Received: 27 January 2024 | Accepted: 1 August 2024

DOI: 10.1111/1365-2478.13595

Geophysical Prospecting | EAGE EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS

ORIGINAL ARTICLE

A probabilistic full waveform inversion of surface waves

Sean Berti[1,2] | **Mattia Aleardi**[1] | **Eusebio Stucchi**[1]

[1]Earth Sciences Department, University of Pisa, Pisa, Italy

[2]Earth Sciences Department, University of Florence, Firenze, Italy

Correspondence
Sean Berti, Earth Sciences Department, University of Pisa, via s. Maria 53, 56126 Pisa, Italy. Email: sean.berti@unifi.it

Abstract

Over the past decades, surface wave methods have been routinely employed to retrieve the physical characteristics of the first tens of meters of the subsurface, particularly the shear wave velocity profiles. Traditional methods rely on the application of the multichannel analysis of surface waves to invert the fundamental and higher modes of Rayleigh waves. However, the limitations affecting this approach, such as the 1D model assumption and the high degree of subjectivity when extracting the dispersion curve, motivate us to apply the elastic full-waveform inversion, which, despite its higher computational cost, enables leveraging the complete information embedded in the recorded seismograms. Standard approaches solve the full-waveform inversion using gradient-based algorithms minimizing an error function, commonly measuring the misfit between observed and predicted waveforms. However, these deterministic approaches lack proper uncertainty quantification and are susceptible to get trapped in some local minima of the error function. An alternative lies in a probabilistic framework, but, in this case, we need to deal with the huge computational effort characterizing the Bayesian approach when applied to non-linear problems associated with expensive forward modelling and large model spaces. In this work, we present a gradient-based Markov chain Monte Carlo full-waveform inversion where we accelerate the sampling of the posterior distribution by compressing data and model spaces through the discrete cosine transform. Additionally, a proposal is defined as a local, Gaussian approximation of the target density, constructed using the local Hessian and gradient information of the log posterior. We first validate our method through a synthetic test where the velocity model features lateral and vertical velocity variations. Then we invert a real dataset from the InterPACIFIC project. The obtained results prove the efficiency of our proposed algorithm, which demonstrates to be robust against cycle-skipping issues and able to provide reasonable uncertainty evaluations with an affordable computational cost.

KEYWORDS
elastics, full-waveform, inversion, seismics

INTRODUCTION

In recent years, full-waveform inversion (FWI) has gained significant popularity owing to its capability in generating high-resolution estimates of subsurface parameters such as compressional and shear wave velocities (Virieux & Operto, 2009).

In FWI, seismic waveforms are exploited to update subsurface model parameters by trying to match the recorded data with the estimated data. Although acoustic FWI is widely employed for imaging complex subsurface structures, its limitation lies in exclusively reconstructing the P-wave velocity model, disregarding the S-wave. However, seismic data often exhibit elastic effects, particularly in media with high-contrast interfaces due to the generation of converted waves. Additionally, in the case of land data, a substantial portion of seismic energy generated by the source propagates as surface waves (Rayleigh or Love waves). In the near-surface context, the elastic inversion of Rayleigh waves assumes particular importance for reconstructing the shear wave velocity model within the initial tens of meters of the subsurface. This significance arises due to the remarkable sensitivity of Rayleigh waves to variations of this elastic parameter.

Conventional methods for surface wave inversion, such as the multichannel analysis of surface waves (MASW) (Aleardi & Stucchi, 2021; Maraschini & Foti, 2010; Socco & Strobbia, 2004; Socco et al., 2010), involve three primary steps:

1. Conversion from the space–time domain to the frequency–velocity domain, often represented as the velocity spectrum.
2. Interpretation of the velocity spectrum to extract the dispersion curve corresponding to the fundamental mode.
3. Inversion of the surface-wave dispersion curve to derive the V_s model, representing the subsurface shear-wave velocity distribution.

However, these methods encounter several limitations, similar to those affecting the shear wave dispersion analysis (Bodin et al., 2012): First, the extraction of the dispersion curve usually relies on user interpretation, thus introducing a degree of subjectivity in the inversion procedure. In addition, picking at very low frequencies can be inaccurate, introducing uncertainties in the depth position and velocity of the deepest layers (Foti et al., 2014, 2018). Another challenge lies in the identification of higher modes that can be more energetic than the fundamental mode, especially in the case of velocity reversals and/or strong vertical velocity contrasts (De Nil, 2005; Gao et al., 2016; Pan et al., 2019). Furthermore, the MASW method hinges on the 1D assumption of the model, which may not hold for many practical applications. Therefore, in the presence of lateral velocity variations in the subsurface, alternative approaches have been proposed. For example, the recorded data can be divided into subsets, and each subset is used to derive a 1D model. Then the derived velocity profiles are merged to construct a pseudo-2D subsurface model (Bergamo & Socco, 2014; Bohlen et al., 2004; O'Neill et al., 2008; Strobbia et al., 2006). Nevertheless, there are still limitations regarding the lateral resolution of the obtained 2D models, given that the profile length of each subset cannot be arbitrarily small.

In this context, employing more advanced techniques like elastic FWI becomes imperative for enhancing the lateral resolution and the fidelity of the estimated velocity field. There are several successful applications of FWI in exploration geophysics and seismology (Fichtner et al., 2009; Brossier et al., 2009; Warner & Gausch, 2016), and in the last few years, numerous studies have also shown the applicability of FWI to image the shallow subsurface. For example, among many others, Groos et al. (2014, 2017), Xing and Mazzotti (2019) and Lamuraglia et al. (2022) demonstrated the efficacy of 2D time-domain elastic FWI in efficiently delineating the shallow subsurface with high resolution, whereas Romdhane et al. (2011) and Tran et al. (2013) have showcased the applicability and reliability of 2D frequency-domain elastic FWI in reconstructing heterogeneous near-surface models.

The FWI of Rayleigh waves is mainly employed to retrieve the V_s parameter because this is the elastic parameter mostly constrained by the data. For this reason, in this work, we primarily describe the estimated S-wave velocity models, whereas the recovered V_p models are briefly illustrated in Appendix A. For all the inversion tests, we considered the V_p/V_s ratios to be greater than $\sqrt{2}$.

Elastic FWI amplifies the typical challenges of acoustic FWI, such as the high computational cost and the high nonlinearity of the inversion problem, in which the error function is characterized by multiple local minima (especially when considering the conventional least-squares misfit function). The high computational cost is mainly related to the very fine grid: At least 20 points per wavelength are needed to properly simulate the wave propagation with a finite-difference forward model (according to Pierini & Stucchi, 2020), without being affected by numerical dispersion issues. Moreover, a very small time-step must be chosen to satisfy the stability condition, expressed by the Courant–Friedrichs–Lewy condition.

For a local optimization approach, an appropriate starting model is crucial to achieve convergence without getting trapped in some local minima of the objective function (Alkhalifah, 2016; Schafer et al., 2014; Xing & Mazzotti, 2019), among others. On the other hand, global optimization methods can alleviate this requirement, but the computational burden required for convergence might be prohibitively long, especially for a large number of unknowns (Aleardi, 2019;

Datta & Sen, 2016; Ely et al., 2018; Ray et al., 2016; Sajeva et al., 2017; Sen & Stoffa, 2013).

This work aims to introduce an inversion approach that addresses some of the challenges posed by the elastic FWI of surface waves for near-surface investigations. Specifically, our goal is to alleviate the stringent requirement for a starting point lying in the global minimum valley and to also provide statistical assessments of the estimated model parameters. In particular, we provide a solution to the elastic FWI problem in a Bayesian inference framework (Gebraad et al., 2020; Mosegaard & Tarantola, 2002) using a Markov Chain Monte Carlo (MCMC) method. This method is designed to efficiently explore the model space, mitigating the cycle-skipping issue and, at the same time, accurately quantifying the uncertainties affecting the recovered solution.

In a Bayesian inversion, the outcome is the so-called posterior probability density (PPD) function in the model space. This function combines prior knowledge about the model parameters with the information provided by the experiment (i.e. the recorded seismic data). Although the PPD can be analytically computed for linear forward operators and Gaussian priors and data likelihood functions, a sampling method is required to estimate the posterior target distribution for more complex scenarios. MCMC methods offer a clever way to construct a Markov chain that produces samples drawn from the PPD (Sambridge & Mosegaard, 2002). Theoretically, MCMC methods can estimate the posterior distribution with sufficiently long Markov chains starting from random points in the model space.

However, the main challenge of these methods lies in their diminished sampling efficiency in high-dimensional problems due to the so-called curse of dimensionality (Curtis & Lomax, 2001). Additionally, their convergence rate to the target PPD relies heavily on the random perturbation applied to the current state of the chain. Traditional sampling strategies, like the random walk Metropolis (Sherlock et al., 2010), may necessitate many forward evaluations before reaching convergence, rendering them computationally unfeasible for high-dimensional problems or when dealing with expensive forward modellings, such as in the case of elastic FWI.

Efforts have been dedicated to mitigating the computational costs associated with the Bayesian approach by employing dimensionality reduction methods making use of compression techniques (Malinverno, 2002; Dejtrakulwong et al., 2012; Aleardi, 2019). Moreover, to enhance the efficiency of sampling the PPD using MCMC methods, it is essential that the proposal distribution reasonably approximates the target posterior density. Recent years have seen the emergence of several MCMC techniques, such as self-adaptive MCMC (Haario et al., 2001), parallel tempering (Sambridge, 2014) and Hamiltonian MCMC (Neal, 2011). In particular, the class of gradient-based MCMC methods has attracted considerable interest in recent years for addressing geophysical inverse problems. The Hamiltonian Monte Carlo (HMC) leverages Hamiltonian's equations to evolve a continuous-time Markov process, a dynamic system defined by potential and kinetic energies (e.g. Aleardi, 2020b; Fichtner & Simutè, 2018; Gebraad et al., 2020; Koch et al., 2020; Sen & Biswas, 2017). This system is practically approximated using multiple iterations of a leapfrog integrator, with the HMC algorithm combining this dynamic with the Metropolis–Hastings acceptance step. A specific case of the HMC method is the Metropolis-adjusted Langevin algorithm (MALA), which involves using only one integration step among sampling proposals. MALA belongs to the Langevin Monte Carlo algorithm family, which integrates gradient information with random noise to guide the sampling process. Another possible solution for solving Bayesian inference problems that is gaining a lot of attention in the latest years is represented by variational inference algorithms (Zhang & Curtis, 2020, 2021).

In this work, we make use of a gradient-based MCMC (GB-MCMC), where the proposal distribution is constructed using the local gradient and the Hessian of the negative log posterior. This method (originally named Stochastic Newton MCMC) was theoretically introduced by Martin et al. (2012) and later applied by Zhao and Sen (2021) to the acoustic FWI, without employing any reparameterization technique to reduce the dimensionality of the problem and only considering the diagonal entries of the Hessian matrix, resulting in a reduced sampling efficiency. Then Berti et al. (2024a, 2024b) decreased the computational cost of the approach when applied to acoustic and elastic FWI by introducing a model and data compression through a discrete cosine transform (DCT) reparameterization. Here, we extend the method to the elastic case by inverting surface wave data. From here on, for simplicity and on the line of Zhao and Sen (2021) and Berti et al. (2024a, 2024b), we refer to our method as the GB-MCMC.

The implemented algorithm significantly accelerates convergence by providing a local approximation of the target PPD using information from the local Hessian and gradient computed around the current state of the chain. Although this approach necessitates derivative computation for each sampled model, which can be computationally intensive in large-dimensional problems, we address this limitation through a compression of models and data spaces using reparameterization techniques such as the DCT, effectively reducing the number of unknowns in our inverse problem and inherently acting as a regularization in the model space (Kotsi et al., 2020; Lochbuhler et al., 2014).

To validate our approach, we apply it to a synthetic near-surface model, with notable lateral and vertical velocity variations. Subsequently, we employ GB-MCMC FWI on a real dataset from the InterPACIFIC project acquired at the Grenoble test site in France (Garofalo et al., 2016). The

application of our algorithm to this dataset is bolstered by the availability of well-log data, providing validation for the obtained results.

METHODS

We will briefly describe the gradient-based Markov Chain Monte Carlo (GB-MCMC) method we employed. Further details can be found in Zhao and Sen (2021) or Berti et al. (2024a) who applied the method to probabilistically solve the acoustic full-waveform inversion. The Bayesian solution of an inverse problem is fully described by the posterior probability density (PPD) in the model space that, following the Bayes theorem, can be written as

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

where the $\mathbf{m}$ and $\mathbf{d}$ vectors represent the model and data parameters, respectively; $p(\mathbf{d}|\mathbf{m})$ is the so-called data likelihood function measuring the misfit between observed and predicted data, whereas $p(\mathbf{m})$ and $p(\mathbf{d})$ represent, respectively, the a priori distribution of model parameters and the evidence term; $p(\mathbf{m}|\mathbf{d})$ is the PPD, which fully quantifies the solution uncertainties.

The PPD can be analytically derived only in the case of Gaussian-distributed data and model parameters and for linear forward operators. In the context of seismic inverse problems, a numerical approach is necessary to estimate the PPD. Sambridge and Mosegaard (2002) proposed the use of an MCMC sampling method to address this need. The fundamental idea behind this approach is to construct a Markov chain, the equilibrium distribution of which is the target PPD. The chain is iteratively updated by proposing new models, according to a well-defined proposal distribution, and the probability γ to move from the current state of the chain $\mathbf{m}_k$ to the next proposed state $\mathbf{m}_{k+1}$ is regulated by the Metropolis–Hasting (M–H) rule:

$$\begin{aligned}\gamma &= p(\mathbf{m}_{k+1}|\,\mathbf{m}_k)\\ &= \min\left[1, \frac{p\left(\mathbf{m}_{k+1}\right)}{p\left(\mathbf{m}_k\right)} \times \frac{p(\mathbf{d}\,|\mathbf{m}_{k+1})}{p(\mathbf{d}\,|\mathbf{m}_k)} \times \frac{q(\mathbf{m}_k\,|\mathbf{m}_{k+1})}{q(\mathbf{m}_{k+1}\,|\mathbf{m}_k)}\right],\end{aligned} \tag{2}$$

where $q(.)$ is the proposal distribution responsible for drawing the model from a probability distribution $q(\mathbf{m}_{k+1}|\mathbf{m}_k)$ only influenced by the current state of the chain $\mathbf{m}_k$ (Hastings, 1970). The proposal ratio term vanishes if symmetric proposals are used (i.e. such as in the classical random walk M–H algorithm). In our case, the proposal is not symmetric, and for this reason, the proposal ratio should be computed. However, because the proposal is Gaussian, both $q(\mathbf{m}_k|\mathbf{m}_{k+1})$ and $q(\mathbf{m}_k|\mathbf{m}_{k+1})$ can be analytically evaluated.

Upon acceptance of $\mathbf{m}_{k+1}$, the current model undergoes an update, and so $\mathbf{m}_k = \mathbf{m}_{k+1}$; conversely, if $\mathbf{m}_{k+1}$ is rejected, $\mathbf{m}_k$ is retained in the chain, and a new state is generated as a random deviation from $\mathbf{m}_k$. To enhance exploration across the parameter space and prevent the sampling from getting trapped in some local maxima of the PPD, several Markov chains are typically employed. This results in various random paths originating from different points in the model space. An established practice in numerically evaluating the posterior density through MCMC algorithms involves discarding early realizations of the chains and beginning to collect samples once the initial state impact is alleviated. This strategy, known as *burn-in*, aims to consider samples only after the Markov chain has sufficiently approached the stationary regime. The samples collected after the *burn-in* period are used for the numerical computation of the statistical properties, such as mean and standard deviation, associated with the target PPD.

If we assume Gaussian distributions for data, noise and model parameters, then:

$$p(\mathbf{m}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{\text{prior}}\right)^{\mathrm{T}}\mathbf{C}_{\mathrm{m}}^{-1}\left(\mathbf{m}-\mathbf{m}_{\text{prior}}\right)\right), \tag{3}$$

$$p(\mathbf{d}\,|\mathbf{m}) \propto \exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{G}(\mathbf{m}))^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}(\mathbf{d}-\mathbf{G}(\mathbf{m}))\right), \tag{4}$$

where $\mathbf{m}_{\text{prior}}$ is the prior model vector; $\mathbf{C}_{\mathrm{d}}$ and $\mathbf{C}_{\mathrm{m}}$ are the data and prior model covariance matrices, respectively, and $\mathbf{G}$ is the forward modelling operator which relates the model to the corresponding data. Now we can derive a Gaussian approximation of the PPD around $\mathbf{m}_{\mathrm{k}}$:

$$\begin{aligned}&p(\mathbf{m}\,|\mathbf{d}) \propto \exp\\ &\left(-\frac{1}{2}\left(\mathbf{m}-\left(\mathbf{m}_k-\mathbf{H}^{-1}\mathbf{g}\right)\right)^{\mathrm{T}}\mathbf{H}\left(\mathbf{m}-\left(\mathbf{m}_k-\mathbf{H}^{-1}\mathbf{g}\right)\right)\right),\end{aligned} \tag{5}$$

where $\mathbf{g}$ represents the gradient vector and $\mathbf{H}$ the Hessian matrix, defined as

$$\mathbf{g} = \mathbf{J}^{\mathrm{T}}\,\mathbf{C}_{\mathrm{d}}^{-1}\Delta\mathbf{d}\left(\mathbf{m}_k\right) + \mathbf{C}_{\mathrm{m}}^{-1}\left(\mathbf{m}_k-\mathbf{m}_{\mathbf{prior}}\right), \tag{6}$$

$$\mathbf{H} \approx \mathbf{J}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{J} + \mathbf{C}_{\mathrm{m}}^{-1}. \tag{7}$$

In the previous equations, $\Delta\mathbf{d}(\mathbf{m}_k) = \mathbf{G}(\mathbf{m}_k)$, where $\mathbf{d}$ is the data misfit vector calculated using the current model $\mathbf{m}_k$ and $\mathbf{J}$ denotes the Jacobian matrix, which expresses the partial derivative of the data with respect to the model parameters and can be used in a local inversion for an approximate, local sen-

sitivity analysis of the inverse problem (Aleardi & Tognarelli, 2016).

We can now define a sampling technique that uses the local approximation of the PPD to define the proposal density:

$$q(\mathbf{m}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{m}-\left(\mathbf{m}_k-\alpha\mathbf{H}^{-1}\mathbf{g}\right)\right)^{\mathrm{T}}\frac{\mathbf{H}}{\beta^2}\left(\mathbf{m}-\left(\mathbf{m}_k-\alpha\mathbf{H}^{-1}\mathbf{g}\right)\right)\right), \tag{8}$$

where $\mathbf{H}^{-1}\mathbf{g}$ is the so-called Newton step (notice that if $\beta = 0$, we obtain the standard gradient descent). Because we are recomputing the Hessian matrix at every iteration and analytically evaluating the proposal ratio (Equation 2) for each considered model, the scaling factor depending on the determinant of the Hessian matrix is properly considered. The adaptable parameters, α and β^2, play a crucial role in determining how far to go along the negative gradient direction and the variance of the applied random perturbation, respectively. These parameters impact the algorithm's balance between exploration and exploitation, thereby influencing the acceptance ratio of the sampling. Furthermore, although the proposal is based on a local Gaussian approximation of the PPD, it can be applied for sampling from any type of posterior model and under any prior assumptions (e.g. non-parametric prior, Aleardi, 2021). Indeed, the estimated posterior is theoretically independent from the adopted proposal, which only affects the sampling efficiency. In this work, we limit to consider a log-Gaussian prior, but, if needed, the method can manage any kind of prior information either analytical or non-parametric (Aleardi et al., 2022).

Algorithm 1 GB-MCMC

1. **Define the prior information → Full space**
2. **Define the starting models**
3. **For** chains = 1: n_chains **do**
4. **While** iter ≤ n_iter **do**
5. Define the current model $\mathbf{m}_k$ → **DCT space**
6. Compute Jacobian matrix **J**, gradient vector **g** and Hessian matrix **H**
7. Draw the proposed model $\mathbf{m}_{k+1}$ from a multivariate Gaussian
8. Compute the forward evaluation to obtain the current data → **Full space**
9. Project the calculated data onto the DCT domain → **DCT space**
10. Metropolis–Hasting rule $\gamma = p(\mathbf{m}_{k+1}|\mathbf{m}_k)$
11. **If** $\mathbf{m}_{k+1}$ is accepted **then**
12. $\mathbf{m}_k = \mathbf{m}_{k+1}$; iter = iter + 1
13. **Else**:
14. $\mathbf{m}_k = \mathbf{m}_k$; iter = iter + 1
15. Repeat steps 7–10
16. Store all the accepted models
17. Compute mean and standard deviation of the PPD → **Full space**

Algorithm 1 presents a schematic overview of the workflow for the proposed sampling method, emphasizing the steps conducted in both the full and discrete cosine transform (DCT) spaces (the arrows on the right show in which domain the various steps are performed). It is evident that numerous forward and inverse DCT transformations are required, but these operations incur negligible computational costs. Furthermore, despite the inversions occur in the reduced DCT space, it is crucial to project the sampled model back into the full domain immediately before the forward modelling phase.

The GB-MCMC method is designed to concentrate sampling efforts on the most favourable regions of the parameter space, characterized by high posterior density values. Additionally, it employs a proposal that leverages the inverse of the Hessian matrix, incorporating information about the local covariance structure of the target density. Notably, the reduced dimensionality of the model space (see next subsection concerning the DCT) enables the consideration of the entire approximate Hessian, not just its diagonal values. This implies that the sampling of the model space is guided not only by data illumination but also by the correlation among model parameters. The incorporation of this information not only accelerates the probabilistic sampling but also enhances sample independence while simultaneously ensuring high acceptance rates (Aleardi et al., 2022).

The major computational requirement in GB-MCMC sampling lies in computing the Jacobian matrix associated with each accepted model, the dimensions of which depend on the number of unknowns and data points. This Jacobian matrix is calculated using a forward finite-difference scheme. In this context, the number of forward evaluations required for its computation grows linearly with the number of model parameters. The advantage is that the computation of each column of the Jacobian matrix can be done independently, facilitating its distribution across different computing units. In addition, the use of the DCT drastically reduces the computational cost for the Jacobian evaluation and the manipulation of the Hessian matrix and gradient vector.

Discrete cosine transform

The DCT is a linear orthogonal transformation that efficiently decomposes a signal (that can be mono or multidimensional) into a combination of cosinusoids oscillating at various frequencies. This transformation has proven to be highly effective in imaging compression standards due to its remark-

able ability to concentrate most of the signal's energy within the low-order coefficients (Ahmed et al., 1974; Aleardi, 2020a; Jain, 1989).

If we consider, for example, a 2D velocity model, most of the spatial variability comes from the first DCT coefficients in both dimensions, corresponding to the first q rows and p columns of the transformed matrix. Therefore, the DCT transformation allows for a reduction of the $(M_y \times M_x)$ full-velocity model space to a $(q \times p)$ DCT compressed parameter space, with $q << M_y$ and $p << M_x$. A similar transformation can also be applied in the data space, to compress the 2D matrices representing each shot gather.

Here, the model space expresses the 2D V_p/V_s and V_s values in the subsurface, whereas the density is not inverted given its minimal impact on the observed Rayleigh waves (Ivanov et al., 2016). The choice of the number of retained DCT coefficients to compress the model and data spaces is a crucial step of our inversion framework and should always involve a trade-off between the desired model and data resolution and the computational cost of the procedure. If needed, a trans-dimensional approach could be employed, in which the number of DCT coefficients is treated as an unknown to be estimated from the inversion (Hawkins & Sambridge, 2015; Ray et al., 2016). Here we use a simpler strategy for the estimation of the optimal number of DCT coefficients to retain in the parameter space: In the synthetic case, we have quantified how the variability of the true V_s model changes as the number of DCT coefficients varies. Here, the variability is computed as the ratio between the variance of the compressed model and the true one (Aleardi, 2021). In the real case, the optimal model compression can be set according to the expected velocity variability in the study area and according to the achieved data matching. If available, borehole information can be employed to facilitate the selection of the DCT coefficients in the model space. For example, in our application, the available borehole data were used to estimate the number of DCT coefficients along the vertical direction. For the horizontal direction, we used the same number as in the synthetic case, considering that they were enough to reconstruct a very strong lateral variation in a model with similar dimensions. We will discuss this aspect more in detail when discussing the field data inversion.

For the estimation of the optimal number of DCT coefficients needed to approximate the data, we can use the same strategy as for the model space, thus analysing the variability of the observed data before and after the compression. However, we have observed that the variability can reach high values using a low number of DCT coefficients in both directions, but the reconstructed seismogram after the compression step is affected by some artefacts. For this reason, we have used a different strategy to estimate the optimal number of DCT coefficients, for both the synthetic and real cases. In particular, we have analysed how the relative percentage error, calculated as the ratio between the L2 norm difference of the observed and compressed data and the L2 norm of the observed data, varies with different combinations of DCT coefficients.

In all the following tests, we are assuming log-Gaussian prior distributions for the V_s and V_p/V_s values. Therefore, the a priori mean vectors and a priori model covariance matrix $\mathbf{C}_m$ can be analytically projected to the DCT space, thanks to the linearity of the DCT transformation (for further details, please refer to Berti et al., 2024a). We also point out that in our applications, the $\mathbf{C}_m$ matrix codes a 2D stationary and Gaussian correlogram expressing the assumed spatial variability of the velocity values along both spatial directions and expresses the covariance for both the V_s and V_p/V_s models. Due to the large number of data points in the full space, we applied a numerical Monte Carlo simulation to project the data covariance matrix $\mathbf{C}_d$ onto the compressed domain (see Aleardi et al., 2022).

RESULTS

Synthetic inversion test

To validate our proposed methodology, we first apply our approach to a 2D near-surface synthetic velocity model, derived from three-component downhole recordings acquired in San Sepolcro and Barga (Tuscany, Italy). The model parameters to be estimated are the V_s and V_p/V_s values, with the density assumed to be perfectly known.

First, we need to derive two very important parameters that play a crucial role in the inversion setting: Δx and Δt, the spatial grid interval and the time sampling, respectively. To avoid numerical dispersion in the finite difference (FD) modelling, the maximum spatial grid interval Δx should satisfy

$$\Delta x < \frac{V_{min}}{n \cdot f_{max}}, \tag{9}$$

where V_{min} is the minimum velocity in the subsurface model, n is the number of points per minimum wavelength and f_{max} is the maximum frequency in the modelling. In the case of body wave modelling, a value of n equal to 5 is generally used (Alford et al., 1974), but when we are dealing with surface wave modelling, a much higher value is recommended (in our work, it has been set to at least 20, following Pierini & Stucchi, 2020). On the other hand, to guarantee stability, the time sampling interval Δt should satisfy

$$\Delta t < C \cdot \frac{\Delta x}{V_{max}}, \tag{10}$$

where V_{max} is the maximum velocity in the subsurface model and C is the Courant number (the value of which is 0.606 in the case of a fourth-order approximation of derivatives along space and a second-order approximation along time).

Taking into consideration these two conditions, the grid size for generating the observed data is set to $680(n_x) \times 120(n_y)$, where n_x and n_y represent the number of grid points in the horizontal and vertical directions, respectively, with a uniform grid spacing of 0.5 m in both directions. Employing a Ricker wavelet with a central frequency of 10 Hz as the source, we have simulated 5 shots evenly distributed along the horizontal direction, and each shot is recorded by 96 receivers with a 3 m receiver interval. All sources and receivers are positioned at the free surface. A time sampling interval of 0.1 ms is utilized for the forward modelling (then resampled to a 2 ms time interval), with a registration time of 600 ms. Uncorrelated Gaussian white noise is added to the observed data, with a standard deviation equal to 20% of the standard deviation of the noise-free data. The resulting signal-to-noise ratio is 14 dB. Elastic forward modelling is executed using SOFI2D (Bohlen, 2002), a viscoelastic forward modelling code that solves the pure elastic or viscoelastic wave equation through an FD scheme in the time domain. To optimize the computational efficiency of calculating the Jacobian matrix using the FD approach, we have parallelized the computation of each column of the Jacobian matrix across different processes. In the test presented here below, we assume a perfectly known wavelet, and all the receivers illuminated by the sources are used for the inversion. However, to prove the robustness of the implemented algorithm, we have performed other tests: deriving the wavelet from the observed data instead of assuming it was known, increasing the level of noise in the data and also considering some missing channels. These results will be shown in Appendix B.

The transformation of data and models into the discrete cosine transform (DCT) space, where Markov Chain Monte Carlo (MCMC) sampling operates, necessitates the selection of the number of retained DCT coefficients for both seismic data and velocity models. For the model space, we can observe that utilizing less than eight coefficients along the two spatial dimensions captures over 98% of the original velocity model's variability (Figure 1). This compression strategy results in a substantial reduction of the $120 \times 680 \times 2 = 163{,}200$-dimensional elastic space to a more manageable $8 \times 4 \times 2 = 64$-dimensional domain, considering an equal number of DCT coefficients for both the V_p/V_s and V_s models. This choice strikes an optimal balance between reducing the unknowns in our inverse problem and preserving the essential features of the velocity model after compression. Parts (a) and (b) of Figure 1 show the comparison between the V_s model before and after the DCT compression, revealing the preservations of all main features, including vertical and lateral variations, with only a minor loss of high-resolution details.

For what concerns seismic data, we have seen that retaining $40 \times 70 = 2800$ coefficients correspond to a relative percentage error below 10% with respect to the observed data. Therefore, the full $300 \times 96 \times 5 = 144{,}000$-dimensional data space (considering the same number of DCT coefficients for all the five shots) has been effectively reduced to a $40 \times 70 \times 5 = 14{,}000$-dimensional domain.

It is important to note that all the matrices involved in the inversion procedure become more tractable in the compressed domain; for instance, the Jacobian matrix, which has a number of rows equal to the number of data parameters and a number of columns equal to the model parameters, has been reduced from a $144{,}000 \times 163{,}200$ matrix in the full domain to a $14{,}000 \times 64$ matrix in the compressed space.

The numerical tests were conducted using a Python implementation running on different servers equipped with Intel Xeon CPU E5-2630 v4 @ 2.20 GHz. Each iteration, which includes computing the Jacobian, the gradient, the Hessian matrix (if the proposed model is accepted) and drawing a sample, requires approximately 300 s of wall clock time, whereas 3.000 iterations of a single chain took approximately 3 days. We have employed five independent chains, each running on a distinct server.

As a prior mean model common to all the five chains, we employed a homogeneous model with a constant velocity of 600 m/s, also used as a starting point for one of the chains (Figure 2a). Both the vertical and lateral Gaussian correlograms used in this inversion test are shown in Figure 2b, where the ranges along the vertical and lateral directions are, respectively, 3 and 8 m.

Figure 3 illustrates that despite initiating from different and rather simple initial models, the posterior mean V_s models independently estimated for two distinct chains are nearly identical and closely reproduce the main features of the true compressed model (Figure 1b). This underscores the algorithm's efficiency in exploring the parameter space without being affected by cycle-skipping issues.

Examining the progression of the square root of the negative log-likelihood in Figure 4a reveals that the chains originate from distinct initial values, reflecting the different starting models employed in the inversions. However, they all reach the same stationary regime after approximately 750 iterations, when they start to oscillate around similar likelihood values. These initial iterations constitute the *burn-in* period, during which the corresponding velocity models are excluded from the numerical estimation of the statistical properties of the posterior probability density (PPD). The different convergence speeds between the chains are influenced by the specific choice for the α and β parameters in Equation (8) (it will be discussed in more detail in the 'Discussion' section). Note

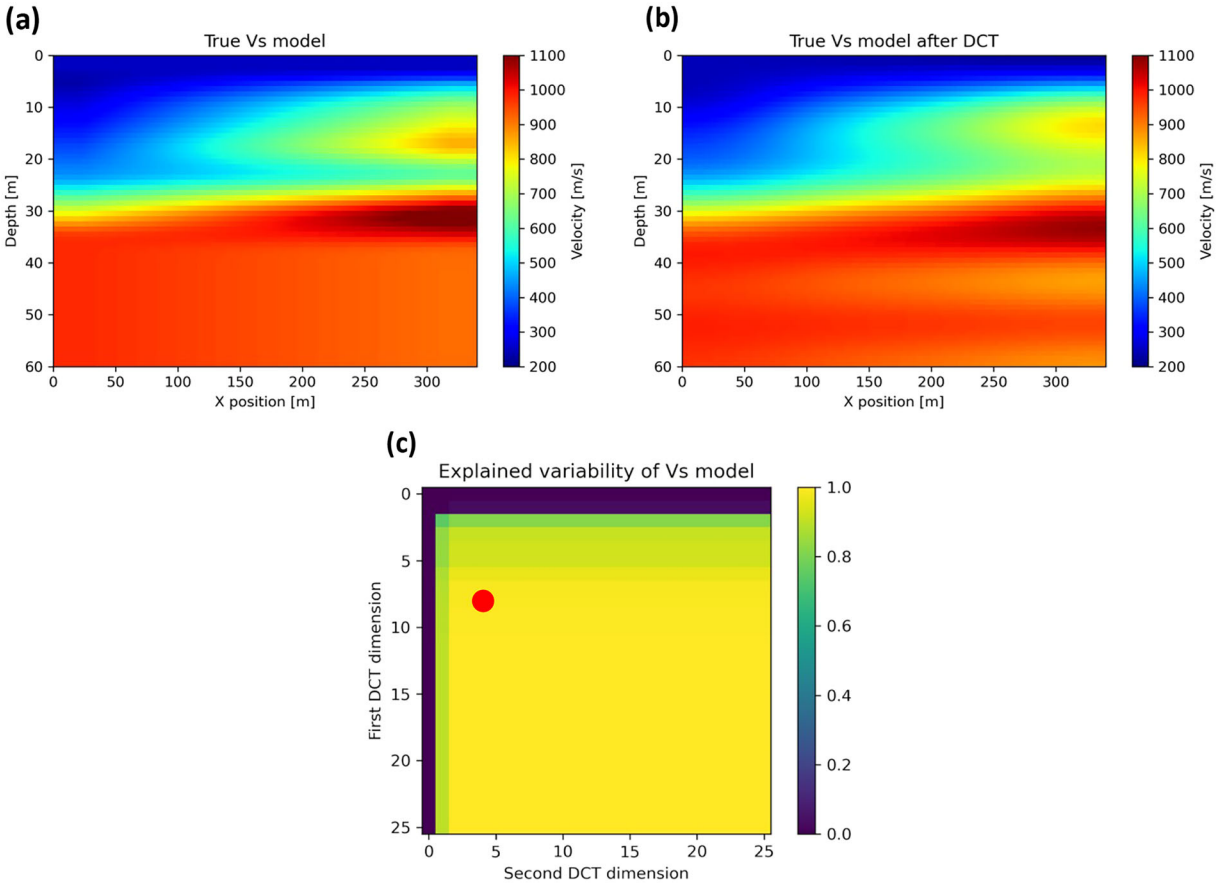

FIGURE 1 (a) True V_s model; (b) true V_s model after the discrete cosine transform (DCT) compression using only eight coefficients along the first dimension (depth) and four along the second dimension (horizontal position); (c) explained model variability with the increasing number of retained coefficients along the two DCT dimensions. The numerical value represented by coordinates (x,y) indicates the accounted variability when compressing the model space using the first y coefficients along the first dimension and x coefficients along the second dimension. The red dot highlights the chosen combination of DCT coefficients in the model space.

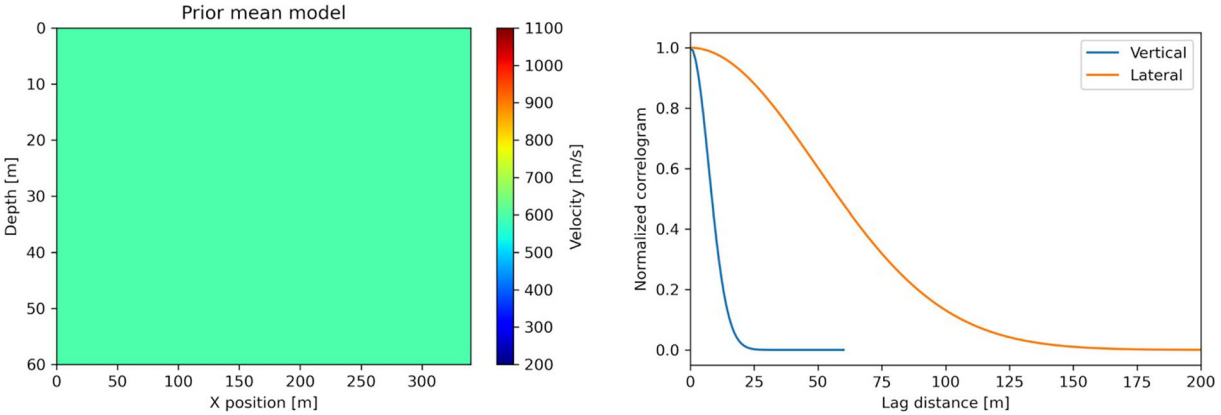

FIGURE 2 On the left, the prior mean model used for all the five chains; on the right, the normalized spatial correlation function associated with the assumed 2D Gaussian variogram model.

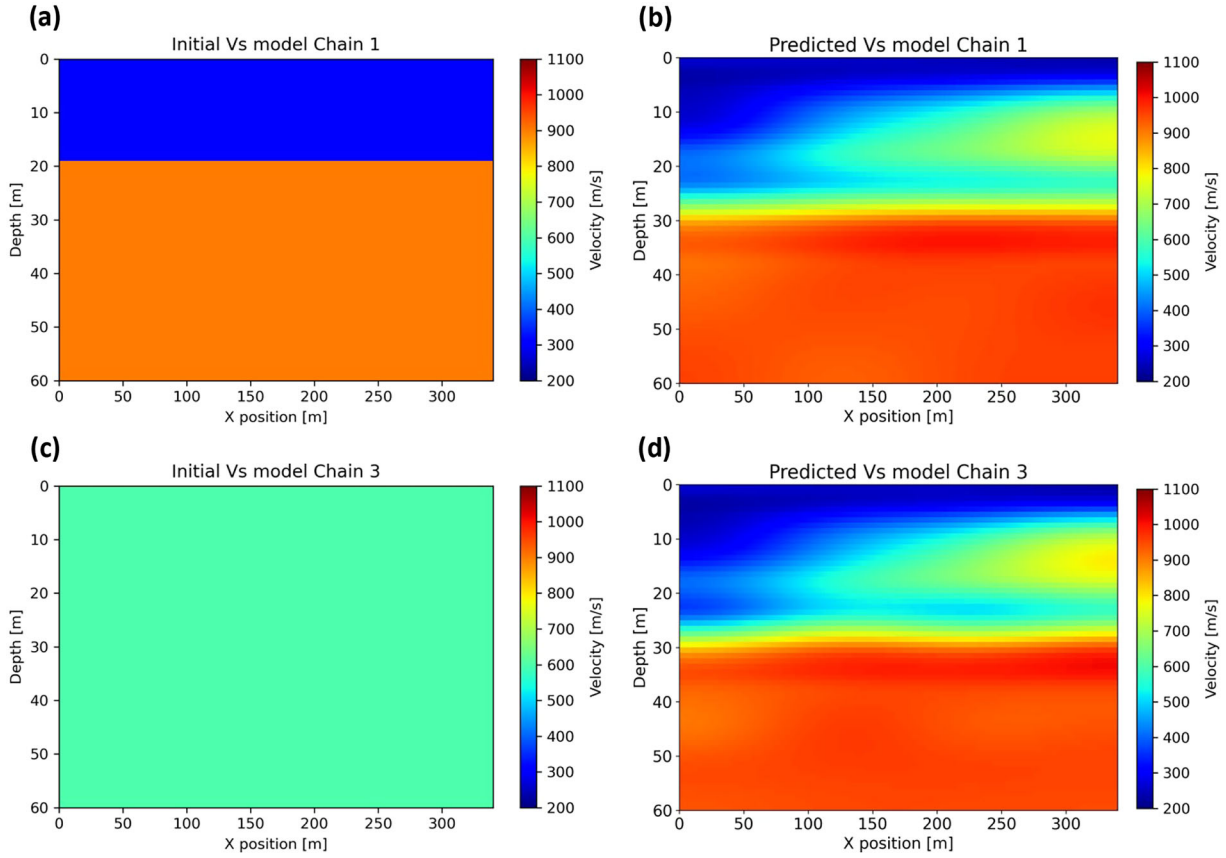

FIGURE 3 Examples of two chains: on the left, the initial V_s models of the gradient-based Markov Chain Monte Carlo (GB-MCMC) inversions ((a) a simple two-layers model; (c) a homogeneous model); on the right, the posterior mean V_s models independently estimated by the two chains ((b) and (d), respectively).

that we show the square root of the negative log-likelihood to better point out the differences of the data misfit values pertaining to the chains in the initial iterations.

Figure 4b illustrates the acceptance ratio for each chain, calculated as the number of accepted models divided by the number of iterations. We can observe that we reach values above 90%, much higher than the ones obtained using standard, gradient-free MCMC algorithms that typically yield values around 25%. Although these high values result in an increased computational cost (notice that for each accepted model we should compute the associated Jacobian matrix, which is the most computationally expensive step of the method), they also prevent unnecessary forward evaluations for velocity models that will not be accepted.

It is important to note that, given these very high acceptance ratios, we implemented a strategy to reduce the overall computational cost of the inversion. Specifically, after the end of the burn-in period, considering that the high probability region of the PPD has been approached, we have calculated the Jacobian matrix every 20 accepted models (as opposed to each one as done during the burn-in period). This significantly speeds up the computing time without compromising the statistical properties of the estimated PPD.

One method to determine if a chain has converged to the equilibrium distribution is by comparing its behaviour with other chains with different starting points. Gelman and Rubin (1992) introduced the potential scale reduction factor (PSRF) for this purpose. This statistic measures the ratio between the average variance within each chain to the variance of the pooled samples across all the chains. As the number of iterations increases, the PSRF value approaches 1. Although the ideal proximity to 1 can vary depending on the problem, a common target is to achieve a PSRF value below 1.2 for most of the model parameters. Figure 4c shows the evolution of the PSRF value for all the V_s model parameters in the DCT domain: The PSRF values for most of them are lower than 1.2, thus proving full convergence for these parameters, and even for the remaining coefficients, the PSRF is very close to the 1.2 threshold. For this reason, we can claim that, from a practical point of view, the sampling has reached a stable PPD estimation.

Figure 5a shows the original, uncompressed, V_s velocity model, whereas in Figure 5b, we illustrate the posterior mean V_s model obtained considering all the five chains. All the main velocity variations of the true model are well reproduced, whereas the lower resolution comes from the DCT compres-

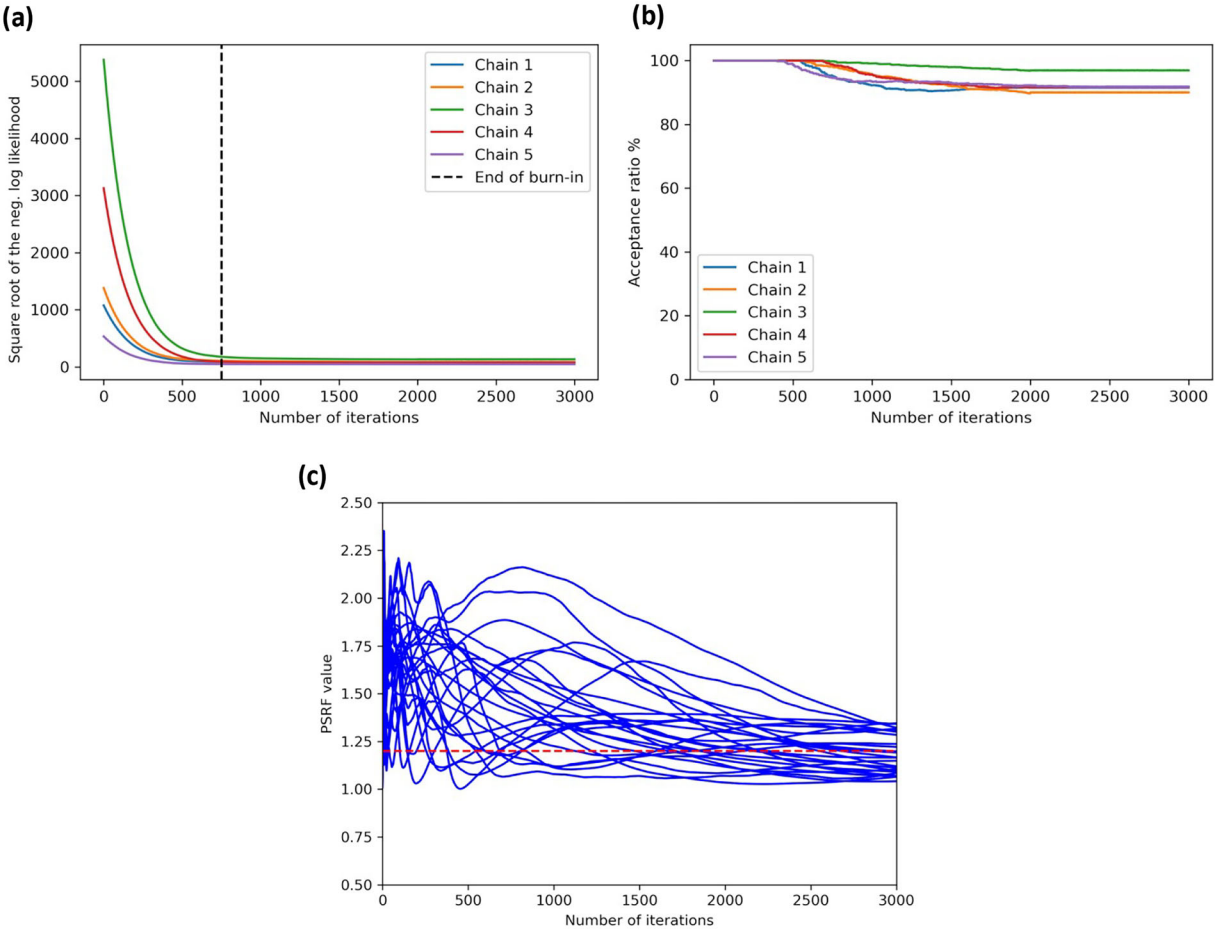

FIGURE 4 (a) Progression of the square root of the negative log-likelihood values for the gradient-based Markov Chain Monte Carlo (GB-MCMC) inversion, with each colour denoting a distinct chain, and the dashed line marking the end of the burn-in period; (b) acceptance ratio computed for each chain; (c) evolution of the potential scale reduction factor (PSRF) value over the number of iterations for the V_s model parameters in the discrete cosine transform (DCT) domain (each blue line corresponds to one of the retained DCT coefficients). The red dotted line highlights the 1.2 threshold for convergence.

sion, as explained before. We can improve the resolution of our predicted model considering a higher number of DCT coefficients in both directions while increasing, at the same time, the overall computational cost of the algorithm. Otherwise, the gradient-based MCMC (GB-MCMC) estimates can be used as a starting model for a subsequent step of local inversion.

The standard deviation map depicted in Figure 5c suggests higher uncertainties corresponding to the high-velocity layer and lateral velocity variations on the right edge of the model. As expected, we get low standard deviation values in the shallow part of the model, where there is substantial data coverage and illumination.

In Figure 6a, we present the middle shot of the noise-contaminated observed data, whereas parts (b) and (c) of Figure 6 show, respectively, the same shot calculated using one of the starting models employed for the inversion (in particular, the one depicted in Figure 3a) and the sample-by-sample difference between observed and initial data.

Subsequently, we illustrate the predicted data computed on the posterior mean model from Figure 5b and the residual difference compared to the observed data (Figure 6d,e). The data matching substantially improves when moving from the starting model towards the estimated posterior mean. The remaining difference between observed and modelled data computed on the GB-MCMC estimation can be partially attributed to the decreased resolution resulting from the DCT compression.

The close-ups of parts (a) and (b) of Figure 7 show two traces corresponding to two different offsets extracted from the middle shot of the data, computed using three different velocity models: the true, uncompressed model of Figure 5a, the predicted model of Figure 5b and one of the starting models of the chains used in the inversion (in particular, the one shown in Figure 3a). Here it is highlighted the severe cycle skipping between the observed and the initial data, and how this issue is significantly attenuated when the final predicted data is considered. This demonstrates that the GB-

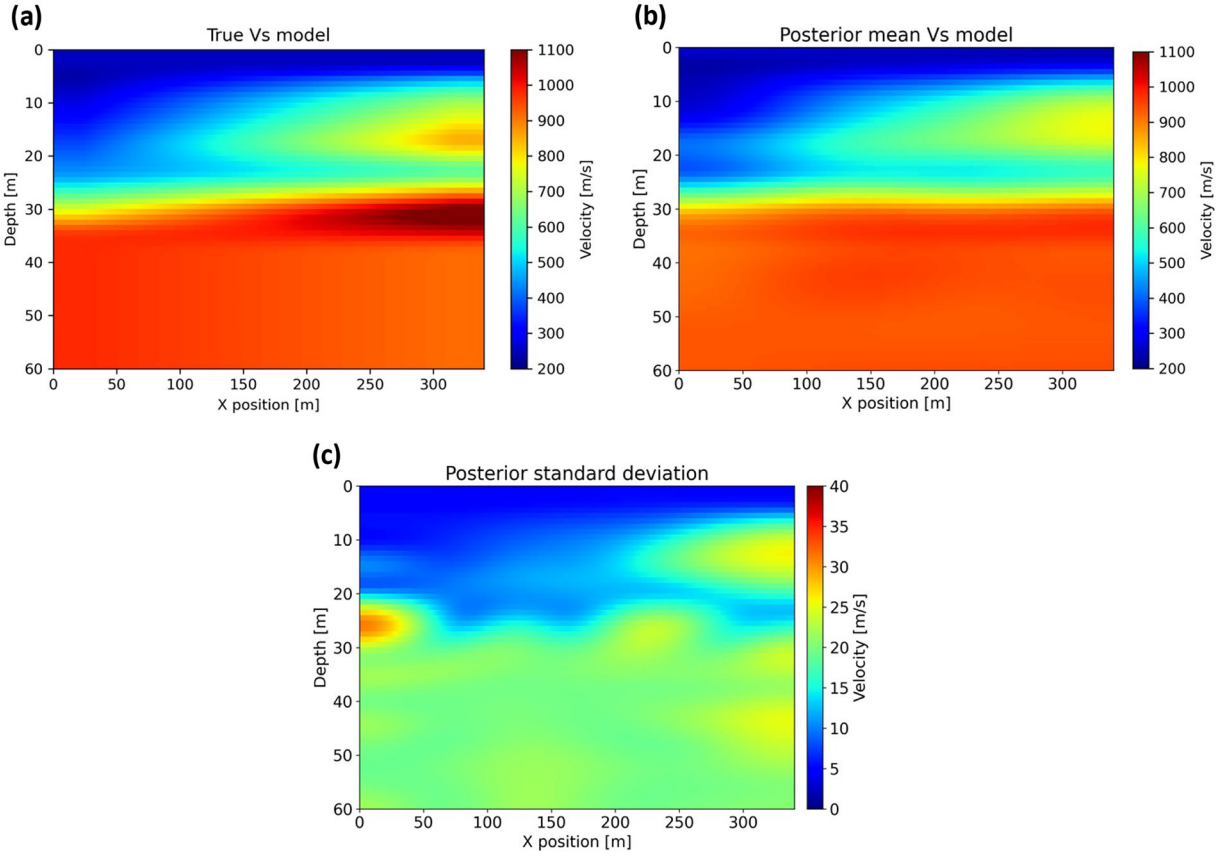

FIGURE 5 (a) The true V_s semi-real model; (b) the posterior mean V_s model and (c) the posterior standard deviation map computed considering all the five chains.

MCMC method is capable of efficiently exploring the model parameter space without being heavily affected by the initial model.

Local inversion

We now illustrate that the standard local inversion method fails to achieve convergence when employing our previously tested initial models. However, we demonstrate that GB-MCMC estimations can serve as effective starting points for a subsequent step of local inversion, aimed at enhancing the model resolution (i.e. minimizing the impact of the DCT compression) and further reducing the misfit between observed and predicted data.

The chosen local inversion algorithm is IFOS2D, an open source 2D elastic full waveform inversion code which finds the solution of this inverse problem using preconditioned conjugate gradient methods where the gradients are calculated in the time domain through the adjoint method (Plessix, 2006). The forward modelling is conducted in the time domain using a FD algorithm with a 2D Cartesian standard staggered grid. We also adopted a two-step frequency marching strategy, using 0–10 and 0–20 Hz as range of frequencies, whereas the normalized L2 norm between modelled and observed data constitutes the error function to be minimized (the misfit is scaled with respect to the energy of the observed seismograms; see Choi & Alkhalifah, 2012).

Parts (a) and (b) of Figure 8 showcase the results of the local inversion starting from, respectively, the same initial model of the first chain used for the GB-MCMC approach (Figure 3a) and the posterior mean model derived from all five chains (Figure 5b). In the first case, the inversion yields reasonable velocity magnitude values but falls short in reconstructing lateral velocity variations in the first layer and the position of the second high-velocity layer. For the second case instead, the inversion improves the resolution of the GB-MCMC result by introducing additional features, particularly along the right edge of the model, such as the high velocity zones between 10–20 and 30–35 m of depth.

This test underscores the superior capability of the GB-MCMC method, compared to classic local inversion, in mitigating cycle-skipping effect when initiating the inversions from very simple starting models. In addition, the reduced resolution of the GB-MCMC estimation can easily be improved through a second step of local inversion.

Additionally, we present the same two traces as in the previous section (Figure 9) extracted from the observed noisy

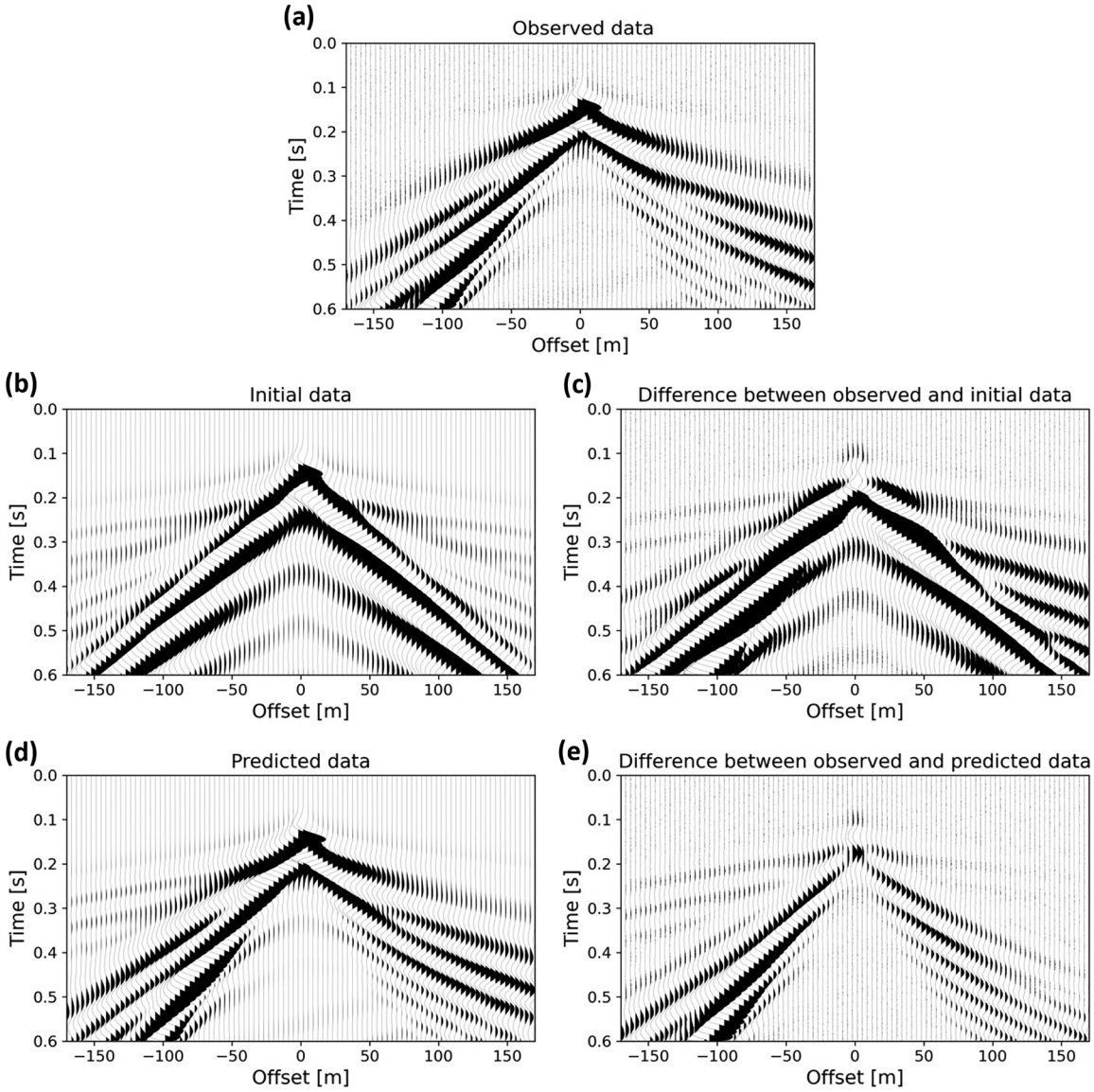

FIGURE 6 (a) The middle shot of the noisy observed data; (b) the same shot generated using the initial model of Figure 3a and (c) their sample-by-sample difference; (d) the middle shot of the predicted data computed using the posterior mean model of Figure 5b and (e) their sample-by-sample difference. All the plots display the raw amplitude data using the same gain, for an easier comparison of the results.

data, the predicted data calculated on the GB-MCMC estimation of Figure 5b, and the data generated from the model shown in Figure 8b. We can observe how a slight reduction in the misfit between observed and predicted data is achieved by the local inversion, if compared with the GB-MCMC predictions.

Real dataset inversion

After validating the proposed approach on synthetic data, we proceed to showcase its effectiveness in inverting field data. This section delineates our approach applied to a dataset acquired in Grenoble, France, located in the French Alps near the Institute Laue Langevin nuclear research facility. As part of the InterPACIFIC project, a borehole was drilled to a depth of 50 m, and the inclusion of three distinct well-log datasets serves as further validation of the obtained results, facilitating direct comparisons with the V_s predictions.

The dataset under examination in this study employs an active source to energize the subsurface. The acquisition surface is flat, featuring a linear arrangement of 48 vertical geophones with a natural frequency of 4.5 Hz and a spacing of 1 m. As the seismic source, an 8 kg sledgehammer was used, resulting in the acquisition of three shot gathers: one split–spread and two off-ends. The split–spread and off-end shot gathers have maximum offsets of 23.5 and 51 m, respectively (Table 1).

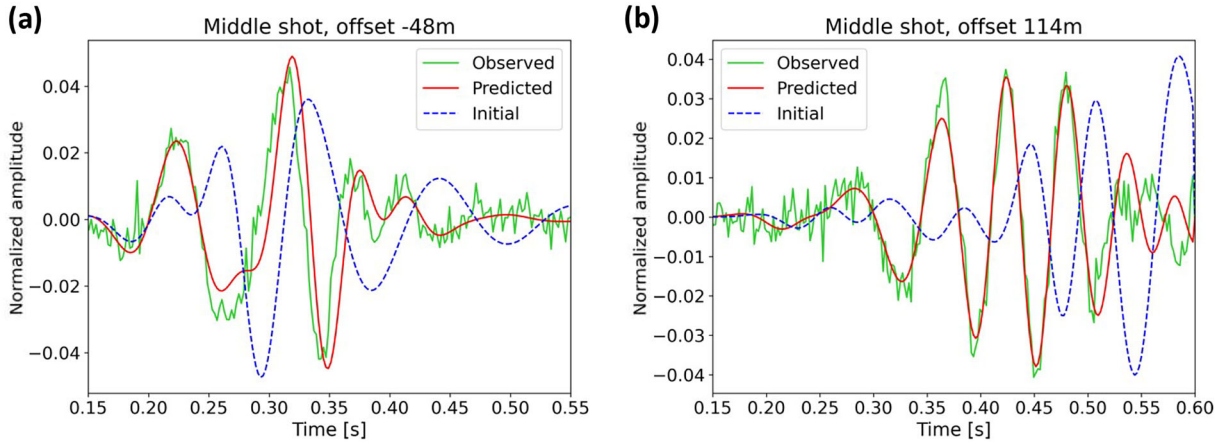

FIGURE 7 Comparison of two single traces corresponding to two different offsets of the middle shot using the true model (green), the posterior mean model of the gradient-based Markov Chain Monte Carlo (GB-MCMC) algorithm (red) and the initial model of Figure 3a (dashed blue).

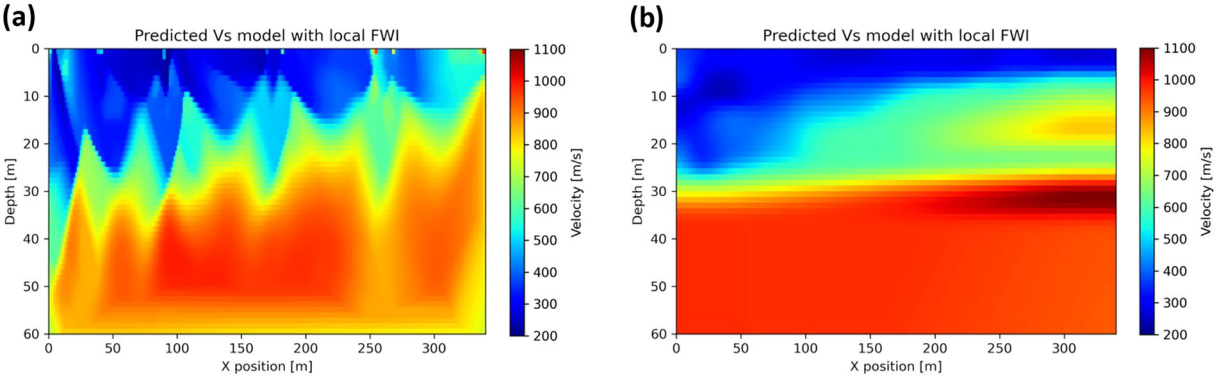

FIGURE 8 (a) Result of the local inversion, starting from the initial model of Figure 3a; (b) result of the local inversion starting from the gradient-based Markov Chain Monte Carlo (GB-MCMC) estimation shown in Figure 5b.

TABLE 1 Acquisition parameters for the Grenoble seismic data.

Seismograph	Geode (geometrics)
Geophones	Vertical geophones (4.5 Hz)
Source	8 kg sledgehammer
Acquisition time (ms)	2000
Sampling rate (ms)	0.25
Receivers spacing (m)	1
Number of channels	48
First channel (m)	3
Last channel (m)	50
Shot locations (m)	0; 26.5; 54

In the synthetic case, we operated under ideal conditions; where the source wavelet was known, the observed data were contaminated with Gaussian noise and computed with the same modelling engine as the predicted data. In this real case, none of these conditions hold true, and so, as an initial step of our processing sequence, we applied a trace-by-trace amplitude normalization to our seismic data and a zero-phase band-pass filter (3–30 Hz), aimed at reducing the background noise and constraining the maximum frequency for the inversion. Then, a 3D-to-2D correction (as proposed by Forbriger et al., 2014) is needed to account for the geometrical spreading between the real case point source and the 2D forward modelling that implicitly uses line sources for the simulations. Figure 10 provides an example of a shot gather before and after applying the processing sequence. The source wavelet was estimated from the near-offset traces by the SVD approach (Biondi et al., 2014; Kirlin & Done, 1999). Notably, the same processing sequence applied to the observed data was also performed for the simulated data. Differently from Lamuraglia et al. (2022), no mute was applied to either the observed or modelled seismograms.

Similar to the synthetic case, the model parameters to be estimated are the V_s and V_p/V_s values, whereas the density is fixed at a constant value of 2 g/cm^3. To generate the predicted data, we established a grid with a size of 276(n_x) × 150(n_y),

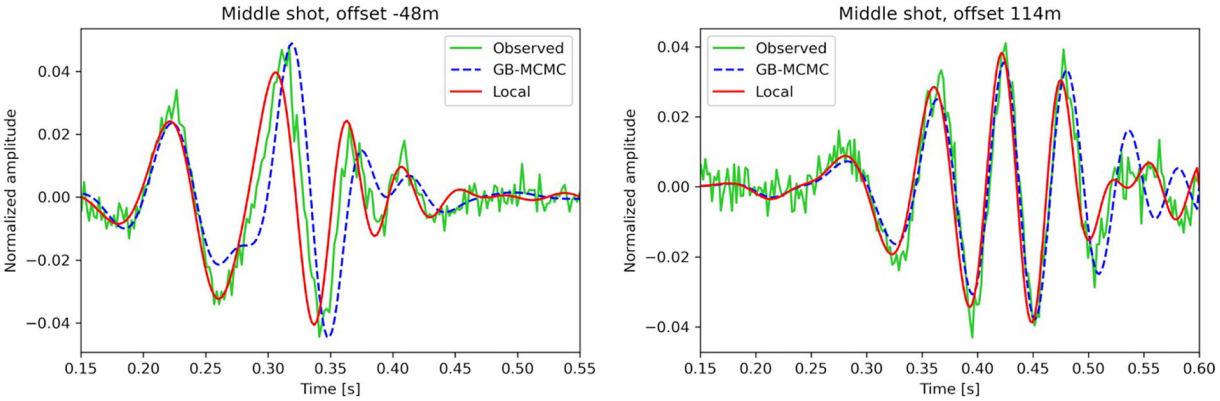

FIGURE 9 Comparison of two single traces corresponding to two different offsets of the middle shot using the true model (green), the posterior mean model of the gradient-based Markov Chain Monte Carlo (GB-MCMC) algorithm (dashed blue) and the predicted model with a second step of local inversion shown in Figure 8b (red).

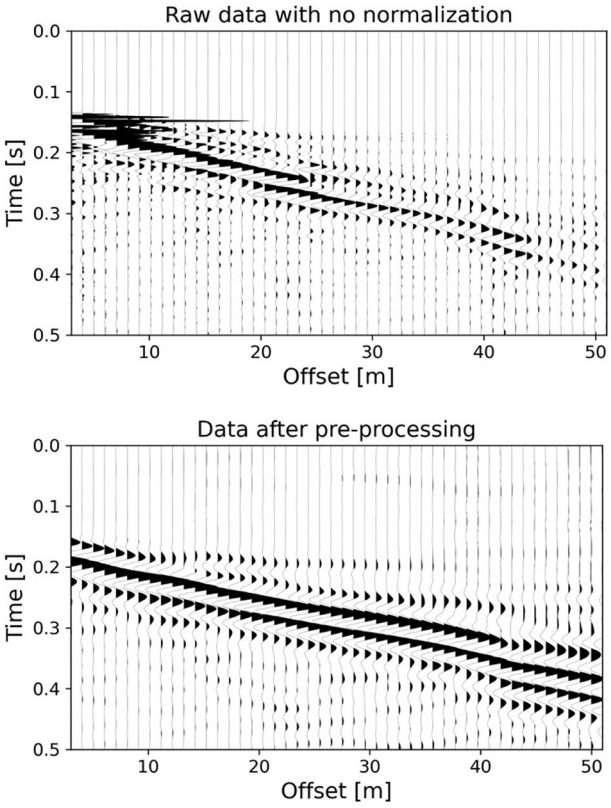

FIGURE 10 Comparison between the leftmost shot of the observed data before and after the application of the pre-processing steps that we have performed.

where n_x and n_y are the number of grid points in the horizontal and vertical directions. The grid spacing is set to 0.25×0.25 m^2, aligning with the dispersion condition outlined in Equation (9), and the absorbing boundaries consist of 30 grid points. Both sources and receivers are positioned at the free surface. For the forward modelling, a time sampling rate of 0.1 ms is employed, with a registration time set at 0.5 s. Both the predicted and observed data undergo resampling using a time interval of 2 ms. The simulation of the shots is once again executed utilizing SOFI2D (Bohlen, 2002).

To alleviate the computational burden associated with computing the Jacobian matrix via the forward FD approach, we distribute the computation of each column of the matrix across multiple processes. As shown in the synthetic test, both the data and models are transformed into the DCT domain, where the MCMC sampling is conducted.

Analysing the seismic data, we note that $60 \times 45 = 2700$ coefficients per shot gather exhibit a relative percentage error with respect to the observed data below 10%. Consequently, the full $250 \times 48 \times 3 = 36{,}000$-dimensional data space (derived from a sampling interval of 2 ms and an equal number of DCT coefficients for all three shots) gets reduced to a more manageable $60 \times 45 \times 3 = 8100$-dimensional domain.

In this initial experiment, we exploit the available borehole information to determine the appropriate number of DCT coefficients in the model space along the vertical (depth) direction.

Figure 11 illustrates the downhole measurements and their reconstruction through 1D DCT compression with varying numbers of coefficients. It is essential to highlight that utilizing fewer than 20 coefficients along the vertical dimension proves insufficient to reconstruct the velocity inversion discernible at depths of 15–20 m in the downhole data (the same analysis was carried out also with the other borehole measurements). Consequently, the well-log data play a pivotal role in guiding the selection of the optimal number of retained DCT coefficients. This aspect is further investigated in the next subsection, where we present the inversion results achieved using a reduced number of DCT coefficient for the model along the vertical direction. In the horizontal direction, we opt for seven coefficients, given the limited expected lateral velocity variations in the study site. As a result, the original $150 \times 276 \times 2 = 82{,}800$-dimensional elastic space undergoes

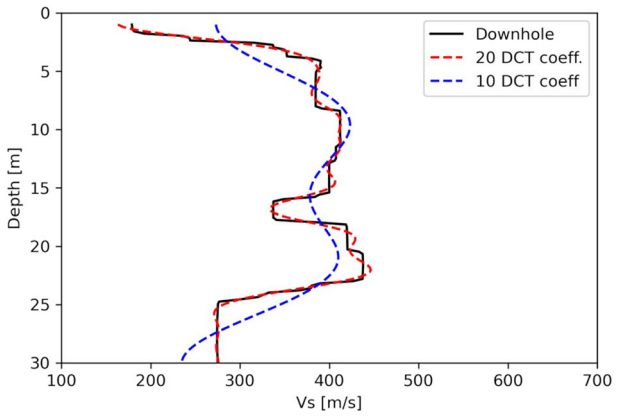

FIGURE 11 Comparison between the downhole measurement (black line) and the same profile after the discrete cosine transform (DCT) compression using, respectively, 20 (dashed red) and 10 (dashed blue) retained coefficients.

a substantial reduction to a $20 \times 7 \times 2 = 280$-dimensional domain (again, as in the previous test, we are considering the same number of DCT coefficients for both the V_p/V_s and V_s models).

To conduct the numerical tests, we have utilized a Python implementation running on the same servers as those employed for the synthetic test. Each iteration, including the computation of the Jacobian matrix (if the proposed model is accepted) and the sampling process, required an average wall clock time of approximately 10 min. Drawing insights from Louboutin et al. (2017), we estimated a computational demand of 14 TFLOPs for a single iteration. Considering that the peak computing performance of the single computer is approximately 40 GFLOPs, we estimated an approximate lower boundary of around 350 s per iteration, to be compared with the measured clock time. Our approach involved employing 5 chains, each comprising 4000 iterations, running on distinct servers. As in the synthetic inversions, the Jacobian is not computed at each iteration after the burn-in period (see the 'Discussion' section). Consequently, each chain concluded its computations in approximately 9 days.

As the prior mean model common to all chains, we have incorporated one of the smoothed two-layered models also used as a starting point (Figure 12a). A Gaussian correlogram is used to impose the spatial continuity to the recovered velocity models. The vertical and lateral Gaussian correlograms are illustrated in Figure 12b, and the ranges along the vertical and lateral directions are equal to 2 and 6 m, respectively.

The starting models for the chains can be realizations of the prior distribution or can be set by the user. In this work, the ones utilized for the five chains can be classified into two groups: The first three chains employ homogeneous or two-layered velocity models, whereas Chains 4 and 5 adopt a highly smoothed version of the 1D model derived from extending the available borehole data.

If we analyse the evolution of the negative log-likelihood (Figure 13a), it becomes evident that all the chains converge to the same stationary regime after approximately 1000 iterations, where the likelihoods begin to oscillate around similar values. This illustrates that despite the very different starting points, all the independent chains converge to models characterized by similar data misfit values.

Each chain starts with a different initial value based on the chosen starting model, exhibiting varying convergence speeds influenced by the specific choices for the α and β parameters in Equation (8) (we adopted different combinations of these hyperparameter values for the various chains, all within the ranges specified in the 'Discussion' section). The initial 1000 iterations are considered the *burn-in* period, and we discard the corresponding samples for the numerical computation of the statistical properties (i.e. mean and standard deviation) of the PPD.

Figure 13b illustrates the acceptance ratios of all the chains, calculated as the number of accepted models over the number of iterations. Notably, we observe values consistently above 65%–70% for all the chains, again a significant improvement over the typical rates achieved by standard, gradient-free MCMC algorithms. As in the synthetic test, given these high acceptance ratios, we adopted the same strategy after the end of the burn-in period to reduce the overall computational cost of the inversion, calculating the Jacobian matrix every 30 accepted models.

To verify the convergence of the chains, we have analysed the evolution of the PSRF values with the number of iterations for every model parameter in the compressed domain. As shown in Figure 13c, we observe that for most of the model parameters (more than 85%), we have reached PSRF values below the 1.2 threshold.

All the models sampled by the five chains after the *burn-in* period are then projected onto the full domain for the computation of the mean and standard deviation of the PPD.

Parts (a) and (b) of Figure 14 show two examples of starting models used for two different chains, related to the two different groups mentioned above. Parts (c) and (d) of Figure 14 represent, respectively, the posterior mean V_s model derived by considering all chains and the corresponding standard deviation map. To validate these outcomes, the available borehole data is utilized for comparison with the 1D velocity profile extracted from the posterior mean V_s model at the spatial position of 10 m, corresponding to the borehole position.

Figure 14d highlights, as expected, the region at 25 m of depth as having the highest standard deviation values, considering that it is located at the bottom part of the model and beneath the high velocity layer. Another portion of the model displaying high standard deviation values can be found at the interface between the first and second layers, which is characterized by the highest velocity variation. Figure 14e

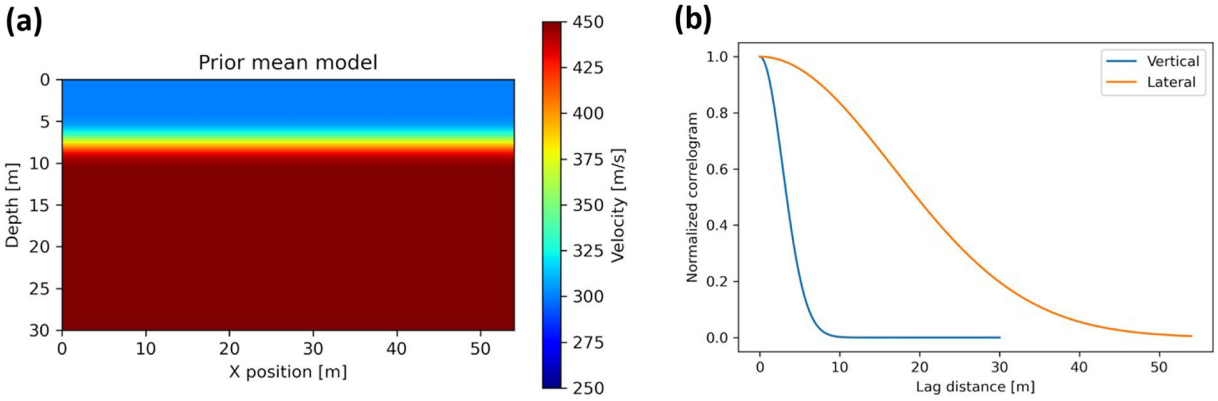

FIGURE 12 (a) Prior mean model used for all the five chains; (b) normalized spatial correlation function associated with the assumed 2D Gaussian variogram model.

demonstrates the accurate reproduction of the main velocity variations, specifically the detection of two high velocity layers at depths 10–15 and 19–25 m with a V_S value around 430 m/s. A thinner layer (approximately 5 m of thickness) characterized by a lower V_S velocity around 370 m/s is also visible. We are also able to identify the velocity inversion at 25 m of depth, at the bottom edge of the model.

In parts (a) and (b) of Figure 15, we display the left-most shot of the observed data, after the pre-processing steps discussed before, compared, respectively, with the data computed using one of the starting models of the chains (specifically, the one shown in Figure 14a) and with the predicted data generated using the posterior mean model of Figure 14c. Despite the fact that the algorithm starts from a very simple model, the inversion achieves a remarkable match between predicted and observed data.

Close-ups in Figure 16 highlight substantial differences between observed data and data calculated using one of the starting models, revealing a reduction in the cycle-skips when considering the finally predicted data.

This again confirms the robustness of the GB-MCMC approach against the cycle-skipping issue and its ability to sample the most promising model space regions (i.e. where velocity models with good data matching are located).

Inversion test with fewer DCT coefficients

This section describes, for the study site and for the employed acquisition layout, the important role played by the available borehole information to successfully recover the velocity inversion occurring in the thin layer located around 15–20 m depth. To illustrate this, we rerun the inversion, considering just 10 retained DCT coefficients along the vertical direction (instead of the 20 used in the previous chapter). In this new experiment, we use just one single MCMC chain starting from the starting model shown in Figure 14a.

This chain reaches a steady state at around 1000 iterations, again fixed as the end of the burn-in period. Figure 17 illustrates the posterior mean V_S model of the considered chain. Notably, the algorithm struggles to capture the thin, low-velocity layer, which was clearly visible in the previous results. However, it successfully reproduces all the other key features of the model with consistent velocity magnitudes. In other terms, the selected combination of DCT coefficients, together with the limited offset range of the available seismic gathers and the resulting poor data illumination of that thin layer, do not allow the inversion to successfully recover the velocity reversal between 15 and 20 m of depth. However, it is able to highlight the velocity inversion that starts at 25 m of depth without any well-log information.

The inspection of the likelihood values after the burn-in period for the two chains, one running with 20 and the other with 10 retained DCT coefficients, starting from the same initial model, can give some indication on the different outcomes. We observe that they both stabilize around similar values, which are just slightly lower for the chain employing 20 coefficients (Figure 18a). This suggests that 20 coefficients yield slightly better data matching, facilitating the accurate estimation of the velocity reversal at 15 m of depth.

The marginal enhancement in the data matching can also be appreciated when comparing two traces of the observed data with the corresponding traces of the predicted data generated from the posterior mean models obtained in the two considered inversion experiments.

For the mid-offset trace (Figure 18b), the traces pertaining to the two predicted data almost perfectly overlap. Instead, only for the farthest offset (Figure 18c), we observe very small differences. This means that a longer offset would probably be needed to better illuminate this velocity reversal occurring in such a thin layer. Table 2 provides a quantitative evaluation of how the chosen combination of retained DCT coefficients in the model space affects the predicted data. The relative percentage error with respect to the observed data, when

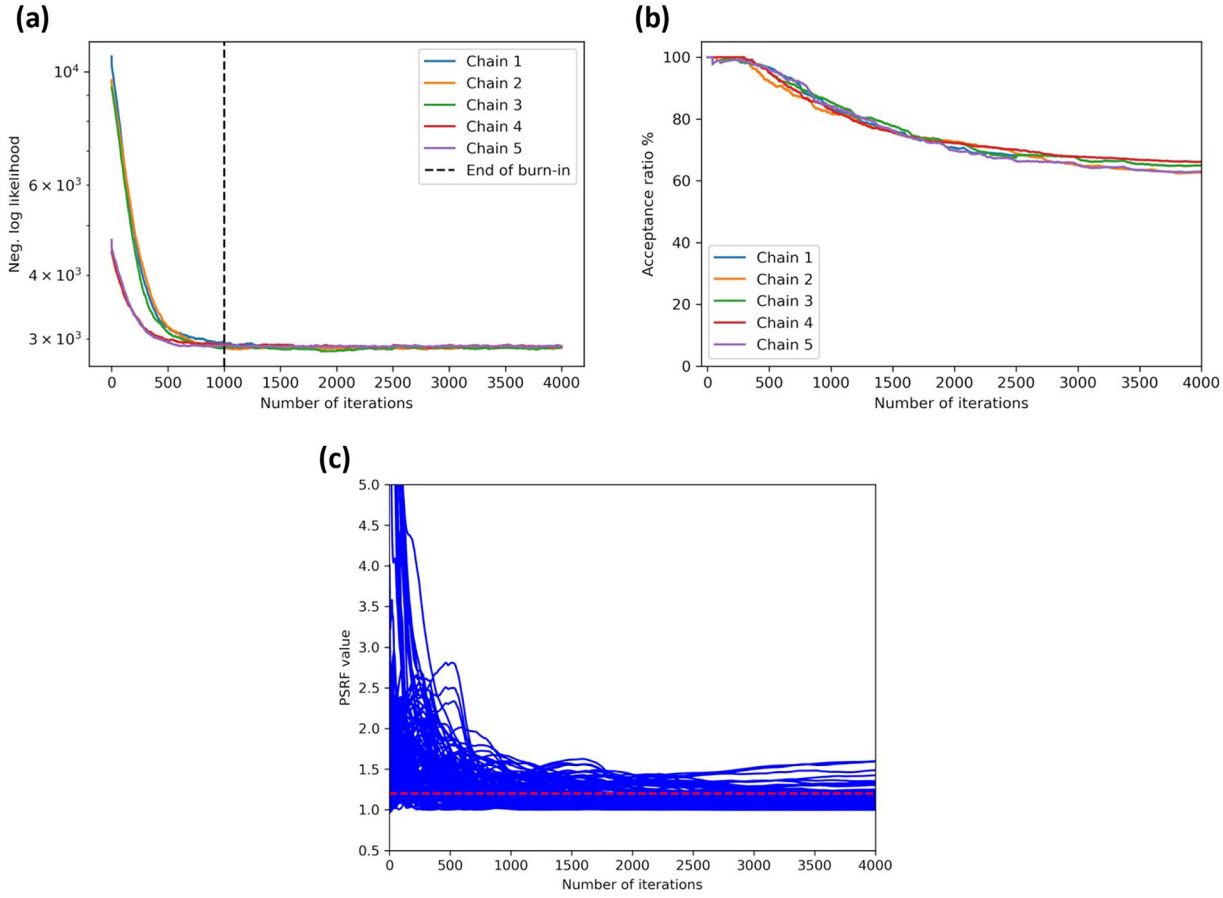

FIGURE 13 (a) The evolution of the negative log-likelihood with the number of iterations for all the considered chains. The dashed black vertical line defines the end of the burn-in period that we have selected; (b) acceptance ratios for all the five chains, measured as the number of accepted models over the number of iterations; (c) evolution of the potential scale reduction factor (PSRF) values for the model parameters in the discrete cosine transform (DCT) compressed domain (each blue line corresponds to one of the retained coefficients). The red dashed line highlights the threshold value for convergence (equal to 1.2).

TABLE 2 Comparison of the relative percentage error between predicted and observed data if we consider a different number of discrete cosine transform (DCT) coefficients along the vertical dimension.

Number of DCT coefficients along vertical dimension	Relative percentage error
10	34.5
20	27.7

employing a larger number of coefficients, is approximately 7% lower than that achieved with 10 coefficients.

Other tests, not shown here for brevity, demonstrate that the velocity reversal is not even recovered by a local inversion triggered from the predictions obtained with the reduced number of coefficients (i.e. from the model shown in Figure 17). Differently, the velocity inversion is recovered by the deterministic approach only if such a characteristic is already contained in the starting model (as we will show in the following section).

Local inversion

In this section, we test the same local inversion algorithm employed in the synthetic case to the real dataset. Specifically, we demonstrate that when initiating the local inversion with the same starting models used in the Bayesian approach, it fails to converge. However, if we start the inversion with the GB-MCMC estimation of Figure 14c, we not only achieve a further reduction in the misfit between predicted and observed data but also obtain a slight enhancement in the resolution of the velocity model, thus mitigating the impact of DCT compression.

The IFOS2D code is once again employed for the inversion, and, in this case, a three-step frequency marching strategy is adopted, covering frequency ranges of 0–10, 0–20 and 0–30 Hz. The same processing steps applied in the Bayesian inversion are implemented for the observed data.

Parts (a) and (b) of Figure 19 showcase the results of the local inversion, using the velocity models presented in parts (a) and (b) of Figure 14 as starting points, respectively. Even when initiating from a highly smoothed version of the velocity

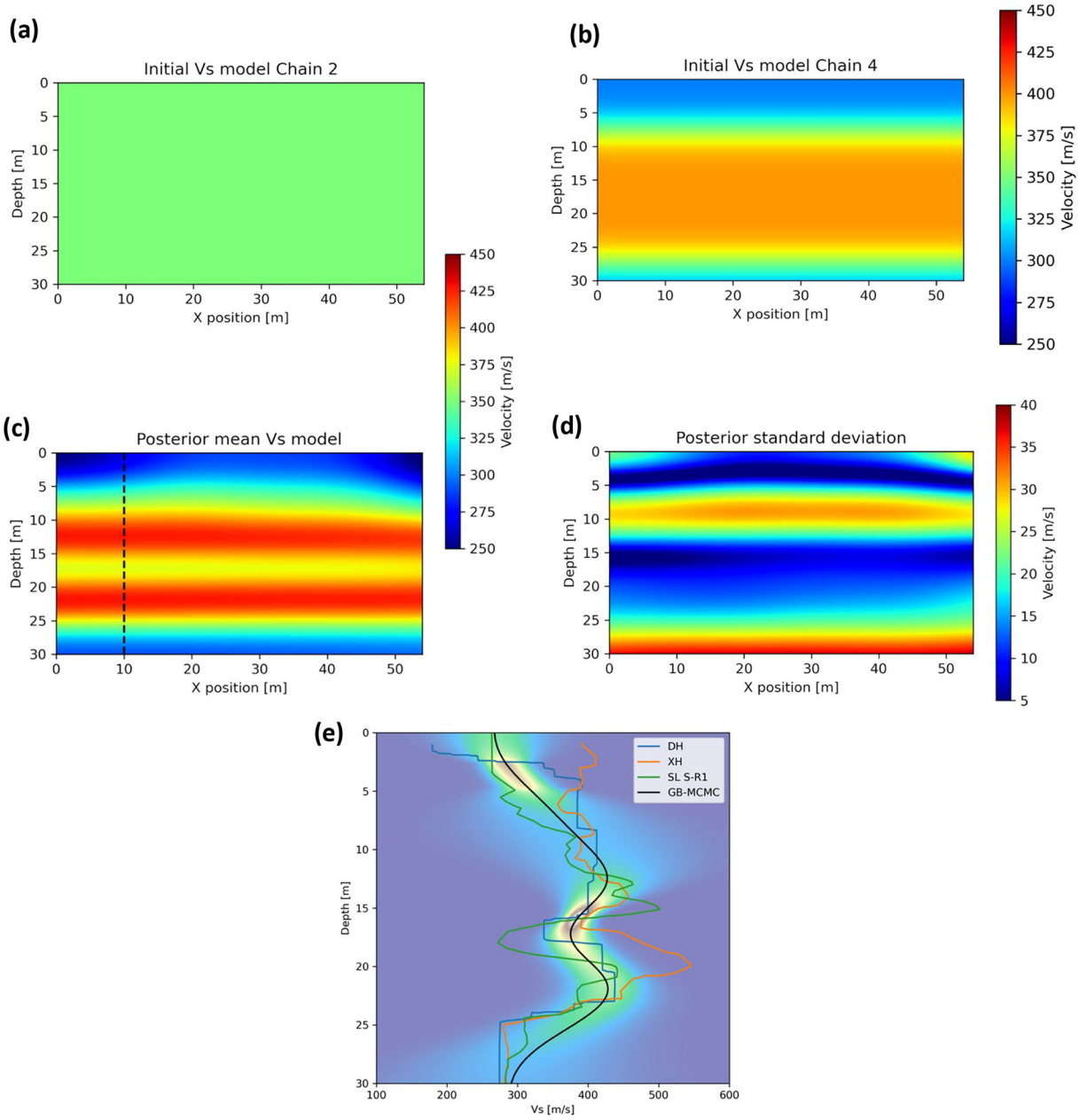

FIGURE 14 (a) Starting model used for Chain 2; (b) starting model used for Chain 4 derived from the available borehole data; (c) gradient-based Markov Chain Monte Carlo (GB-MCMC) estimation considering all the five chains with the dashed black line highlighting the position of the borehole data; (d) posterior standard deviation map computed using all the chains; (e) marginal distribution at the spatial position of 10 m (blue corresponds to zero probability, whereas hot colours correspond to high probability) together with the available borehole data (the blue, orange and green lines represent, respectively, the downhole, cross-hole and P–S suspension logging measurements), whereas the 1D velocity profile extracted from the posterior mean V_s model is displayed in black.

model derived from the borehole data, the local inversion fails to achieve convergence, getting stuck in some local minima of the error function and yielding predicted models significantly far from those obtained through the GB-MCMC approach.

Conversely, the GB-MCMC estimation from Figure 14c demonstrates to be a suitable starting point for the local inversion as we achieved convergence in fewer than 10 iterations. In particular, Figure 20 illustrates the final velocity model obtained in this test, in which additional features such as slight increases in velocity values in the upper part of the model and within the high velocity layers at depths of 10–15 and 19–25 m are added to the results of the GB-MCMC inversion.

The differences between the seismogram calculated with the model in Figure 20 after the local inversion and the one in Figure 14c obtained with the GB-MCMC are hardly discernible in a display of the whole shot. Instead, Table 3

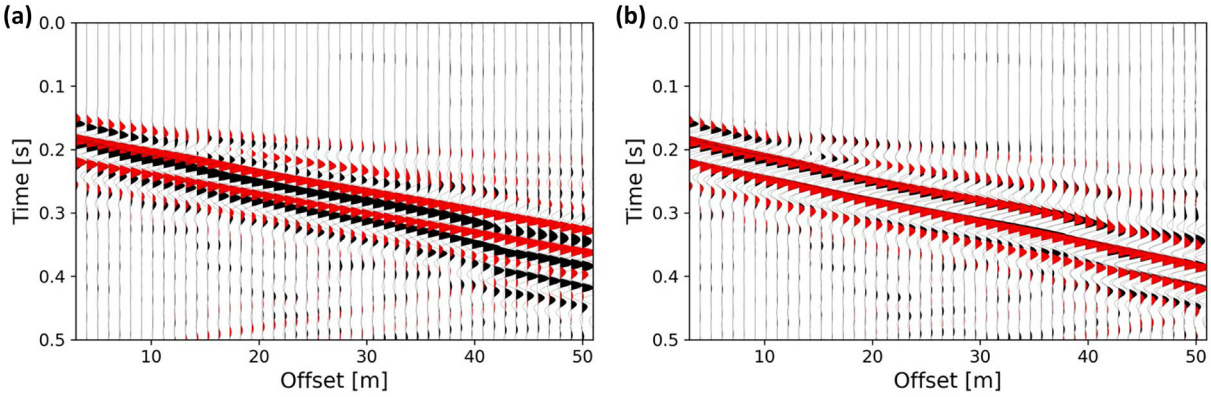

FIGURE 15 (a) Comparison between the leftmost shot of the observed data (in black) and the initial data, calculated using the starting model of Figure 14a (in red); (b) comparison between the same shot of the observed data (in black) and the predicted data, calculated using the gradient-based Markov Chain Monte Carlo (GB-MCMC) estimation of Figure 14c (in red).

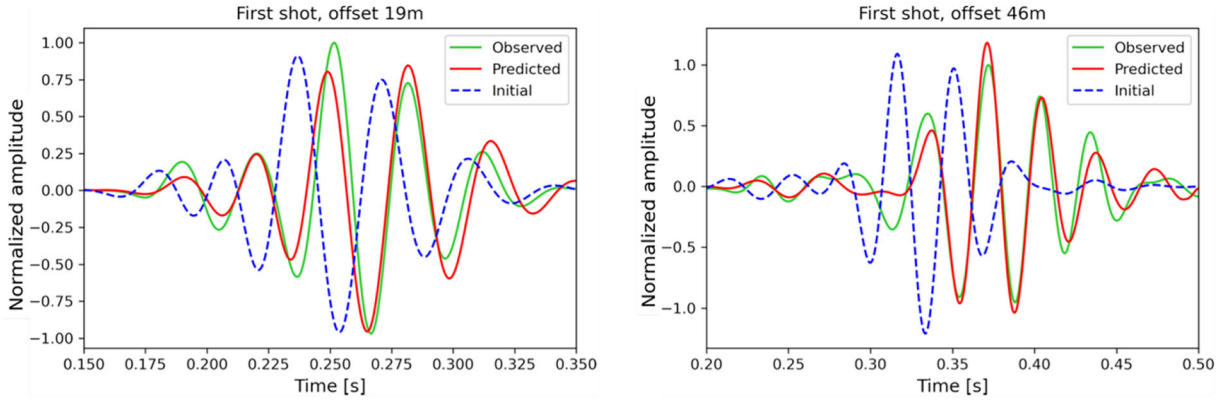

FIGURE 16 Close-ups on two traces of the first shot corresponding to mid and far offset, extracted from the observed data (green), the predicted data (red) and the initial data (dashed blue).

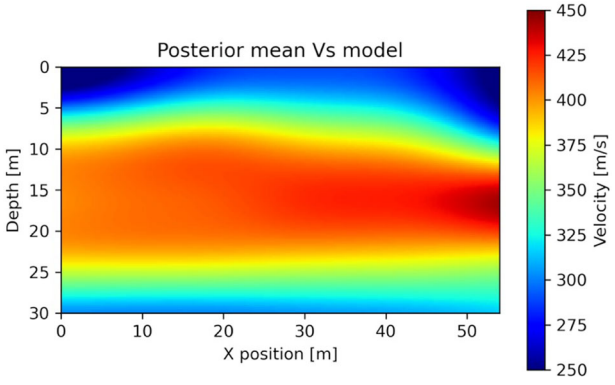

FIGURE 17 Posterior mean model obtained considering only 1 single-chain and 10 discrete cosine transform (DCT) coefficients along the vertical dimension.

provides a more quantitative comparison computing the relative percentage error with respect to the observed data for the GB-MCMC inversion and the combined GB-MCMC and local inversion. It results in a further reduction in the L2 norm data error of approximately 6%.

TABLE 3 Comparison of the relative percentage error between predicted and observed data using as inversion procedure the gradient-based Markov Chain Monte Carlo (GB-MCMC) algorithm and the combination of the GB-MCMC and local approaches.

Inversion method	Relative percentage error
GB-MCMC	27.7
GB-MCMC + LOCAL	21.4

DISCUSSION

In this work, we have proposed a novel probabilistic approach for inverting surface waves, which overcomes some of the limitations of the local full-waveform inversion (FWI) inversion. In particular, our method is devoted to attenuating the depen-

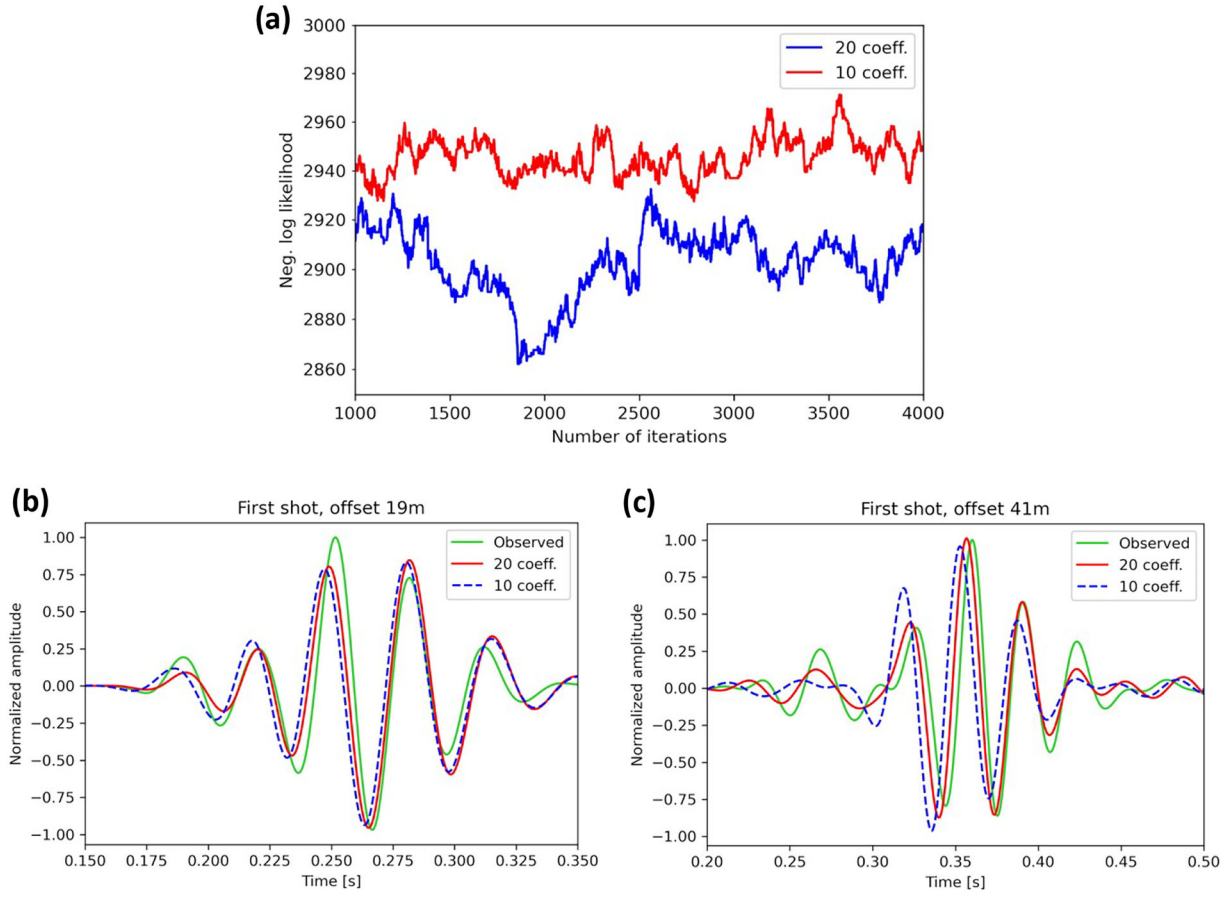

FIGURE 18 (a) Negative log-likelihood values after the burn-in period for the two chains using a different number of discrete cosine transform (DCT) coefficients along the vertical direction; (b and c) close-ups on two different traces extracted from the observed data (green), the predicted data calculated on the posterior mean model of Figure 14c (red) and the predicted data calculated on the posterior mean model of Figure 17 (dashed blue).

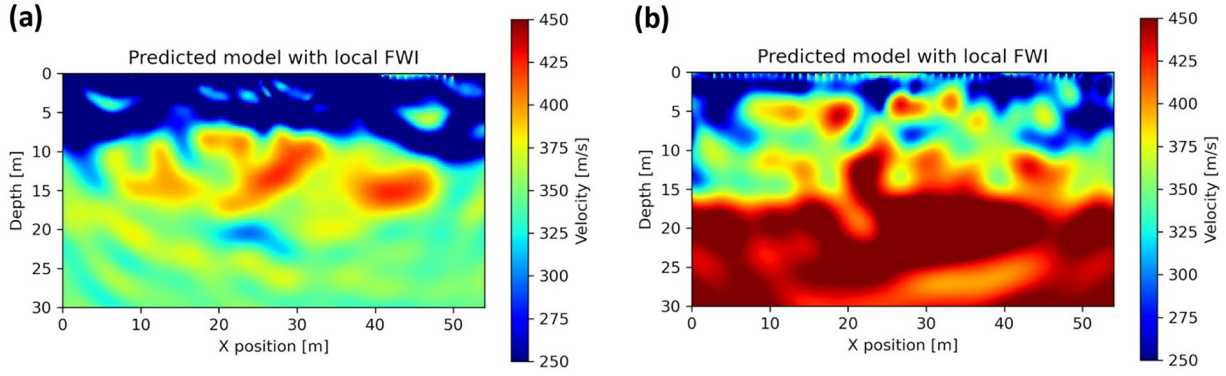

FIGURE 19 (a) Resulting model after 30 iterations of local inversion, using as starting point the model of Figure 14a; (b) resulting model after 20 iterations of local inversion, using as starting point the model of Figure 14b.

dency on the starting model and also providing an estimation of the uncertainties affecting the recovered velocity model.

Probabilistic FWI suffers from the curse of dimensionality in the case of high-dimensional problems and is computationally demanding in the case of expensive forward modellings. To overcome these two critical challenges, we have combined the efficiency of gradient-based Markov Chain Monte Carlo (GB-MCMC) sampling with the discrete cosine transform (DCT) compression of model and data spaces. The implemented MCMC algorithm exploits the local information brought by the gradient and the Hessian to define a proposal distribution as an approximation of the target posterior prob-

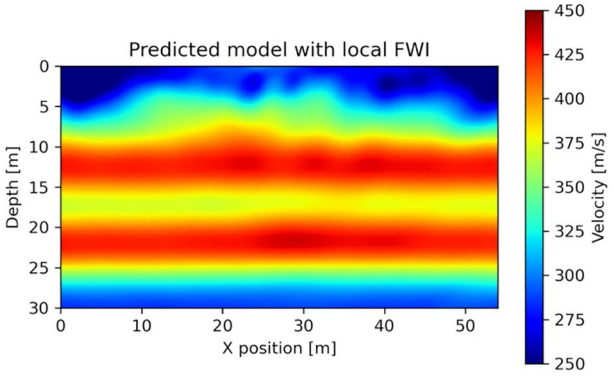

FIGURE 20 Predicted model of the local inversion after nine iterations, using as starting model the posterior mean of all the five chains provided by the gradient-based Markov Chain Monte Carlo (GB-MCMC) approach (Figure 14c).

ability density (PPD) around the currently sampled model. This significantly reduces the time requested by the algorithm to reach the steady state, compared to standard gradient-free MCMC methods. In addition, the random perturbation drawn from the proposal distribution and the use of multiple chains allow to reduce entrapments in local maxima of the posterior density, mitigating the cycle-skipping issue that affects local inversion methods. The compression of model and data spaces is a fundamental step of our algorithm, considering that it reduces the time for the Jacobian computation and the dimensions (and so the required storage memory) of all the matrices involved in the inversion procedure.

The reduced resolution of the velocity models due to the DCT effect can be improved using a higher number of retained coefficients, although increasing the computational cost of the inversion, or by choosing as a starting model for a second step of local FWI the outcomes of the MCMC inversion. The implemented method can also be used as a convenient strategy to define an initial model for the classical local inversion. Theoretically, the method can be extended to the 3D case because the DCT can be applied to signals with arbitrary dimensions. However, in this case, we must consider the extra computational cost needed by the forward modelling. From our side, additional tests are needed to assess the feasibility of the 3D inversion, with a particular focus on the computational efficiency of 3D forward modelling.

To further reduce the computational cost of our algorithm, we have adopted a strategy after the burn-in period. At these iterations, each chain has reached the steady state, and all the sampled models are very similar between each other. For this reason, we have decided to calculate the Jacobian matrix every approximately 20 or 30 accepted models (via a trial-and-error procedure) instead of for each one during the burn-in period. This strategy does not affect the estimated PPD but greatly reduces the time needed for each iteration considering that, without the Jacobian computation, we only need a number of forward evaluations equal to the number of shot gathers we are considering in the inversion procedure.

The efficacy of our approach depends on the precise calibration of the α and β^2 values, shown in Equation (8). These parameters can be promptly assessed during the initial iterations of the MCMC sampling. In this study, we set these parameters to achieve a favourable acceptance ratio (>60%), much higher than that usually associated with gradient-free MCMC methods. Additionally, our inversion tests revealed that multiple combinations of α and β^2 values can yield the desired acceptance ratio values, emphasizing that an appropriate calibration of these hyperparameters is not complicated to find. For instance, in both our synthetic and field data tests, all the α values within the [0.001, 0.05] interval and β^2 spanning $[10\alpha, 200\alpha]$ guaranteed efficient sampling.

In the context of GB-MCMC FWI, given that the inversion is primarily influenced by the V_s models, one possible approach to reduce the computational burden of the method could involve using different combinations of DCT coefficients for the V_s and V_p/V_s parameters. Specifically, opting for a reduced number of retained coefficients for the V_p/V_s model could reduce its resolution but, at the same time, the number of unknowns in our inverse problem. This adjustment aims to mitigate computational expenses without significantly compromising the inversion results in terms of data matching.

Opportunities exist to further alleviate the computational burden of probabilistic FWI, such as leveraging more powerful computing resources or implementing more efficient codes in lower level programming languages. In particular, other tests we carried out with a different forward modelling code running on the GPU (Deepwave; Richardson, 2022) have demonstrated a substantial reduction of the overall computational cost of the inversion. With this different forward modelling code, the synthetic and field data inversions run in 16 and 31 h, respectively, to be compared with the 85 and 200 h needed when using the SOFI algorithm running on the CPU. Additionally, following the works of Wu and Lin (2019), Wu and McMechan (2019) and Zhang and Alkhalifah (2022), our research is now focusing on using a supervised data-driven model prediction deep-learning approach to generate a reliable data-driven initial model that could significantly decrease the number of iterations needed to reach the end of the burn-in period (i.e. the number of Jacobian computations) and so further reduce the overall computational cost of the GB-MCMC inversion (Rincon et al., 2023).

CONCLUSIONS

We presented a probabilistic full-waveform inversion (FWI) of surface waves that integrates gradient-based Markov Chain

Monte Carlo (GB-MCMC) sampling with a compression of model and data spaces through the discrete cosine transform (DCT). The utilization of Hessian and gradient information of the posterior density directs the sampling towards the most favourable regions of the model space. Simultaneously, introducing random perturbations to the proposed models prevents entrapment in local maxima of the posterior density, significantly alleviating the cycle-skipping issue inherent in local FWI methods. The primary role of the DCT is to reduce the number of model parameters, resulting in a compression of the dimensions of the gradient vector, Hessian and Jacobian matrices. This not only addresses the ill-conditioning of the inverse problem but also reduces the computational burden and mitigates the curse of dimensionality affecting traditional MCMC samplings. The results obtained in the synthetic case validate the applicability and reliability of our approach, which, unlike conventional local methods, provides a comprehensive assessment of uncertainties affecting the estimated solution. When applied to the real dataset, we achieved an optimal misfit between predicted and observed data, with the predicted V_S model that accurately reproduces the main velocity variations evident in the available borehole data.

Our proposed GB-MCMC inversion exhibits convergence towards reliable and plausible solutions in both the synthetic and real tests, even when initiated from models significantly far from the optimal solution, highlighting the effectiveness of this approach to avoid cycle-skipping issues. Our upcoming research focuses on trying to further reduce the computational costs by incorporating machine learning techniques.

ACKNOWLEDGEMENTS

We would like to thank Matteo Ravasi (KAUST) and the Associate Editor for their many insightful comments and suggestions that significantly enhanced the quality of the manuscript. We would also like to thank, in particular, reviewer Anandaroop Ray for his constructive comments.

FUNDING INFORMATION

We received no fundings for this research.

DATA AVAILABILITY STATEMENT

Data are available on reasonable request to the corresponding author. The inversion codes are developed internally at the Earth Sciences Department of the University of Pisa and are not currently available. The SOFI2D package used in this work is freely available and can be found at the following link: https://git.scc.kit.edu/GPIAG-Software/SOFI2D.

ORCID

Sean Berti https://orcid.org/0009-0003-7894-5450
Mattia Aleardi https://orcid.org/0000-0003-1433-0281

REFERENCES

Alford, R.M., Kelly, K.R. & Boore, D.M. (1974) Accuracy of finite-difference modelling of the acoustic wave equation. *Geophysics*, 39(6), 834–842.

Ahmed, N., Natarajan, T. & Rao, K. (1974) Discrete cosine transform. *IEEE Transactions on Computers*, C-23(1), 90–93.

Aleardi, M. (2020a) Discrete cosine transform for parameter space reduction in linear and non-linear AVA inversions. *Journal of Applied Geophysics*, 179, 104106.

Aleardi, M. (2020b) Combining discrete cosine transform and convolutional neural networks to speed up the Hamiltonian Monte Carlo inversion of pre-stack seismic data. *Geophysical Prospecting*, 68(9), 2738–2761.

Aleardi, M. (2021) A gradient-based Markov chain Monte Carlo algorithm for elastic pre-stack inversion with data and model space reduction. *Geophysical Prospecting*, 69(3), 926–948. https://doi.org/10.1111/1365-2478.13081

Aleardi, M. & Stucchi, E. (2021) A hybrid residual neural network-Monte Carlo approach to invert surface wave dispersion data. *Near Surface Geophysics*, 19(4), 397–414.

Aleardi, M., Vinciguerra, A., Stucchi, E. & Hojat, A. (2022) Machine learning-accelerated gradient-based Markov chain Monte Carlo inversion applied to electrical resistivity tomography. *Near Surface Geophysics*, 20(4), 440–461.

Aleardi, M. (2019) Using orthogonal Legendre polynomials to parameterize global geophysical optimizations: applications to seismic-petrophysical inversion and 1D elastic full-waveform inversion. *Geophysical Prospecting*, 67(2), 331–348.

Aleardi, M. & Tognarelli, A. (2016) The limits of narrow and wide-angle AVA inversions for high Vp/Vs ratios: an application to elastic seabed characterization. *Journal of Applied Geophysics*, 131, 54–68.

Alkhalifah, T.A. (2016) *Full waveform inversion in an anisotropic world: where are the parameters hiding?* Houten, The Netherlands: EAGE.

Bergamo, P. & Socco, L.V. (2014) Detection of sharp lateral discontinuities through the analysis of surface-wave propagation. *Geophysics*, 79(4), EN77–EN90.

Berti, S., Aleardi, M. & Stucchi, E. (2024a) A computationally efficient Bayesian approach to full-waveform inversion. *Geophysical Prospecting*, 72, 580–603.

Berti, S., Aleardi, M. & Stucchi, E. (2024b) A Bayesian approach to elastic full-waveform inversion: application to two synthetic near surface models. *Bulletin of Geophysics and Oceanography*, 65(2), 291–308.

Biondi, E., Stucchi, E. & Mazzotti, A. (2014) Nonstretch normal moveout through iterative partial corrections and deconvolution. *Geophysics*, 79(4), V131–V141.

Bodin, T., Sambridge, M., Tkalcic, H., Arroucau, P., Gallagher, K. & Rawlinson, N. (2012) Transdimensional inversion of receiver functions and surface wave dispersion. *Journal of geophysical Research: Solid Earth*, 117, B02301.

Bohlen, T. (2002) Parallel 3-D viscoelastic finite difference seismic modelling. *Computers & Geosciences*, 28(8), 887–899.

Bohlen, T., Kugler, S., Klein, G. & Theilen, F. (2004) 1.5D inversion of lateral variation of Scholte wave dispersion. *Geophysics*, 69(2), 330–344.

Brossier, R., Operto, S. & Virieux, J. (2009) Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics*, 74(6), WCC105–WCC118.

Choi, Y. & Alkhalifah, T. (2012) Application of multi-source waveform inversion to marine streamer data using the global correlation. *Geophysical Prospecting*, 60, 748–758.

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

Datta, D. & Sen, M.K. (2016) Estimating a starting model for full-waveform inversion using a global optimization method. *Geophysics*, 81, R211–R223.

Dejtrakulwong, P., Mukerji, T. & Mavko, G. (2012) Using kernel principal component analysis to interpret seismic signatures of thin shaly-sand reservoirs. In *SEG Technical Program Expanded Abstracts 2012*, Society of Exploration Geophysicists.

De Nil, D. (2005) Characteristics of surface waves in media with significant vertical variations in elasto-dynamic properties. *Journal of Environmental and Engineering Geophysics*, 10(3), 263–274.

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

Fichtner, A. & Simutė, S. (2018) Hamiltonian Monte Carlo inversion of seismic sources in complex media. *Journal of Geophysical Research: Solid Earth*, 123(4), 2984–2999.

Forbriger, T., Groos, L. & Schafer, M. (2014) Line-source simulation for shallow-seismic data. Part 1: theoretical background. *Geophysical Journal International*, 198(3), 1387–1404.

Foti, S., Lai, C.G., Rix, G.J. & Strobbia, C. (2014) *Surface wave methods for near-surface site characterization*. Boca Raton, FL: CRC Press.

Foti, S., Hollender, F., Garofalo, F., Albarello, D., Asten, M., Bard, P.Y. et al. (2018) Guidelines for the good practice of surface wave analysis: a product of the InterPACIFIC project. *Bulletin of Earthquake Engineering*, 16(6), 2367–2420.

Gao, L., Xia, J., Pan, Y. & Xu, Y. (2016) Reason and condition for mode kissing in MASW method. *Pure Applied Geophysics*, 173(5), 1627–1638.

Garofalo, F., Foti, S., Hollender, F., Bard, P.Y., Cornou, C., Cox, B. et al. (2016) InterPACIFIC project: comparison of invasive and non-invasive methods for seismic site characterization. Part I: intra-comparison of surface wave methods. *Soil Dynamics and Earthquake Engineering*, 82, 222–240.

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

Gelman, A. & Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–511.

Groos, L., Schafer, M., Forbriger, T. & Bohlen, T. (2014) The role of attenuation in 2D full waveform inversion of shallow seismic body and Rayleigh waves. *Geophysics*, 79(6), R247–R261.

Groos, L., Schafer, M., Forbriger, T. & Bohlen, T. (2017) Application of a complete workflow for 2D elastic full-waveform inversion to recorded shallow-seismic Rayleigh waves. *Geophysics*, 82(2), R109–R117.

Haario, H., Saksman, E. & Tamminen, J. (2001) An adaptive Metropolis algorithm. *Bernoulli*, 7(2), 223–242.

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

Hawkins, R. & Sambridge, M. (2015) Geophysical imaging using trans-dimensional trees. *Geophysical Journal International*, 203, 972–1000.

Ivanov, J., Tsoflias, G., Miller, R.D., Peterie, S., Morton, S. & Xia, J. (2016) Impact of density information on Rayleigh surface wave inversion results. *Journal of Applied Geophysics*, 35, 43–54.

Jain, A.K. (1989) *Fundamentals of digital image processing*. Hoboken, NJ: Prentice Hall.

Kirlin, R.L. & Done, W.J. (1999) *Covariance analysis for seismic signal processing*. SEG Books.

Koch, M.C., Fujisawa, K. & Murakami, A., (2020) Adjoint Hamiltonian Monte Carlo algorithm for the estimation of elastic modulus through the inversion of elastic wave propagation data. *International Journal for Numerical Methods in Engineering*, 121(6), 1037–1067.

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

Lamuraglia, S., Stucchi, E. & Aleardi, M. (2022) Application of a global-local full-waveform inversion of Rayleigh wave to estimate the near-surface shear wave velocity model. *Near Surface Geophysics*, 21, 1–18.

Lochbuhler, T., Breen, S.J., Detwiler, R.L., Vrugt, J.A. & Linde, N. (2014) Probabilistic electrical resistivity tomography of a CO2 sequestration analog. *Journal of Applied Geophysics*, 80, 92.

Louboutin, M., Lange, M., Herrmann, F.J., Kukreja, N. & Gorman, G. (2017) Performance prediction of finite-difference solvers for different computer architectures. *Computer & Geosciences*, 105, 148–157.

Malinverno, A. (2002) Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophysical Journal International*, 151, 675–688.

Maraschini, M. & Foti, S. (2010) A Monte Carlo multimodal inversion of surface waves. *Geophysical Journal International*, 182(3), 1557–1566.

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 Journal on Scientific Computing*, 34, A1460–A1487.

Mosegaard, K. & Tarantola, A. (2002) Probabilistic approach to inverse problems. *International Geophysics Series*, 81, 237–268.

Neal, R. (2011) MCMC using Hamiltonian dynamics. In: *Handbook of Markov chain Monte Carlo*. 116 62, Boca Raton, FL: Chapman Hall/CRC.

O'Neill, A., Campbell, T. & Matsuoka, T. (2008) Lateral resolution and lithological interpretation of surface-wave profiling. *The Leading Edge*, 27, 1550–1563.

Pan, Y., Gao, L. & Bohlen, T. (2019) High-resolution characterization of near-surface structures by surface-wave inversions: from dispersion curve to full waveform. *Survey in Geophysics*, 40, 167–195.

Pierini, S. & Stucchi, E. (2020) Points per wavelength analysis in global elastic FWI of surface waves: a synthetic case study. In 26th European Meeting of Environmental and Engineering Geophysics. Utrecht, the Netherlands, European Association of Geoscientists & Engineers. pp. 1–5.

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

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

Richardson, A. (2022) *Deepwave*. Available at: https://ausargeo.com/deepwave/

Rincon, F., Berti, S., Aleardi, M. & Stucchi, E. (2023) Supervised neural network for surface waves data-driven Vs-model prediction. In: NSG2023 29th European Meeting of Environmental and Engineering Geophysics. Utrecht, the Netherlands, European Association of Geoscientists & Engineers. pp. 1–5.

Romdhane, G., Grandjean, G., Brossier, R., Rejiba, F., Operto, S. & Virieux, J. (2011) Shallow-structure characterization by 2D elastic full-waveform inversion. *Geophysics*, 76(3), R81–R93.

Sajeva, A., Aleardi, M. & Mazzotti, A. (2017) Genetic algorithm full-waveform inversion: uncertainty estimation and validation of the results. *Bulletin of Geophysics and Oceanography*, 58(4), 395–414.

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

Sambridge, M. (2014) A parallel tempering algorithm for probabilistic sampling and multimodal optimization. *Geophysical Journal International*, 196(1), 357–374.

Schafer, M., Groos, L., Forbriger, T. & Bohlen, T. (2014) Line-source simulation for shallow-seismic data. Part2: full-waveform inversion—a synthetic 2D case study. *Geophysical Journal International*, 198(3), 1405–1418.

Sen, M.K. & Stoffa, P.L. (2013) *Global optimization methods in geophysical inversion*. Cambridge: Cambridge University Press.

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

Sherlock, C., Fearnhead, P. & Roberts, G.O. (2010) The random walk metropolis: linking theory and Practice through a case study. *Statistical Science*, 25(2), 172–190.

Socco, L.V. & Strobbia, C. (2004) Surface-wave method for near-surface characterization: a tutorial. *Near Surface Geophysics*, 2(4), 165–185.

Socco, L.V., Foti, S. & Boiero, D. (2010) Surface-wave analysis for building near-surface velocity models – established approaches and new perspectives. *Geophysics*, 75(5), A83–A102.

Tran, K.T., McVay, M., Faraone, M. & Horhota, D. (2013) Sinkhole detection using 2D full seismic waveform tomography. *Geophysics*, 78(5), 175–183.

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

Warner, M. & Guasch, L. (2016) Adaptive waveform inversion: theory. *Geophysics*, 81(6), R429–R445.

Wu, Y. & McMechan, G.A. (2019) Parametric convolutional neural network-domain full-waveform inversion. *Geophysics*, 84(6), R881–R896.

Wu, Y. & Lin, Y. (2019) InversionNet: an efficient and accurate data-driven full waveform inversion. *IEEE Transactions on Computational Imaging*, 6, 419–433.

Xing, Z. & Mazzotti, A. (2019) Two-grid full-waveform Rayleigh-wave inversion via a genetic algorithm-part 1: method and synthetic examples. *Geophysics*, 84(5), R805–R814.

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

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

Zhang, Z. & Alkhalifah, T. (2022) Regularized elastic full-waveform inversion using deep learning. In: *Advances in subsurface data analytics*. Amsterdam, the Netherlands: Elsevier, pp. 219–250.

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

How to cite this article: Berti, S., Aleardi, M. & Stucchi, E. (2024) A probabilistic full waveform inversion of surface waves. *Geophysical Prospecting*, 1–26. https://doi.org/10.1111/1365-2478.13595

APPENDIX A

In this appendix, we present V_p estimates derived from Bayesian inversion in both synthetic and real cases, along with their corresponding uncertainty assessments. For both the inversions, we have considered the V_p/V_s ratios as model parameters, from which we can derive the V_p estimations. The initial V_p/V_s models and prior mean models for both scenarios were set to a homogeneous model with a constant value of 2.5.

For the synthetic case, the posterior mean V_p model in Figure A.1b (again computed using all the sampled models from the five chains, after the burn-in period) has reasonable velocity magnitude values compared to the true V_p model shown in Figure A.1a, and the high-velocity layer below 30 m of depth is well reconstructed. However, in the shallow portion of the model, we can see how the predicted model closely reproduces all the features of the V_s model shown in Figure 5b, maintaining a scaling factor very close to the starting one. The associated posterior standard deviation map (Figure A.1c) exhibits magnitude values similar to the ones associated with the V_s model shown in Figure 5c.

In the field dataset inversion, the posterior mean V_p model (Figure A.1d) can be seen as a scaled version of the predicted V_s model shown in Figure 14c, with a scaling factor approximately matching the one used for the initial model. The posterior standard deviation map (Figure A.1e) in this case indicates higher uncertainties compared to the one associated with the V_s model, shown in Figure 14d.

In this appendix, we will show how our proposed method performs under different noise levels and consider some data quality issues, such as missing channels and a not perfectly known source wavelet. We have used the same starting model shown in Figure 3c for these inversions, keeping the same acquisition geometry.

Parts (a) and (b) of Figure B.1 show the posterior mean models obtained when the signal-to-noise ratio of the observed data is decreased from 14 dB (as in the previously discussed synthetic experiments) to 10 and 5 dB, respectively. We can notice that all the main features of the original V_s

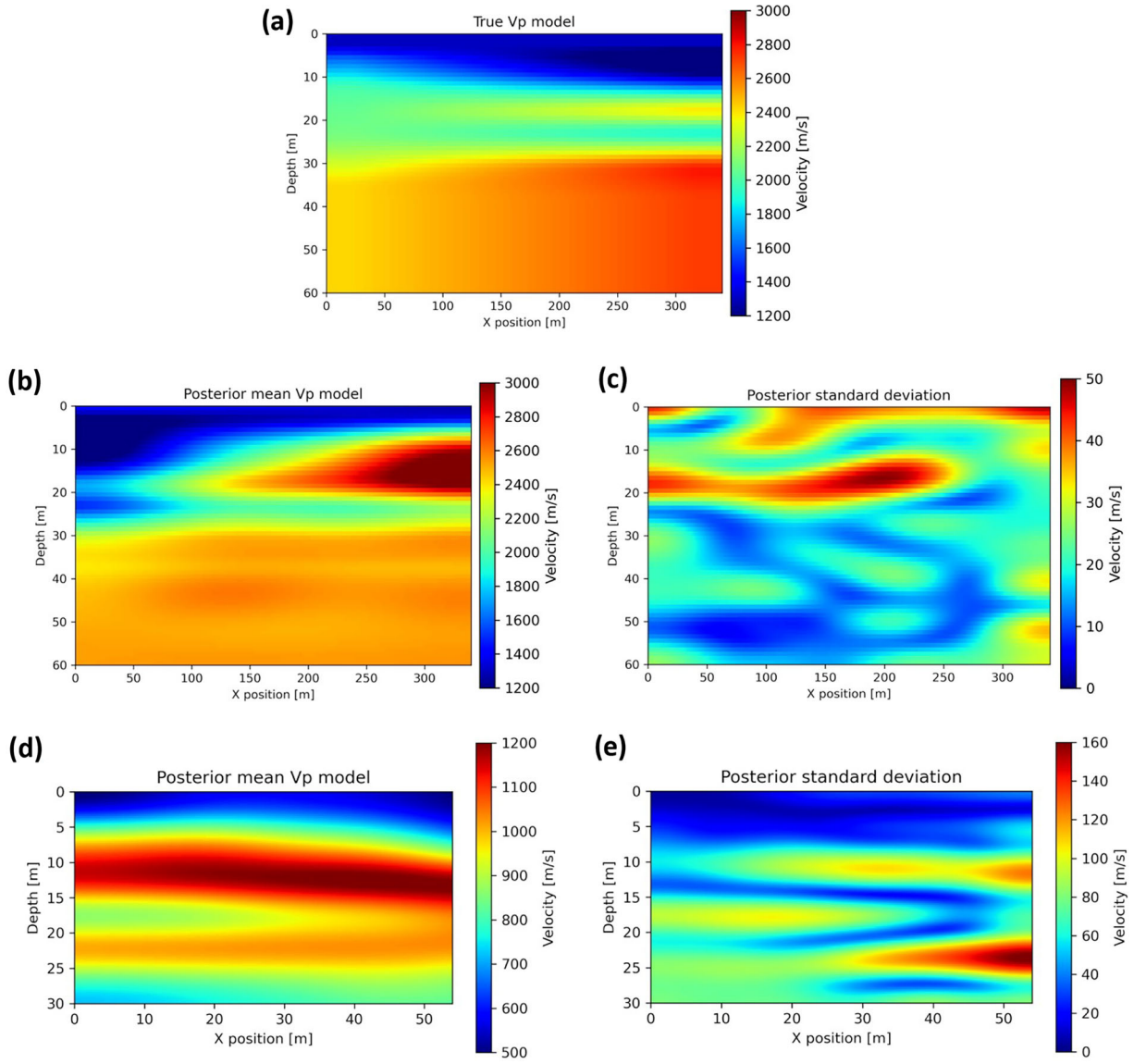

FIGURE A.1 (a) True V_p model for the synthetic case; (b) posterior mean V_p model and (c) its associated standard deviation map; (d) posterior mean V_p model for the field dataset inversion and (e) its associated standard deviation map. For both cases, we have considered the same chains and the same burn-in period as in the previous tests.

are well reconstructed, although some minor artefacts start to appear in the least illuminated portion of the model, especially in the deepest part.

For what concerns the data quality, we have performed two tests, eliminating, respectively, 5% and 10% of the total traces in both the observed and simulated data. The posterior mean models achieved for these two inversions are illustrated in parts (b) and (c) of Figure B.1. As in the previous test, we can see some artefacts in the deepest portion of the study area, but all the main characteristics of the original model are clearly visible.

The last test was performed using the source wavelet to compute the seismograms, which were estimated from the observed data using the SVD approach (the same method employed in the field data application).

Figure B.2 shows the comparison between the two wavelets, the estimated and the true Ricker used for computing the observed data and the posterior mean model estimated in this test. The similarity of the estimated posterior mean with the ground truth proves the robustness of the approach against errors in the estimated source wavelet.

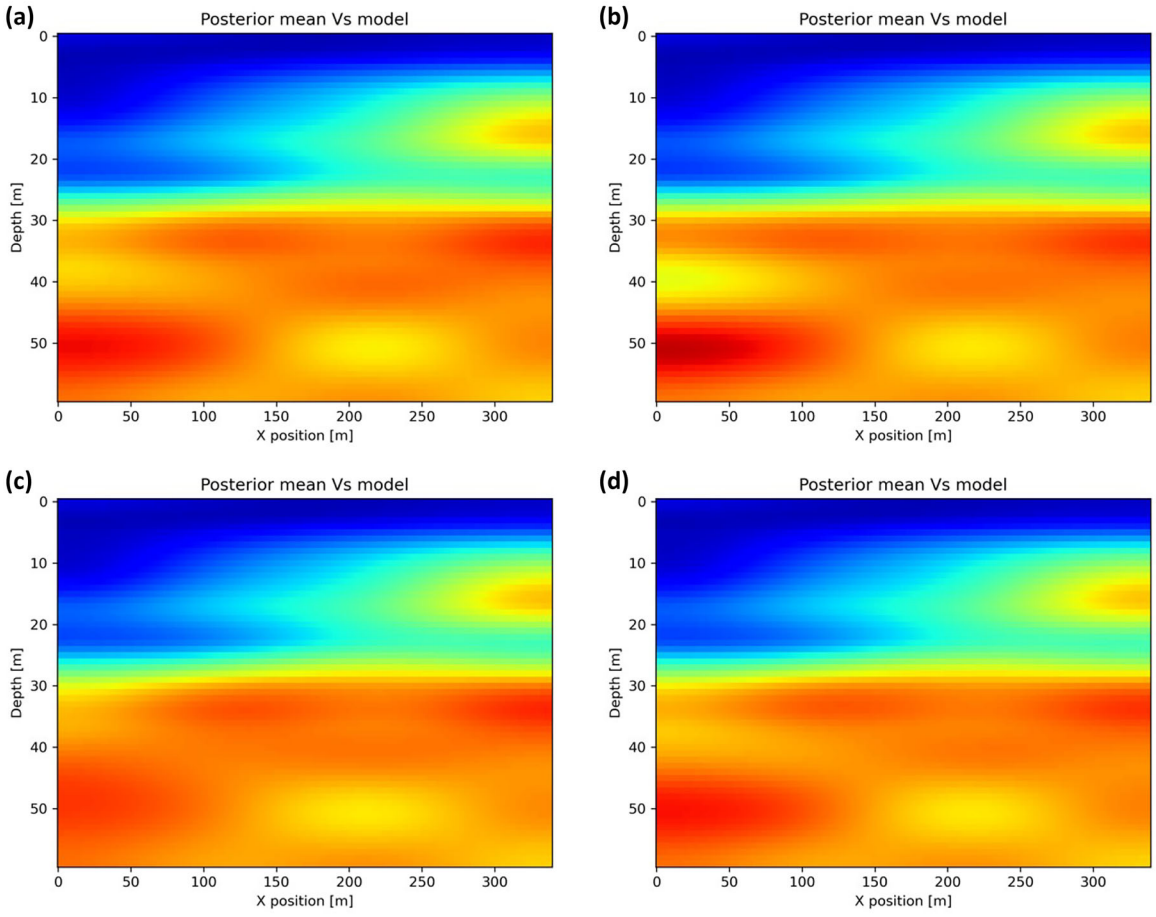

FIGURE B.1 (a and b) Posterior mean V_s models obtained considering different noise levels in our observed data, resulting in a signal-to-noise ratio (SNR) of 10 and 5 dB, respectively; (c and d) posterior mean V_s models obtained considering some missing channels in our observed data, in particular 5% and 10% of the total number of receivers, respectively.

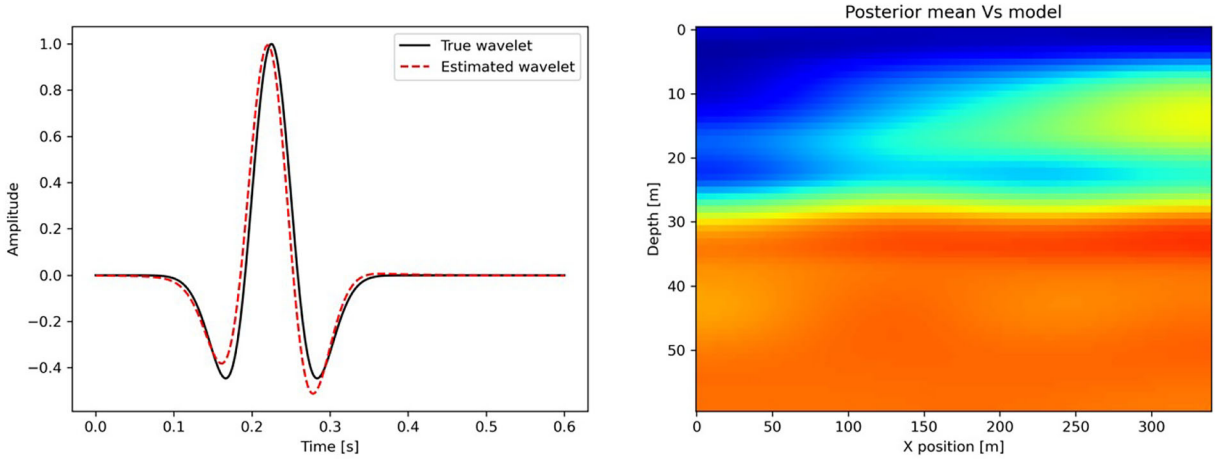

FIGURE B.2 On the left, a comparison between the true Ricker wavelet employed when generating the observed data (black) and the estimated wavelet from the observed data using the SVD approach (dashed red). On the right, the posterior mean V_s model obtained considering the estimated wavelet for generating the simulated data.