Geophysical Journal International

Geophys. J. Int. (2024) **237,** 1624–1638
Advance Access publication 2024 April 5
GJI Seismology

https://doi.org/10.1093/gji/ggae129

Bayesian variational time-lapse full waveform inversion

Xin Zhang [1,2] and Andrew Curtis[2]

[1]*School of Engineering and Technology, China University of Geosciences, Beijing, Haidian District* 100083, *China. E-mail: xzhang@cugb.edu.cn*
[2]*School of GeoSciences, University of Edinburgh, Edinburgh EH*8 9*YL, United Kingdom*

Accepted 2024 April 2. Received 2024 February 6; in original form 2023 August 23

SUMMARY

Time-lapse seismic full-waveform inversion (FWI) provides estimates of dynamic changes in the Earth's subsurface by performing multiple seismic surveys at different times. Since FWI problems are highly non-linear and non-unique, it is important to quantify uncertainties in such estimates to allow robust decision making based on the results. Markov chain Monte Carlo (McMC) methods have been used for this purpose, but due to their high computational cost, those studies often require a pre-existing accurate baseline model and estimates of the locations of potential velocity changes, and neglect uncertainty in the baseline velocity model. Such detailed and accurate prior information is not always available in practice. In this study we use an efficient optimization method called stochastic Stein variational gradient descent (sSVGD) to solve time-lapse FWI problems without assuming such prior knowledge, and to estimate uncertainty both in the baseline velocity model and the velocity change over time. We test two Bayesian strategies: separate Bayesian inversions for each seismic survey, and a single joint inversion for baseline and repeat surveys, and compare the methods with standard linearized double difference inversion. The results demonstrate that all three methods can produce accurate velocity change estimates in the case of having fixed (exactly repeatable) acquisition geometries. However, the two Bayesian methods generate significantly more accurate results when acquisition geometries changes between surveys. Furthermore, joint inversion provides the most accurate velocity change and uncertainty estimates in all cases tested. We therefore conclude that Bayesian time-lapse inversion using a joint inversion strategy may be useful to image and monitor subsurface changes, in particular where variations in the results would lead to different consequent decisions.

Key words: Inverse theory; Probability distributions; Waveform inversion.

1 INTRODUCTION

A wide variety of academic and practical applications require that we detect property changes in the subsurface in order to understand dynamic processes in the Earth's interior. Time-lapse seismic monitoring provides an important tool for this purpose. This involves conducting multiple seismic surveys acquired at the same site at different times (Lumley 2001). Changes in certain subsurface properties are estimated by computing the difference between models constructed for surveys at different times (the first survey being called the baseline survey and subsequent surveys being called monitoring surveys). In order to assess reliability of the property changes and interpret the results with appropriate levels of confidence, it is also necessary to quantify the uncertainty in such estimates.

Seismic full-waveform inversion (FWI) is a method which uses full seismic recordings to characterize properties of the Earth's interior (Tarantola 1984, 1988; Pratt 1999; Tromp *et al.* 2005; Plessix 2006), and has been applied at industrial scale (Virieux & Operto 2009; Prieux *et al.* 2013), regional scale (Tape *et al.* 2009; Fichtner *et al.* 2009) and global scale (French & Romanowicz 2014; Lei *et al.* 2020). Because of its high spatial resolution, the method has been extended to time-lapse studies to image changes in the subsurface. For example, a range of different schemes have been proposed for this purpose, such as parallel FWI (Plessix *et al.* 2010), sequential FWI (Asnaashari *et al.* 2015), double difference FWI (Watanabe *et al.* 2004; Zheng *et al.* 2011) and joint FWI (Maharramov & Biondi 2014; Yang *et al.* 2014). However, all of these studies used linearised methods to solve their respective inverse problems and therefore cannot provide accurate uncertainty estimates. As a result, it becomes difficult to interpret the estimated property changes, and to use those estimates for subsequent applications.

Bayesian inference methods provide a variety of different ways to solve inverse problems and can produce accurate uncertainty estimates. In Bayesian inference, prior information is represented

by a probability density function (pdf) called the *prior* pdf, which describes information about the parameters of interest prior to conducting the inversion. Bayes' theorem updates the prior pdf with new information contained in the data to construct a so-called *posterior* pdf which describes the total state of information about the parameters post inversion. The updating process is referred to as Bayesian inference.

Markov chain Monte Carlo (McMC) is one method that is commonly used to solve Bayesian inference problems. The method generates a set (chain) of successive samples from the posterior probability distribution by taking a structured random-walk through parameter space (Brooks *et al.* 2011). Those samples can thereafter be used to calculate statistics of the posterior pdf, for example the mean and standard deviation. The Metropolis-Hastings algorithm is one such method (Metropolis & Ulam 1949; Hastings 1970; Green 1995) and has been applied to a range of geophysical applications (Mosegaard & Tarantola 1995; Sambridge & Mosegaard 2002; Bodin & Sambridge 2009; Galetti *et al.* 2015; Zhang *et al.* 2018b), including full-waveform inversion (Ray *et al.* 2017; Sen & Biswas 2017; Guo *et al.* 2020). However, due to its random-walk behaviour, the method becomes inefficient in high dimensional parameter spaces (e.g. >1000 dimensions are commonly encountered in geophysical imaging problems). To reduce this issue, a variety of more advanced methods have been introduced to geophysics, such as Hamiltonian Monte Carlo (Duane *et al.* 1987; Fichtner *et al.* 2018; Gebraad *et al.* 2020), Langevin Monte Carlo (Roberts *et al.* 1996; Siahkoohi *et al.* 2020a), stochastic Newton McMC (Martin *et al.* 2012; Zhao & Sen 2019) and parallel tempering (Hukushima & Nemoto 1996; Dosso *et al.* 2012; Sambridge 2013). These allow convergence to be accelerated by assuming specific information about the structure of the problem to be solved. Nevertheless, all of these methods still incur high computational costs and are therefore difficult to use in time-lapse full-waveform inversion. To enable Bayesian inference in time-lapse studies, Kotsi *et al.* (2020b) exploited a fast, local solver together with the discrete cosine transform to solve time-lapse FWI problems using the Metropolis–Hastings algorithm, and directly imaged the velocity change by assuming a known baseline model. To further improve the efficiency, Hamiltonian Monte Carlo was used to solve the problem but with a regular grid parametrization (Kotsi *et al.* 2020a). However, these methods require a known and accurate baseline model and prior knowledge about the location of potential velocity change zones, which are not always available in practice and which therefore restricts their practical applications.

Variational inference solves Bayesian inference problems in a different way: the method seeks an optimal approximation to the posterior pdf within a predefined (simplified) family of probability distributions, by minimizing the difference between the approximating pdf and the posterior pdf (Bishop 2006; Blei *et al.* 2017). A typical measure used to quantify this difference is the Kullback–Leibler (KL) divergence (Kullback & Leibler 1951); since this measure is minimized the method solves Bayesian inference problems using optimisation, rather than stochastic sampling as used in McMC methods. Consequently for some classes of problems variational inference can be computationally more efficient and provide better scaling to higher dimensional problems (Blei *et al.* 2017; Zhang *et al.* 2018a). The method also allows us to take advantage of stochastic and distributed optimisation (Robbins & Monro 1951; Kubrusly & Gravier 1973) by dividing large data sets into small minibatches. In addition, variational inference can often be parallelized at the individual sample level, which makes the method even more efficient in real time. By contrast, in McMC one cannot use small minibatches as they break the detailed balance property required by most McMC methods (O'Hagan & Forster 2004), and McMC does not allow parallelization at the sample level as each sample in McMC depends on the previous sample.

A range of different variational inference methods have been introduced to geophysics. These include mean-field variational inference (Bishop 2006; Nawaz & Curtis 2018, 2019; Nawaz *et al.* 2020), automatic differential variational inference (Kucukelbir *et al.* 2017; Zhang & Curtis 2020a; Zhang & Chen 2022), normalizing flow variational inference (Rezende & Mohamed 2015; Siahkoohi *et al.* 2020b; Zhao *et al.* 2021) and Stein variational gradient descent (SVGD: Liu & Wang 2016; Zhang & Curtis 2020a; Smith *et al.* 2022). Because of its high accuracy in lower dimensional problems, the SVGD method has been used to solve FWI problems (Zhang & Curtis 2020b, 2021). More recently, Zhang *et al.* (2023) introduced a variant of the SVGD method called stochastic SVGD (sSVGD) to solve 3-D FWI problems and demonstrated that the method can provide more accurate results than ADVI and the original SVGD in such high dimensional problems.

Based on the results of these studies we chose to test the sSVGD method for the solution of time-lapse FWI problems. In particular, we do not assume prior knowledge about an accurate baseline model, nor about locations of potential velocity change zones, and we estimate uncertainty for both the baseline velocity and the time-lapse velocity change. To solve time-lapse FWI problems, we consider two Bayesian strategies, namely separate Bayesian inversion for baseline and monitor surveys, and joint Bayesian inversion for both surveys together, and compare the results with those from standard double difference inversion. In addition, we perform studies with both fixed and perturbed acquisition geometries between the baseline and monitoring surveys to test the robustness of each method to typical variations in survey design that may occur in practical applications.

In the following section we first describe the two Bayesian inversion strategies and double difference inversion. In Section 3, we apply the suit of methods to a time-lapse FWI problem and compare the results and their computational costs. We use the results to demonstrate that sSVGD can be used to solve Bayesian time-lapse FWI problems and produce accurate velocity change estimates as well as associated uncertainties. We conclude by defining particular contexts in which Bayesian time-lapse FWI provides an important tool to image and monitor subsurface property changes.

2 METHODS

2.1 Standard time-lapse FWI

The standard way to perform FWI is to minimize a misfit function between observed data $\mathbf{d}$ and model predicted data $\mathbf{u}(\mathbf{m})$ plus a regularization term:

$$L(\mathbf{m}) = \frac{1}{2}|\mathbf{d} - \mathbf{u}(\mathbf{m})|^2 + \lambda|\mathbf{D}\mathbf{m}|^2 \tag{1}$$

where $\mathbf{m} \in \mathbb{R}^r$ is the parametrized earth model, $\mathbf{D}$ typically represents a finite-difference derivative matrix and λ controls the magnitude of regularization. The most straightforward implementation of time-lapse FWI is to perform the above minimization for each data set from individual surveys; differences between the obtained models are regarded as estimates of the time-lapse change. In this mode of implementation either both inversions can be performed

using the same starting model, or the model obtained from inversion of the baseline survey can be used as the starting model for the monitoring survey inversion.

A more efficient method is so-called double difference FWI which uses differential data between the two sets of data obtained in the baseline and monitoring surveys (Watanabe *et al.* 2004; Denli & Huang 2009; Zheng *et al.* 2011). The misfit function for double difference FWI is:

$$L(\mathbf{m}_2) = \frac{1}{2}|(\mathbf{d}_2 - \mathbf{d}_1) - (\mathbf{u}(\mathbf{m}_2) - \mathbf{u}(\mathbf{m}_1))|^2 + \lambda|\mathbf{D}\mathbf{m}_2|^2 + \mu|\mathbf{m}_2 - \mathbf{m}_1|^2, \quad (2)$$

where we used subscript 1 and 2 to denote variables of the baseline inversion and monitoring inversion, respectively, λ and μ are parameters that control the strength of regularization of model $\mathbf{m}_2$ and of the difference between $\mathbf{m}_2$ and $\mathbf{m}_1$, respectively. The above equation can be reformulated as:

$$\begin{aligned} L(\mathbf{m}_2) &= \frac{1}{2}|(\mathbf{d}_2 - \mathbf{d}_1 + \mathbf{u}(\mathbf{m}_1)) - \mathbf{u}(\mathbf{m}_2)|^2 + \lambda|\mathbf{D}\mathbf{m}_2|^2 + \mu|\mathbf{m}_2 - \mathbf{m}_1|^2 \\ &= \frac{1}{2}|\mathbf{d}_2^{'} - \mathbf{u}(\mathbf{m}_2)|^2 + \lambda|\mathbf{D}\mathbf{m}_2|^2 + \mu|\mathbf{m}_2 - \mathbf{m}_1|^2, \end{aligned} \quad (3)$$

where $\mathbf{d}_2^{'} = \mathbf{d}_2 + [\mathbf{u}(\mathbf{m}_1) - \mathbf{d}_1]$ can be regarded as a new data set adjusted by the residual data of the baseline inversion. This adjustment allows unexplained data (those not fit by the earth model) in the baseline survey to be disregarded in inversions of monitoring survey data. With this definition one can use standard FWI algorithms to minimize eq. (3) and obtain the optimal model $\mathbf{m}_2$. The time-lapse change can finally be obtained by $\delta\mathbf{m} = \mathbf{m}_2 - \mathbf{m}_1$.

2.2 Bayesian time-lapse full waveform inversion

Bayesian inference solves inverse problems by updating a prior pdf $p(\mathbf{m})$ with new information contained in the data to construct the posterior pdf $p(\mathbf{m}|\mathbf{d})$. According to Bayes' theorem,

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d})}, \quad (4)$$

where $p(\mathbf{d}|\mathbf{m})$ is the *likelihood* which represents the probability of observing data $\mathbf{d}$ given model $\mathbf{m}$, and $p(\mathbf{d})$ is a normalization factor called the *evidence*. A Gaussian distribution is usually used to represent data uncertainties in the likelihood function, so

$$p(\mathbf{d}|\mathbf{m}) \propto \exp\left[-\frac{1}{2}(\mathbf{d} - \mathbf{u}(\mathbf{m}))^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{d} - \mathbf{u}(\mathbf{m}))\right], \quad (5)$$

where $\mathbf{\Sigma}$ is a covariance matrix which is often assumed to be diagonal in practice.

Similarly as in standard time-lapse full waveform inversion, one can perform Bayesian inversion for each data set $\mathbf{d}_1$ and $\mathbf{d}_2$ separately, and calculate the probability distribution $p(\delta\mathbf{m})$ using the results obtained. This can be achieved by randomly generating or selecting pairs of samples from the two posterior distribution $p(\mathbf{m}_1|\mathbf{d}_1)$ and $p(\mathbf{m}_2|\mathbf{d}_2)$, and computing the difference between each pair which can then be regarded as a sample of distribution $p(\delta\mathbf{m})$. The two inversions can be performed independently, or one can use the posterior samples obtained in the baseline inversion as the starting point for the monitoring inversion. Since in most Bayesian inference methods in theory the results do not depend on starting models (Brooks *et al.* 2011), the two methods should produce the same results when both of them have converged. However, by using the second strategy the burn-in period required in McMC-like methods (also sSVGD) can be significantly reduced. In this study we regard both of the above two methods as *separate Bayesian inversions*, and adopt the second method in our examples below to reduce the computational cost. According to the general Bienaymé's identity, the mean and variance of $\delta\mathbf{m}$ have the following form:

$$\mathrm{mean}(\delta\mathbf{m}^i) = \mathrm{mean}\left(\mathbf{m}_2^i\right) - \mathrm{mean}(\mathbf{m}_1^i) \quad (6a)$$

$$\mathrm{var}\left(\delta\mathbf{m}^i\right) = \mathrm{var}\left(\mathbf{m}_2^i\right) + \mathrm{var}\left(\mathbf{m}_1^i\right) - 2\mathrm{cov}\left(\mathbf{m}_1^i, \mathbf{m}_2^i\right) \quad (6b)$$

where superscript i represents the i^{th} element of a vector. Because the two inversions are conducted separately and the effect of starting models decreases with iteration number in McMC-like methods, the covariance between $\mathbf{m}_1$ and $\mathbf{m}_2$ becomes negligible when a Markov chain converges. The above equation can therefore be approximately expressed as:

$$\mathrm{mean}(\delta\mathbf{m}^i) = \mathrm{mean}(\mathbf{m}_2^i) - \mathrm{mean}(\mathbf{m}_1^i) \quad (7a)$$

$$\mathrm{var}(\delta\mathbf{m}^i) = \mathrm{var}(\mathbf{m}_2^i) + \mathrm{var}(\mathbf{m}_1^i) \quad (7b)$$

The uncertainty of $\delta\mathbf{m}$ obtained using separate Bayesian inversions is thus higher than that of model $\mathbf{m}_1$ or $\mathbf{m}_2$ itself. This seems unlikely to be realistic: any particular earth structure at the time of the baseline survey estimated by model $\mathbf{m}_1$, is likely to affect uncertainties in $\mathbf{m}_2$ in the sense that we would expect these uncertainties to change if a different baseline earth structure was true and estimated in $\mathbf{m}_1$. If this correlation was taken into account we would expect the overall uncertainty on the model differences between the two surveys to decrease. Given that the magnitude of time-lapse change is usually much smaller than that of either model, we would expect that the uncertainty estimate using the above method would be less valuable in practice.

Instead of performing the two inversions separately, one can invert the two data sets simultaneously to obtain the joint distribution of model $\mathbf{m}_1$ and $\mathbf{m}_2$, that is

$$p(\mathbf{m}_1, \mathbf{m}_2|\mathbf{d}_1, \mathbf{d}_2) = \frac{p(\mathbf{d}_1, \mathbf{d}_2|\mathbf{m}_1, \mathbf{m}_2)p(\mathbf{m}_1)p(\mathbf{m}_2)}{p(\mathbf{d}_1, \mathbf{d}_2)} \quad (8)$$

This equation is still consistent with $\mathbf{m}_1$ and $\mathbf{m}_2$ being independent since in that case $p(\mathbf{d}_1, \mathbf{d}_2|\mathbf{m}_1, \mathbf{m}_2)$ can be written as $p(\mathbf{d}_1, \mathbf{d}_2|\mathbf{m}_1, \mathbf{m}_2) = p(\mathbf{d}_1|\mathbf{m}_1)p(\mathbf{d}_2|\mathbf{m}_2)$. However, given that $\mathbf{m}_2 = \mathbf{m}_1 + \delta\mathbf{m}$ for a change in the Earth structure $\delta\mathbf{m}$, we can instead invert for the joint distribution of $\mathbf{m}_1$ and $\delta\mathbf{m}$:

$$p(\mathbf{m}_1, \delta\mathbf{m}|\mathbf{d}_1, \mathbf{d}_2) = \frac{p(\mathbf{d}_1, \mathbf{d}_2|\mathbf{m}_1, \delta\mathbf{m})p(\mathbf{m}_1)p(\delta\mathbf{m})}{p(\mathbf{d}_1, \mathbf{d}_2)} \quad (9)$$

where $p(\delta\mathbf{m})$ is the prior distribution of $\delta\mathbf{m}$. In this way, one can impose prior information on $\delta\mathbf{m}$ by taking into account the fact that time-lapse changes are often small in practice, which therefore correlates estimates of $\mathbf{m}_1$ and $\mathbf{m}_2$ and potentially produces more accurate model change and uncertainty estimates. We refer to this method as *joint Bayesian inversion*. If we assume that the baseline model $\mathbf{m}_1$ is known, the above equation reduces to a form where we solve for the posterior distribution of $\delta\mathbf{m}$ only (Kotsi *et al.* 2020b). We note that a similar idea of joint inversion has been used in a linearized inversion framework with a regularization term directly imposed on model differences (Maharramov & Biondi 2014). From a Bayesian perspective, regularization can be regarded as prior information, so that method is conceptually a linearised form of the joint Bayesian inversion strategy. However, linearized inversion methods cannot provide accurate uncertainty estimates in non-linear problems, and inappropriate regularization may conceal useful information available in the data as demonstrated by

Maharramov & Biondi (2014). In this study we therefore investigate characteristics of the joint Bayesian inversion strategy.

2.3 Stochastic Stein variational gradient descent (sSVGD)

To solve Bayesian inverse problems in eq. (4) or (9), we use a specific method called stochastic Stein variational gradient descent (sSVGD) which combines Monte Carlo and variational inference methods (Gallego & Insua 2018). The method simulates a Markov process using a stochastic differential equation (SDE):

$$\mathrm{d}\mathbf{z} = \mathbf{f}(\mathbf{z})\mathrm{d}t + \sqrt{2\mathbf{D}(\mathbf{z})}d\mathbf{W}(t), \tag{10}$$

where $\mathbf{z} \in \mathbb{R}^m$, $\mathbf{f}(\mathbf{z})$ is called the *drift*, $\mathbf{W}(t)$ is a Wiener process and $\mathbf{D}(\mathbf{z})$ is a positive semidefinite diffusion matrix. If we denote the posterior distribution of interest from either eq. (4) or (9) as $p(\mathbf{z})$, Ma *et al.* (2015) proposed a specific form of eq. (10) which gives an SDE that converges to distribution $p(\mathbf{z})$:

$$\mathbf{f}(\mathbf{z}) = [\mathbf{D}(\mathbf{z}) + \mathbf{Q}(\mathbf{z})]\nabla \log p(\mathbf{z}) + \Gamma(\mathbf{z}), \tag{11}$$

where $\mathbf{Q}(\mathbf{z})$ is a skew-symmetric curl matrix, $\Gamma(\mathbf{z}) = \sum_{j=1}^{m} \frac{\partial}{\partial \mathbf{z}_j}(\mathbf{D}_{ij}(\mathbf{z}) + \mathbf{Q}_{ij}(\mathbf{z}))$, and $\nabla \log p(\mathbf{z})$ represents derivatives of $\log p(\mathbf{z})$ with respect to all variables in $\mathbf{z}$. By choosing different matrices $\mathbf{D}$ and $\mathbf{Q}$, different methods can be obtained (Ma *et al.* 2015). For example, if we choose $\mathbf{D} = \mathbf{I}$ and $\mathbf{Q} = \mathbf{0}$ we obtained the stochastic gradient Langevin dynamics algorithm (Welling & Teh 2011). If we construct an augmented space $\bar{\mathbf{z}} = (\mathbf{z}, \mathbf{x})$ by concatenating $\mathbf{z}$ and a moment term $\mathbf{x}$, and set $\mathbf{D} = \mathbf{0}$ and $\mathbf{Q} = \begin{bmatrix} \mathbf{0} & -\mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix}$, we obtain the stochastic Hamiltonian Monte Carlo method (Chen *et al.* 2014).

The above process can be simulated numerically by discretizing eq. (10) with eq. (11) over time variable t using the Euler–Maruyama discretization:

$$\mathbf{z}_{t+1} = \mathbf{z}_t + \epsilon_t \left[(\mathbf{D}(\mathbf{z}_t) + \mathbf{Q}(\mathbf{z}_t))\nabla \log p(\mathbf{z}_t) + \Gamma(\mathbf{z}_t)\right] + N(\mathbf{0}, 2\epsilon_t \mathbf{D}(\mathbf{z}_t)) \tag{12}$$

where ϵ_t is a small step, and $N(\mathbf{0}, 2\epsilon_t \mathbf{D}(\mathbf{z}_t))$ is a Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $2\epsilon_t \mathbf{D}(\mathbf{z}_t)$. Since $p(\mathbf{z}_t)$ represents the posterior distribution in eq. (4), it depends implicitly on observed data $\mathbf{d}$. The gradient $\nabla \log p(\mathbf{z}_t)$ can be calculated using either the full data set or uniformly randomly selected minibatch data sets in each step in t, and in either case the process converges to the posterior distribution $p(\mathbf{z})$ when $\epsilon_t \to 0$ and $t \to \infty$. In practice because of the limitation of computational resources, we cannot use an infinitesimal ϵ_t. As a result, the ϵ_t should be chosen to be small enough to ensure accuracy of the simulation and large enough so that the process is computationally tractable. One practical option is to use a large step size at the beginning and exponentially decay the size so that a small step size is used after the burn-in peirod to achieve a high accuracy.

sSVGD uses a set of models called particles since sSVGD moves them through parameter space. Define the set of particles as $\{\mathbf{m}_i : i = 1,..., n\}$ where $\mathbf{m}_i \in \mathbb{R}^r$, and construct an augmented space $\mathbf{z} = (\mathbf{m}_1, \mathbf{m}_2, ..., \mathbf{m}_n) \in \mathbb{R}^{nr}$ by concatenating the n particles. Using eq. (12) we construct a sampler that runs n multiple interacting chains:

$$\mathbf{z}_{t+1} = \mathbf{z}_t + \epsilon_t[(\mathbf{D}(\mathbf{z}_t) + \mathbf{Q}(\mathbf{z}_t))\nabla \log p(\mathbf{z}_t) + \Gamma(\mathbf{z}_t)] + N(\mathbf{0}, 2\epsilon_t \mathbf{D}(\mathbf{z}_t)), \tag{13}$$

where $\mathbf{D}, \mathbf{Q} \in \mathbb{R}^{nr \times nr}$ and $\nabla \log p, \Gamma \in \mathbb{R}^{nr}$. Define a matrix $\mathbf{K}$:

$$\mathbf{K} = \frac{1}{n}\begin{bmatrix} k(\mathbf{m}_1, \mathbf{m}_1)\mathbf{I}_{r\times r} & \dots & k(\mathbf{m}_1, \mathbf{m}_n)\mathbf{I}_{r\times r} \\ \vdots & \ddots & \vdots \\ k(\mathbf{m}_n, \mathbf{m}_1)\mathbf{I}_{r\times r} & \dots & k(\mathbf{m}_n, \mathbf{m}_n)\mathbf{I}_{r\times r} \end{bmatrix}, \tag{14}$$

where $k(\mathbf{m}_i, \mathbf{m}_j)$ is a kernel function and $\mathbf{I}_{r\times r}$ is an identity matrix. Note that $\mathbf{K}$ is positive definite according to the definition of kernel functions (Gallego & Insua 2018). By setting $\mathbf{D} = \mathbf{K}$ and $\mathbf{Q} = \mathbf{0}$, eq. (13) becomes:

$$\mathbf{z}_{t+1} = \mathbf{z}_t + \epsilon_t[\mathbf{K}\nabla \log p(\mathbf{z}_t) + \nabla \cdot \mathbf{K}] + N(\mathbf{0}, 2\epsilon_t \mathbf{K}) \tag{15}$$

This defines a Markov process that converges to the posterior distribution $p(\mathbf{z}) = \prod_{i=1}^{n} p(\mathbf{m}_i|\mathbf{d})$ asymptotically for any number of particles n. Note that if we eliminate the noise term $N(\mathbf{0}, 2\epsilon_t \mathbf{K})$ in eq. (15), the method becomes Stein variational gradient descent (SVGD). The sSVGD algorithm is therefore a stochastic gradient McMC method that uses SVGD gradients (Gallego & Insua 2018).

Eq. (15) requires that we generate samples from the distribution $N(\mathbf{0}, 2\epsilon_t \mathbf{K})$, which can be computationally expensive because the matrix $\mathbf{K}$ is potentially large. To perform this more efficiently we define a block diagonal matrix $\mathbf{D}_\mathbf{K}$

$$\mathbf{D}_\mathbf{K} = \frac{1}{n}\begin{bmatrix} \overline{\mathbf{K}} & & \\ & \ddots & \\ & & \overline{\mathbf{K}} \end{bmatrix}, \tag{16}$$

where $\overline{\mathbf{K}}$ is a $n \times n$ matrix with $\overline{\mathbf{K}}_{ij} = k(\mathbf{m}_i, \mathbf{m}_j)$. Note that the matrix $\mathbf{D}_\mathbf{K}$ can be constructed from $\mathbf{K}$ using $\mathbf{D}_\mathbf{K} = \mathbf{P}\mathbf{K}\mathbf{P}^\mathrm{T}$ where $\mathbf{P}$ is a permutation matrix

$$\mathbf{P} = \begin{bmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ & & \ddots & & & & & \\ & & & 1 & & & & \\ \hline 1 & & & & & & & \\ & & 1 & & & & & \\ & & & \ddots & & & & \\ & & & & 1 & & & \\ \hline & \ddots & & \ddots & & \ddots & & \ddots \\ \hline & & 1 & & & & & \\ & & & 1 & & & & \\ & & & & \ddots & & & \\ & & & & & 1 & & \end{bmatrix} \tag{17}$$

The action of $\mathbf{P}$ on a vector $\mathbf{z}$ rearranges the order of the vector elements from the basis where particles are concatenated sequentially to the basis where the first coordinates of all the particle are listed, then the second, etc. With this definition, a sample $\boldsymbol{\eta}$ can be generated more efficiently from $N(\mathbf{0}, 2\epsilon_t \mathbf{K})$ using

$$\begin{aligned} \boldsymbol{\eta} &\sim N(\mathbf{0}, 2\epsilon_t \mathbf{K}) \\ &\sim \sqrt{2\epsilon_t}\mathbf{P}^\mathrm{T}\mathbf{P}N(\mathbf{0}, \mathbf{K}) \\ &\sim \sqrt{2\epsilon_t}\mathbf{P}^\mathrm{T}N(\mathbf{0}, \mathbf{D}_\mathbf{K}) \\ &\sim \sqrt{2\epsilon_t}\mathbf{P}^\mathrm{T}\mathbf{L}_{\mathbf{D}_\mathbf{K}}N(\mathbf{0}, \mathbf{I}), \end{aligned} \tag{18}$$

where $\mathbf{L}_{\mathbf{D}_\mathbf{K}}$ is the lower triangular Cholesky decomposition of matrix $\mathbf{D}_\mathbf{K}$, which can be calculated easily as only the lower triangular Cholesky decomposition of matrix $\overline{\mathbf{K}}$ is required by eq. (16). In practice the number of particles n is usually sufficiently modest that the decomposition of $\overline{\mathbf{K}}$ is computationally negligible. We can

thus use eq. (15) to generate samples of the posterior distribution. In geophysics sSVGD has already been used to solve 3-D FWI problems (Zhang *et al.* 2023); in this study we test the method in the context of solving time-lapse imaging problems by sampling the distributions in eqs (4) and (9).

To illustrate the sSVGD algorithm, we use the method to generate samples from a bivariate Gaussian distribution (blue area in Fig. 1), and compare the results with those obtained using SVGD. For both methods we use 20 particles which are initially generated from a standard bivariate normal distribution $N(\mathbf{0}, \mathbf{I})$ (red dots in Fig. 1a). In SVGD these particles are updated for 1000 iterations using eq. (15) without the Gaussian noise term, and Fig. 1(b) shows the final particles. For sSVGD we perform the method for the same number of iterations using eq. (15) and retain all particles generated in the process, which results in a total of 20 000 samples (red dots in Fig. 1c). No burn-in period is used for this example, the stepsize ϵ_t is set to 0.01 for both methods, and a radial basis function kernel is used to construct the matrix $\mathbf{K}$. The main practical difference between the results is that sSVGD generates many more samples of the distribution than SVGD, since the particle values from every iteration (potentially after some burn-in period) constitute valid samples.

3 RESULTS

3.1 Experimental setup

To understand the robustness and behaviour of each method, we set up a synthetic time-lapse experiment using a part of the Marmousi model to represent the true baseline model (Fig. 2a, Martin *et al.* 2006). To represent the true time-lapse model we reduce the velocity of a small square area in the baseline model by 2 per cent (Figs 2a and c). We choose a square area since this discriminates geometrically correct imaging results from errors, since some of the latter are shown below to follow geological strata and hence to look realistic. Both baseline and time-lapse models are parameterized using a regular 200 × 120 grid with a spacing of 20 m. Ten sources are located at 20 m water depth (red stars in Fig. 2), and 200 equally spaced receivers are located on the seabed at 360 m water depth across the horizontal extent of the model. Since it is not possible to repeat exactly the same acquisition geometries in time-lapse seismic surveys, and attempts to do so usually incur significant cost (Beasley *et al.* 1999; Yang *et al.* 2015; Calvert 2005), we study performance of the different methods when the source locations are repeated and when they are perturbed by 100 m in the monitoring survey (yellow stars in Fig. 2). In both cases we assume that the locations of the source positions used in each survey are known. All waveform data are simulated using a time-domain finite difference method with a Ricker wavelet of 10 Hz central frequency, and we added 1 per cent uncorrelated Gaussian noise to the data. For all inversions the gradients of the misfit (likelihood) function with respect to wave velocity in each cell are calculated using the adjoint method (Tarantola 1988; Tromp *et al.* 2005; Fichtner *et al.* 2006; Plessix 2006).

For the prior information on absolute seismic velocity we use a Uniform distribution over an interval of 2 km s^{-1} at each depth (Fig. 2b). To ensure that the rock velocity is higher than the velocity in water we impose an additional lower bound of 1.5 km s^{-1}. Given that time-lapse changes in seismic velocity are usually much smaller than the velocity itself, we use a Uniform distribution between −0.2 and 0.2 km s^{-1} (Fig. 2d) for the prior information $p(\delta\mathbf{m})$ in eq. (9). Note that no prior information is imposed directly on the model difference $\delta\mathbf{m}$ in the separate Bayesian inversion strategy.

3.2 Exactly repeated acquisition geometry

We first perform time-lapse studies with the acquisition geometry repeated identically in the baseline and monitor surveys. The standard double difference FWI method requires a good baseline model to obtain accurate velocity changes (Asnaashari *et al.* 2015). To attempt to achieve this we adopt a multiscale FWI strategy (Bunks *et al.* 1995) in the baseline inversion: we first invert for a long wavelength model using low frequency data simulated using a Ricker wavelet of 4 Hz central frequency. The initial model in this low frequency inversion is set to be laterally-constant with velocity equal to the average velocity of the prior distribution in Fig. 2(b), and the range of models in Fig. 2(b) is also imposed as a set of constraints on velocities at each depth. The resulting long wavelength model serves as the starting model for the inversion using higher frequency data (10 Hz wavelet). For both inversions we use the LBFGS method (Liu & Nocedal 1989) to minimize misfit functions as in eq. (1), in which the control parameter of the regularization term is selected by trial and error. Fig. 3(a) shows the obtained baseline model which provides an accurate estimate of the true model. Note that because of the low data sensitivity around the bottom and edges of the model caused by acquisition geometry limitations, velocity structures in these areas exhibit larger errors.

We then use this baseline model to conduct double difference FWI by minimizing the misfit function in eq. (3). The obtained time-lapse changes are shown in Fig. 3(b). The results demonstrate that the double difference method can obtain reasonably accurate estimates of the true velocity changes, as found in previous studies (Watanabe *et al.* 2004; Denli & Huang 2009; Zheng *et al.* 2011; Asnaashari *et al.* 2015; Yang *et al.* 2015). However, the method cannot provide accurate uncertainty estimates for those velocity changes since it only accounts for linearized physics relating model parameters and data (Smith 2013; Zhang *et al.* 2018b).

To quantify uncertainties in velocity changes we perform time-lapse studies using the above two Bayesian methods. For the separate Bayesian inversion we adopt the strategy which uses the particles of the baseline inversion as the starting point for the monitoring survey inversion as this 'warm start' procedure has been demonstrated to be more effective than two independent inversions in linearised methods (Zheng *et al.* 2011; Asnaashari *et al.* 2015), and was shown to be similarly effective when data of increasingly high frequency were added to an FWI solution found by SVGD (Zhang & Curtis 2021). The prior distributions are set to be the same for both baseline and monitoring inversions, equal to the Uniform distribution in Fig. 2(b). In the baseline inversion we randomly generate 20 samples from the prior distribution as the initial particles, which are then updated using eq. (15) for 4000 iterations after an additional burn-in period of 2000. The step size is set to be 0.2 at the beginning and exponentially decayed to 0.01 at the burn-in period which is then fixed for the rest of iterations. For the kernel function in eq. (14) we use a commonly used radial basis function,

$$k(\mathbf{m}_1, \mathbf{m}_2) = \exp\left[-\frac{\|\mathbf{m}_1 - \mathbf{m}_2\|^2}{2h^2}\right], \tag{19}$$

where h is a scale factor that controls the intensity of interaction between two particles based on their distance apart. As suggested by previous studies (Liu & Wang 2016; Zhang & Curtis 2020a), we choose h to be $\tilde{d}/\sqrt{2\log n}$ where $\tilde{d}$ is the median of pairwise

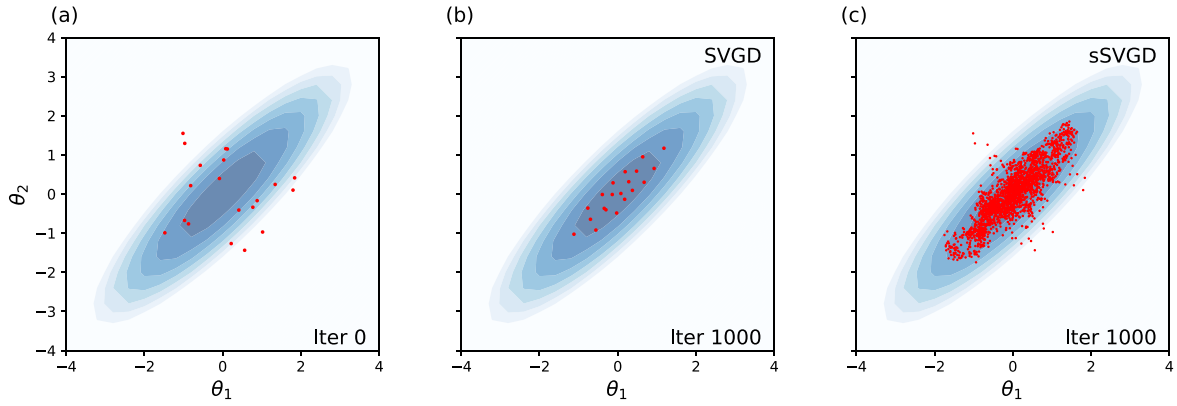

Figure 1. An example which uses SVGD and sSVGD to sample a bivariate Gaussian distribution (blue shades). Red dots show (a) the initial 20 particles, and the samples obtained using (b) SVGD and (c) sSVGD after 1000 iterations.

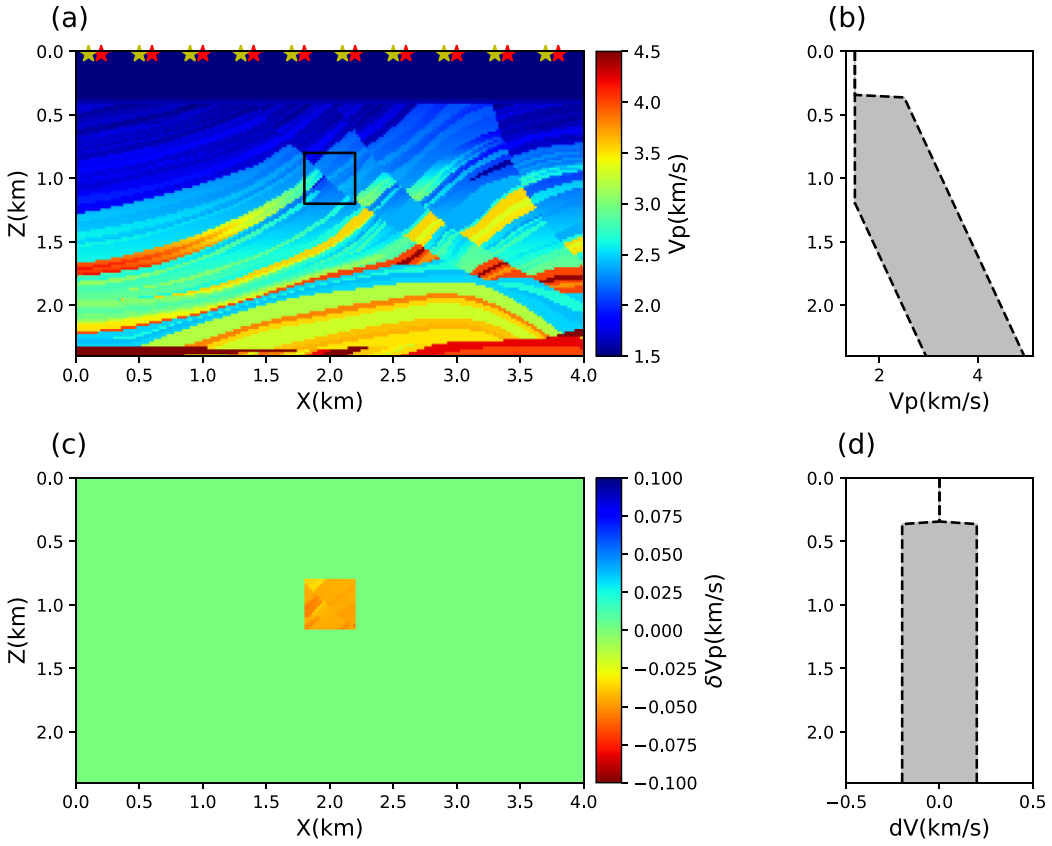

Figure 2. (a) The true velocity model at the time of the baseline survey, and the acquisition geometry used in this study. Red stars denote the source locations in the baseline survey while yellow stars show the perturbed locations in the monitoring survey. 200 receivers are equally spaced at the seabed at 360 m depth (not shown). (b) The prior distribution of velocity: a Uniform distribution with an interval of 2 km s^{-1}, other than above 1.2 km depth where an extra lower bound of 1.5 km s^{-1} is also imposed to ensure rock velocity is higher than the velocity in water. (c) The true time-lapse velocity change in the monitoring survey. (d) The prior distribution of velocity change which is set to be a Uniform distribution over an interval of $\pm$0.2 km s^{-1}.

distances between all particles and n is the number of particles. To reduce the memory and storage cost we only retain every tenth sample after a burn-in period of 2000 iterations, which results in a total of 8000 samples. Those samples are then used to calculate statistics (mean and standard deviation) of the posterior distribution.

Figs 4(a) and (b) show the mean and standard deviation models obtained in the baseline inversion. Although the inversion is performed using the high frequency data directly, the mean model still provides an accurate estimate of the true model, similarly to the linearised inversion which uses the extra low frequency data set described above. Again similarly to the linearised inversion, the structure in the bottom and edges differ from the true model because of low sensitivity. Note that the mean model shows pixel-scale randomness which reflects the uncertainty of neighbouring pixels since there is nothing in the problem setup that prefers smooth models, and neither does seismic waveform data. For example, a model that

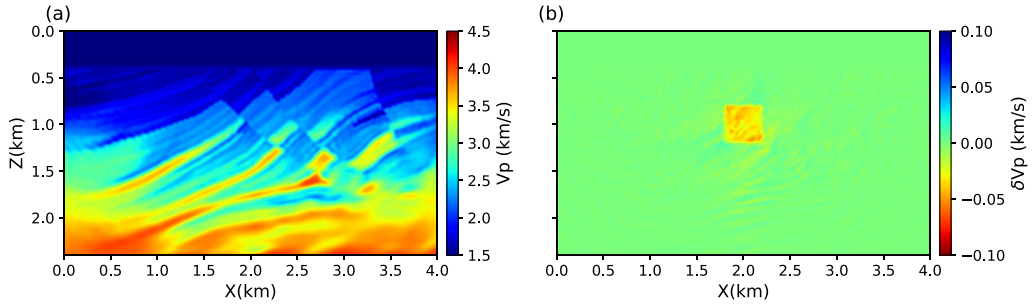

Figure 3. (a) Baseline velocity model obtained using standard linearised FWI. (b) Time-lapse velocity change obtained using the double difference method with identical source locations in the baseline and monitoring survey.

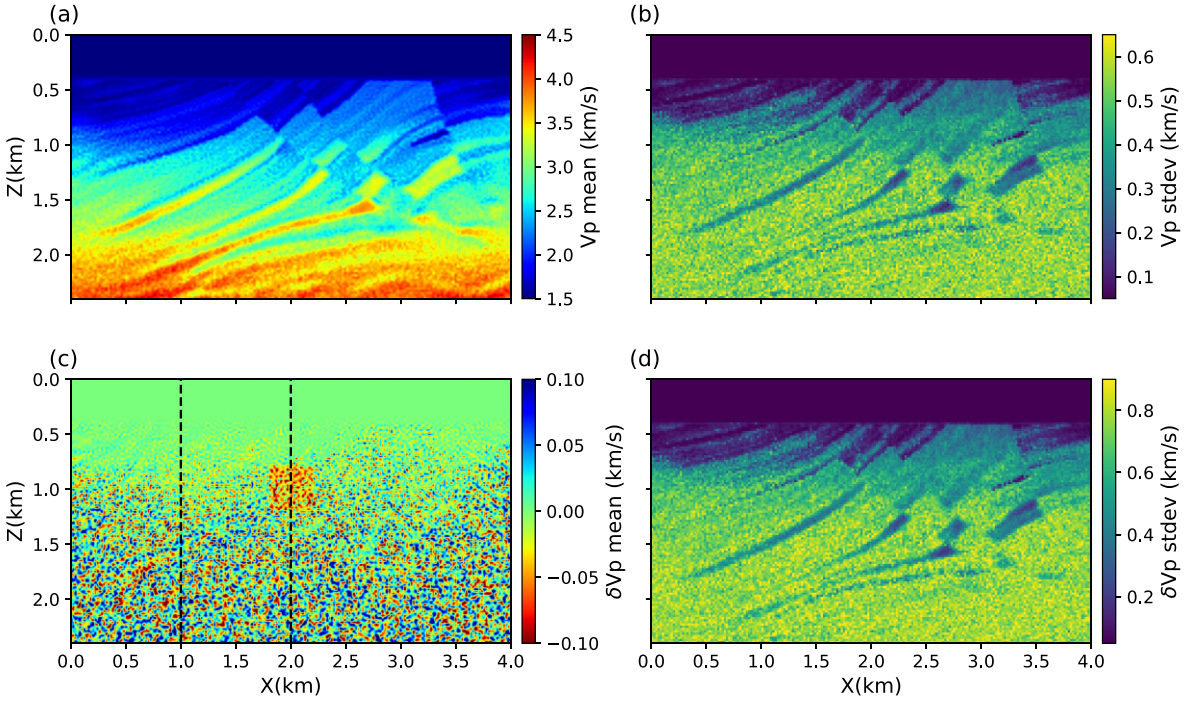

Figure 4. (a) The mean and (b) standard deviation of velocity obtained using sSVGD in the baseline survey. (c) The mean and (d) standard deviation of velocity change obtained using separate Bayesian inversions with identical source locations in the baseline and monitoring survey. The dashed black lines show well log locations referred to in the main text. Abbreviation stdev stands for standard deviation.

contains significant power in short spatial scale structures can generate identical low frequency data to those generated by a smoother model (Lyu *et al.* 2021). When inverting such data the parameter uncertainty should therefore contain models with small scale structure. As a result, McMC methods usually generate samples with pixel-scale randomness if no prior information of smoothness is imposed (Gebraad *et al.* 2020; Zhang & Curtis 2020b; Zhang *et al.* 2023). The model obtained using the linearised method is much smoother because smoothness was imposed as additional regularization. Overall the standard deviation map shows similar geometries to the mean, which has also been found in previous studies (Zhang & Curtis 2021; Gebraad *et al.* 2020; Zhang & Curtis 2020b). In addition, the results show higher uncertainties at large depths (>1.2 km) because of reduced data sensitivity, which is also consistent with results obtained using the SVGD method (Zhang & Curtis 2021).

For the monitoring inversion we restart the sampling in the above sSVGD from the final 20 particles using the new data set $\mathbf{d}_2$, and continue for another 2000 iterations. No burn-in period is specified for this inversion as the starting models are supposed to be close to the true model. In addition we only retain every fifth sample of particle values so that the total number of samples used is the same as that in the baseline inversion. To obtain samples of the time-lapse change, we randomly select pairs of samples from the two sets of model samples obtained in the baseline and monitoring inversion, and calculate the time-lapse change using $\delta\mathbf{m} = \mathbf{m}_2 - \mathbf{m}_1$. The statistics of the posterior distribution of time-lapse change can then be computed.

Figs 4(c) and (d) show the mean and standard deviation maps of time-lapse changes. As in the double difference inversion, the mean map clearly shows the outline of true velocity change. However, there are additional small scale structures (a few pixels in size) in the results, which may reflect the uncertainty in velocity differences, or may exist because of the stochastic nature of the algorithm. To further understand this, we conduct two independent inversions for the baseline survey using identical starting particles and the same number of iterations and step size, but a different pseudo-random number seed, and calculate the difference between

the obtained mean models (Fig. 5a). In addition, we also compute the difference between two mean models calculated using the first 5000 samples and the last 5000 samples from a single inversion (Fig. 5b). Fig. 5(a) shows that the difference obtained from two independent inversions have similar magnitudes of small scale structure as those in Fig. 4. By contrast, although the mean difference obtained from a single inversion (Fig. 5b) shows similar small scale structures, the magnitudes of those structures are much smaller. This comparison therefore demonstrates that those small scale structures in Fig. 4 probably reflect the stochastic nature of the algorithm. Note that this stochastic nature may exist because the algorithm has not fully converged given that this is a high dimensional problem ($r = 24\,000$), or may be caused by the discretization error in eq. (12). Either way, those structures do not significantly impact on the overall interpretation of the results from a geological point of view since the image speckle can largely be ignored. The standard deviation map shows almost the same structure as that obtained in the baseline inversion except that the magnitude is much higher. This is because the two inversions are conducted separately, and the variance of the time-lapse change is the summation of the variances of velocity obtained in each inversion (eq. 7b). As a result, the standard deviation model is not particularly useful in practice as the magnitude of uncertainty is far higher than that of the time-lapse change itself.

In the joint Bayesian inversion we use the same prior distribution for velocity (Fig. 2b) and add additional prior information about the velocity change (Fig. 2c). Similarly to above, we generate 20 particles from the prior distribution and update them using eq. (15) for 4000 iterations with an additional burn-in period of 4000. Only every tenth sample is retained which results in a total of 8000 samples. Other settings of the sSVGD method are kept the same as in the separate inversion strategy above. Finally, statistics of the posterior distribution $p(\mathbf{m}_1, \delta\mathbf{m}|\mathbf{d}_1, \mathbf{d}_2)$ are computed using the samples obtained.

Fig. 6 shows the results obtained in the joint inversion. The mean and standard deviation of velocity obtained for the baseline period (Figs 6a and b) show almost the same structures as those obtained in the separate Bayesian inversions (Figs 4a and b). For example, the mean also represents a good estimate of the true model and the standard deviation shows similar geometrical features to those on the mean map. Thus, although the number of parameters is doubled in the joint inversion compared to the separate inversion strategy, the method still provides good estimates of the baseline model.

Similarly to the results obtained above, the mean model of velocity change provides a reasonable estimate of the true velocity change (Fig. 6c). There are also small scale structures in the mean model as in the separate Bayesian inversion, which probably have similar origins. However, the magnitudes of those structures are much smaller than those from the separate Bayesian inversions. Note that there are also some negative velocity changes at the edges around the depth of 0.6 km and some geological structures close to the bottom around depth of 2.0 km that are associated with similar structures in the velocity model. This is probably because the data sets cause velocity and velocity change to be correlated with each other, and consequently uncertainty in velocity can introduce uncertainties to the velocity change. The standard deviation model indicates that the uncertainty estimates from joint inversion are almost an order of magnitude smaller than those obtained using separate inversions because of the additional prior information imposed on the velocity change. Similarly to the standard deviation of velocity, the uncertainty of velocity change is smaller in shallow parts ($<$1.0 km) and larger in deeper parts ($>$1.0 km) of the model. Note that because of coupling between velocity and velocity change, the magnitude of uncertainty in the deeper part is actually similar to that of the prior distribution (0.12 km s^{-1}). This indicates that the uncertainty in the baseline model can have a high impact on the uncertainty in velocity changes. Nevertheless, compared to the results obtained using separate inversions, the joint inversion strategy produces more accurate velocity changes and more reliable uncertainty estimates.

3.3 Perturbed acquisition geometry

We now study the performance of the suite of methods in the case in which source positions are perturbed in the monitoring survey (yellow stars in Fig. 2a). Although it is possible to use time-lapse binning or data interpolation to emulate repeatable data acquisition (Asnaashari *et al.* 2015), in this study none of these procedures is performed because of the sparseness of the source positions which make these procedures inaccurate, and also because our purpose is to study the reliability of different inversion methods under different geometries. In the double difference inversion we use the same baseline model as above and follow the same procedure to minimize the misfit function in eq. (3), accounting for the different source locations in baseline and monitoring surveys (i.e. assuming that these are known). The obtained time-lapse change is shown in Fig. 7. Although the shape of the true velocity change can be observed in the results, there are many additional structures which have similar magnitudes to the true velocity change but do not represent any real changes. Since these structures follow geological strata, they can certainly bias dynamic interpretations of the observed changes. We therefore conclude that double difference FWI generates significantly biased results in the case of perturbed acquisition geometries, even if the perturbed source locations are known, a result that has also been found in previous studies (Asnaashari *et al.* 2015; Yang *et al.* 2015).

For the separate Bayesian inversion strategy we follow the same procedure as described in the corresponding section above, except that different source locations and data are used in the monitoring inversion. The results are shown in Fig. 8. Overall the mean velocity change (Fig. 8c) shows similar features to that obtained in the case of fixed acquisition geometries. For example, the true velocity change can be observed clearly in the mean map and there are also many small scale random structures across area. Again the magnitude of standard deviation of velocity change is higher than that of velocity in the baseline inversion because of independence of the two inversions. A novel feature of these results is the set of dipping, slightly negative anomalies at depths $<$1 km, which are therefore attributed to the perturbation in source locations.

For joint Bayesian inversion we conduct the inversion in the same way as above to invert for the baseline model and velocity change simultaneously. Overall the results show almost the same mean and standard deviation maps to those obtained in the case of fixed acquisition geometry for both the baseline velocity and velocity change (Fig. 9).

The above results show that the Bayesian methods are more stable with respect to variations in the acquisition geometry than the traditional double-difference algorithm. Furthermore, compared to the results obtained using separate Bayesian inversions, the results obtained using joint inversion show more accurate velocity changes and more reliable uncertainties because of the additional prior information imposed.

To further understand the results, in Fig. 10 we show marginal distributions of velocity change obtained using the two Bayesian

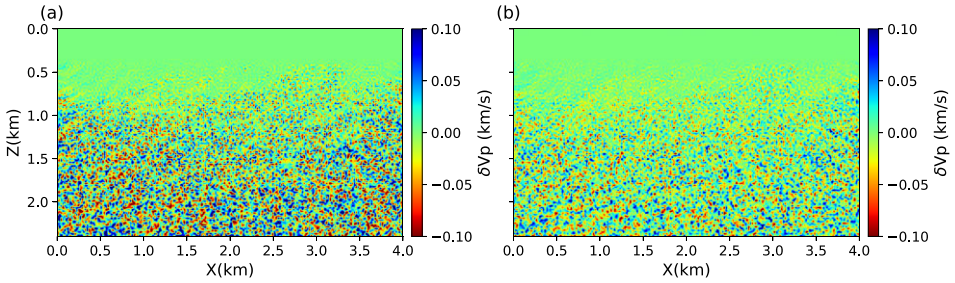

Figure 5. The difference between mean models obtained (a) from two independent inversions and (b) from the first 5000 samples and the last 5000 samples after the burn-in period in a single inversion.

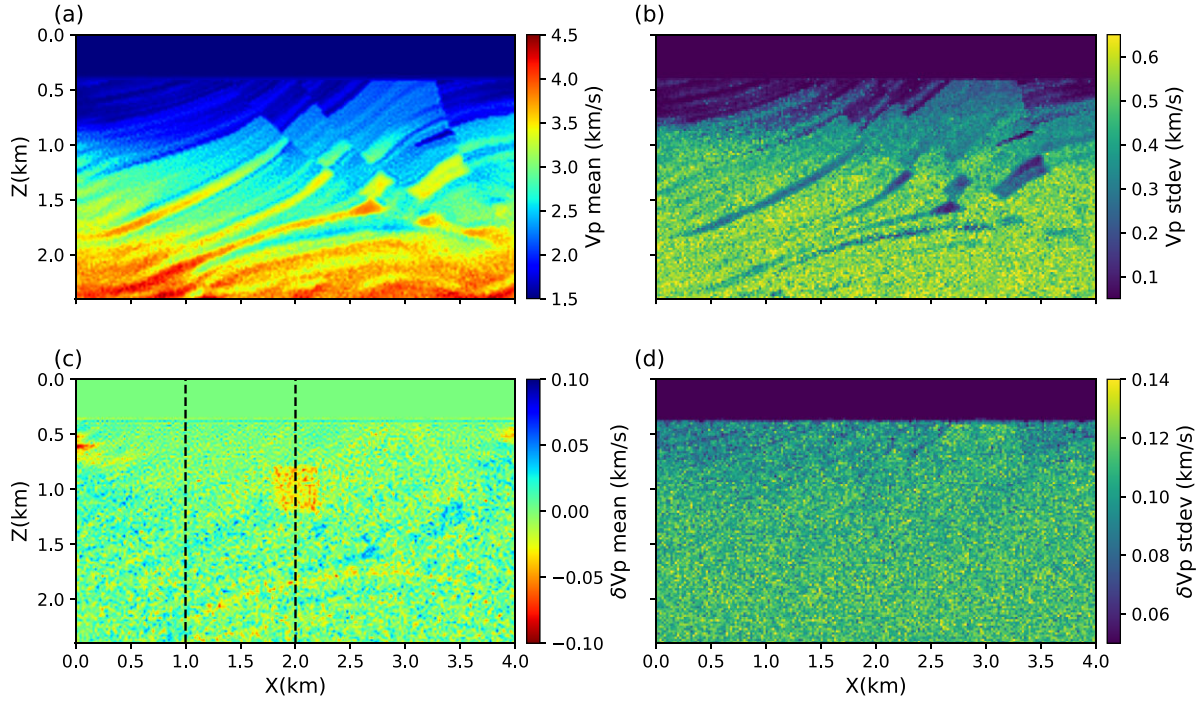

Figure 6. The mean and standard deviation of velocity (top panels) and velocity change (bottom panels) obtained using the joint Bayesian inversion strategy. Key as in Fig. 4.

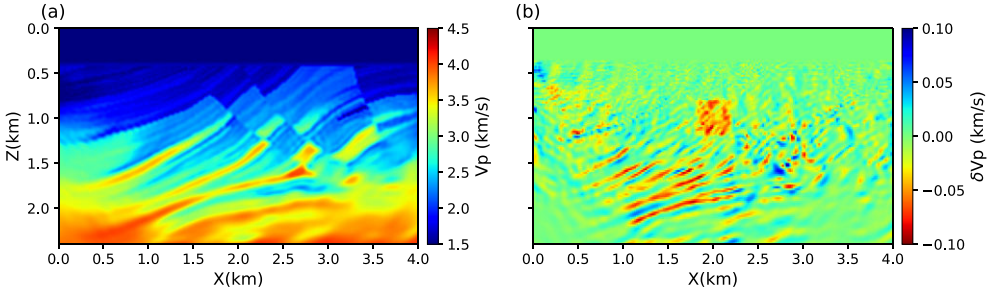

Figure 7. (a) The baseline velocity model obtained using the standard linearized method. (b) The time-lapse velocity change obtained using the double difference method with perturbed source locations in the monitoring survey (yellow stars in Fig. 2a).

methods in the different cases along two vertical profiles whose locations are denoted by black dashed lines in Figs 4, 6, 8 and 9. Similarly to above, the results obtained using separate Bayesian inversions show significantly broader distributions than those obtained using joint inversion because of the assumed independence between baseline and monitoring inversions in the former inversion strategy. For all results, the shallow part (<1.0 km) has lower uncertainty than deeper parts. Although the standard deviation models obtained using the joint inversion do not show lower uncertainty within the zone of velocity changes (Figs 6d and 9d), the marginal distributions clearly reflect lower velocity within the area (Figs 10b and d) which suggests that the velocity change is well constrained by the data and the prior information. By contrast, it is difficult to notice lower velocity from the distributions obtained using the separate inversion strategy because of their high uncertainty. Note that the marginal distributions obtained using the joint inversion show high

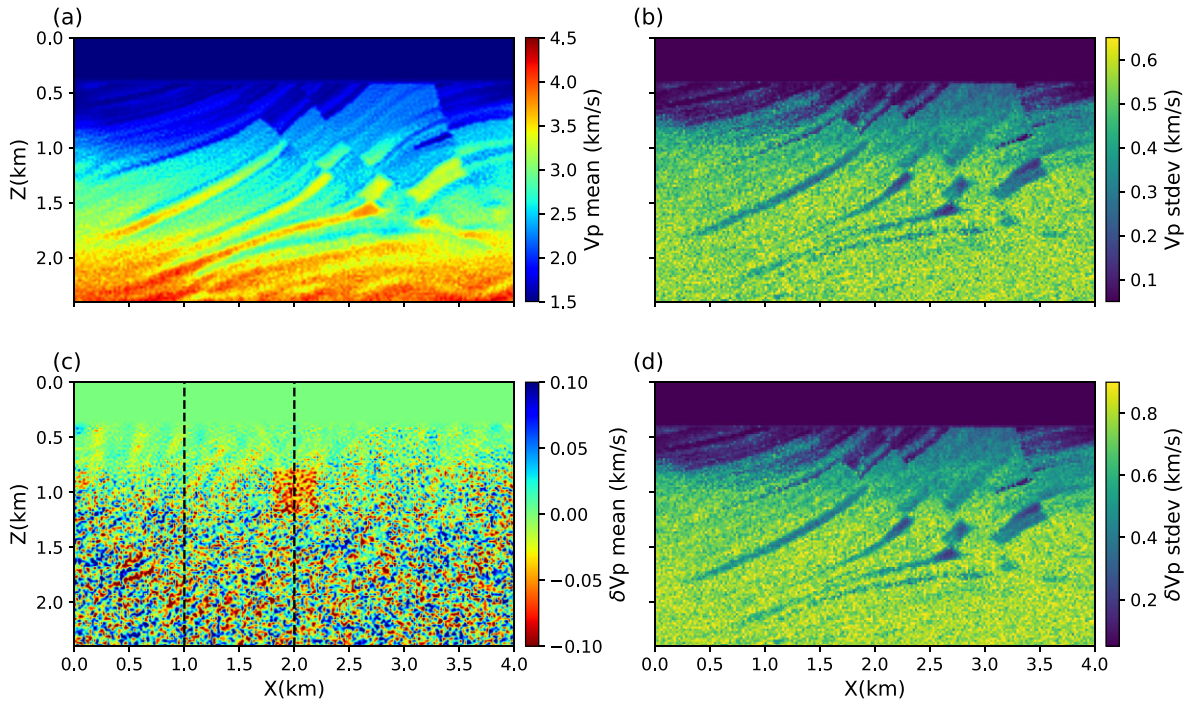

Figure 8. The mean and standard deviation of velocity (top panels) and velocity change (bottom panels) obtained using the separate Bayesian inversion strategy with perturbed source locations. Key as in Fig. 4.

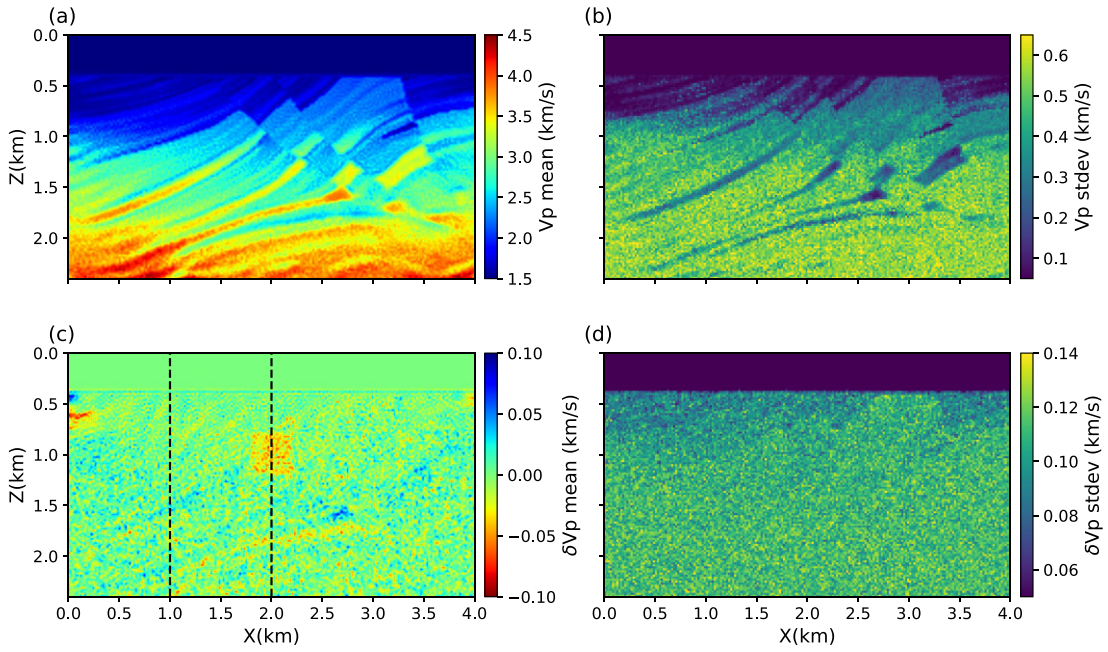

Figure 9. The mean and standard deviation of velocity (top panels) and velocity change (bottom panels) obtained using the joint Bayesian inversion strategy with perturbed source locations. Key as in Fig. 4.

probability density values at the boundaries of the prior distribution. This may be because the velocity change would have higher uncertainties if weaker prior information was imposed (i.e. the change is not well constrained by the data itself). As a result, when tight Uniform prior distributions are imposed, the mass of marginal distributions that would otherwise lie outside of the support of the Uniform distribution, concentrates close to its boundaries. It may also be possible that this is caused by biases of the algorithm itself, for example, the finite step we used in eq. (15); in practice the step length is always restricted by available computation power, and we have used the smallest size that was feasible.

3.4 Computational cost

We summarize the number of simulations required by each method in Table 1. This provides a good metric of the overall computational cost because the forward and adjoint simulations are the

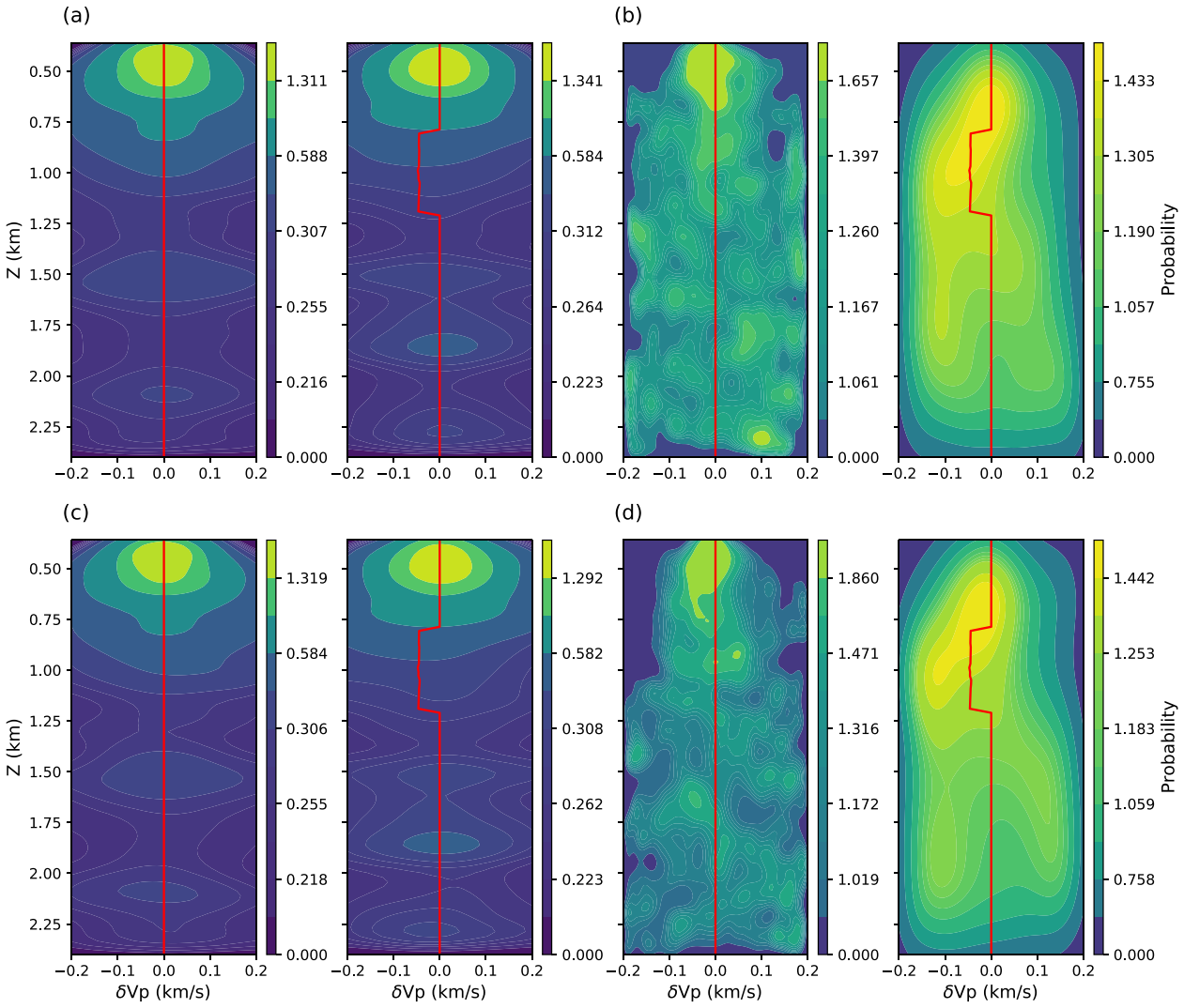

Figure 10. The marginal distributions of velocity change at two well locations (black dashed line in Figs 4, 6, 8 and 9) obtained using (a, c) separate Bayesian inversions and (b, d) joint Bayesian inversion in the case of fixed (top panels) and perturbed (bottom panels) acquisition geometry. The distributions are estimated using the kernel density estimate method from posterior samples (Parzen 1962). Contours are plotted at the same level of probability mass in the different panels. Red lines denote the true velocity change.

Table 1. A comparison of computational cost for the suite of inversion methods.

Method	Number of simulations	
Double difference inversion	baseline:	163
	monitoring:	50
Separate Bayesian inversion	baseline:	6000
	monitoring:	2000
Joint Bayesian inversion		16 000

most time-consuming components for each method. Note that because the inversions for fixed and perturbed acquisition geometries are conducted in the same way which results in the same number of simulations, we do not discriminate between the two cases in the table. Apparently the traditional double difference inversion is the most efficient method, but it cannot produce accurate uncertainty estimates, and it provides biased estimates when the acquisition geometry changes between surveys. The two Bayesian methods require significantly more computation than the double difference method. In addition, because in the joint inversion we simulate the baseline and monitoring data together at each iteration, the required number of simulations (16 000) is twice that required in the separate Bayesian inversion (8000) even though the two inversions are conducted using the same number of iterations. However, the separate Bayesian inversion strategy does not provide useful uncertainty estimates for the velocity change due to the assumed independence of baseline and monitoring inversions. By contrast, the results obtained using joint inversion provide more accurate and useful uncertainty estimates because the method can take advantage of additional prior information on the velocity change itself. In addition, compared to the double difference inversion, both Bayesian methods provide stable and accurate mean velocity change estimates in the case of either fixed or perturbed acquisition geometries.

Note that the above comparison depends on subjective assessments of convergence for each method, so the absolute computational time required by each method may not be entirely accurate. Nevertheless the comparison at least provides a reasonable insight into the efficiency of each method. To give an overall idea of the time required by the two Bayesian methods, the above inversions

required 65 hr and 111 hr in wall time for the separate and joint inversions, respectively, both of which are parallelized using 40 AMD EPYC CPU cores.

4 DISCUSSION

We demonstrated that Bayesian methods (separate Bayesian inversions of baseline and monitoring surveys, and a joint Bayesian inversion) can be used to detect velocity changes and quantify uncertainty for time-lapse inversions, and that they provide more accurate results than the traditional double difference method in the case where acquisition geometries were changed between the two surveys, even when the locations of sources and receivers were known exactly in each survey. This is because in double difference inversion the unexplained events of the baseline survey data are not compensated by the residual term $r(\mathbf{m}_1) = \mathbf{u}(\mathbf{m}_1) - \mathbf{d}_1$ in the new data $\mathbf{d}_2'$ due to the change of source locations. As a result, those unexplained events can still affect the final time-lapse results. If the baseline model is perfect, that is, there are no unexplained events in the baseline survey data, the time-lapse change can be detected clearly even with perturbed acquisition geometries (Asnaashari *et al.* 2015). By contrast, Bayesian methods characterize the full Bayesian posterior distributions of seismic velocity in the baseline and monitoring surveys, or the full posterior distribution of velocity change between the two surveys. In either case, the obtained distribution contains information of time-lapse changes, regardless of whether there are perturbations in the acquisition geometries.

Bayesian methods are therefore particularly valuable when high repeatability of acquisition geometry is difficult to achieve or emulate by interpolation, for example, when source or receiver geometries are sparse. When using dense acquisition systems, time-lapse binning and data interpolation are usually applied to improve acquisition repeatability in standard double difference inversions, and the same can be applied in the Bayesian methods. In addition, the standard double difference inversion method demands an accurate baseline model, which may require more effort to build than in the Bayesian methods. For example, in the above study we inverted an extra low frequency data set in order to build an accurate baseline model, which may not always be available in practice.

In this study we used a Uniform prior distribution on seismic velocities with a relatively large support (2 km s^{-1}), which leads to high uncertainty for velocity and consequently high uncertainty for velocity change. In practice where more knowledge about the subsurface is available, one can use a more informative prior distribution for the velocity. This will produce more accurate models and lower uncertainty for both velocity and velocity change. Note that when conducting separate Bayesian inversions the obtained uncertainty for velocity change is always larger than that obtained for velocity because of the implicit assumption of independence of baseline and monitoring inversions. In the joint inversion the velocity and velocity change are explicitly coupled, so strong prior information on velocity can also improve the accuracy of velocity change estimates. And of course, if an accurate baseline model is available and can be fixed during the inversion, one can also use the differential data between monitoring and baseline surveys and invert for the velocity change directly in the Bayesian inversion as in the standard double difference inversion (Kotsi *et al.* 2020b). In addition, one can add some prior regularization as in linearised inversions, for example, by imposing prior information on the difference between neighbouring cells. Gaussian process may also be used to inject prior information into the inference scheme (Ray & Myer 2019), or one can use neural networks to encode geological information as prior distributions (Laloy *et al.* 2017; Mosser *et al.* 2020).

We estimated the velocity change for the entire model area. This requires full model simulations during the inversion which can be computationally inefficient. If knowledge about locations of potential velocity change zones are available, one can also perform target oriented time-lapse inversions by assuming that the rest of the model is known. Then a local solver can be used to increase efficiency (Asnaashari *et al.* 2015; Kotsi *et al.* 2020b). Alternatively, if it is not possible to perform target oriented inversion in practice, one can also use other faster, approximate forward modelling methods to improve efficiency, for example, neural network based modelling methods (Sirignano & Spiliopoulos 2018; Moseley *et al.* 2021).

Although Bayesian inversion can produce more accurate results than the standard double difference inversion and can quantify uncertainty, it is also significantly more computational expensive. To improve efficiency of the methods, one might exploit high order gradient information, for example using a Hessian kernel function (Wang *et al.* 2019) or the stochastic Stein variational Newton method (Leviyev *et al.* 2022). In addition, one can also use stochastic inversion by dividing the whole data set into minibatches to reduce the computation cost as demonstrated by Zhang *et al.* (2023).

The results obtained here may contain biases. For example, the small random structures in the velocity change model obtained using separate Bayesian inversions and those structures in the deeper part of the model obtained using joint inversion may constitute genuine biases due to lack of convergence of the algorithm. To further improve accuracy of the results, one may run the sSVGD algorithm longer. In addition, the discretization used in equation (eq. 12) may cause errors and biases in results; a Metropolis–Hastings correction step can be added at each iteration (Metropolis & Ulam 1949; Hastings 1970).

Although in this study we only applied Bayesian methods to 2-D time-lapse change problems, the method should also be applicable to 3-D cases since the sSVGD algorithm has already been used to solve 3-D Bayesian FWI problems (Zhang *et al.* 2023). However, because of the extremely high dimensional parameter space, it may not be easy for sSVGD to converge sufficiently, and consequently the time-lapse change may be difficult to obtain. In such cases stronger prior information on velocity might be required in order to detect velocity changes since this will reduce the computational complexity of converging to the solution. Alternatively one may try to reduce the dimensionality of the problem itself. For example, other parametrizations which use fewer parameters to represent the model may be used such as Voronoi tessellation (Bodin & Sambridge 2009; Zhang *et al.* 2018b), Delaunay and Clough–Tocher parametrization (Curtis & Snieder 1997), wavelet parameterization (Hawkins & Sambridge 2015), discrete cosine transform (Kotsi *et al.* 2020a; Urozayev *et al.* 2022) and neural network parameterization (Laloy *et al.* 2017; Mosser *et al.* 2020; Bloem *et al.* 2022). Other methods which project high dimensional spaces into lower dimension space may also be used to improve efficiency of the methods, for example, slice SVGD (Gong *et al.* 2020) or projected SVGD (Chen & Ghattas 2020).

5 CONCLUSION

In this study, we explored two Bayesian inversion strategies: separate Bayesian inversions for baseline and monitoring surveys, and

a joint Bayesian inversion of both survey data sets to solve time-lapse FWI problems. We compared the results to those obtained using standard double difference inversion. The results show that all methods can provide accurate velocity change estimate in the case of fixed acquisition geometries, but in the case of perturbed acquisition geometries the two Bayesian methods produce significantly more accurate results than double difference inversion. In addition, Bayesian methods provide uncertainty estimates that account for the full non-linearity of the model-data relationships, and any form of prior probability and data uncertainty distributions, which cannot be obtained using double difference inversion. However, when using the separate Bayesian inversion strategy the assumed independence between baseline and monitoring inversions causes the magnitude of the uncertainty estimate for velocity change to be higher than that for velocity itself, which makes the results less useful in practice. By contrast, the uncertainty estimates for velocity change from a single, joint Bayesian inversion are almost an order of magnitude smaller than those obtained from separate inversions because of additional prior information that can be imposed on the velocity change. This demonstrates that the joint inversion provides more accurate uncertainty estimates as the magnitude of velocity change is usually much smaller than that of velocity. We therefore conclude that Bayesian time-lapse FWI, especially joint Bayesian inversion, can be used to detect velocity change and to quantify associated uncertainties in time-lapse inversion and monitoring. We also note that the method can be used to target the region of interest in results obtained using convectional double difference inversion.

DATA AVAILABILITY

The model used in this article is available in SEG Wiki at https://wiki.seg.org/wiki/AGL_Elastic_Marmousi. The code underlying this article will be shared on reasonable request to the corresponding author.

ACKNOWLEDGMENTS

The authors thank the Edinburgh Imaging Project sponsors (BP and Total), National Natural Science Foundation of China (42204055) and the Fundamental Research Funds for the Central Universities for supporting this research. This work has made use of the resources provided by the Edinburgh Compute and Data Facility (http://www.ecdf.ed.ac.uk/).

REFERENCES

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P. & Virieux, J., 2015. Time-lapse seismic imaging using regularized full-waveform inversion with a prior model: Which strategy?, *Geophys. Prospect.,* **63**(1), 78–98.

Beasley, C.J., Chambers, R., Workman, R., Craft, K. & Meister, L., 1999. Repeatability of 3-D ocean bottom cable seismic surveys, in *Proceedings of the 6th International Congress of the Brazilian Geophysical Society,* European Association of Geoscientists & Engineers.

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

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

Bloem, H., Curtis, A. & Tetzlaff, D., 2022. Introducing conceptual geological information into bayesian tomographic imaging, *Basin Research* **36**(1), e12811.

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

Brooks, S., Gelman, A., Jones, G. & Meng, X.-L., 2011. *Handbook of Markov Chain Monte Carlo,* CRC Press.

Bunks, C., Saleck, F., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics,* **60**(5), doi:10.1190/1.1443880.

Calvert, R., 2005. 4d technology: where are we, and where are we going?, *Geophys. Prospect.,* **53**(2), 161–171.

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

Chen, T., Fox, E. & Guestrin, C., 2014. Stochastic gradient Hamiltonian Monte Carlo, in *Proceedings of the International Conference on Machine Learning,* pp. 1683–1691, PMLR.

Curtis, A. & Snieder, R., 1997. Reconditioning inverse problems using the genetic algorithm and revised parameterization, *Geophysics,* **62**(5), 1524–1532.

Denli, H. & Huang, L., 2009. Double-difference elastic waveform tomography in the time domain, in *Proceedings of the 2009 SEG Annual Meeting,* OnePetro.

Dosso, S.E., Holland, C.W. & Sambridge, M., 2012. Parallel tempering for strongly nonlinear geoacoustic inversion, *J. acoust. Soc. Am.,* **132**(5), 3030–3040.

Duane, S., Kennedy, A.D., Pendleton, B.J. & Roweth, D., 1987. Hybrid Monte Carlo, *Phys. Lett. B,* **195**(2), 216–222.

Fichtner, A., Bunge, H.-P. & Igel, H., 2006. The adjoint method in seismology: I. Theory, *Phys. Earth planet. Inter.,* **157**(1–2), 86–104.

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, *J. geophys. Int.,* **179**(3), 1703–1725.

Fichtner, A., Zunino, A. & Gebraad, L., 2018. Hamiltonian Monte Carlo solution of tomographic inverse problems, *J. geophys. Int.,* **216**(2), 1344–1363.

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

Galetti, E., Curtis, A., Meles, G.A. & Baptie, B., 2015. Uncertainty loops in travel-time tomography from nonlinear wave physics, *Phys. Rev. Lett.,* **114**(14), doi:10.1103/PhysRevLett.114.148501.

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

Gebraad, L., Boehm, C. & Fichtner, A., 2020. Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo, *J. geophys. Res.,* **125**(3), e2019JB018428.

Gong, W., Li, Y. & Hernández-Lobato, J.M., 2020. Sliced kernelized stein discrepancy, preprint (arXiv:2006.16531).

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Byesian model determination, *Biometrika,* **82**(4), 711–732.

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

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

Hawkins, R. & Sambridge, M., 2015. Geophysical imaging using trans-dimensional trees, *J. geophys. Int.,* **203**(2), 972–1000.

Hukushima, K. & Nemoto, K., 1996. Exchange Monte Carlo method and application to spin glass simulations, *J. Phys. Soc. Jpn.,* **65**(6), 1604–1608.

Kotsi, M., Malcolm, A. & Ely, G., 2020a. Time-lapse full-waveform inversion using Hamiltonian Monte Carlo: a proof of concept, in *Proceedings of the SEG Technical Program Expanded Abstracts 2020,* pp. 845–849, Society of Exploration Geophysicists. Online

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

Kubrusly, C. & Gravier, J., 1973. Stochastic approximation algorithms and applications, in *Proceedings of the 1973 IEEE Conference on Decision and Control Including the 12th Symposium on Adaptive Processes,* pp. 763–766, IEEE.

Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A. & Blei, D.M., 2017. Automatic differentiation variational inference, *J. Mach. Learn. Res.,* **18**(1), 430–474.

Kullback, S. & Leibler, R.A., 1951. On information and sufficiency, *Ann. Math. Stat.,* **22**(1), 79–86.

Laloy, E., Hérault, R., Lee, J., Jacques, D. & Linde, N., 2017. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network, *Adv. Water Resour.,* **110,** 387–405.

Lei, W. *et al.*, 2020. Global adjoint tomography—model glad-m25, *J. geophys. Int.,* **223**(1), 1–21.

Leviyev, A., Chen, J., Wang, Y., Ghattas, O. & Zimmerman, A., 2022. A stochastic stein variational newton method, preprint (arXiv:2204.09039).

Liu, D.C. & Nocedal, J., 1989. On the limited memory bfgs method for large scale optimization, *Math. Program.,* **45**(1-3), 503–528.

Liu, Q. & Wang, D., 2016. Stein variational gradient descent: a general purpose Byesian inference algorithm, in *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS 2016),* Barcelona, Spain, pp. 2378–2386.

Lumley, D.E., 2001. Time-lapse seismic reservoir monitoring, *Geophysics,* **66**(1), 50–53.

Lyu, C., Capdeville, Y., Al-Attar, D. & Zhao, L., 2021. Intrinsic non-uniqueness of the acoustic full waveform inverse problem, *J. geophys. Int.,* **226**(2), 795–802.

Ma, Y.-A., Chen, T. & Fox, E., 2015. A complete recipe for stochastic gradient MCMC, in *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems,* Vol. **2,** pp. 2917–2925.

Maharramov, M. & Biondi, B., 2014. Joint full-waveform inversion of time-lapse seismic data sets, in *SEG Technical Program Expanded Abstracts 2014,* pp. 954–959, Society of Exploration Geophysicists.

Martin, G.S., Wiley, R. & Marfurt, K.J., 2006. Marmousi2: an elastic upgrade for Marmousi, *Leading Edge,* **25**(2), 156–166.

Martin, J., Wilcox, L.C., Burstedde, C. & Ghattas, O., 2012. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion, *SIAM J. Sci. Comput.,* **34**(3), A1460–A1487.

Metropolis, N. & Ulam, S., 1949. The Monte Carlo method, *J. Am. Stat. Assoc.,* **44**(247), 335–341.

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

Moseley, B., Markham, A. & Nissen-Meyer, T., 2021. Finite basis physics-informed neural networks (fbpinns): a scalable domain decomposition approach for solving differential equations, *Adv. Comput. Math.,* **49**(4), 62.

Mosser, L., Dubrule, O. & Blunt, M.J., 2020. Stochastic seismic waveform inversion using generative adversarial networks as a geological prior, *Math. Geosci.,* **52**(1), 53–79.

Nawaz, M.A. & Curtis, A., 2018. Variational Bayesian inversion (VBI) of quasi-localized seismic attributes for the spatial distribution of geological facies, *J. geophys. Int.,* **214**(2), 845–875.

Nawaz, M.A. & Curtis, A., 2019. Rapid discriminative variational Byesian inversion of geophysical data for the spatial distribution of geological properties, *J. geophys. Res.,* **124**(6), 5867–5887.

Nawaz, M.A., Curtis, A., Shahraeeni, M.S. & Gerea, C., 2020. Variational Bayesian inversion of seismic attributes jointly for geological facies and petrophysical rock properties, *Geophysics,* **85**(4), doi:10.1190/geo2019-0163.1.

O'Hagan, A. & Forster, J.J., 2004. *Kendall's Advanced Theory of Statistics,* Vol. 2B: Bayesian Inference, Arnold.

Parzen, E., 1962. On estimation of a probability density function and mode, *Ann. Math. Stat.,* **33**(3), 1065–1076.

Plessix, R.-E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *J. geophys. Int.,* **167**(2), 495–503.

Plessix, R.-E., Michelet, S., Rynja, H., Kuehl, H., Perkins, C., de Maag, J. & Hatchell, P., 2010. Some 3d applications of full waveform inversion, in *Proceedings of the 72nd EAGE Conference and Exhibition-Workshops and Fieldtrips,* European Association of Geoscientists & Engineers.

Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain. Part 1: theory and verification in a physical scale model, *Geophysics,* **64**(3), 888–901.

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, *J. geophys. Int.,* **194**(3), 1640–1664.

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

Ray, A., Kaplan, S., Washbourne, J. & Albertin, U., 2017. Low frequency full waveform seismic inversion within a tree based Byesian framework, *J. geophys. Int.,* **212**(1), 522–542.

Rezende, D.J. & Mohamed, S., 2015. Variational inference with normalizing flows, *Proceedings of the 32nd International Conference on Machine Learning,* **37,** 1530–1538. Lille, France. PMLR.

Robbins, H. & Monro, S., 1951. A stochastic approximation method, *Ann. Math. Stat.,* **22**(3), 400–407.

Roberts, G.O., Tweedie, R.L. *et al.*, 1996. Exponential convergence of Langevin distributions and their discrete approximations, *Bernoulli,* **2**(4), 341–363.

Sambridge, M., 2013. A parallel tempering algorithm for probabilistic sampling and multimodal optimization, *J. geophys. Int.,* **196**(1), doi:10.1093/gji/ggt342.

Sambridge, M. & Mosegaard, K., 2002. Monte carlo methods in geophysical inverse problems, *Rev. Geophys.,* **40**(3), 3–1.

Sen, M.K. & Biswas, R., 2017. Transdimensional seismic inversion using the reversible jump Hamiltonian Monte Carlo algorithm, *Geophysics,* **82**(3), R119–R134.

Siahkoohi, A., Rizzuti, G. & Herrmann, F.J., 2020a. Uncertainty quantification in imaging and automatic horizon tracking—a Bayesian deep-prior based approach, in *Proceedings of the SEG Technical Program Expanded Abstracts 2020,* pp. 1636–1640, Society of Exploration Geophysicists.

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

Sirignano, J. & Spiliopoulos, K., 2018. Dgm: A deep learning algorithm for solving partial differential equations, *J. Comput. Phys.,* **375,** 1339–1364.

Smith, J.D., Ross, Z.E., Azizzadenesheli, K. & Muir, J.B., 2022. Hyposvi: hypocentre inversion with stein variational inference and physics informed neural networks, *J. geophys. Int.,* **228**(1), 698–710.

Smith, R.C., 2013. *Uncertainty Quantification: Theory, Implementation, and Applications,* Vol. **12,** SIAM.

Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the southern California crust, *Science,* **325**(5943), 988–992.

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

Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, in *Scattering and Attenuations of Seismic Waves, Part I,* pp. 365–399, Springer.

Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *J. geophys. Int.,* **160**(1), 195–216.

Urozayev, D., Ait-El-Fquih, B., Hoteit, I. & Peter, D., 2022. A reduced-order variational Bayesian approach for efficient subsurface imaging, *J. geophys. Int.,* **229**(2), 838–852.

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

Wang, D., Tang, Z., Bajaj, C. & Liu, Q., 2019. Stein variational gradient descent with matrix-valued kernels, in *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019),* Vancouver, Canada., pp. 7836–7846.

Watanabe, T., Shimizu, S., Asakawa, E. & Matsuoka, T., 2004. Differential waveform tomography for time-lapse crosswell seismic data with application to gas hydrate production monitoring, in *Proceedings of the SEG Technical Program Expanded Abstracts 2004,* pp. 2323–2326, Society of Exploration Geophysicists, Denver, Colorado, USA.

Welling, M. & Teh, Y.W., 2011. Bayesian learning via stochastic gradient Langevin dynamics, in *Proceedings of the 28th International Conference on Machine Learning (ICML-11),* pp. 681–688, Citeseer.

Yang, D., Malcolm, A. & Fehler, M., 2014. Time-lapse full waveform inversion and uncertainty analysis with different survey geometries, in *Proceedings of the 76th EAGE Conference and Exhibition 2014,* Vol. **2014,** pp. 1–5, EAGE Publications BV.

Yang, D., Meadows, M., Inderwiesen, P., Landa, J., Malcolm, A. & Fehler, M., 2015. Double-difference waveform inversion: Feasibility and robustness study with pressure data, *Geophysics,* **80**(6), M129–M141.

Zhang, C. & Chen, T., 2022. Bayesian slip inversion with automatic differentiation variational inference, *J. geophys. Int.,* **229**(1), 546–565.

Zhang, C., Bütepage, J., Kjellström, H. & Mandt, S., 2018a. Advances in variational inference, *IEEE Trans. Pattern Anal. Mach. Intell.,* **41**(8), 2008–2026.

Zhang, X. & Curtis, A., 2020a. Seismic tomography using variational inference methods, *J. geophys. Res.,* **125**(4), e2019JB018589.

Zhang, X. & Curtis, A., 2020b. Variational full-waveform inversion, *J. geophys. Int.,* **222**(1), 406–411.

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

Zhang, X., Curtis, A., Galetti, E. & de Ridder, S., 2018b. 3-D Monte Carlo surface wave tomography, *J. geophys. Int.,* **215**(3), 1644–1658.

Zhang, X., Lomas, A., Zhou, M., Zheng, Y. & Curtis, A., 2023. 3-D Bayesian variational full waveform inversion, *J. geophys. Int.,* **234**(1), 546–561.

Zhao, X., Curtis, A. & Zhang, X., 2021. Bayesian seismic tomography using normalizing flows, *J. geophys. Int.,* **228**(1), 213–239.

Zhao, Z. & Sen, M.K., 2019. A gradient based MCMC method for FWI and uncertainty analysis, in *Proceedings of the SEG Technical Program Expanded Abstracts 2019,* pp. 1465–1469, Society of Exploration Geophysicists.

Zheng, Y., Barton, P. & Singh, S., 2011. Strategies for elastic full waveform inversion of time-lapse ocean bottom cable (OBC) seismic data, in *Proceedings of the SEG Technical Program Expanded Abstracts 2011,* pp. 4195–4200, Society of Exploration Geophysicists.