DOI 10.4430/bgo00442

A Bayesian approach to elastic full-waveform inversion: application to two synthetic near surface models

S. Berti[1,2], M. Aleardi[1] and E. Stucchi[1]

[1] *Earth Sciences Department, University of Pisa, Pisa, Italy*
[2] *Earth Sciences Department, University of Florence, Florence, Italy*

(Received: 28 May 2023; accepted: 12 November 2023; published online: 16 February 2024)

ABSTRACT Imaging of the first metres of the subsurface with seismic methods constitutes a key challenge for several applications. In this context, the analysis of Rayleigh waves can reveal information about the S-wave velocity structure in the first metres of the subsurface. The waves recorded can be inverted using several techniques, of which the most widely used is the multichannel analysis of surface waves, where dispersion curves are picked on the velocity-frequency spectrum. A full-waveform inversion of surface waves has been implemented, offering the possibility to exploit the complete information content of the recorded seismograms. This method has only recently been tested with elastic approximation on synthetic data, as the application in near-surface scenarios is very challenging due to the high nonlinearity of the problem and the considerable computational costs. This paper presents a gradient-based Markov chain Monte Carlo elastic full-waveform inversion method, where posterior sampling is accelerated by compressing data and model spaces through the discrete cosine transform and, also, by defining a proposal that is a local, Gaussian approximation of the target posterior probability density. The applicability of the approach is demonstrated by performing two synthetic inversion tests on two different near-surface models: a two-layered model with lateral velocity variations, and a four-layered model with velocity inversions.

Key words: full-waveform inversion, near-surface, elastic, seismic waves.

1. Introduction

Full-waveform inversion (FWI), a high-resolution technique that has become very popular in recent years due to the rapid development of computational power (Virieux and Operto, 2009), is used in geophysics to build consistent physical models of the Earth's subsurface, based on the observations of complete seismic wave propagation.

Acoustic FWI is widely used to image complex subsurface structures, although acoustic approximation only reconstructs the P-wave velocity model and ignores the S-wave structure and propagation. However, most seismic data show elastic effects (especially true for media containing high contrast interfaces) which are not only due to the generation of converted waves, but, above all, to the fact that most of the seismic energy generated by the source propagates as surface waves (Rayleigh or Love waves). Within the near surface context (and more specifically in the first metres of the subsurface), the elastic inversion of Rayleigh waves takes on great importance in the reconstruction of shear wave velocity models for subsoil characterisation,

© 2024 - OGS

given their high sensitivity to this elastic parameter. Methods established for the inversion of surface waves, such as the multichannel analysis of surface waves [MASW: Bohlen *et al*. (2004), Socco and Strobbia (2004), Maraschini and Foti (2010), Socco *et al*. (2010), and Aleardi and Stucchi (2021)], have many limitations, the main being the 1D assumption of the model. However, the picking of the dispersion curves, at very low frequencies, is usually inaccurate and, therefore, also causes uncertainties in the depth position and velocity of the deepest layers (Foti *et al*., 2014, 2018). Moreover, the choice of the intended final model can be dictated by personal preferences and previous assumptions about the subsurface structure rather than by information contained in the seismic data. Over the past few years, several approaches have been proposed seeking to overcome these limitations and recover pseudo-2D subsurface models, though strong assumptions are required to regularise the inversion (Strobbia and Foti, 2006; O'Neill *et al*., 2008; Bergamo and Socco, 2014) and, in the case of heterogeneous geologic settings where no continuous dispersion curves can be extracted from the seismic data, these approaches fail.

Conversely, elastic FWI has the potential to build high-resolution subsurface models also for 2D structures. There have been several successful FWI applications in exploration geophysics and seismology (Fichtner *et al*., 2009), yet there are few studies that apply FWI to image the shallow subsurface. Romdhane *et al*. (2011) and Tran *et al*. (2013) demonstrated the applicability and reliability of 2D frequency-domain elastic FWI in reconstructing heterogeneous near-surface models, while Groos *et al*. (2014, 2017), Xing and Mazzotti (2019), and Lamuraglia *et al*. (2022) proved that 2D time-domain elastic FWI could efficiently delineate shallow subsurfaces with high resolutions.

The elastic FWI method adds a further set of challenges to practical FWI applications, due to the complex nature of the elastic wavefields. Typical issues encountered with the acoustic FWI consist in the considerable computational costs and the highly non-linear and non-convex optimisation, including multiple local minima with the possible entrapment of the inversion, especially for the conventional least-squares misfit (i.e. the so-called cycle skipping issue). Compared to the acoustic method, the elastic method also entails a higher computational cost [up to 20 points per wavelength must be used in the finite difference (FD) modelling step to satisfy the dispersion conditions, thus implying a very small time step in order to fulfil the Courant-Friedrichs-Lewy condition (Pierini and Stucchi, 2020)] and unknowns that must be determined, i.e. V_p and density. An adequate initial model should be provided for the inversion algorithm when a local optimisation method is applied (Schafer *et al*., 2014; Alkhalifah, 2016; Xing and Mazzotti, 2019), though this may not be available in all cases. On the contrary, if a global optimisation method is adopted, this very strict requirement can be eliminated (Sajeva *et al*., 2017; Aleardi *et al*., 2019). The downside of this approach is that, when the number of unknowns is high, the computational time is usually too long to be sustainable.

The aim of this work is to address issues affecting the FWI of elastic waves and, in particular: i) to attenuate the need for a starting model close to the true subsurface solution, and ii) to provide statistical assessments on the estimated model parameters. For this reason, the elastic FWI solution is sought for in a Bayesian inference framework (Mosegaard and Tarantola, 2002; Gebraad *et al*., 2020) using a Markov chain Monte Carlo (MCMC) method that, theoretically, results in an inversion process able to explore the model space, which, on one hand, overcomes the cycle skipping issue and, on the other, could accurately estimate the model uncertainties.

The final result of a Bayesian inversion is the so-called posterior probability density (PPD) function in the model space that fully quantifies the uncertainties affecting the recovered solution. The PPD is constructed from the combination of two separate pieces of information:

the prior knowledge on the model parameters, and the information provided by the experiment, represented by the recorded seismic data.

In case of linear forward modellings and Gaussian uncertainties, the probabilistic formalism provides the same closed-form solution as the conventional least-squares approach. In general, however, the resulting high-dimensional integrals cannot be computed. MCMC methods provide a knowledgeable way to construct a Markov chain that produces samples drawn from the posterior target distribution (Sambridge and Mosegaard, 2002). In theory, a MCMC method is able to estimate the posterior distribution using sufficiently long Markov chains with random starting points.

One of the difficulties in implementing MCMC methods lies in the fact that the convergence rate towards the target PPD critically depends on the random perturbation applied to the current state of the chain. In particular, probabilistic sampling is maximally efficient when the proposal is a reasonable approximation of the target posterior. In addition, the sampling ability of these methods is known to dramatically narrow down to large-dimensional problems due to the so-called curse of dimensionality (Curtis and Lomax, 2001). Ultimately, traditional sampling strategies [such as the random walk Metropolis (Sherlock *et al*., 2010)] may require billions of forward evaluations before reaching convergence. A traditional sampling strategy, thus, becomes computationally unfeasible for expensive forward modellings or high-dimensional problems (as in the case of elastic FWI).

To tackle the computational complexity of the MCMC methods, when applied to FWI, in this paper, a sampling algorithm where the proposal distribution is built using the local gradient and the Hessian of the negative log posterior, is introduced and, in addition, both data and model spaces are reduced through the discrete cosine transform (DCT) reparameterisation.

The algorithm used here is named the gradient-based Markov chain Monte Carlo (GB-MCMC). A similar approach was proposed by Zhao and Sen (2021) for the acoustic FWI (without model and data compression) by only taking into consideration the diagonal entries of the Hessian matrix, so as to reduce the sampling efficiency of the method.

The use of the local Hessian and gradient information, computed around the current state of the chain, greatly increases the convergence of probabilistic sampling, as the proposed model is drawn from a local approximation of the target PPD. The downside of this approach is that derivatives must be calculated for each model sampled, and that, in large model and data spaces, the Hessian matrix and the gradient vector become computationally intractable. A suitable strategy to reduce the computational complexity of this type of inverse problem is to compress the model and data spaces through adequate reparameterisation techniques (e.g. the DCT), thus reducing the number of data and model parameters, and, consequently, the dimensions of both the Hessian matrix and the gradient vector. In this paper, this strategy is applied to two different synthetic models: a two layered model with lateral velocity variations, and a four layered model.

2. Method

2.1. The implemented MCMC inversion

Initially, the gradient-based MCMC method employed is briefly described. Further details can be found in Zhao and Sen (2021) and Berti *et al*. (2023), and Aleardi *et al*. (2022a), who applied the method to solve the electrical resistivity tomography in a probabilistic manner.

A Bayesian inversion is designed to estimate the full posterior distribution in the model space,

and can be written as:

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

where the **m** and **d** vectors represent the model and data parameters, respectively; $p(\mathbf{m}|\mathbf{d})$ is the PPD, $p(\mathbf{d}|\mathbf{m})$ is the so-called data likelihood function, which measures the matching between the observed and predicted data, while $p(\mathbf{m})$ and $p(\mathbf{d})$ represent the *a priori* distributions of model and data parameters, respectively.

If the PPD cannot be expressed in a closed form, an MCMC algorithm can be applied for a numerical evaluation of the posterior model. Within this framework, the probability, γ, to migrate from the current state of the chain, $\mathbf{m}_k$, to the next proposed state, $\mathbf{m}_{k+1}$, is determined according to the Metropolis-Hasting (M-H) rule:

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

where q is the proposal distribution that defines the new state, $\mathbf{m}_{k+1}$, as a random deviation drawn from a probability distribution, $q(\mathbf{m}_{k+1}|\mathbf{m}_k)$, conditioned only by the current state, $\mathbf{m}_k$ (Hastings, 1970).

If $\mathbf{m}_{k+1}$ is accepted, the current model is updated, and so is $\mathbf{m}_k = \mathbf{m}_{k+1}$, conversely, $\mathbf{m}_k$ is retained in the chain, and another state is generated as a random deviation from $\mathbf{m}_k$.

Usually, several Markov chains are used, and so various random paths are performed starting from different areas of the model space, in order to increase the exploration of the parameter space, and to prevent the sampling from being confined to some local PPD maxima.

A common practice to numerically evaluate the posterior density is to discard the early realisations in the chains and start collecting samples only once the effect of the initial value is mitigated. The main reason behind this strategy, known as 'burn-in', is to consider the samples only after the Markov chain has come sufficiently close to the stationary regime. The collection of samples after the 'burn-in' period is used to numerically compute the statistical properties of the target posterior probability, such as mean and standard deviation.

One way to check if a chain has converged to the equilibrium distribution is to compare its behaviour with other randomly initialised chains. For this purpose, Gelman and Rubin (1992) proposed the so-called potential scale reduction factor (PSRF). This factor measures the ratio of the average variance of samples within each chain to the variance of the pooled samples across the chains. It converges to one as the number of iterations approaches infinity. The recommended proximity to one varies with each problem, yet a general rule is to achieve a PSRF that is less than 1.2 for most of the model parameters.

Assuming Gaussian distributions for data, noise, and model parameters, the following equations can be expressed:

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

$$p(\mathbf{d}|\mathbf{m}) \propto exp\left\{-\frac{1}{2}[\mathbf{d}-\mathbf{G}(\mathbf{m})]^T \mathbf{C}_{\text{d}}^{-1}[(\mathbf{d}-\mathbf{G}(\mathbf{m})]\right\} \quad (4)$$

where $\mathbf{m}_{prior}$ is the prior model vector; $\mathbf{C}_d$ and $\mathbf{C}_m$ are the data and prior model covariance matrices, respectively, and **G** is the forward modelling operator, which relates the model to the corresponding data; we are, then, able to retrieve the local, Gaussian approximation of the PPD around the current model, $\mathbf{m}_k$:

$$p(\mathbf{m}|\mathbf{d}) \propto exp\left\{-\frac{1}{2}[\mathbf{m}-(\mathbf{m}_k-\mathbf{H}^{-1}\mathbf{g})]^T\mathbf{H}[\mathbf{m}-(\mathbf{m}_k-\mathbf{H}^{-1}\mathbf{g})]\right\} \quad (5)$$

where **g** represents the gradient vector, and **H** the Hessian matrix, defined as:

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

and:

$$\mathbf{H} \approx \mathbf{J}^T\mathbf{C}_d^{-1}\mathbf{J} + \mathbf{C}_m^{-1}. \quad (7)$$

In the previous equations, $\Delta\mathbf{d}(\mathbf{m}_k) = \mathbf{G}(\mathbf{m}_k) - \mathbf{d}$ is the data misfit vector calculated using the current model $\mathbf{m}_k$, and **J** identifies the Jacobian matrix, which expresses the partial data derivative with respect to the model parameters.

After constructing this posterior density approximation, a sampling technique, which uses a proposal density to draw the perturbation to be applied to the current model, can be defined as:

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

where $\mathbf{H}^{-1}\mathbf{g}$ is the so-called Newton step (if $1/\beta^2 = 0$, the standard gradient descent is obtained).

The M-H rule, previously described, is used to decide whether to accept or not each proposed model. α and β^2 values are adaptable parameters that determine how far to continue along the negative gradient direction and to which extent the model is randomly perturbed, respectively. The optimal setting of these parameters can be found by inspecting the acceptance ratio and the convergence rate of the sampling towards the stationary regime of the chain. Moreover, although the proposal assumes a local Gaussian approximation of the PPD, it can be used for sampling from any type of posterior model and under any prior assumption (e.g. non-parametric prior).

The GB-MCMC method does not only focus on the sampling around the most favourable areas of the parameter space (i.e. those characterised by high posterior density values), but it also uses a proposal that, by exploiting the inverse of the Hessian matrix, incorporates information on the local covariance structure of the target density. As a matter of fact, the reduced dimension of the model parameter also enables considering the full approximate Hessian matrix and not only its diagonal values. This means that model space sampling is also guided by model parameter correlation, and not only by data illumination. Including this information not only accelerates probabilistic sampling, but also maximises sample independence while maintaining high acceptance probabilities (Aleardi *et al*., 2022b).

The two major computational requirements of GB-MCMC sampling are the computation of the Jacobian matrix, associated with each accepted model, and the manipulation of the Hessian matrix and gradient vector, whose dimensions depend on the number of unknowns and data

points. The Jacobian matrix is, here, calculated via a forward FD scheme, and in such context, the number of forward evaluations needed for its computation increases linearly with the number of model parameters. The advantage is that each column of the Jacobian matrix can be independently computed, and, therefore, its computation can be easily distributed across different computing units. A possible approach to reduce the dimensions of the Jacobian matrix, Hessian matrix, and gradient vector, is to reduce both data and model spaces with an appropriate compression strategy.

2.2. Discrete cosine transform

The DCT of a signal indicates its distribution in the frequency domain spectrum. Usually, most of the energy of the signal is expressed by low-order DCT coefficients and, consequently, this mathematical transformation can be used for model compression, achieved by zeroing the coefficients of the base function terms beyond a certain threshold. As for any Fourier-related transform, DCTs express a function or a signal in terms of a sum of cosinusoids with different frequencies and amplitudes (Britanak *et al*., 2010).

The choice of the number of DCT coefficients, to compress the model and data space, should always constitute a compromise between the desired data, the model resolution, and the resulting computational cost of the inversion procedure. To estimate the optimal number of DCT coefficients needed to approximate the model and data spaces, how the variability of the true model and the observed data change as the number of DCT coefficients changes, when variability is defined as the ratio of the variance of the compressed model and data to the corresponding variances before compression (Aleardi, 2021), are quantified in this paper.

In the case of a 2D velocity model, most of the spatial variability derives from the low-order DCT coefficients, corresponding to the first q rows and p columns of the transformed matrix.

Consequently, the DCT transformation allows the ($N_y \times N_x$) 2D full velocity model space to be reduced to a ($q \times p$) 2D DCT compressed parameter space with $p < N_x$ and $q < N_y$. The same transformation can be applied to compress the 2D matrix representing each shot gather. As previously mentioned, there is a trade-off between the model resolution and the number of coefficients employed, considering that the spatial resolution of the recovered model increases as the number of retained DCT coefficients increases.

In all the following inversion tests, log-Gaussian prior distributions are assumed for the velocity, and, due to the linearity of the DCT transformation, the assumed *a priori* mean vectors and the *a priori* model covariance matrix can be analytically projected onto the DCT space [see Aleardi (2020) for further details]. Due to the large number of data points in the entire data space, a numerical Monte Carlo simulation has been applied to project the data covariance matrix, $\mathbf{C}_d$, in the compressed DCT domain (Aleardi, 2021).

3. Results

3.1. Synthetic inversion: model 1

The first step consisted in applying the proposed GB-MCMC elastic FWI method to a 2D near surface model, with a number of lateral velocity variations.

The model parameters to be estimated are the V_s values, a homogeneous constant model for the density is considered, and the V_p value is set to be proportional to the V_s value.

The grid size for generating the observed data is 290(nx_0) × 50(nz_0), where nx_0 and nz_0 are the number of grid points in the horizontal and vertical directions, respectively, with a grid spacing of 0.2×0.2 m^2, so 14,500 parameters form the full model space. Actually, this is an oversampling of the model space resulting from the requirements of the dispersion relation. A Ricker wavelet, with a peak frequency of 12 Hz, is used as source signature. Five shots, equally spaced along the horizontal axis, were simulated and each shot is recorded by 58 evenly-spaced receivers with a one-metre receiver interval. The time sampling is of 0.1 ms and the registration time is 0.5 s. Source and receivers are all placed at the free surface. Uncorrelated Gaussian noise, with a standard deviation equal to 20% of the standard deviation of the noise-free data, is added to the observed data.

The simulation of the shots is performed using SOFI2D (Bohlen, 2002), a visco-elastic forward modelling code that solves the pure elastic or visco-elastic wave equation by means of an FD scheme in the time domain.

In order to reduce the computational cost of the calculation of the Jacobian matrix, using the forward FD approach, the computation of each column of the Jacobian matrix is parallelised across different servers.

The data and models are projected onto the DCT domain, where the MCMC sampling operates. We observe that less than 10 coefficients along the two DCT dimensions explain more than 93% of the variability of the original model. Therefore, the DCT enables operation in a model space of 6×8 = 48-D dimensions. The same analysis is realised on the seismic data of this research, and, in this case 50×45 = 2,250 coefficients in the data space (for each shot gather) basically explain the variability of the original seismogram. For this reason, the full 250×58×5 = 72,500-D data space (considering a data sampling rate of 2 ms) is reduced to a 50×45×5 = 11,250-D domain. For real data, the number of DCT coefficients can be estimated for the model space compression, by considering the variability of a number of velocity models drawn from prior models, as done by Aleardi (2021). For the data, instead, the variability of the observed seismograms can still be used to decide the optimal number of DCT coefficients.

The DCT compression is also extremely useful in reducing the total number of forward evaluations needed for the Jacobian computation, via the FD scheme. In fact, the 72,500×14,500 Jacobian matrix in the full domain becomes an 11,250×48 matrix in the compressed space.

A Python implementation, running on different servers equipped with Intel® Xeon® CPU E5-2630 v4 at 2.20 GHz, has been used to perform the numerical tests. Each iteration, including Jacobian matrix computation and sample drawing, takes approximately 150 s wall-clock time. Following the work of Louboutin *et al*. (2017), 1.4 TFLOPs needed for one single iteration were estimated, and, considering that the computer peak computing performance is approximately 40 GFLOPs, a lower boundary of 35 s per iteration, to be compared with the measured wall-clock time, was obtained. Five chains were used, each with 4,000 iterations, running in parallel on different servers, and each ending in 60 hours.

The chain starting models are homogeneous models with a constant velocity value, and the prior V_s model used is also a homogeneous model with a constant velocity value of 200 m/s.

In the case of real data inversion, information of the study area that is available can be incorporated in the inversion framework as a prior model.

In Fig. 1, the posterior mean models, independently estimated for two different chains, are practically shown to be equal and, in both cases, they reproduce the features of the true model very well. Therefore, the two inversions have reached similar posterior mean estimations although they start from different initial states. This indicates the ability of the implemented method to explore the parameter space and to converge towards a model that is close to the true one, without being affected by cycle skipping issues.

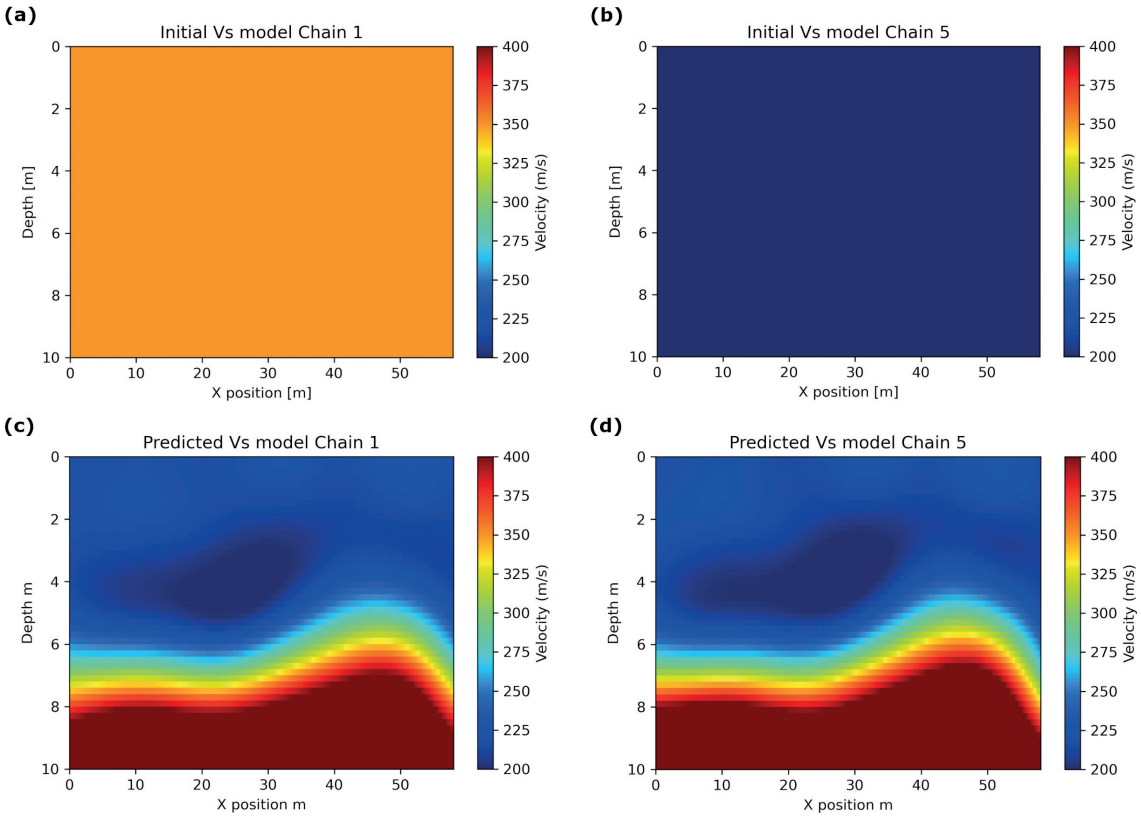

Fig. 1 - Example of two chains: on the top, panels a and b with the starting models of the GB-MCMC inversion; on the bottom, the posterior mean models independently estimated by the two chains (panels c and d, respectively).

By analysing the evolution of the negative log-likelihood (Fig. 2a), the first 1,000 iterations are considered as the 'burn-in' period and the corresponding samples are discarded. Depending on the starting model, every chain presents a different convergence speed, however after 'burn-in' they have all reached the stationary regime, where all the likelihoods oscillate around similar and stable values.

By observing the acceptance ratio of each chain (Fig. 2b), i.e. the ratio of the number of accepted models to the number of iterations, we can see that it is higher (approximately 70-80%) than the acceptance ratios usually obtained for standard, gradient-free MCMC algorithms (generally approximately 20-30%). These high acceptance ratios enable avoiding forward evaluations for models that, most likely, will not be accepted. However, this also increases the computational cost of the algorithm as the Jacobian matrix needs to be evaluated anew for each model accepted.

A possible strategy to reduce the computational effort of the inversion is computing the Jacobian, in the post-'burn-in' phase, every 20 models accepted. This procedure does not affect the statistical properties of the estimated PPD but could only slightly affect the efficiency (the rate of convergence towards a stable posterior) of the probabilistic sampling.

Fig. 2c shows the progression of the PSRF for all the model parameters in the DCT domain. This plot suggests that the GB-MCMC reaches convergence for most of the model parameters within approximately 1,500 iterations.

The models sampled by all the chains after the 'burn-in' phase are projected onto the full domain to numerically compute the mean and standard deviation of the PPD (Fig. 3).

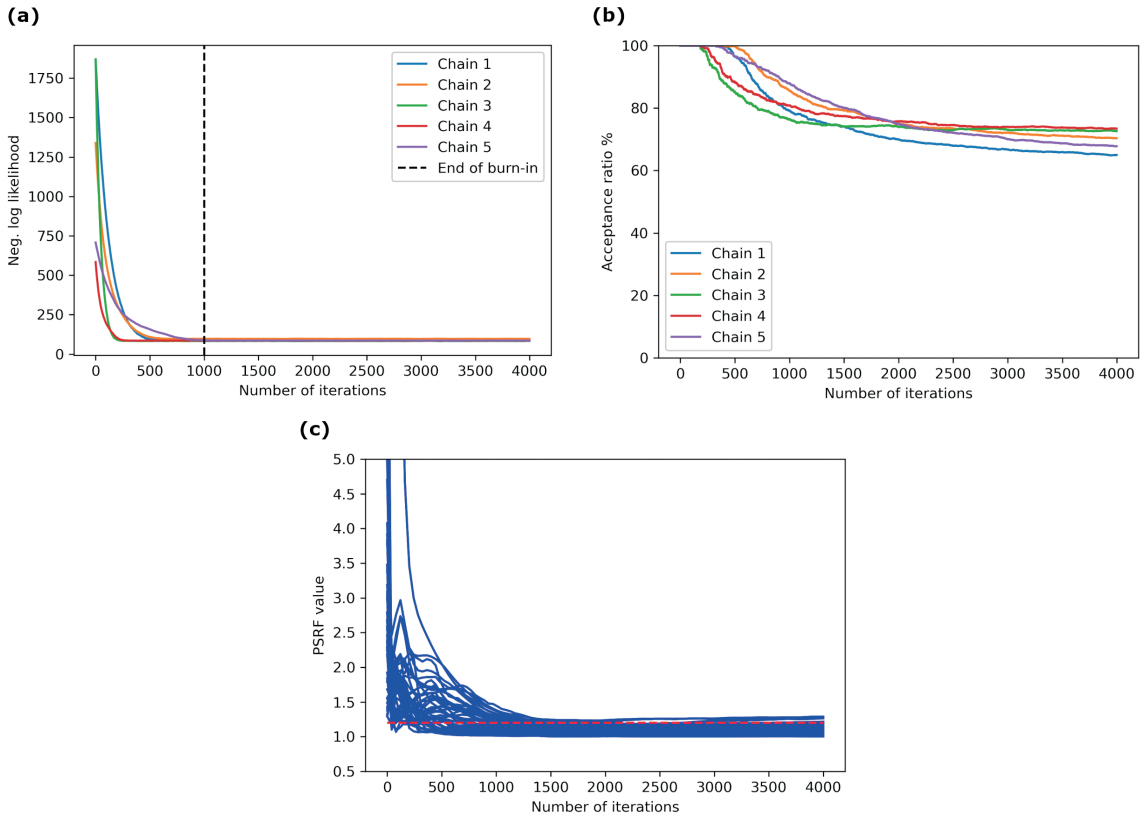

Fig. 2 - a) Evolution of the negative log-likelihood values for the GB-MCMC inversion where each colour represents a different chain, and the dashed line defines the 'burn-in' period; b) acceptance ratio, defined as the number of accepted models to the number of iterations, for each chain; c) PSRF values for the model parameters in the compressed domain, calculated for each iteration. Each blue line represents the evolution of the PSRF value over the number of iterations for each model parameter in the compressed space. The red dashed line represents the threshold of convergence (i.e. a PSRF value of 1.2).

Fig. 3a shows the original, uncompressed portion of the true model considered in this test, whereas Fig. 3b represents the posterior mean model computed considering all the chains. If a higher resolution in the predicted model is required, a higher number of DCT coefficients in the model space can be considered although this will result in an increased computational cost.

The standard deviation map (Fig. 3c), calculated using all the models sampled by the five chains after the 'burn-in', suggests very low uncertainties in the shallower parts of the model, with good data coverage and good illumination. These small deviations indicate that the predicted velocity values are well constrained in these areas, and that the inversion results are, then, affected by lower uncertainties. Conversely, the deeper parts and the edges of the model are associated with higher posterior uncertainties (15 m/s and greater) due to the reduced illumination of the parameters.

Fig. 4 displays the leftmost shot gather, calculated using different models. In the left column, the observed noise data is observed; in the centre, the predicted data calculated using the posterior mean model of Fig. 3b is illustrated together with the same shot generated from the starting model represented in Fig. 1a; in the right column, their sample-by-sample difference. The predicted data matches the observed data better compared to the initial data generated on one of the starting models, as expected. The remaining difference is associated with the lower

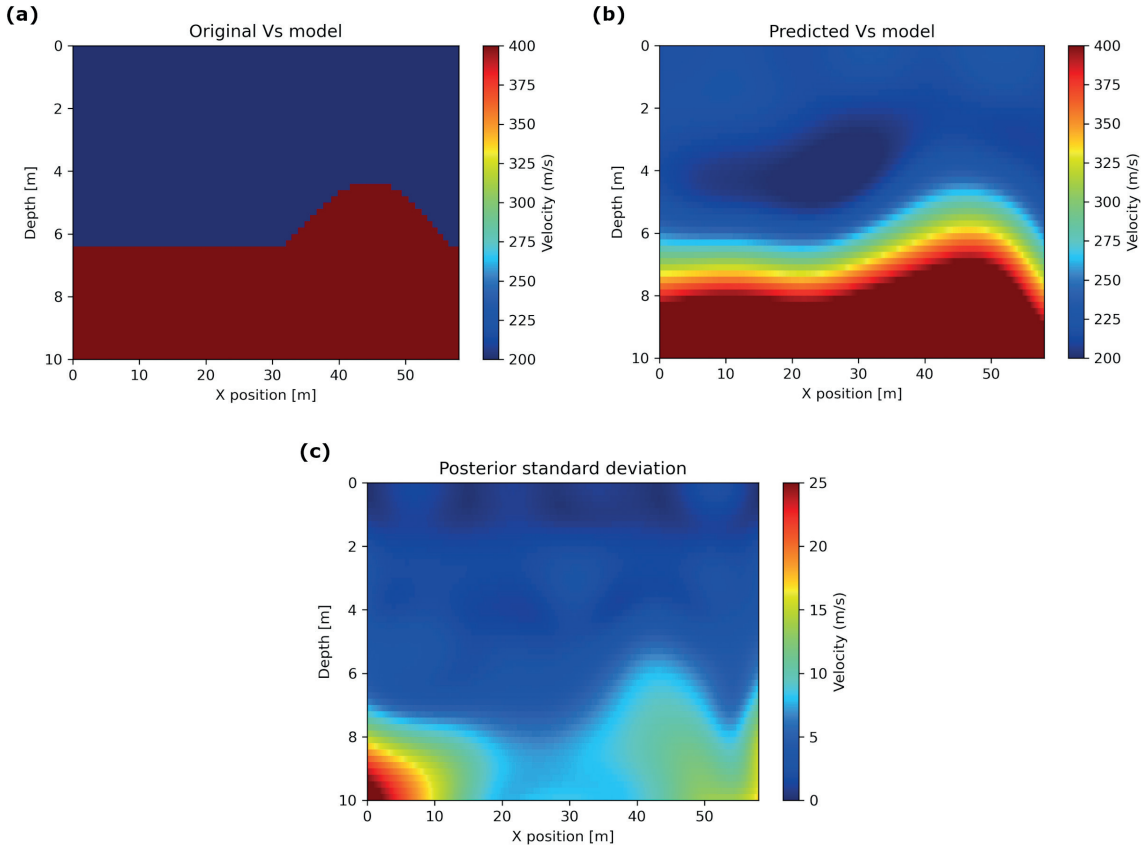

Fig. 3 - a) The original V_s model considered in this paper; b) the posterior mean V_s model and c) standard deviation computed considering all the chains.

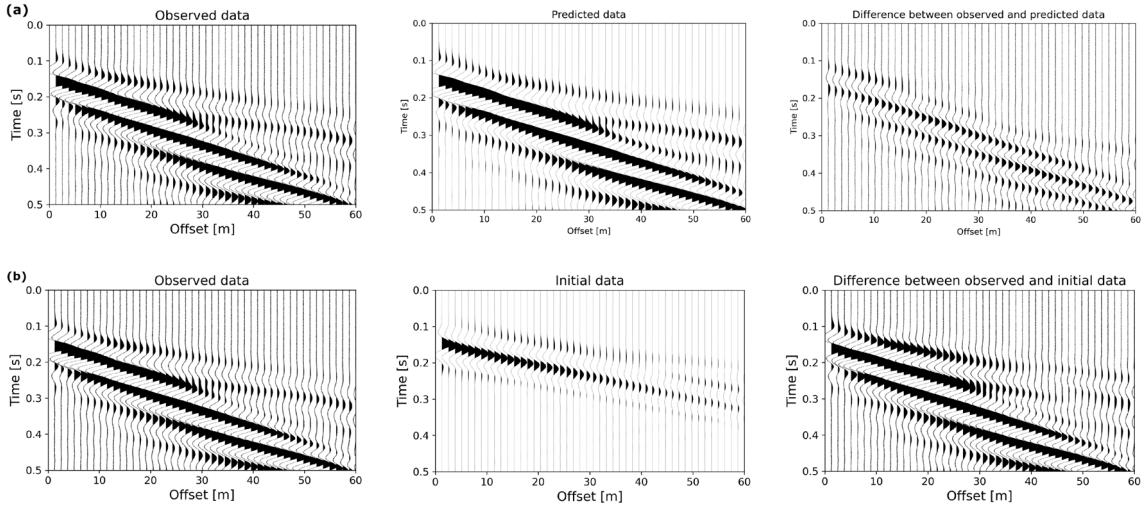

Fig. 4 - a) The observed data of the first shot on the left of the model, the predicted data generated by the posterior mean model of Fig. 3b in the centre, and their difference on the right; b) the same shot of the observed data on the left, the data calculated using the initial model of Fig. 1a in the centre, and their difference on the right. All the plots display the raw amplitude data using the same gain.

resolution of the posterior mean model, due to the DCT compression. Some inversion tests, using a higher number of DCT coefficients in both directions, were performed, and resulted in predicted velocity models with an increased resolution (as well as an increased computational time of the inversion procedure), and a further reduction of the remaining difference between observed and predicted data.

Fig. 5 shows a close-up on two traces extracted from the first shot, in which the observed, predicted and initial data are superimposed. It is worth noting the severe cycle skipping between the observed and initial data, which vanishes when considering the data generated on the posterior mean model. This proves the robustness of the implemented method against the cycle skipping issue, and its ability to exhaustively explore the model space.

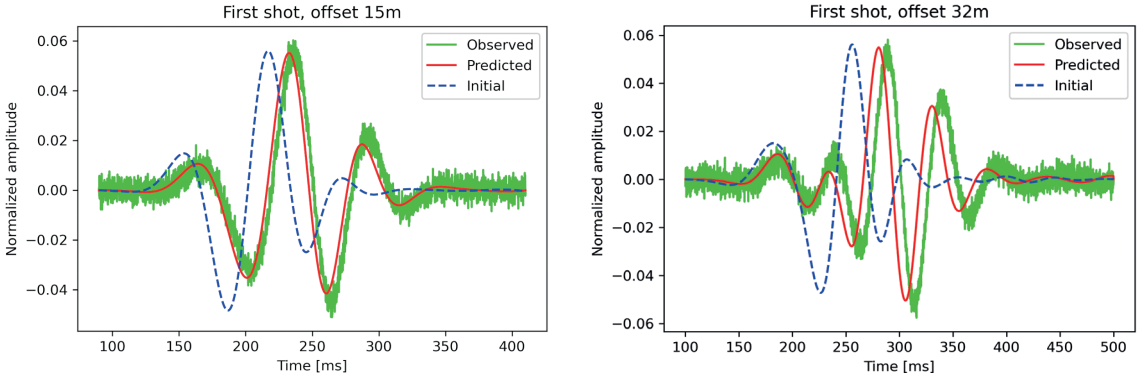

Fig. 5 - Comparison of two traces at different offsets of the first shot computed using the true model (green), the GB-MCMC predicted posterior mean model (red), and the starting model of Fig. 1a (blue).

3.2. Synthetic inversion: model 2

The proposed method is also applied to another near-surface velocity model: a four-layered model with velocity inversions but with no lateral variations. The same grid of the previous test has been used, resulting in a 14,500-D full model space. Similarly to the previous example, a Ricker wavelet, with a peak frequency of 12 Hz, is used as source signature and the same acquisition parameters were adopted. Uncorrelated Gaussian noise, with a standard deviation equal to 20% of the total standard deviation of the noise-free data, is added to the observed data.

The simulation of the shots is again performed using SOFI2D.

Again, the data and models are projected onto the DCT domain, where the MCMC sampling operates. We observe that less than 10 coefficients along the two DCT dimensions explain more than 93% of variability of the original model. In particular, nine and six coefficients are selected along the vertical and horizontal directions, respectively. Therefore, the compression enables operating in a 9×6 = 54-D domain. The same analysis is carried out on seismic data, and, in this case, 50×45 = 2,250 retained coefficients in the data space explain almost all the variability of the original seismogram. As in the previous experiment, a Python implementation, running on different servers equipped with Intel® Xeon® CPU E5-2630 v4 at 2.20 GHz, was used to perform the numerical tests. Each iteration, considering the Jacobian computation and sample drawing, requires approximately 180 s wall-clock time. Four chains were run, each with 4,000 iterations and each finishing the run within 65 hours.

The chains start from homogeneous models with a constant velocity value (Fig. 6b), once

again used also as prior information (in this test, a V_s model with a constant value of 350 m/s).

Figs. 6a to 6c compare the true model after the DCT compression and the posterior mean computed by considering all the samples collected by the four chains after the 'burn-in' period. The GB-MCMC estimation is, yet again, observed to closely reproduce the main features of the original model and the correct velocity values of all the layers.

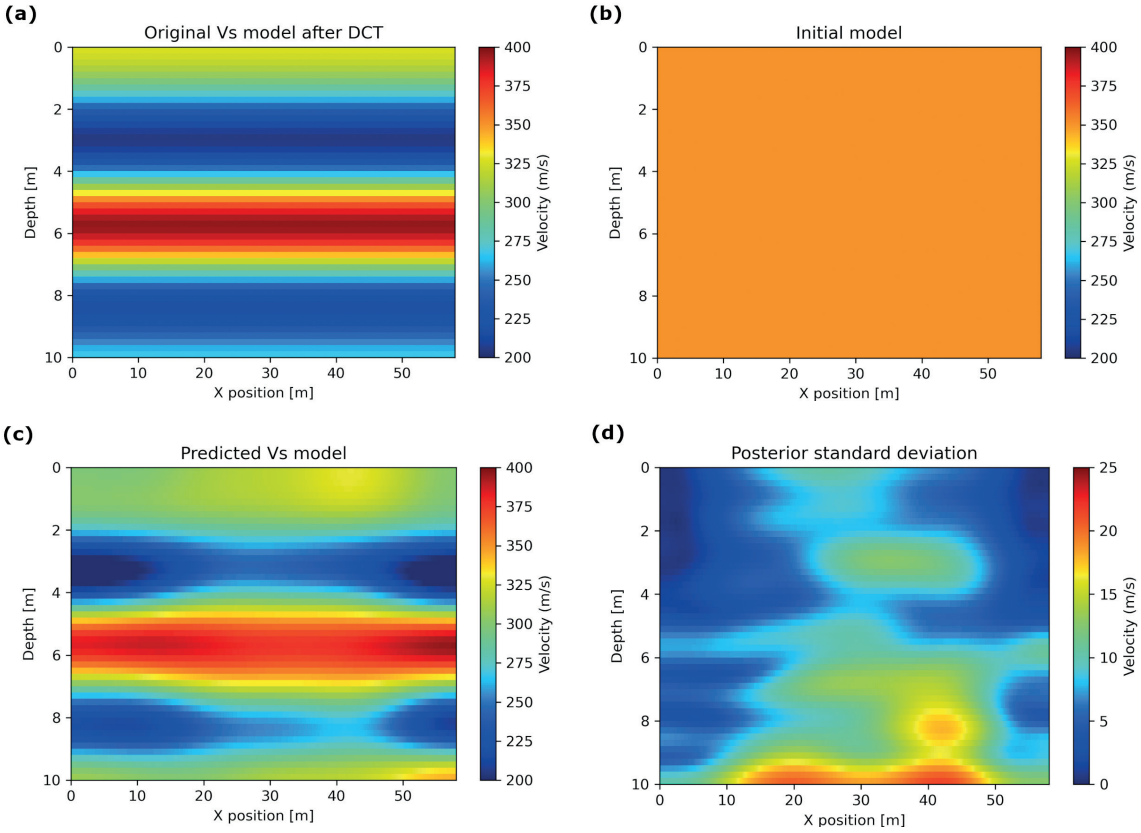

Fig. 6 - a) The original, uncompressed, portion of the synthetic model taken in consideration; b) the starting model for one of the chains; c) the posterior mean model considering all the four chains; d) the posterior standard deviation estimated by the GB-MCMC algorithm when all the chains are considered.

More specifically, the high velocity layer is well reproduced, while the two surrounding layers (in particular the bottom layer) present a lower horizontal continuity, as in the presence of a velocity inversion. The horizontal continuity, as well as the spatial resolution of the entire model, can be improved by considering a higher number of DCT coefficients along the horizontal and vertical directions and by increasing the computational inversion time.

The standard deviation map (Fig. 6d), calculated using all the models sampled by the four chains after the 'burn-in' period, suggests very small deviations for all the models (lower than 25 m/s). As expected, also larger uncertainties are observed in the deeper parts of the model with less illumination and data coverage, in particular under the high velocity layer, and also in correspondence of the layer edges.

From the inspection of the evolution of the negative log-likelihood (Fig. 7a), we can notice that after a certain number of iterations the model reaches the stationary regime, where it

starts to oscillate around stable values. In this example, the first 2,000 iterations are considered as the 'burn-in' period and the corresponding samples are discarded, so that only the following 2,000 samples of each chain are used to generate the mean and standard deviation models (as previously shown in Fig. 6). The acceptance ratio of each chain (Fig. 7b) is once again higher than the one expected by standard gradient-free MCMC algorithms. This indicates that the proposed method is of greater efficiency in sampling from the target posterior, however it also increases the computational cost needed for the Jacobian computation of the accepted models. The same strategy used in the previous experiment to reduce the computational inversion cost was adopted by calculating the Jacobian computation every 10 models accepted after the 'burn-in' period. Fig. 7c shows the evolution of the PSRF values for all the model parameters in the compressed domain. PSRF values lower than 1.2 illustrate that convergence is attained for most model parameters.

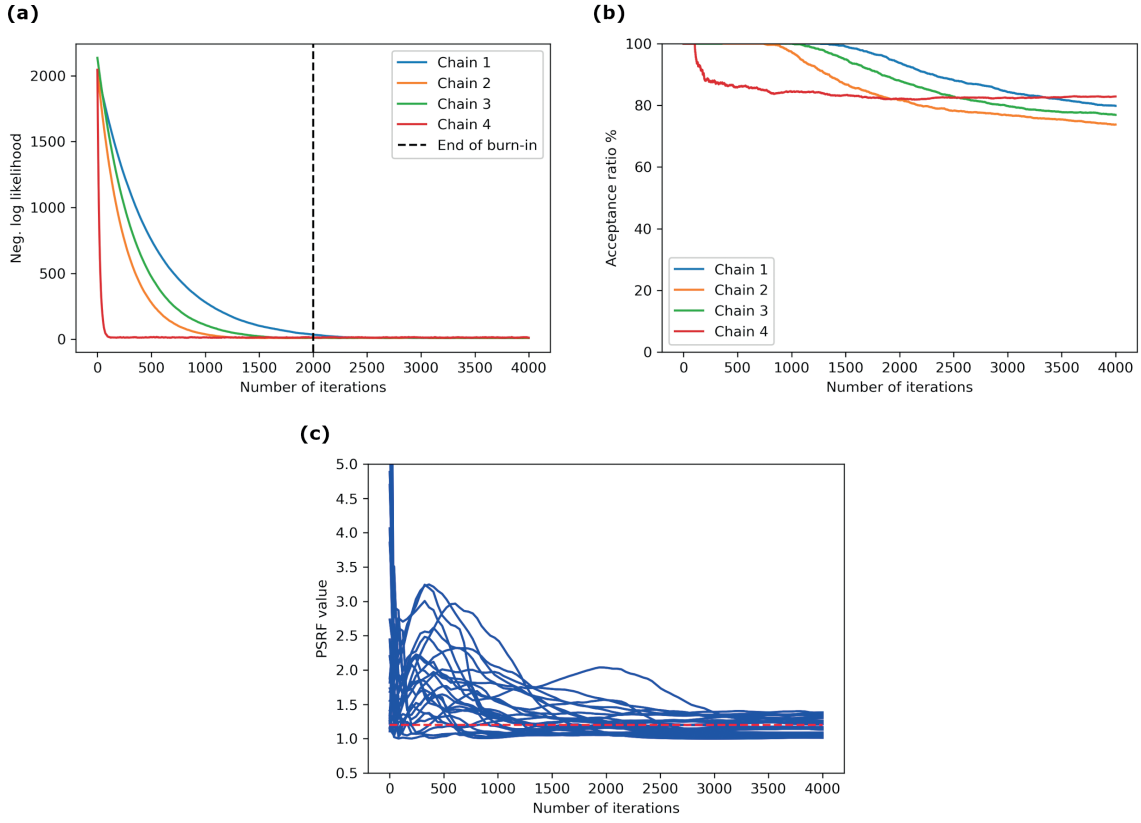

Fig. 7 - a) Evolution of the negative log likelihood for all the chains, the dashed vertical line defines the 'burn-in' period; b) the acceptance ratio calculated for all four chains as the ratio of the number of accepted models to the number of iterations; c) evolution of the PSRF values for each model parameter in the compressed space. The red dashed line represents the convergence threshold fixed at 1.2.

In Fig. 8, the observed leftmost shot gather is compared with the data generated on the starting model of Fig. 6b and with the shot computed on the posterior mean model of Fig. 6c. Despite the simplicity of the starting models, the inversion achieved a good match between observed and predicted data, and significantly improved data fitting with respect to the data calculated on the starting model. The remaining difference (Fig. 8c) is associated with the lower resolution of the posterior mean model, due to the DCT compression and to the choice of the

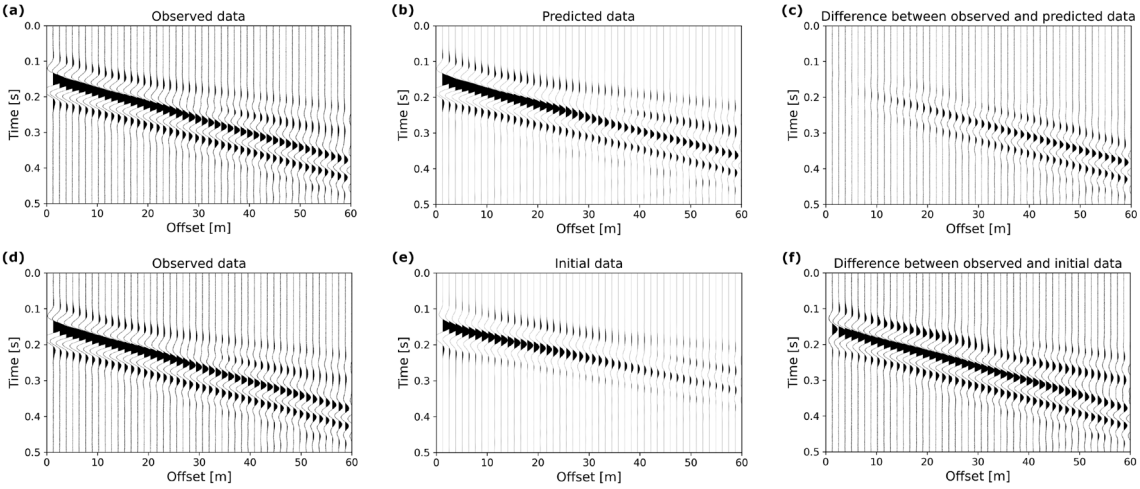

Fig. 8 - Panels a and d: the leftmost shot of the observed data; b) the same shot calculated using the posterior mean model of Fig. 6c; c) their difference; e) the same shot generated by the starting model of Fig. 6b; f) sample-by-sample difference with the observed data. All the plots display the raw amplitude using the same gain.

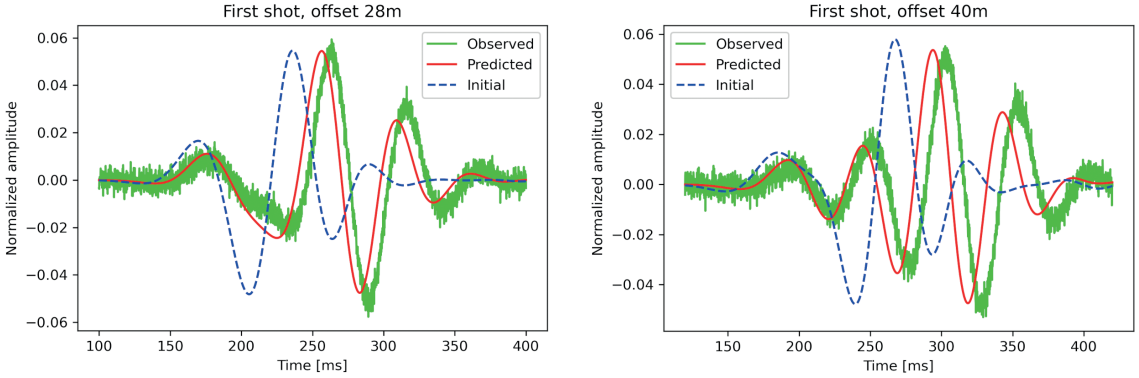

Fig. 9 - Comparison of two seismic traces of the leftmost shot computed from the original model (green), the GB-MCMC posterior mean (red), and the starting model of Fig. 6b (blue).

number of DCT coefficients to be considered.

From a close-up of Fig. 9, the significant differences between the observed and initial data can be observed. Also of note is the fact that the cycle skipped events are clearly evident in both cases. Such cycle skips almost disappear when considering the predicted data calculated on the posterior mean model, thus demonstrating the robustness of the GB-MCMC approach.

To further reduce the remaining shift between the predicted and the observed data, a second step of local FWI can be implemented, starting from the final result of the GB-MCMC inversion. This can also be useful to improve the spatial resolution of the predicted model, thus reducing the DCT compression effect.

3.3. Synthetic inversions: effect of attenuation

The visco-elastic FWI (Bohlen, 2002) requires much longer computational time compared to the elastic FWI, due to the requirement of more complex modelling equations and to the introduction of the shear and compressional wave quality factors (Qs and Qp, respectively).

Groos *et al*. (2014) concluded that accounting for the quality factors in the inversion process was important for obtaining correct surface wave FWI results.

The influence of attenuation on the GB-MCMC inversion algorithm was tested on both synthetic models previously analysed. During the inversion procedure, the authors did not consider the quality factors as unknowns, instead chose a fixed value of 10 for both the Qs and Qp. Attenuation is considered in computing the observed data, while the forward modelling used, during the inversion, is the 2D elastic one. Some effects on the elastic FWI inversion results are expected due to the differences between the visco-elastic and elastic data resulting from the very low quality factors used.

The inversions are carried out using the same parameters employed in the previous cases for synthetic models 1 and 2; in particular, the same starting models and prior information were used (Figs. 1a and 6b).

In Figs. 10a and 10b, the predicted V_s model and the relative posterior standard deviation for model 1 can be observed. The predicted model reproduces the main features of the true model to a good degree, in particular, the vertical and lateral velocity variations. The velocity of the first layer is almost the same as that of the real model, while the velocity of the second layer has been underestimated by the inversion, although, its thickness and shape are well predicted. As expected, the posterior standard deviation values are higher than the ones shown in Fig. 3c, where no attenuation was considered in the observed data, in accordance with the less accurate result obtained for the predicted model with respect to the one of Fig. 3b. In particular, there are higher uncertainties in correspondence of the second layer where the velocity value is

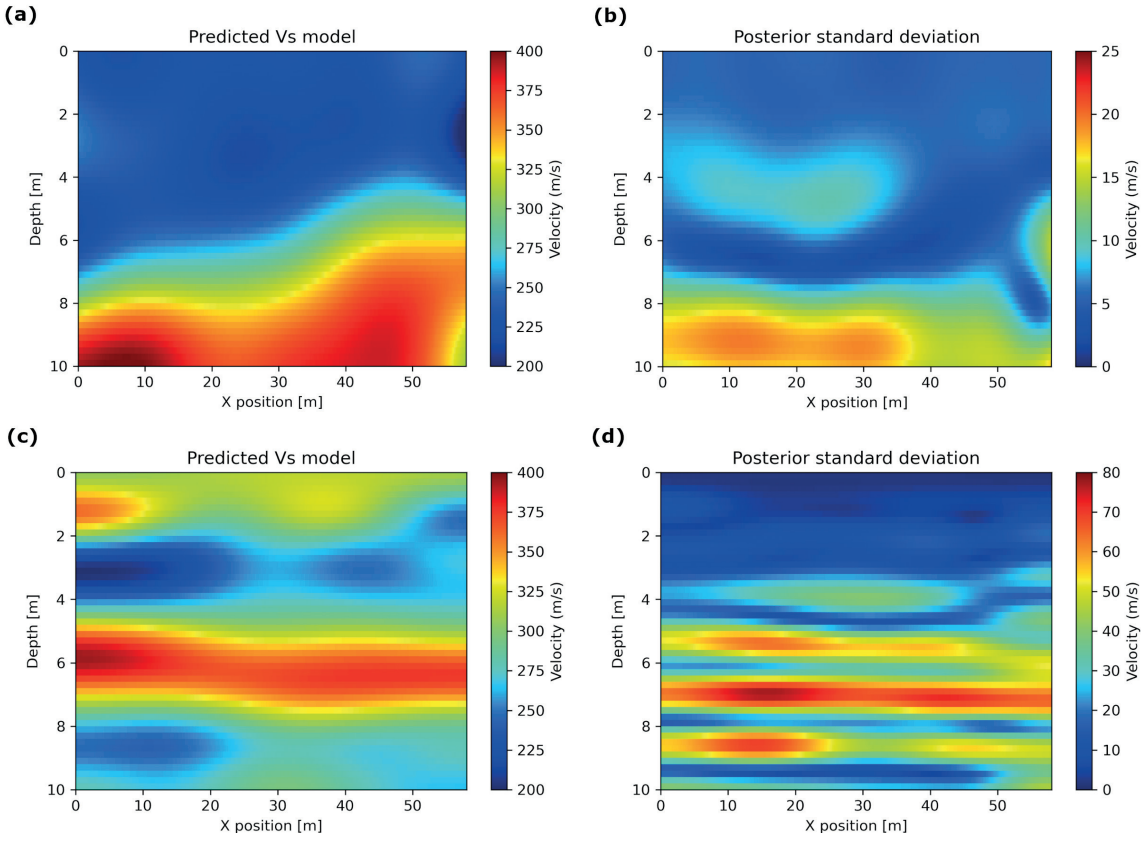

Fig. 10 - On the left, panels a and c the predicted *Vs* models for both the inversion tests and, on the right, panels b and d their posterior standard deviation, when attenuation is considered in the observed data.

underestimated, while in Fig. 3c higher values can be observed only at the bottom edges of the model, due to the lower illumination.

Figs. 10c and 10d show the same results (posterior mean and standard deviation, respectively) for model 2. In this case, it is again possible to observe that the predicted V_s model reproduces the main features of the true model, in particular the high velocity layer, but the result is worse, in terms of predicted velocity values and horizontal continuity of the four layers, than that of Fig. 6c.

In particular, when no attenuation is considered, the velocity values are very similar for the first two layers, whereas considering attenuation in the observed data there is a more significant difference in the upper part of the model (i.e. up to 50 m/s). As expected, the posterior standard deviation presents higher values in the bottom half of the model, due to the lower illumination and data coverage. It should be noted that the highest values of the standard deviation are localised in correspondence of the highest velocity layer and velocity inversions, which, also due to the limited offset extension of the model, considerably complicate the resolution of the inversion problem (note that the maximum value of the two colour bars of the standard deviation maps of Figs. 6d and 10d are very different).

Fig. 11 illustrates the comparison between the observed data, calculated using the 2D visco-elastic FD modelling, and the predicted data, calculated on the two velocity models shown in Figs. 10a and 10c, using the elastic version of the code. The results obtained are quite satisfactory, considering the high attenuation affecting the observed data in these two experiments due to the very low quality factors chosen (notice the difference between the observed data in Figs. 11a to 11c, and that of Figs. 4a and 8a, respectively, where no attenuation was considered).

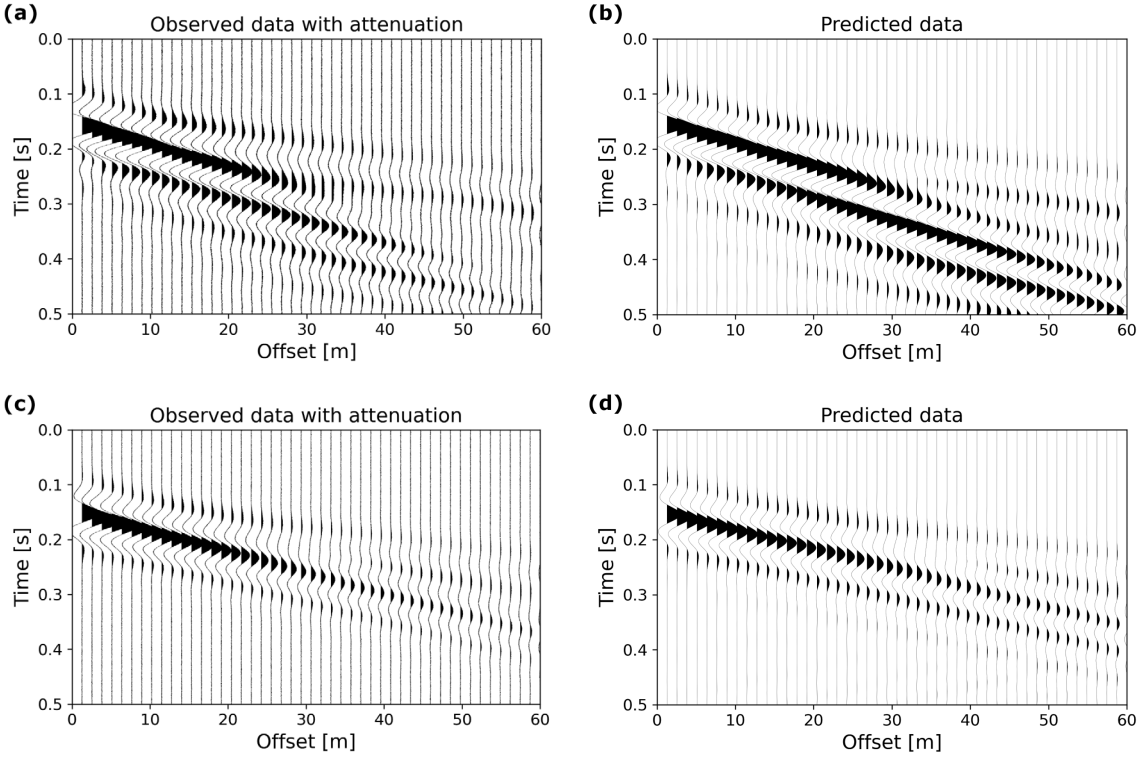

Fig. 11 - On the left, a) and c), the observed data for both the inversion tests, computed by means of a 2D visco-elastic FD modelling; on the right, b) and d), the predicted data for the two inversions. All the plots display the raw amplitude data using the same gain.

4. Conclusions

In this paper, an elastic FWI, combining a gradient-based MCMC sampling, with a DCT compression of model and data spaces, was presented. The Hessian and gradient information of the posterior density rapidly guides the sampling towards the most promising regions of the model space, while the random perturbation applied to the proposed models avoids having to deal with some local maxima of the posterior density, thus, significantly mitigating the cycle skipping issue affecting the elastic FWI. The DCT was primarily used to reduce the number of model parameters and observations, thus, compressing the dimensions of the gradient vector, and of the Hessian and Jacobian matrices, and also mitigated the curse of the dimensionality issue. The results obtained with the two near-surface experiments confirmed the applicability and reliability of the approach implemented, which, unlike conventional local approaches, provides a complete assessment of the uncertainties affecting the estimated solution as the final results. The proposed GB-MCMC inversion converged to stable and plausible solutions even when the probabilistic sampling started from initial models, which were very far from the optimal solution, also considering the effect of attenuation in the simulation of the observed data. The next step of our research is the application of the implemented algorithm to field data.

REFERENCES

Aleardi M.; 2020: *Discrete cosine transform for parameter space reduction in linear and non-linear AVA inversions*. J. Appl. Geophys., 179, 104106, 42 pp., doi: 10.1016/j.jappgeo.2020.104106.

Aleardi M.; 2021: *A gradient-based Markov chain Monte Carlo algorithm for elastic pre-stack inversion with data and model space reduction*. Geophys. Prospect., 69, 926-948, doi: 10.1111/1365-2478.13081.

Aleardi M. and Stucchi E.; 2021: *A hybrid residual neural network-Monte Carlo approach to invert surface wave dispersion data*. Near Surf. Geophys., 19, 397-414.

Aleardi M., Pierini S. and Sajeva A.; 2019: *Assessing the performances of recent global search algorithms using analytic objective functions and seismic optimization problems*. Geophys., 84, R767-R781.

Aleardi M., Vinciguerra A., Stucchi E. and Hojat A.; 2022a: *Machine learning-accelerated gradient-based Markov chain Monte Carlo inversion applied to electrical resistivity tomography*. Near Surf. Geophys., 20, 440-461.

Aleardi M., Vinciguerra A., Stucchi E. and Hojat A.; 2022b: *Stochastic electrical resistivity tomography with ensemble smoothers and deep convolutional autoencoders*. Near Surf. Geophys., 20, 160-177.

Alkhalifah T.A.; 2016: *Full waveform inversion in an anisotropic world: where are the parameters hiding?* EAGE Publications, Houten, The Netherlands, 197 pp.

Bergamo P. and Socco L.V.; 2014: *Detection of sharp lateral discontinuities through the analysis of surface-wave propagation*. Geophys., 79, EN77-EN90.

Berti S., Aleardi M. and Stucchi E.; 2023: *A computationally efficient bayesian approach to full-waveform inversion*. Geophys. Prospect., 24 pp., doi: 10.1111/1365-2478.13437.

Bohlen T.; 2002: *Parallel 3-D viscoelastic finite difference seismic modelling*. Comput. Geosci., 28, 887-899.

Bohlen T., Kugler S., Klein G. and Theilen F.; 2004: *1.5D inversion of lateral variation of Scholte wave dispersion*. Geophys., 69, 330-344.

Britanak V., Yip P. and Rao K.R.; 2010: *Discrete cosine and sine transforms: general properties, fastalgorithms and integer approximations*. Academic Press Inc., Elsevier Science, Amsterdam, The Netherlands, 353 pp.

Curtis A. and Lomax A.; 2001: *Prior information, sampling distributions and the curse of dimensionality*. Geophys., 66, 372-378.

Fichtner A., Kennett B.L.N., Igel H. and Bunge H.P.; 2009: *Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods*. Geophys. J. Int., 179, 1703-1725.

Foti S., Lai C.G., Rix G.J. and Strobbia C.; 2014: *Surface wave methods for near-surface site characterization 1st ed*. CRC Press, London, UK, 487 pp., doi: 10.1201/b17268.

Foti S., Hollender F., Garofalo F., Albarello D., Asten M., Bard P.Y., Comina C., Cornou C., Cox B., Di Giulio G., Forbriger T., Hayashi K., Lunedei E., Martin A., Mercerat D., Ohrnberger M., Poggi V., Renalier F., Sicilia D. and Socco V.; 2018: *Guidelines for the good practice of surface wave analysis: a product of the InterPacific project*. Bull. Earthquake Eng., 16, 2367-2420, doi: 10.1007/s10518-017-0206-7.

Gebraad L., Boehm C. and Fichtner A.; 2020: *Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo*. J. Geophys. Res.: Solid Earth, 125, e2019JB018428, 18 pp., doi: 10.1029/2019JB018428.

Gelman A. and Rubin D.B.; 1992: *Inference from iterative simulation using multiple sequences.* Stat. Sci., 7, 457-511.

Groos L., Schafer M., Forbriger T. and Bohlen T.; 2014: *The role of attenuation in 2D full-waveform inversion of shallow-seismic body and Rayleigh waves*. Geophys., 79, R247-R261.

Groos L., Schafer M., Forbriger T. and Bohlen T.; 2017: *Application of a complete workflow for 2 elastic full-waveform inversion to recorded shallow-seismic Rayleigh waves*. Geophys., 82, R109-R117.

Hastings W.K.; 1970: *Monte Carlo sampling methods using Markov chains and their applications*. Biometrika, 57, 97-109.

Lamuraglia S., Stucchi E. and Aleardi M.; 2022: *Application of a global-local full-waveform inversion of Rayleigh wave to estimate the near-surface shear wave velocity model*. Near Surf. Geophys., 21, 21-38, doi: 10.1002/nsg.12243.

Louboutin M., Lange M., Herrmann F.J., Kukreja N. and Gorman G.; 2017: *Performance prediction of finite-difference solvers for different computer architectures*. Comput. Geosci., 105, 148-157.

Maraschini M. and Foti S.; 2010: *A Monte Carlo multimodal inversion of surface waves*. Geophys. J. Int., 182, 1557-1566.

Mosegaard K. and Tarantola A.; 2002: *Probabilistic approach to inverse problems*. Int. Geophys. Ser., 81, 237-268.

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

Pierini S. and Stucchi E.; 2020: *Points per wavelength analysis in global elastic FWI of surface waves: a synthetic case study*. In: Proc. NSG2020 26th European Meeting of Environmental and Engineering Geophysics, European Association of Geoscientists & Engineers, Belgrade, Serbia, vol. 2020, pp. 1-4.

Romdhane G., Grandjean G., Brossier R., Rejiba F., Operto S. and Virieux J.; 2011: *Shallow-structure characterization by 2D elastic full-waveform inversion*. Geophys., 76, R81-R93.

Sajeva A., Aleardi M. and Mazzotti A.; 2017: *Genetic algorithm full-waveform inversion: uncertainty estimation and validation of the results*. Boll. Geof. Teor. Appl., 58, 395-414, doi: 10.4430/bgta0199.

Sambridge M. and Mosegaard K.; 2002: *Monte Carlo methods in geophysical inverse problems*. Rev. Geophys., 40, 3-1-3-29, doi: 10.1029/2000RG000089.

Schafer M., Groos L., Forbriger T. and Bohlen T.; 2014: *Line-source simulation for shallow-seismic data. Part 2: full-waveform inversion - a synthetic 2D case study*. Geophys. J. Int., 198, 1405-1418.

Sherlock C., Fearnhead P. and Roberts G.O.; 2010: *The random walk metropolis: linking theory and practice through a case study*. Stat. Sci., 25, 172-190, doi: 10.1214/10-STS327.

Socco L.V. and Strobbia C.; 2004: *Surface-wave method for near-surface characterization: a tutorial*. Near Surf. Geophys., 2, 165-185.

Socco L.V., Foti S. and Boiero D.; 2010: *Surface-wave analysis for building near-surface velocity models - Established approaches and new perspectives*. Geophys., 75, A83-A102.

Strobbia C. and Foti S.; 2006: *Multi-offset phase analysis of surface wave data (MOPA)*. J. Appl. Geophys., 59, 300-313.

Tran K.T., McVay M., Faraone M. and Horhota D.; 2013: *Sinkhole detection using 2D full seismic waveform tomography*. Geophys., 78, R175-R183.

Virieux J. and Operto S.; 2009: *An overview of full-waveform inversion in exploration geophysics*. Geophys., 74, WCC1-WCC26, doi: 10.1190/1.3238367.

Xing Z. and Mazzotti A.; 2019: *Two-grid full-waveform Rayleigh-wave inversion via a genetic algorithm - Part 1: method and synthetic examples*. Geophys., 84, R805-R814.

Zhao Z. and Sen M.K.; 2021: *A gradient-based Markov chain Monte Carlo method for full-waveform inversion and uncertainty analysis*. Geophys., 86, R15-R30, doi: 10.1190/geo2019-0585.1.

Corresponding author: Sean Berti
Dipartimento di Scienze della Terra, Facoltà di Scienze Naturali, Matematica e Fisica,
Università degli Studi di Pisa
Via Santa Maria 53, Pisa
Phone: +39 334 269 0517; e-mail: sean.berti@unifi.it