Wang, 2016), and stochastic SVGD (sSVGD) (Gallego and Insua, 2018). The inversion settings in this example are identical to those used in Zhang et al. (2023), allowing us to compare the four sets of results directly. We set the initial value of PSVI to be a standard Gaussian distribution, as suggested by Kucukelbir et al. (2017). This results in a laterally homogeneous initial mean velocity model, with velocity values at different depths being the mean of the uniform prior distribution shown in Figure 1b.

Figure 2 shows the inversion results of the vertical slice shown in Figure 1c, obtained using different methods indicated in the title of each panel. From top to bottom, each row shows the posterior mean velocity, standard deviation, and relative error maps, respectively, where the relative error is the absolute difference between the true and mean velocities divided by the standard deviation at each point. Each mean velocity map recovers the main features of the true velocity model. However, we also observe that some structures are inverted incorrectly, such as those inside the dashed black boxes in Figure 2, which are supposed to be horizontal according to the true velocity model displayed in Figure 1c. This is possibly because all four inversions fail to recover the correct low-wavenumber components of the true model due to 2π phase shifts in forward-modeled waveform data that fit the observed data equally well — a phenomenon often referred to as cycle skipping in FWI. Mean field ADVI provides the lowest posterior uncertainties due to its theoretical assumption of an uncorrelated Gaussian posterior PDF. Uncertainty values from SVGD and PSVI are larger than those from mean field ADVI, but smaller than those from sSVGD. Most of the relative errors in the four plots in the bottom row remain relatively high because they all underestimate

the posterior uncertainties to some extent for such a high-dimensional inverse problem.

To justify the preceding statements and to better compare the four sets of results, in Figure 3, we compare the posterior marginal distributions obtained from the four methods along two vertical profiles at horizontal locations of 1 km (top row) and 3 km (bottom row), respectively. The locations of these two profiles are displayed by dashed red lines in Figure 1c. In Figure 3, red lines show the true velocity profile. Overall, the marginal distributions in Figure 3c and 3d shows similar features and are broader than those in Figure 3a and 3b, indicating that PSVI and sSVGD provide better posterior uncertainty estimates, on average, compared with the other two methods.

However, the preceding four results, especially those obtained using PSVI and sSVGD, show some consistent but incorrect patterns. First, the posterior marginals in the first row in Figure 3 provide biased phase information in which the true velocity profile is excluded from the posterior PDFs, as marked by white arrows. PSVI uncertainties are significantly narrower in that section of the profile, accentuating the error, presumably because the algorithm is stuck at a local minimum of the variational optimization problem that happens to have low values for the velocity standard deviations. This section corresponds to the results highlighted by the dashed black boxes in Figure 2, due to the incorrectly inverted low-wavenumber component (cycle skipping). Second, in Figure 3, we observe roughly the same magnitude of posterior uncertainties (i.e., similar width of those marginal PDFs) in the shallower and deeper parts of the model. Given that the sensitivity of the surface seismic data should often be higher in the near surface than in deeper parts of the Earth, we would expect to observe higher uncertainties at depth. This is caused by additional prior information (spatial smoothness) imposed on the inversion.

In the following section, we solve these two issues to improve the inversion results.

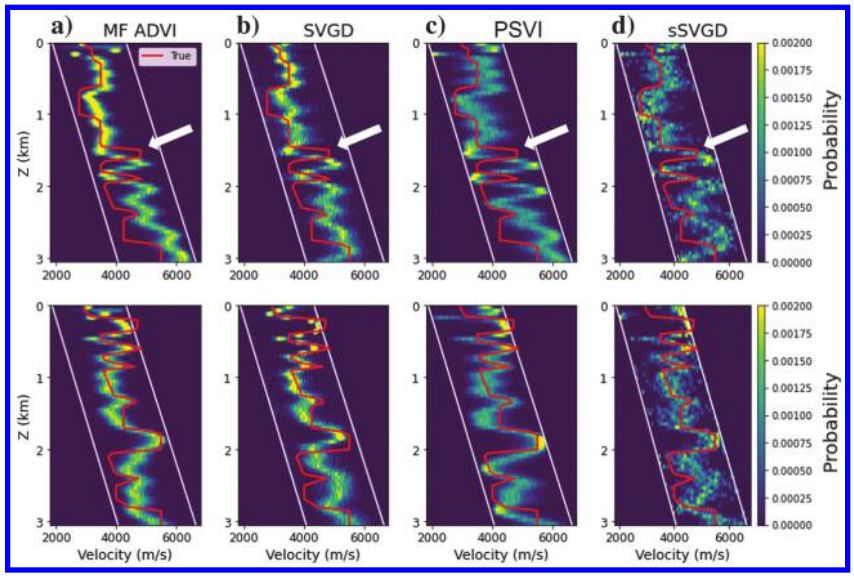

Figure 3. Posterior marginal PDFs obtained using different methods, extracted at two locations marked by the dashed red lines in Figure 1c. The red lines show the true velocity profile. The white arrows highlight the posterior marginal PDFs that fail to find the correct solution, due to cycle skipping.

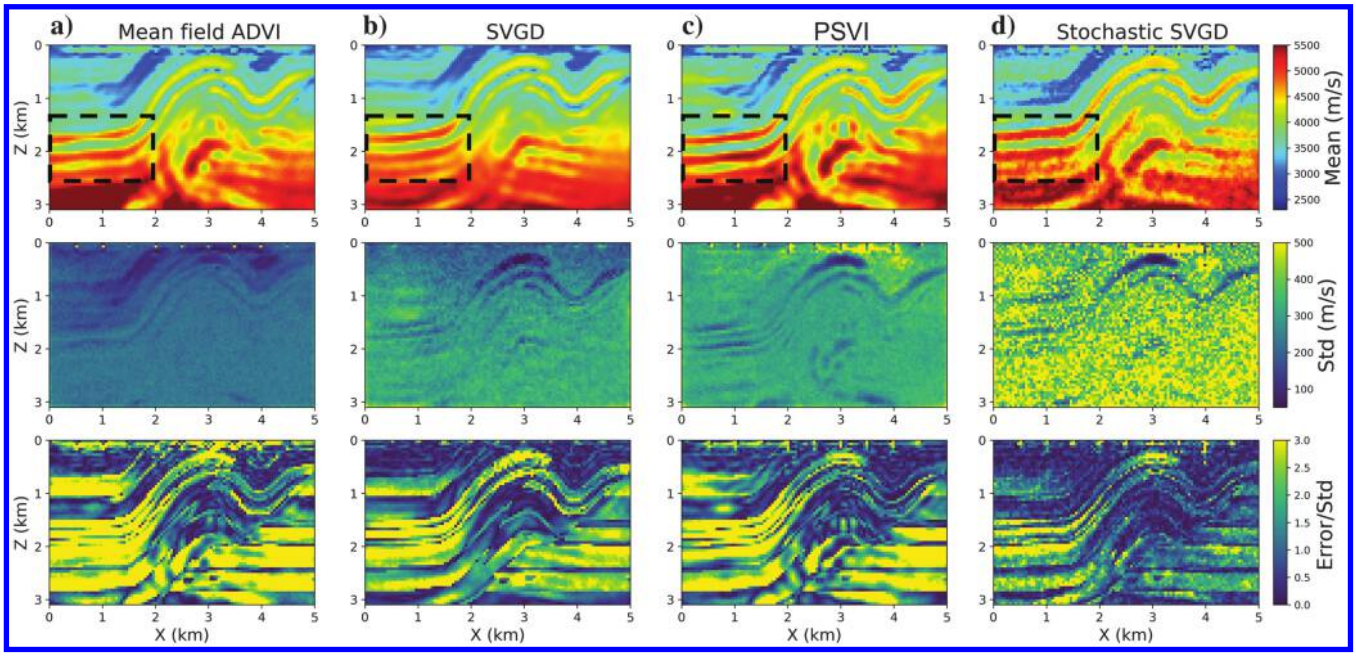

Figure 2. Single frequency band inversion results obtained using four different methods, at the vertical section Y = 2.5 km. The true velocity model is shown in Figure 1c. From top to bottom, each row shows the posterior mean velocity, standard deviation (std), and the relative error maps, respectively. The relative error is the absolute difference between the true and mean velocities (the error) divided by the standard deviation at each point. The dashed black boxes highlight structures that are not correctly inverted by any method due to cycle skipping.

DOI:10.1190/geo2024-0774.1

Improved inversion results

In this section, we show how to improve the 3D FWI results. To avoid cycle skipping, similarly to the common strategy in linearized FWI problems, we use a multiscale inversion approach by inverting waveform data from low frequencies to high frequencies (Bunks et al., 1995). We consider three frequency bands referred to as the low- (1–6 Hz), intermediate- (1–10 Hz), and high-frequency band (1–13 Hz). Variational optimizations using data in intermediate and high-frequency bands used results from the previous band as a starting point. The starting point for the low-frequency band variational inversion is set to be a standard Gaussian distribution within an unconstrained space, as shown in equation 3 and suggested by Kucukelbir et al. (2017). In preparation for subsequent sections, we refer to the final solution using high-frequency data as $q_{\text{old}}(\mathbf{m})$.

Any regularization (a form of prior information) is somewhat arbitrary, as we will never know whether the imposed prior information is consistent with the true Earth or not a priori, and thus whether the regularization will spuriously bias the inversion results. We therefore remove all regularizations (note that spatial smoothing is imposed in the previous tests) and use the uniform PDF displayed in Figure 1b as prior information for FWI. As long as the posterior solution is not unduly constrained by the boundaries of the uniform distribution, we assume that the inversion is mainly constrained by the likelihood function, and hence, by observed waveform data. Note that this will increase the effective support of the posterior PDF (the hyper-volume over which it is significantly greater than zero), thereby increasing the complexity of this FWI problem, compared with more strongly regularized problems (e.g., with tighter uniform bounds). Due to limited computational resources, we only perform PSVI in this test. After obtaining reasonable inversion results using this noninformative, broad prior distribution, we efficiently analyze different prior assumptions in the next section using the VPR methodology.

Figure 4 shows the final inversion results obtained using high-frequency band data; interim results using low and intermediate frequency bands are presented in Appendix A. The top two rows display two vertical sections at Y = 2.5 km and X = 2.5 km. The bottom row illustrates one horizontal section at a depth Z = 1.25 km. The three sections are marked by dashed black lines in Figure 1a. From left to right, Figure 4a–4d shows the three true velocity sections, the average velocity, the standard deviation, and the relative error maps of the posterior probability distribution, where the relative error is the absolute difference between the true and mean velocities divided by the standard deviation at each point.

The three mean sections resemble the true velocity structures (note that even for an entirely correct solution, this need not be the case in probabilistic inversions because the mean of the models is a statistic, not a model in itself). This is especially true in the shallow subsurface above 2 km, where high-resolution structures are accurately inverted. However, the mean and true values diverge at deeper levels, possibly due to reduced data sensitivity, as the standard deviations broadly increase with depth. The overall relative errors are less than three, indicating that the differences between the true and inverted mean models are within three standard deviations, as would be expected of the true solution to this Bayesian problem. Note that in Figure 4d, higher relative errors (shown in yellow) are present near the surface because, in fact, the true model deliberately lies outside of the prior bounds shown in Figure 1b, to assess the method's behavior in such circumstances (Zhang et al., 2023). Overall, the inverted images and

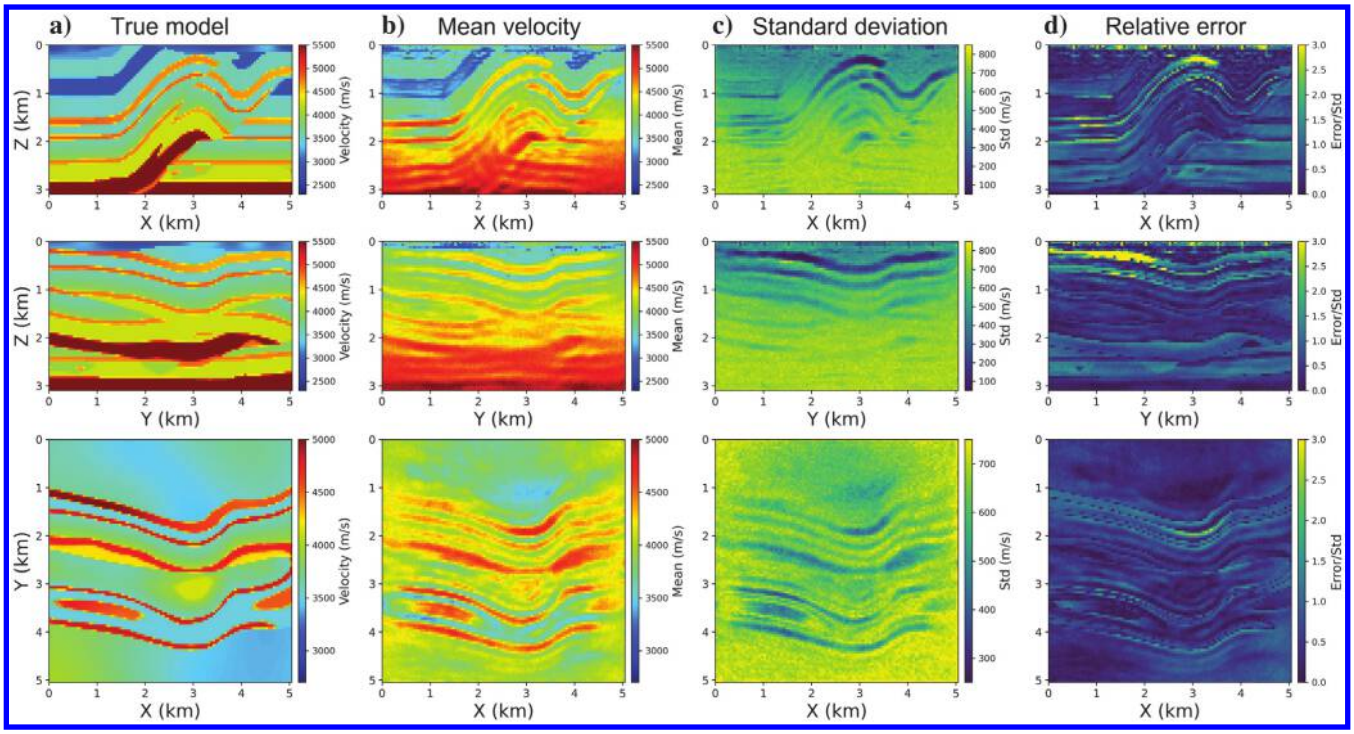

Figure 4. High frequency 3D variational Bayesian FWI results obtained using PSVI and the uniform prior distribution displayed in Figure 1b. The top two rows show vertical sections at Y = 2.5 km and X = 2.5 km, and the bottom row shows a horizontal section at Z = 1.25 km. (a)–(d) The true velocity model, the mean, the standard deviation, and the relative error (calculated by the absolute difference between the true and mean velocities divided by the standard deviation at each point) maps of the posterior distribution, respectively.

DOI:10.1190/geo2024-0774.1

the posterior uncertainties in Figure 4 are significantly more accurate than those displayed in Figure 2.

Testing different prior hypotheses

In addition to the uniform prior distribution, we now consider a set of smoothed prior distributions. Define a second-order finite-difference operator **S** to calculate the curvature of the model parameter values between adjacent grid cells in a spatial 3D velocity model represented by a model vector **m** (Zhao and Curtis, 2024c). This is further applied to **m** to impose spatial smoothness, by imposing a Gaussian distribution on the product **Sm**

$$p(\mathbf{Sm}) \propto \exp\left(-\frac{(\mathbf{Sm})^T(\mathbf{Sm})}{2\sigma_s^2}\right), \tag{6}$$

where σ_s is a hyperparameter that controls the strength of the spatial smoothness assumed (a smaller σ_s causes the probability of any particular rough model to become lower). Equation 6 can be interpreted as applying a Tikhonov matrix **S** to **m** (Golub et al., 1999). Then, the smoothed model prior PDF can be expressed as

$$p_{\text{smooth}}(\mathbf{m}) \propto p(\mathbf{Sm})p_{\text{old}}(\mathbf{m}), \tag{7}$$

where $p_{\text{old}}(\mathbf{m})$ is the (old) uniform prior distribution defined in the previous section. We consider seven smoothed prior PDFs with different σ_s values: 2000, 1000, 500, 250, 125, 62.5, and 31.25, respectively. In the subsequent part of this paper, we also denote the uniform prior distribution as $\sigma_s = \infty$, indicating no smoothing. These eight prior PDFs are used to mimic a real situation in which we have different prior hypotheses and wish to analyze their consequences.

The posterior PDFs, using each of the seven smoothed priors, are obtained by replacing the uniform prior PDF in the inversion results $p_{\text{old}}(\mathbf{m}|\mathbf{d}_{\text{obs}})$ from the previous section (Figure 4) by the respective smoothed prior PDFs, using VPR. Thus, no additional inversion is performed to obtain these results. In Appendix B, we verify the results of VPR by comparing them to those obtained by performing independent Bayesian inversions (the verification of VPR in 2D FWI was also conducted by Zhao and Curtis [2024c]). Figure 5 shows the results using all eight prior PDFs on the vertical section at Y = 2.5 km. From left to right, each column shows one posterior sample, the mean, the standard deviation, and the relative error maps of the posterior distribution, respectively. From top to bottom, each row presents the posterior PDF obtained using the σ_s value indicated on the left of the figure.

From Figure 5a to 5h, as the magnitude of smoothness increases, stronger prior information favoring spatially smooth structures is injected into the inversion results. Consequently, both the posterior samples and the mean velocity maps become smoother with fewer spatial variations. The standard deviations decrease (note that different colorbar scales are used in the standard deviation maps in the third column); this leads to increased relative errors, as (true) large velocity contrasts between neighboring cells are precluded by the prior information for small σ_s.

Figure 6 shows the posterior marginal PDFs of two vertical velocity profiles at locations X = 1 km (the top row of Figure 6) and X = 3 km (the bottom row), marked by dashed red lines in Figure 1c. Figure 6a shows the posterior marginal PDFs obtained from the uniform prior distribution, and Figure 6b–6h shows those from the smoothed prior PDFs — the smoothness values σ_s are noted in the title. In each figure, the red line shows the true velocity profile, and the black line displays the inverted mean velocity profile. We observe that with the increase in the magnitude of smoothness injected by the prior information, the posterior marginal PDFs become smoother, with fewer spatial variations (and therefore, lower uncertainties), similar to the conclusions drawn earlier.

By comparing each set of results with the true velocity model in Figure 1c, we find that the posterior PDFs become more tightly constrained around the true values as we increase the magnitude of the smoothness up to $\sigma_s = 500$ in Figure 5d, especially in the deeper parts of the model. Interestingly, in the third column of Figure 5c and 5d, we observe higher uncertainties at the boundaries between strata below a depth of 1.5 km, capturing the expected "uncertainty-loop" characteristic of the posterior uncertainties proposed by Galetti et al. (2015). However, Figure 5 shows that uncertainty loops appear because smoothness is applied to models and do not appear when the models are allowed to be rough. This is consistent with the results of Galetti et al. (2015), which were obtained using a trans-dimensional Monte Carlo travel-time tomography method that imposed strict smoothness (indeed, constant velocity) over large areas of the model. We, therefore, qualify the proposition of Galetti et al. (2015) by hypothesizing that posterior uncertainty loops are, in fact, a product of the smoothness constraints in tomographic problems (including FWI).

The inversion results become worse in Figure 5e–5h, as the effect of prior information overpowers the information in the observed data, resulting in biased posterior solutions and overly smoothed structures. Not surprisingly, most of the relative errors are then greater than three. We conclude that the smoothed prior distributions with $\sigma_s = 1000$ or $\sigma_s = 500$ (Figure 5c and 5d) are relatively good choices for this FWI problem.

However, in reality, we never know the true velocity structure, so we cannot calculate or use relative errors to choose between prior distributions, except in synthetic tests, such as those shown in Figures 5 and 6. In deterministic FWI, an order of preference is often constructed by calculating the data misfit values arising from the different inversion results (referred to as an L-curve). Following a similar approach, we compare the eight posterior distributions by drawing $N = 100$ samples from each posterior PDF and performing forward simulations to calculate the corresponding misfit values between the observed and synthetic data. This results in eight approximate probability distributions of misfit values from the inversion results using different priors, shown in Figure 7a. As expected, the overall misfit values increase as more smoothness is injected into the inversion.

The mean and standard deviation of each distribution are shown in Figure 7b by black dots and red error bars, connected by a red line, which we call a *Bayesian L-curve*. Moving from left to right along the curve, the magnitude of prior information (relative smoothness in this case) decreases from $\sigma_s = 31.25$ (the top-left point) to $\sigma_s = \infty$ (no smoothing — the bottom-right point). An arrow marks the case with $\sigma_s = 1000$, and we observe that the data misfits start to decrease significantly if we further increase the roughness value σ_s of the prior information beyond this point, implying that this value provides a balance between information from the smoothing distribution (used to narrow the possible parameter space) and the observed data (used to achieve good data fits).

From a Bayesian point of view, if we do not know the smoothness of the true Earth, then we have little reason to choose any one of

these posterior distributions — they are all part of the posterior uncertainty. We may have sufficient prior intuition to suspect that posterior models are too rough in Figure 5a and too smooth in Figure 5h, but it is difficult to establish a probability distribution over levels of smoothness *a priori*. Hierarchical Bayesian methods might be used to estimate this distribution *a posteriori*. However,

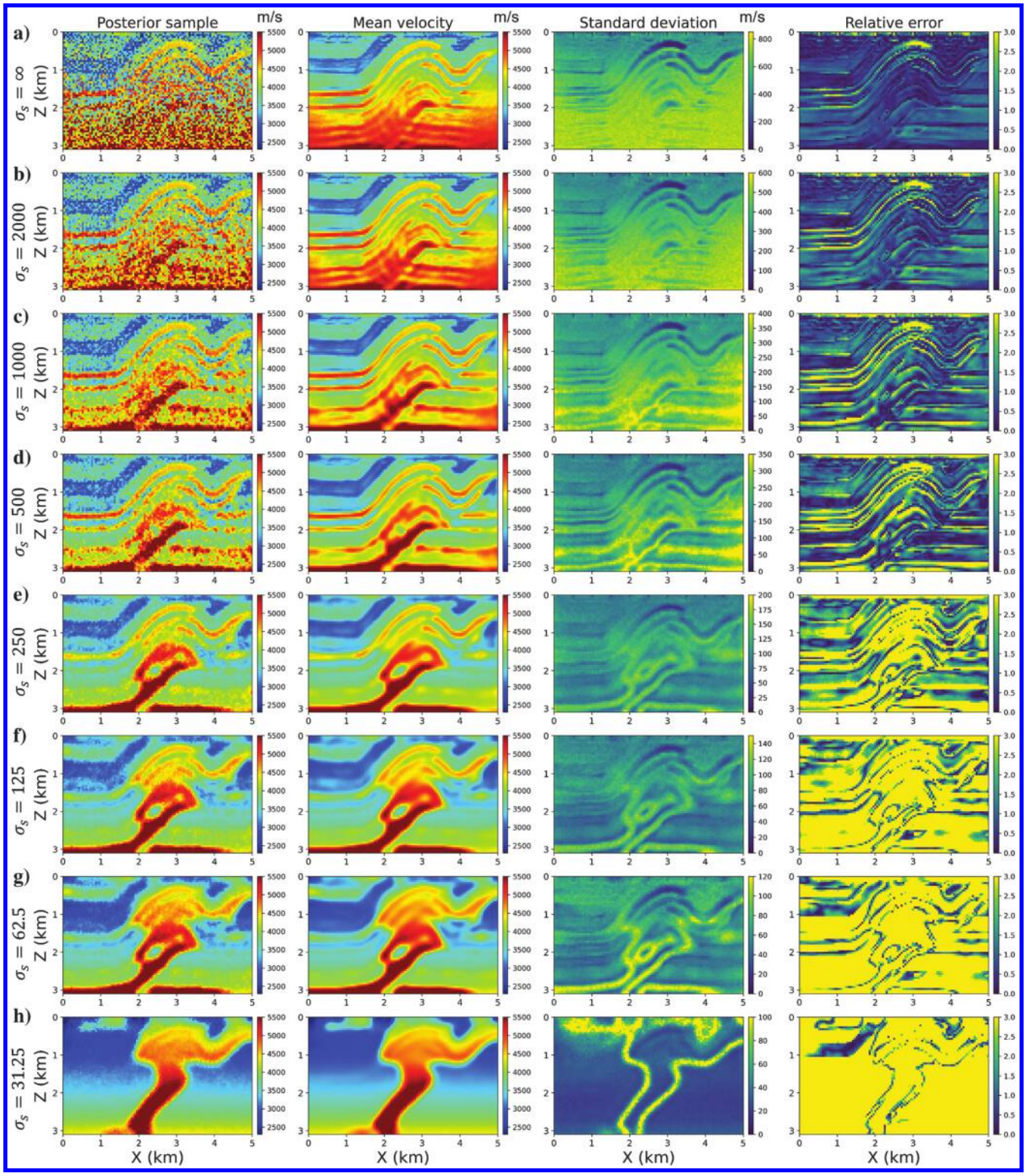

Figure 5. Inversion results obtained using (a) the uniform prior PDF and (b)–(h) the seven smoothed priors with decreasing values of σ_s (an equivalently increasing smoothness), in a vertical section Y = 2.5 km. From left to right, they represent one posterior sample, the mean velocity, the standard deviation, and the relative error maps.

DOI:10.1190/geo2024-0774.1

such an approach introduces physical inconsistency into Bayesian inversion results (Mosegaard and Curtis, 2024). Alternatively, in a departure from Bayesian methods, we might decide to consider only a subset of these various results, for example, those around $\sigma_s = 1000$.

More generally, many different types of prior hypotheses can be represented by a hypothesis space $\mathcal{H}$, as schematically represented in Figure 7c. Given a rather uninformative uniform prior distribution in the old posterior PDF, we can, in principle, inject any additional prior assumptions into the inversion; each would have different effects on the inversion results, between which we may wish to discriminate. For example, we consider another prior assumption from real-geology. The geology-informed prior information is obtained by subsampling a set of realistic geologic images using a predefined window size. These subimages are used to calculate a local correlation matrix between the parameter pairs that lie within that window size, thereby defining prior correlations between pairs of model parameters used in this test (Zhao and Curtis, 2024c). We consider five such geologic prior distributions constructed using different local cubic windows with edge lengths of two, four, six, eight, and 10 cells, respectively. Again, the corresponding posterior distributions are calculated using VPR by replacing the old uniform prior distribution with each geologic prior distribution in turn.

Figure 8 shows the corresponding posterior PDFs, where the sizes of the correlation windows used are denoted on the left side of Figure 8. Figure 8a shows the results obtained using the uniform prior distribution. We observe that larger prior correlation windows generally yield better inversion results that more closely match the true velocity model (Figure 1c), as a larger window can provide more accurate prior correlation information, which improves the inversion results. However, note that a larger window also has higher memory requirements.

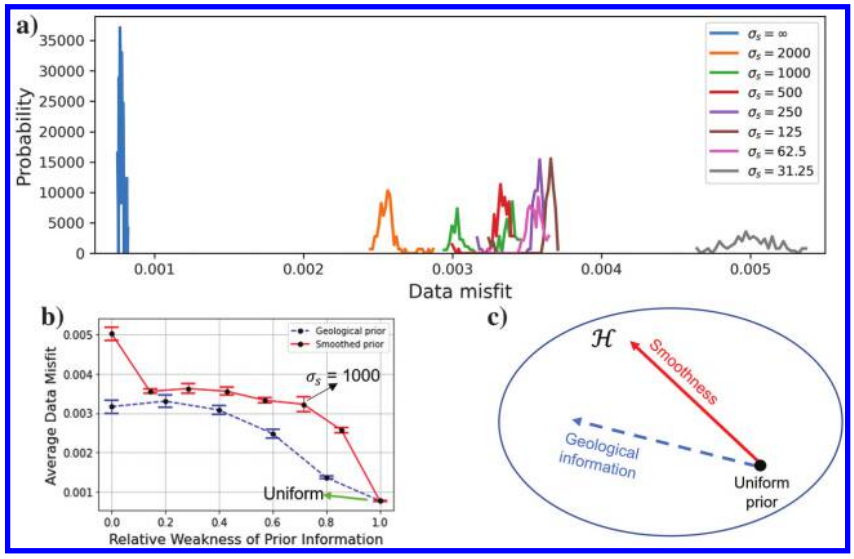

Figure 7. (a) Probability density functions for data misfit values calculated from posterior PDFs displayed in Figure 5 using differently smoothed priors. (b) Bayesian L-curves representing the average data misfits and their standard deviations (error bars) with respect to different prior hypotheses. The horizontal axis represents the normalized (relative) weakness of prior information applied by each prior distribution; a value of 1 represents the noninformative, uniform prior PDF represented in Figure 1b. The red line denotes the L-curve found by requiring smoothness in the prior PDFs, and the dashed blue line denotes the equivalent curve when requiring consistency with geologic prior information. (c) A schematic diagram of the space $\mathcal{H}$ containing all possible prior hypotheses. The black dot represents a noninformative uniform prior distribution, and we denote different prior assumptions, such as smoothed and geologic prior PDFs, by different paths (arrows) through $\mathcal{H}$ in this study.

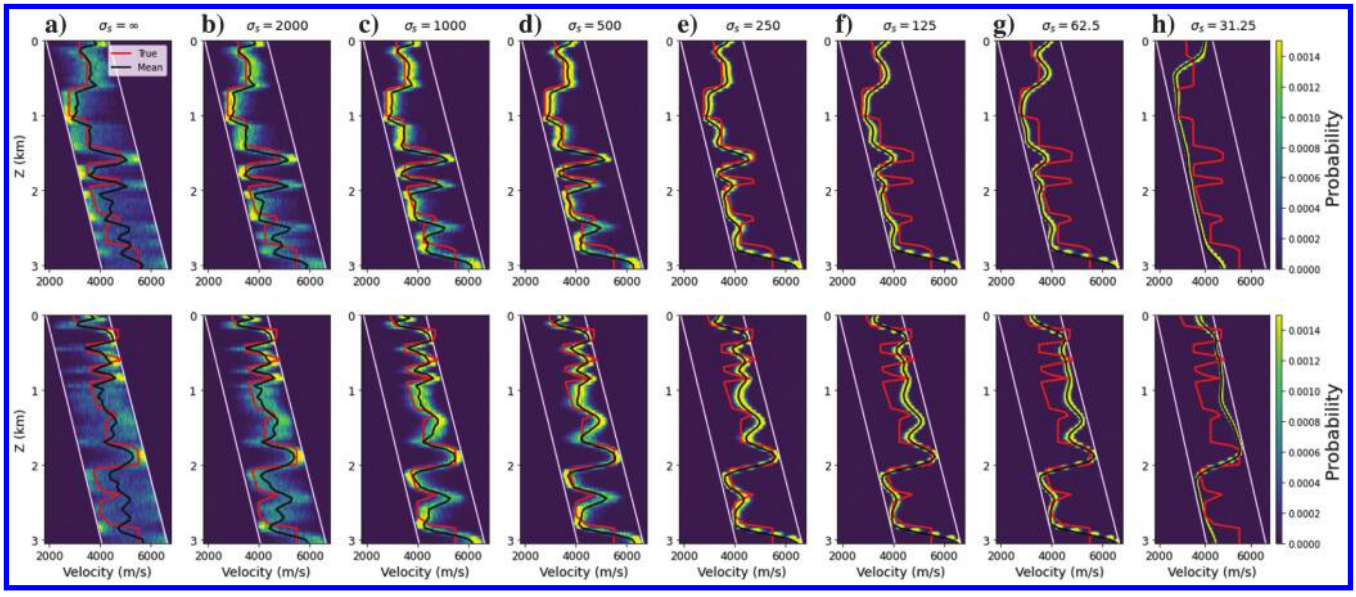

Figure 6. Posterior marginal PDFs at two vertical locations marked by the dashed red lines in Figure 1c. (a) Results obtained using the uniform prior PDF, and (b)–(h) results using the smoothed prior PDFs with different σ_s values.

These inversion results are then used to construct a Bayesian L-curve, as displayed by the dashed blue line in Figure 7b. From left to right along this dashed blue line, the correlation window size decreases from $10 \times 10 \times 10$ (the top-left point) to $0 \times 0 \times 0$ (the bottom-right point), with the latter being the uniform prior PDF with no prior correlation. This indicates that the strength of geologic prior information decreases from left to right. The data misfits are lower than those from smoothing priors in red because geologically informed priors impose less smoothness vertically than horizontally (which is correct in the true model and often in the Earth).

Computational cost

The inversion using the uniform prior distribution (Figure 4) is performed within three frequency bands, each involving 200, 300, and 800 iterations for variational inference (optimization) in the low, intermediate, and high-frequency bands, respectively. In each iteration, four random samples are used to estimate expectations in the EBLO$[q(\mathbf{m})]$ in equation 2. Therefore, a total of 5200 samples, including 5200 forward and adjoint simulations, are used to obtain the inversion results displayed in Figure 4. To further reduce the computational cost, we use a minibatch of 36 shot gathers, randomly selected from a total of 81 shots in each FWI simulation (Zhang et al., 2023). For comparison, the same test conducted in Zhang et al. (2023) was obtained using 4000, 400,000, and 80,000 samples (forward and adjoint simulations) when using automatic differentiation variational inference ADVI (ADVI) (Kucukelbir et al., 2017), SVGD (Liu and Wang, 2016), and stochastic SVGD (sSVGD) (Gallego and Insua, 2018), respectively. SVGD and sSVGD are far more expensive than PSVI; ADVI is computationally cheaper but provides biased inversion results with strongly underestimated posterior uncertainties (Zhang et al., 2023).

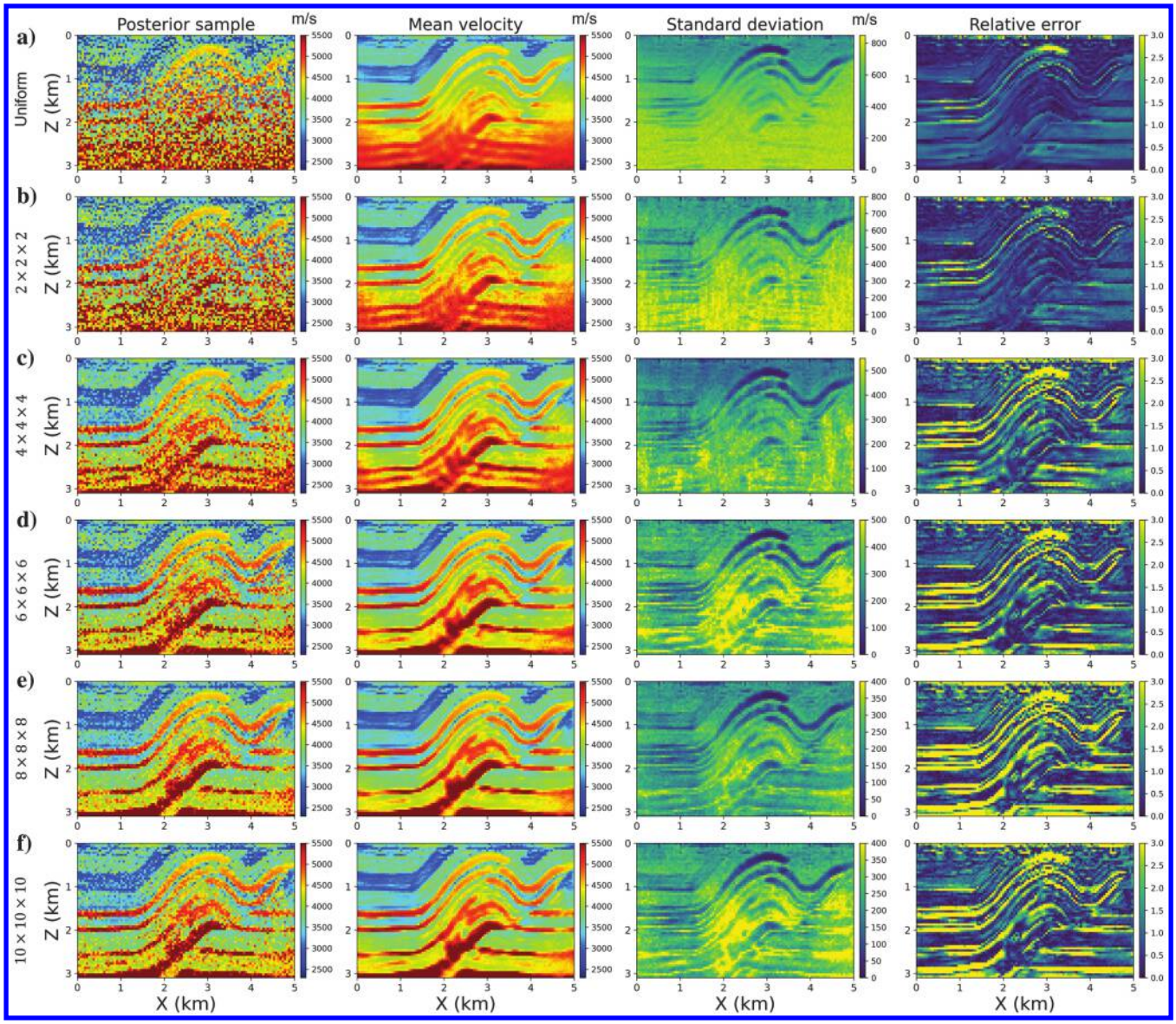

Figure 8. Inversion results on a vertical section Y = 2.5 km obtained using (a) the uniform prior PDF and (b)–(f) the geologic prior PDFs with different sizes of local correlation windows, denoted on the left side of each row. From left to right, they represent one posterior sample, the mean velocity, the standard deviation, and the relative error maps.

DOI:10.1190/geo2024-0774.1

Note that in linearized, deterministic FWI, one typically performs ~100 iterations to obtain a reasonable result, and a full batch of 81 shots is normally used at each iteration to provide accurate gradient information. Because forward simulation using 81 shot gathers would require 2.25 times more computation than using a minibatch of 36 shots, the computational cost of 100 iterations implemented in deterministic FWI is equivalent to performing 225 forward evaluations in this study.

With PSVI, we therefore obtain reasonable probabilistic uncertainty estimates using only an order of magnitude more computation than deterministic FWI, from which no robust uncertainty information can be obtained. Compared with three to four orders of magnitude more computation that would be expected using conventional Monte Carlo sampling methods, the reduction in computational cost is already significant. In addition, this single set of variational inversion results allows multiple posterior distributions corresponding to different prior PDFs to be produced using VPR, without any further FWI simulations, even in three dimensions. For example, imagine that we wish to test 10 different prior hypotheses. The computational cost using our method remains roughly the same as that reported previously. However, for deterministic FWI methods, we need to solve 10 inversion problems. In this case, our method employs the same order of computation to solve the inversion problem and provide full uncertainty estimates, compared to deterministic FWI.

DISCUSSION

Within a Bayesian framework, we solve FWI (or any other inverse/inference problems) probabilistically — i.e., all information is represented by probability distributions, and both unknown parameters and observations are represented by random variables. In conventional deterministic FWI, unknowns are defined as a model vector to be optimized during inversion. In practice, optimization begins with an initial model obtained from, for example, migration velocity analysis, travel time tomography, or surface wave inversion. This initial model must be sufficiently accurate so that the iterative optimization converges toward the correct solution, which must, therefore, be close to the true model. In our implementation of PSVI (variational FWI), we also need a starting point (initial value) for optimizing the variational distribution; however, this starting point is a probability distribution rather than a velocity model (mathematically, a vector). For example, in our test, we used a standard Gaussian distribution (in the unconstrained space) as the starting point of PSVI. Future work might consider starting with a Gaussian distribution that has a mean vector defined by results from migration velocity analysis, travel time tomography, or surface wave inversion, which will speed up the variational optimization.

We also need to note that in (variational) Bayesian inversion, the initial distribution is different from prior information (hypotheses). Given the complexity of 3D FWI, we still need to test different prior hypotheses and select the optimal one to constrain the inversion best, even when we have a good initial distribution. This is equivalent to the situation in deterministic FWI, in which a good initial velocity model cannot replace regularization terms. No matter what initial models we use, we still need additional regularization terms to stabilize the inversion process.

All of the posterior PDFs with different priors were obtained from only a single Bayesian 3D FWI using a uniform prior distribution. However, to translate uncertainties from the model parameter space into the data space, hundreds of additional forward simulations were performed to obtain the data misfit values, which were used to calculate the Bayesian L-curves displayed in Figure 7. Alternatively, boosting variational inference implicitly provides a small number (tens) of representative samples that represent major components of the uncertainties (Zhao and Curtis, 2024a). These representative samples might be used to construct Bayesian L-curves at a significantly reduced computational cost compared to the approach used here.

A common way to discriminate between prior hypotheses in Bayesian inference is to calculate the evidence term

$$p(\mathbf{d}_{\text{obs}}) = \int_{\mathbf{m}} p(\mathbf{d}_{\text{obs}}|\mathbf{m})p(\mathbf{m})d\mathbf{m} \tag{8}$$

for each hypothesis, which requires an integral over the entire model parameter space to be evaluated. Although this is computationally possible for some low-dimensional problems (Strutz and Curtis, 2024), it is infeasible for Bayesian 3D FWI due to the high dimensionality of the problem. Hence, the approach used here is the first to allow such models and prior hypotheses to be compared in a probabilistic manner and with manageable computational cost in three dimensions.

Although we used prior distributions with varying degrees of smoothness to simulate different prior hypotheses, more realistic prior distributions that encode geologic information can be used in Bayesian inversion (Mosser et al., 2020; Levy et al., 2022; Bloem et al., 2024; Sun and Williamson, 2024). By using the VPR framework, we can calculate the corresponding posterior distributions and analyze different prior hypotheses at a relatively low cost.

Rather than approximating the true solution to the desired inverse problem directly, the VPR approximation of each new posterior PDF includes only information about the true solution that is already included within the old posterior PDF. It therefore carries over an imprint of any errors in the latter (Zhao and Curtis, 2024c). VPR results may therefore differ from independent Bayesian inversion results. This effect is potentially nonnegligible due to the higher dimensionality and complexity of 3D Bayesian FWI problems, as shown in Appendix B; qualitatively, the results are not as good as those obtained for 2D FWI in Zhao and Curtis (2024c). To improve the results, we can use a relatively lower cost calculation to refine (fine-tune) the outcomes obtained from VPR using observed data. We do this by invoking Bayes' rule again (performing another variational Bayesian inversion using PSVI), but with a smaller number of iterations that start from the VPR output because the VPR solution should already be reasonably close to the true posterior PDF. For example, we fine-tuned the preceding VPR results by performing variational inversion with an additional 200 iterations using four samples per iteration. This costs an extra 800 forward and adjoint simulations, and the results are shown in Figure 9. These are improved compared with the initial VPR results and are more similar to independent inversion results (more details can be found in Appendix B). If we wanted to reduce this additional computation further, we might again depart from formal Bayesian analysis by using the Bayesian L-curve obtained from VPR to select a smaller subset of prior hypotheses to consider, then fine-tune only those VPR posterior PDFs.

In our presented example, the waveform data we used have a lowest frequency of 1 Hz. In exploration seismology applications, it is often the case that we do not have high-quality data below 3–5 Hz.

DOI:10.1190/geo2024-0774.1

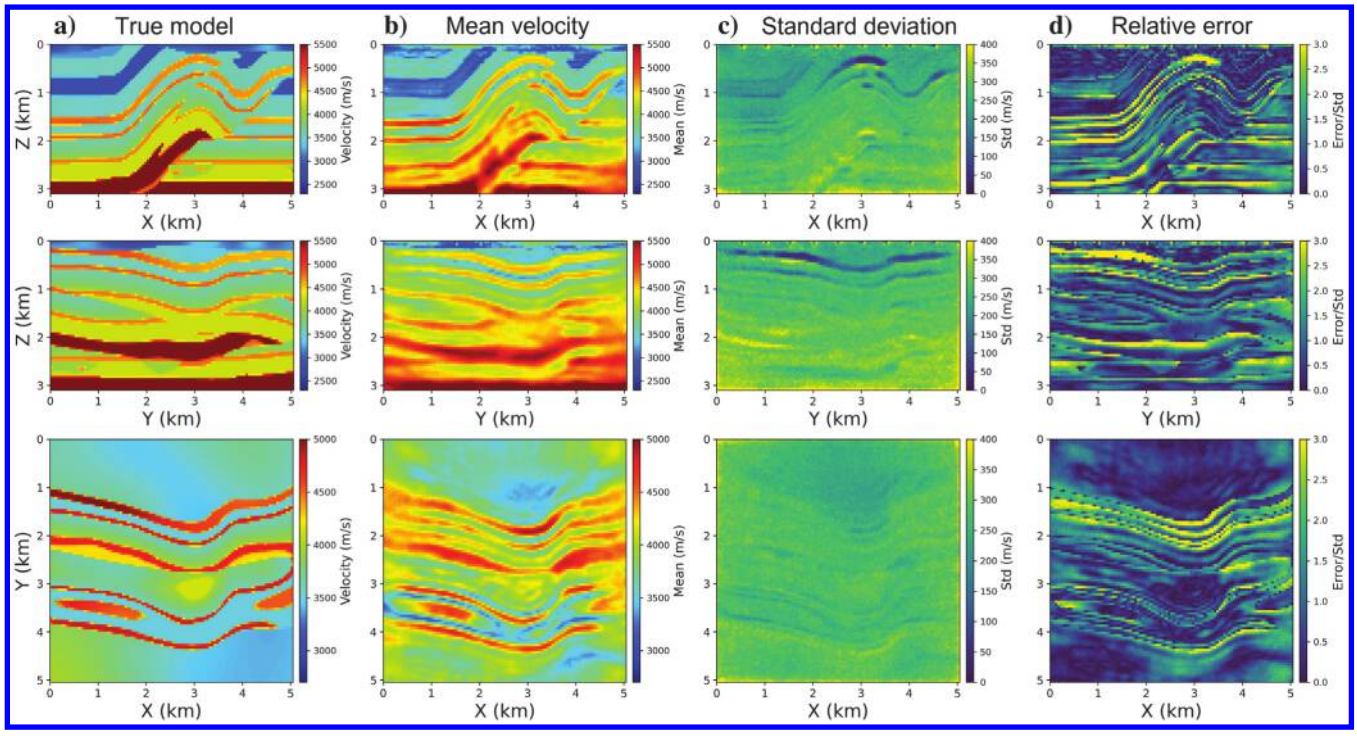

Figure 9. Inversion results after refining the VPR results, which are more similar to independent variational Bayesian inversion results (presented in Appendix B).

Recently, several methods have been developed to handle this issue and improve the performance of deterministic 3D FWI, including low-frequency waveform data extrapolation using high-frequency data (Li and Demanet, 2016; Fang et al., 2020). These techniques can also be applied to our variational Bayesian FWI framework to avoid problems such as cycle skipping.

CONCLUSION

We solve a 3D Bayesian FWI problem using PSVI, wherein a transformed Gaussian distribution is used to approximate the posterior distribution of the fully nonlinear inverse problem. Using a noninformative uniform prior distribution, we obtain a posterior solution with reasonable uncertainty estimates. The inversion progresses through low-, intermediate-, and high-frequency bands, with a total computational cost only an order of magnitude higher than that of deterministic FWI. In addition, we explore multiple different prior distributions and calculate the corresponding posterior distributions using VPR, without the need for additional FWI simulations. These results are used to analyze different prior hypotheses through the construction of Bayesian L-curves. We thus demonstrate that accurate Bayesian FWI and analysis, as well as the choice of prior hypotheses, are, in principle, now feasible in three dimensions.

ACKNOWLEDGMENTS

We thank the Edinburgh Imaging Project (EIP) (https://blogs.ed.ac.uk/imaging/) sponsors (BP and TotalEnergies) for supporting this research. We also thank X. Zhang at the China University of Geosciences in Beijing for providing insightful discussions on three variational inversion methods (ADVI, SVGD, and sSVGD) and for supplying the corresponding inversion results.

DATA AVAILABILITY

The code underlying this article is confidential and cannot be released.

DATA AND MATERIALS AVAILABILITY

Data associated with this research are confidential and cannot be released.

APPENDIX A

INVERSION RESULTS USING LOW AND INTERMEDIATE FREQUENCY WAVEFORM DATA

Figures A-1 and A-2, respectively, show the inversion results obtained using low- and intermediate-frequency waveform data, performed using PSVI (Zhao and Curtis, 2024b). Similar to the inversion results displayed in Figure 4 in the main text, the top two rows display two vertical sections at locations Y = 2.5 km and X = 2.5 km, whereas the bottom row illustrates one horizontal section at Z = 1.25 km. From left to right, each column shows three true velocity sections (marked by the dashed black lines in Figure 1a), the corresponding inverted average velocity, the standard deviation, and the relative error maps of the posterior distribution, wherein the relative error is the absolute difference between the true and mean values divided by the standard deviation at each point. Comparing these results with those using high-frequency data (Figure 4), we observe that as the frequency of the waveform data used for inversion increases, the spatial resolution improves, and the velocity estimates become more accurate. The overall relative errors are within three standard deviations, which is expected of the true posterior probability distribution.

DOI:10.1190/geo2024-0774.1

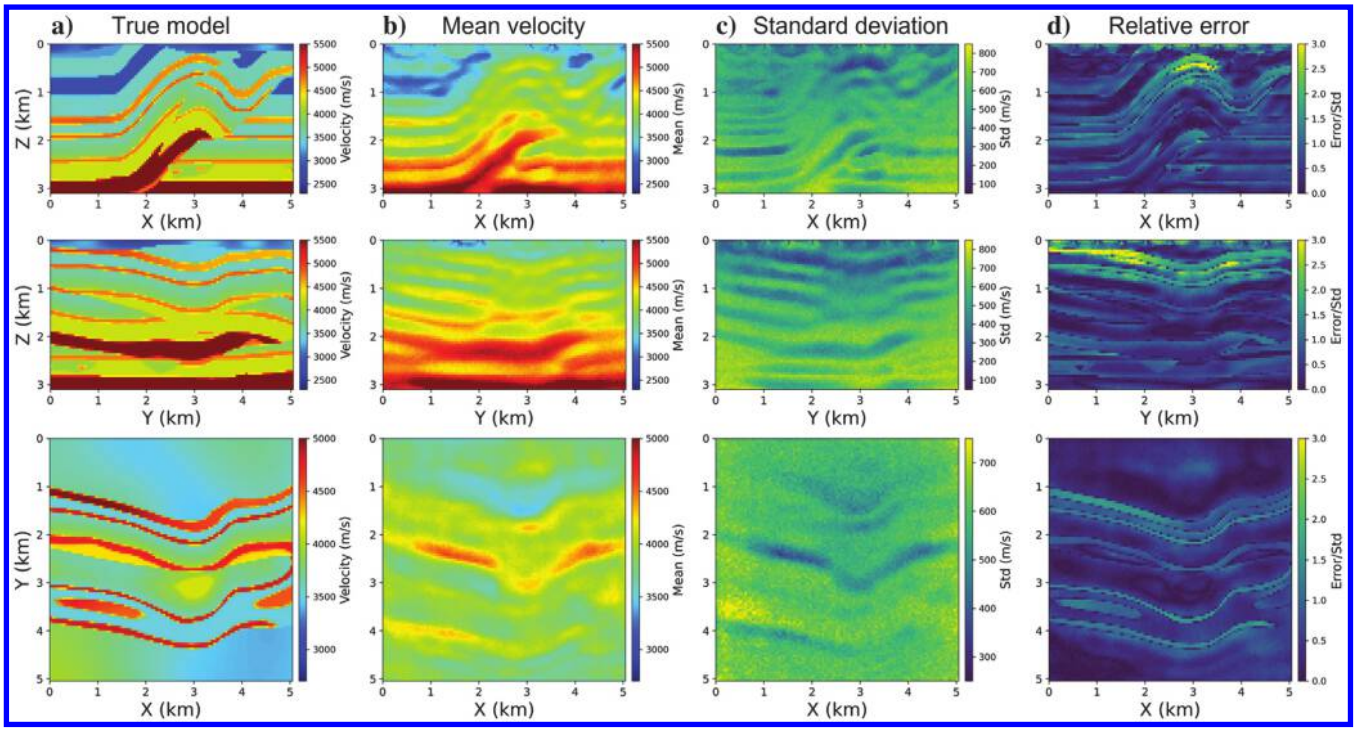

Figure A-1. Low frequency 3D variational Bayesian FWI results obtained using PSVI and a uniform prior distribution. The top two rows show two vertical sections at Y = 2.5 km and X = 2.5 km, whereas the bottom row shows a horizontal section at Z = 1.25 km. (a)–(d) The true velocity model, the mean, the standard deviation, and the relative error maps of the posterior distribution.

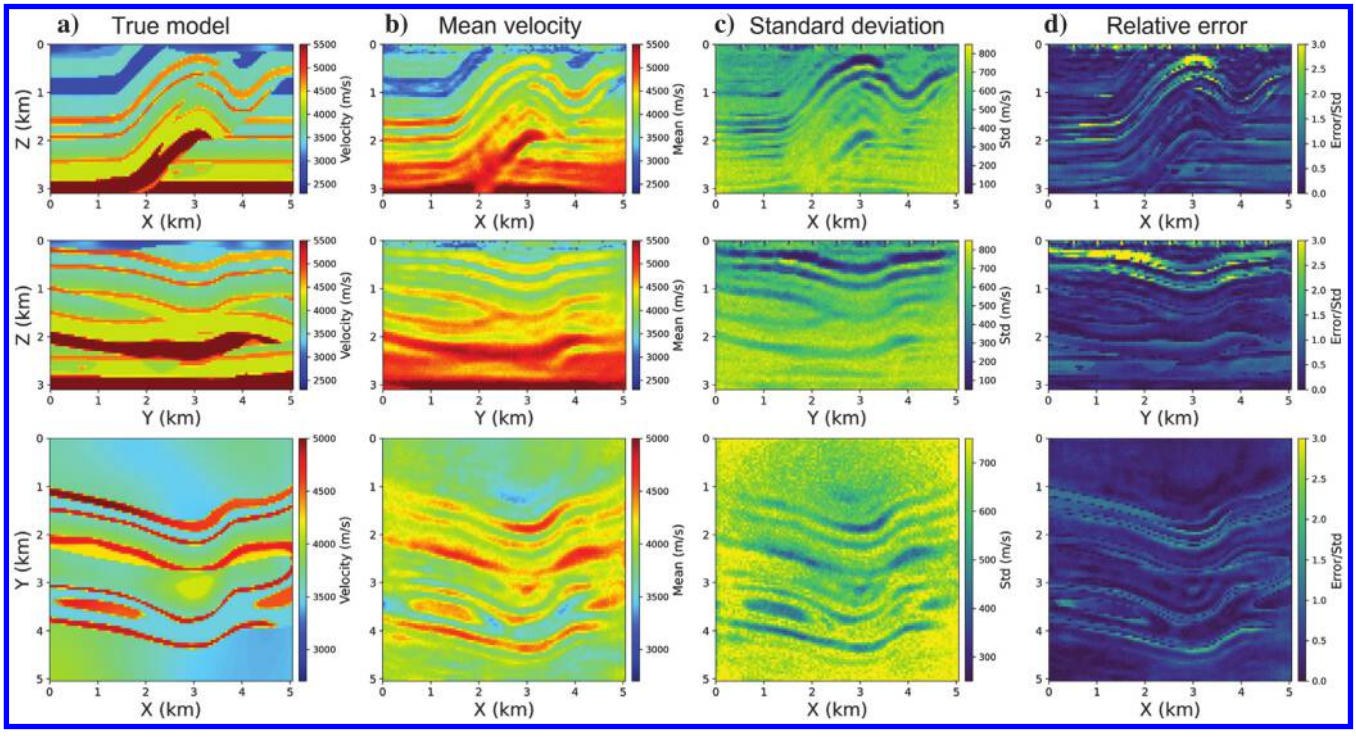

Figure A-2. The intermediate frequency 3D variational Bayesian FWI results.

APPENDIX B

THE VERIFICATION OF VARIATIONAL PRIOR REPLACEMENT

In this appendix, we test the effectiveness and accuracy of VPR (Zhao and Curtis, 2024c) for 3D Bayesian FWI by comparing the posterior solutions obtained using VPR to those obtained from independent Bayesian inversions. For VPR, we consider the uniform prior distribution defined in the main text as the old prior PDF, which is then removed from the old posterior distribution (displayed in Figure 4 in the main text) and replaced by a smoothed prior distribution with $\sigma_s = 1000$ (see details in the main text). The cor-

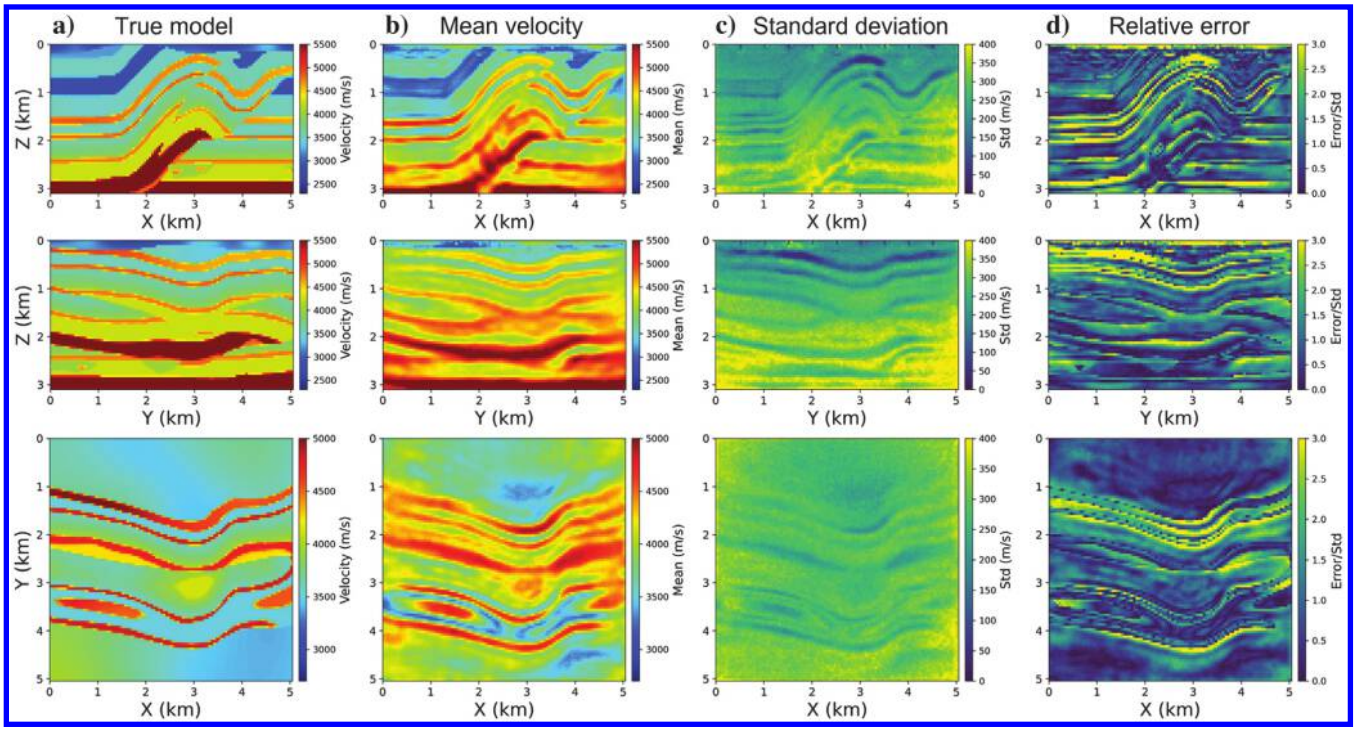

Figure B-1. The VPR results obtained by removing the uniform prior information from the old posterior PDF and imposing a smoothed prior PDF with a smoothness value of $\sigma_s = 1000$.

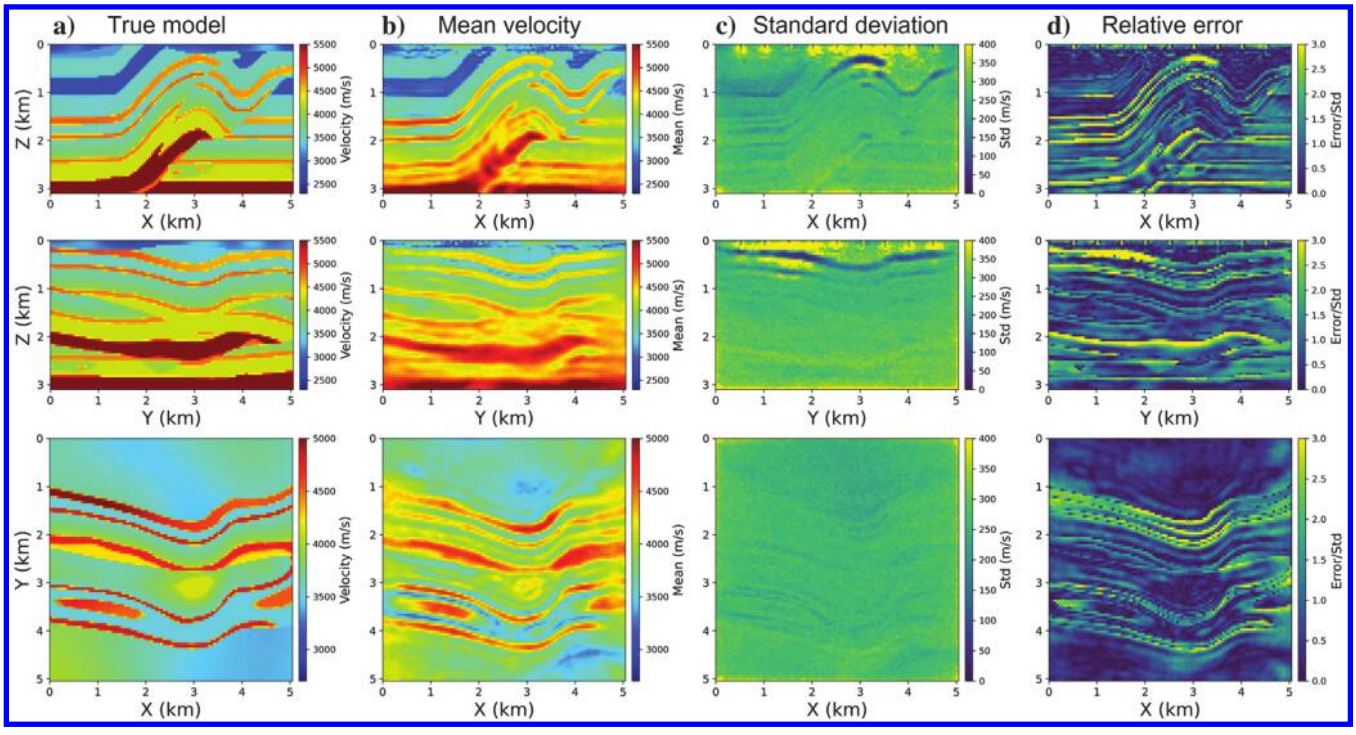

Figure B-2. Independent variational Bayesian inversion results obtained using the smoothed prior distribution with a smoothness value of $\sigma_s = 1000$. This is used to verify the VPR results displayed in Figure B-1.

responding results are shown in Figure B-1. We then perform an independent variational Bayesian inversion using the same smoothed prior distribution. Similar to the uniform prior case, the inversion is performed using the same three frequency bands of waveform data with the same number of forward simulations. Figure B-2 shows the results. The main features of these two results are roughly consistent, indicating that VPR can replace the old prior information with the new one in the inversion results without having to solve a Bayesian inverse problem from scratch repeatedly. Some small discrepancies between these two sets of results remain, especially in the deeper parts of the model. This is partly because VPR introduces a variational distribution $q_{\text{new}}(\mathbf{m})$ to approximate the new posterior PDF $p_{\text{new}}(\mathbf{m}|\mathbf{d}_{\text{obs}})$, rather than calculating the true new posterior distribution, as expressed in the fifth line of equation 5 in the main text (Walker and Curtis, 2014; Zhao and Curtis, 2024c). In addition, this also indicates that VPR might not be as accurate for this extremely high-dimensional and complex 3D Bayesian inverse problem, compared with previous tests in 2D (Zhao and Curtis, 2024c). Nevertheless, this result is obtained almost for free, compared with the huge computational cost typically involved in 3D forward simulation and Bayesian inversion (Zhang et al., 2023). In the main text, we introduce a fine-tuning scheme to improve the inversion results displayed in Figure B-1.

REFERENCES

Aghamiry, H., A. Gholami, and S. Operto, 2018, Hybrid Tikhonov + total-variation regularization for imaging large-contrast media by full-waveform inversion: 88th Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2018-2996968.1.

Aminzadeh, F., P. Weimer, and T. Davis, 1996, 3-D salt and overthrust seismic models: Studies in Geology, **42**, 247–256.

Arnold, R., and A. Curtis, 2018, Interrogation theory: Geophysical Journal International, **214**, 1830–1846, doi: 10.1093/gji/ggy248.

Asnaashari, A., R. Brossier, S. Garambois, F. Audebert, P. Thore, and J. Virieux, 2013, Regularized seismic full waveform inversion with prior model information: Geophysics, **78**, R25–R36, doi: 10.1190/geo2012-0104.1.

Bates, O., L. Guasch, G. Strong, T. C. Robins, O. Calderon-Agudo, C. Cueto, J. Cudeiro, and M. Tang, 2022, A probabilistic approach to tomography and adjoint state methods, with an application to full waveform inversion in medical ultrasound: Inverse Problems, **38**, 045008, doi: 10.1088/1361-6420/ac55ee.

Berti, S., M. Aleardi, and E. Stucchi, 2023, A computationally efficient Bayesian approach to full-waveform inversion: Geophysical Prospecting, **72**, 580–603, doi: 10.1111/1365-2478.13437.

Biswas, R., and M. K. Sen, 2022, Transdimensional 2D full-waveform inversion and uncertainty estimation: arXiv preprint arXiv:2201.09334.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe, 2017, Variational inference: A review for statisticians: Journal of the American Statistical Association, **112**, 859–877, doi: 10.1080/01621459.2017.1285773.

Bloem, H., A. Curtis, and D. Tetzlaff, 2024, Introducing conceptual geological information into Bayesian tomographic imaging: Basin Research, **36**, e12811, doi: 10.1111/bre.12811.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic wave-form inversion: Geophysics, **60**, 1457–1473, doi: 10.1190/1.1443880.

Curtis, A., and A. Lomax, 2001, Prior information, sampling distributions, and the curse of dimensionality: Geophysics, **66**, 372–378, doi: 10.1190/1.1444928.

Ely, G., A. Malcolm, and O. V. Poliannikov, 2018, Assessing uncertainties in velocity models and images with a fast nonlinear uncertainty quantification method: Geophysics, **83**, R63–R75, doi: 10.1190/geo2017-0321.1.

Etgen, J. T., and S. Brandsberg-Dahl, 2009, The pseudo-analytical method: Application of pseudo-Laplacians to acoustic and acoustic anisotropic wave propagation: 79th Annual International Meeting, SEG, Expanded Abstracts, 2552–2556.

Evensen, G., 1994, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics: Journal of Geophysical Research: Oceans, **99**, 10143–10162, doi: 10.1029/94JC00572.

Fang, J., H. Zhou, Y. Elita Li, Q. Zhang, L. Wang, P. Sun, and J. Zhang, 2020, Data-driven low-frequency signal recovery using deep-learning predictions in full-waveform inversion: Geophysics, **85**, A37–A43, doi: 10.1190/geo2020-0159.1.

Galetti, E., A. Curtis, G. A. Meles, and B. Baptie, 2015, Uncertainty loops in travel-time tomography from nonlinear wave physics: Physical Review Letters, **114**, 148501, doi: 10.1103/PhysRevLett.114.148501.

Gallego, V., and D. R. Insua, 2018, Stochastic gradient MCMC with repulsive forces: STAT, **1050**, 30, doi: 10.48550/arXiv.1812.00071.

Gebraad, L., C. Boehm, and A. Fichtner, 2020, Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo: Journal of Geophysical Research: Solid Earth, **125**, e2019JB018428, doi: 10.1029/2019JB018428.

Golub, G. H., P. C. Hansen, and D. P. O'Leary, 1999, Tikhonov regularization and total least squares: SIAM Journal on Matrix Analysis and Applications, **21**, 185–194, doi: 10.1137/S0895479897326432.

Hoffmann, A., R. Brossier, L. Métivier, and A. Tarayoun, 2024, Local uncertainty quantification for 3-D time-domain full-waveform inversion with ensemble kalman filters: application to a North Sea OBC data set: Geophysical Journal International, **237**, 1353–1383, doi: 10.1093/gji/ggae114.

Hu, S., M. K. Sen, Z. Zhao, A. Elmeliegy, and S. Zhang, 2024, Efficient bayesian full-waveform inversion using a deep convolutional autoencoder prior: Fourth International Meeting for Applied Geoscience & Energy, Society of Exploration Geophysicists and American Association of Petroleum, 993–997.

Izzatullah, M., A. Alali, M. Ravasi, and T. Alkhalifah, 2024, Physics-reliable frugal local uncertainty analysis for full waveform inversion: Geophysical Prospecting, **72**, 2718–2738, doi: 10.1111/1365-2478.13528.

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

Kullback, S., and R. A. Leibler, 1951, On information and sufficiency: The Annals of Mathematical Statistics, **22**, 79–86, doi: 10.1214/aoms/1177729694.

Levy, S., E. Laloy, and N. Linde, 2022, Variational Bayesian inference with complex geostatistical priors using inverse autoregressive flows: Computers & Geosciences, 105263.

Li, Y. E., and L. Demanet, 2016, Full-waveform inversion with extrapolated low-frequency data: Geophysics, **81**, R339–R348, doi: 10.1190/geo2016-0038.1.

Liu, Q., and D. Wang, 2016, Stein variational gradient descent: A general purpose Bayesian inference algorithm: Advances in Neural Information Processing Systems, 2378–2386.

Lomas, A., S. Luo, M. Irakarama, R. Johnston, M. Vyas, and X. Shen, 2023, 3D probabilistic full waveform inversion: Application to Gulf of Mexico field data: 84th EAGE Annual Conference & Exhibition, EAGE, 1–5.

Mosegaard, K., and A. Curtis, 2024, Inconsistency and acausality in Bayesian inference for physical problems: arXiv preprint arXiv:2411.13570.

Mosser, L., O. Dubrule, and M. J. Blunt, 2020, Stochastic seismic waveform inversion using generative adversarial networks as a geological prior: Mathematical Geosciences, **52**, 53–79, doi: 10.1007/s11004-019-09832-6.

Mulder, W., and B. Kuvshinov, 2024, Accelerating target-oriented multiparameter elastic full-waveform uncertainty estimation by reciprocity: Geophysical Prospecting, **73**, 38–48, doi: 10.1111/1365-2478.13650.

Ray, A., A. Sekar, G. M. Hoversten, and U. Albertin, 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**, 915–937, doi: 10.1093/gji/ggw061.

Rezende, D., and S. Mohamed, 2015, Variational inference with normalizing flows: Proceedings of the 32nd International Conference on Machine Learning, 1530–1538.

Ryan, H., 1994, Ricker, ormsby; klander, bntterwo-a choice of wavelets: CSEG Recorder, **19**.

Sambridge, M., and K. Mosegaard, 2002, Monte Carlo methods in geophysical inverse problems: Reviews of Geophysics, **40**, 3–1, doi: 10.1029/2000RG000089.

Scales, J. A., 2005, Uncertainties in seismic inverse calculations: Inverse Methods: Interdisciplinary Elements of Methodology, Computation, and Applications, 79–97, doi: 10.1007/BFb0011766.

Siahkoohi, A., G. Rizzuti, and F. J. Herrmann, 2022, Deep Bayesian inference for seismic imaging with tasks: Geophysics, **87**, S281–S302, doi: 10.1190/geo2021-0666.1.

Strutz, D., and A. Curtis, 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**, 1309–1331, doi: 10.1093/gji/ggad492.

Sun, C., A. Malcolm, R. Kumar, and W. Mao, 2024, Enabling uncertainty quantification in a standard full-waveform inversion method using normalizing flows: Geophysics, **89**, R493–R507, doi: 10.1190/geo2023-0755.1.

Sun, Y., and P. Williamson, 2024, Invertible neural networks for uncertainty quantification in refraction tomography: The Leading Edge, **43**, 358–366, doi: 10.1190/tle43060358.1.

Tsai, V. C., C. Huber, and C. A. Dalton, 2023, Towards the geological parametrization of seismic tomography: Geophysical Journal International, **234**, 1447–1462, doi: 10.1093/gji/ggad140.
Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26, doi: 10.1190/1.3238367.
Walker, M., A. Crosby, A. Lomas, E. Kazlauskas, R. Biswas, P. Paramo, K. Wolf, and M. Vyas, 2024, Novel approaches to uncertainty estimation in seismic subsurface characterization: The Leading Edge, **43**, 347–357, doi: 10.1190/tle43060347.1.
Walker, M., and A. Curtis, 2014, Varying prior information in Bayesian inversion: Inverse Problems, **30**, 065002, doi: 10.1088/0266-5611/30/6/065002.
Wang, W., G. A. McMechan, and J. Ma, 2023, Re-weighted variational full waveform inversions: Geophysics, **88**, N1–N19, doi: 10.1190/geo2021-0311.1.
Xie, Y., H. Chauris, and N. Desassis, 2024, Stochastic full waveform inversion with deep generative prior for uncertainty quantification: arXiv preprint arXiv:2406.04859.
Yin, Z., R. Orozco, M. Louboutin, and F. J. Herrmann, 2024, WISE: full-waveform variational inference via subsurface extensions: Geophysics, **89**, WA1–WA11, doi: 10.1190/geo2023-0080.1.
Zhang, X., and A. Curtis, 2020, Seismic tomography using variational inference methods: Journal of Geophysical Research: Solid Earth, **125**, e2019JB018589, doi: 10.1029/2019JB018589.
Zhang, X., and A. Curtis, 2021, Bayesian full-waveform inversion with realistic priors: Geophysics, **86**, A45–A49, doi: 10.1190/geo2021-0118.1.
Zhang, X., and A. Curtis, 2022, Interrogating probabilistic inversion results for subsurface structural information: Geophysical Journal International, **229**, 750–757, doi: 10.1093/gji/ggab496.
Zhang, X., A. Lomas, M. Zhou, Y. Zheng, and A. Curtis, 2023, 3D Bayesian variational full waveform inversion: Geophysical Journal International, **234**, 546–561, doi: 10.1093/gji/ggad057.
Zhao, X., and A. Curtis, 2024a, Bayesian inversion, uncertainty analysis and interrogation using boosting variational inference: Journal of Geophysical Research: Solid Earth, **129**, e2023JB027789, doi: 10.1029/2023JB027789.
Zhao, X., and A. Curtis, 2024b, Physically structured variational inference for Bayesian full waveform inversion: Journal of Geophysical Research: Solid Earth, **129**, e2024JB029557, doi: 10.1029/2024JB029557.
Zhao, X., and A. Curtis, 2024c, Variational prior replacement in Bayesian inference and inversion: Geophysical Journal International, **239**, 1236–1256, doi: 10.1093/gji/ggae334.
Zhao, X., and A. Curtis, 2025, On the design of ultra-sparse seismic surveys for monitoring subsurface CO2 storage sites using full waveform inversion: International Journal of Greenhouse Gas Control, **146**, 104433, doi: 10.1016/j.ijggc.2025.104433.
Zhao, X., A. Curtis, and X. Zhang, 2021, Bayesian seismic tomography using normalizing flows: Geophysical Journal International, **228**, 213–239, doi: 10.1093/gji/ggab298.
Zhao, X., A. Curtis, and X. Zhang, 2022, Interrogating subsurface structures using probabilistic tomography: an example assessing the volume of Irish Sea basins: Journal of Geophysical Research: Solid Earth, **127**, e2022JB024098, doi: 10.1029/2022JB024098.
Zhao, Z., and M. K. Sen, 2021, A gradient-based Markov chain Monte Carlo method for full-waveform inversion and uncertainty analysis: Geophysics, **86**, R15–R30, doi: 10.1190/geo2019-0585.1.
Zhdanov, M. S., 2002, Geophysical inverse theory and regularization problems: Elsevier.
Zunino, A., L. Gebraad, A. Ghirotto, and A. Fichtner, 2023, HMCLab: a framework for solving diverse geophysical inverse problems using the Hamiltonian Monte Carlo method: Geophysical Journal International, **235**, 2979–2991, doi: 10.1093/gji/ggad403.

Biographies and photographs of the authors are not available.