

Bayesian full waveform inversion of surface waves with annealed stein variational gradient descent

Sean Berti,^{1,2} Matteo Ravasi,³ Mattia Aleardi¹ and Eusebio Stucchi¹

¹*Earth Sciences Department, University of Pisa, via S. Maria 53, I-56126 Pisa, Italy E-mails: sean.berti@unifi.it*

²*Earth Sciences Department, University of Florence, via G. la Pira 4, I-50121 Firenze, Italy*

³*Earth Science and Engineering Program, Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia*

Accepted 2025 February 16. Received 2025 January 27; in original form 2024 November 7

SUMMARY

Elastic full-waveform inversion has recently been utilized to estimate the physical properties of the upper tens of metres of the subsurface, leveraging its capability to exploit the complete information contained in recorded seismograms. However, due to the nonlinear and ill-posed nature of the problem, standard approaches typically require an optimal starting model to avoid producing non-physical solutions. Additionally, conventional optimization methods lack a robust uncertainty quantification, which is essential for subsequent informed decision-making. Bayesian inference offers a framework for estimating the posterior probability density function through the application of Bayes' theorem. Methods based on Markov Chain Monte Carlo processes use multiple sample chains to quantify and characterize the uncertainty of the solution. However, despite their ability to theoretically handle any form of distribution, these methods are computationally expensive, limiting their usage in large-scale problems with computationally expensive forward modellings, as in the case of full-waveform inversion. Variational inference provide an alternative approach to estimating the posterior distribution through a parametric or non-parametric proposal distribution. Among this class of methods, stein variational gradient descent stands out for its ability to iteratively refine a set of samples, usually referred to as particles, to approximate the target distribution through an optimization process. However, mode and variance-collapse issues affect this approach when applied to high-dimensional inverse problems. To address these challenges, in this work we propose to utilize an annealed variant of the stein variational gradient descent algorithm and apply this method to solve the elastic full-waveform inversion of surface waves. We validate our proposed approach with a synthetic test, where the velocity model is characterized by significant lateral and vertical velocity variations. Then, we invert a field data set from the InterPACIFIC project, proving that our method is robust against cycle-skipping issues and can provide reasonable uncertainty estimations with a limited computational cost.

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

INTRODUCTION

Reconstructing near-surface elastic models is essential in geophysical and geotechnical applications. Over the past two decades, advancements in surface-wave analysis and inversion have established these methods as popular non-invasive techniques for estimating near-surface structures. Their effectiveness is largely attributed to the dominance of surface waves in the shallow seismic wavefield, resulting in a high signal-to-noise ratio in field recordings. The study of shallow-seismic surface waves began with the spectral analysis of surface waves (SASW; Nazarian *et al.* 1983) and gained

significant attention with the introduction of multichannel analysis of surface waves (MASW; Park *et al.* 1999; Xia *et al.* 1999), which enhanced the efficiency of surface-wave surveys by employing a multistation approach. Both techniques exploit the dispersion characteristics of surface waves to assess near-surface structures. Specifically, MASW follows two main steps: extracting and inverting the surface-wave dispersion curve to reconstruct *S*-wave velocity profiles as a function of depth. However, the method faces limitations due to the assumption in the forward calculation of a 1-D layered model and plane waves. Additionally, accurately identifying multimodal dispersion patterns remains challenging and influ-

enced by human interpretation (Zhang & Chan 2003; Boaga *et al.* 2013).

With the rapid growth of computational power, full-waveform inversion (FWI; Lailly 1984; Tarantola 1984) has become feasible for resolving subsurface models by directly fitting observed waveforms. However, in shallow seismic studies, the presence of surface waves invalidates the acoustic approximation commonly used in exploration seismology and increases the nonlinearity of FWI (Gelis *et al.* 2007; Brossier *et al.* 2009). Numerical examples have demonstrated that FWI is a promising method for quantitatively imaging near-surface structures (Groos *et al.* 2014, 2017; Xing & Mazzotti 2019) and field applications have further demonstrated FWI's high resolution and effectiveness in characterizing near-surface properties (Romdhane *et al.* 2011; Tran *et al.* 2013; Pan *et al.* 2019). However, shallow-seismic FWI is an ill-posed problem and when solved with a deterministic approach can easily converge to a local minimum of the objective function due to the cycle skipping issue, especially when the inversion is started from a poor initial model (Virieux & Operto 2009). In addition, such a deterministic framework fails in providing accurate uncertainty quantification. From the one hand, the local minima issue can be mitigated through strategic designs of the objective function (Bunks *et al.* 1995; Perez Solano *et al.* 2014; Wu *et al.* 2014). An alternative to mitigate the dependence on the initial model is the use of global optimization approaches, though these are often associated with prohibitive computational costs, especially in high-dimensional spaces (Sen & Stoffa 2013; Sajeva *et al.* 2014; Datta & Sen 2016; Ray *et al.* 2016; Aleardi *et al.* 2019; Lamuraglia *et al.* 2022). From the other hand, a Bayesian inference framework (Mosegaard & Tarantola 2002) is theoretically capable to solve the cycle skipping issue and allows for a comprehensive uncertainty assessment of the inversion results. In this context, the solution is represented by the posterior probability density (PPD) function in the model space, which integrates prior knowledge about the model parameters with information provided by the recorded seismic data. Markov chain Monte Carlo (MCMC) methods are often employed to sample the target posterior distribution (Sambridge & Mosegaard 2002), using sufficiently long Markov chains with random starting points. A significant challenge with MCMC methods is that convergence to the target PPD heavily depends on the random perturbation applied to the current state of the chain. This perturbation, commonly referred to as the proposal distribution, must closely approximate the target posterior to ensure computational efficiency. Moreover, the efficiency of these methods significantly decreases in high-dimensional problems due to the so-called curse of dimensionality (Curtis & Lomax 2001). Traditional sampling strategies, such as the random walk Metropolis, might require billions of forward evaluations before reaching convergence, making them computationally impractical when dealing with expensive forward operators, as in the case of FWI.

Recent studies have aimed at reducing the dimensionality of the problem using reparametrization techniques (Malinverno 2002; Aleardi 2020; Berti *et al.* 2023) or by treating the number of model parameters as an additional unknown (reversible-jump MCMC; Ray *et al.* 2016; Guo *et al.* 2020). To enhance efficiency, several alternative MCMC techniques have been proposed and, among these, gradient-based MCMC methods have gained significant attention. One prominent method is Hamiltonian Monte Carlo (HMC; Fichtner & Simutè 2018; Aleardi & Salusti 2020; Gebraad *et al.* 2020), which can improve the convergence rate over non-gradient-based MCMC (Zunino *et al.* 2022). Another approach is represented by the Stochastic Newton MCMC, which constructs a proposal distribution using the local gradient and the Hessian of the negative

log posterior. Originally introduced by Martin *et al.* (2012), it was later applied by Zhao & Sen (2021) to the acoustic FWI. Additionally, Berti *et al.* (2023, 2024a and 2024b) combined this MCMC approach with a discrete cosine transform (DCT) reparametrization of both model and data spaces for both acoustic and elastic FWI (Rincon *et al.* 2025).

In recent years, variational inference (VI) has been introduced to geophysics as an alternative to MCMC techniques, offering improved efficiency for Bayesian inference in certain types of problems. Variational methods define a family of simple probability distributions, known as the variational family, and seek the optimal member of this family that best approximates the true (unknown) PPD function (pdf; Blei *et al.* 2017). This approach narrows the search space compared to Monte Carlo methods, which often assume fully unknown posterior structures. The optimal solution in VI is derived by minimizing the Kullback–Leibler (KL) divergence (Kullback & Leibler 1951) measuring the difference between the posterior and variational distribution.

Advancements in computational power and modern deep learning frameworks, such as PYTORCH (Paszke *et al.* 2019), have facilitated the development of advanced variational algorithms and their application to various geophysical problems. For instance, Nawaz & Curtis (2018) applied the mean-field variational inference to invert for subsurface geological facies distributions. While computationally efficient, the mean-field approximation disregards correlations between parameters, restricting their applicability to specific problem classes.

To broaden the applicability of VI to more general inverse problems, Kucukelbir *et al.* (2017) introduced a Gaussian variational family to develop a method called automatic differential variational inference (ADVI), that has been then applied to seismic traveltimes tomography (Zhang & Curtis 2020a), and earthquake slip inversion (Zhang & Chen 2022). Another development is the normalizing flow variational inference, proposed by Rezende & Mohamed (2015), which employs a sequence of invertible and differential transforms (called normalizing flows) to convert a simple initial distribution into a complex one that closely approximates the target posterior (Kumar *et al.* 2021; Siahkoobi *et al.* 2021, 2022; Zhao *et al.* 2022).

The effectiveness of these variational inference methods hinges on the complexity and flexibility of the chosen variational family. Expanding the capacity of the variational family to approximate the posterior distribution generally increases the complexity of the optimization task. In many variational approaches, the pre-defined family is rigidly constrained, potentially limiting its ability to accurately approximate the true posterior pdf. Moreover, the target posterior pdf may not be faithfully represented due to the finite basis or parametrization typically used in solving inverse problems, which can be arbitrarily chosen.

Recently, a range of particle based VI methods has emerged to bridge the gap between VI and MCMC techniques. These methods employ a specific number of samples, often called particles, to model the approximating distribution (like MCMC) and update these particles through an optimization process (like VI). These approaches offer greater non-parametric flexibility compared to VI, while also being more particle efficient than traditional MCMC methods, as they leverage interactions between particles more effectively.

One notable method in this category is stein variational gradient descent (SVGD). SVGD is a deterministic particle-based inference algorithm that updates particles iteratively by minimizing the KL divergence, thereby aligning the final particle density with the PPD.

Zhang & Curtis (2020b) successfully applied SVGD to transmission FWI, where earthquake-like sources are positioned beneath the target velocity structure, with receivers on the surface. This method has also been employed in 2-D FWI with realistic priors (Zhang & Curtis 2021; Izzatullah *et al.* 2023) and in 3-D acoustic FWI using both synthetic (Zhang *et al.* 2023) and field data (Lomas *et al.* 2023).

However, empirical evidence suggests that SVGD encounters degeneracy issues under finite particle conditions, causing the particles to collapse into a small number of modes (known as the mode collapse issue; Zhuo *et al.* 2018). Another challenge affecting SVGD is the so-called variance collapse phenomenon, where the estimated variance can be significantly smaller than the variance of the target distribution as the problem dimensionality increases. This is a particular concern in high-dimensional geophysical inverse problems like FWI (Ba *et al.* 2022).

To tackle these challenges, we adopt a variant of the traditional SVGD approach known as annealed SVGD (A-SVGD), introduced by D'Angelo & Fortuin (2021). This adaptation draws inspiration from strategies employed to alleviate mode-collapse issues in MCMC methods and resembles the damped SVGD proposed by Ba *et al.* (2022). For acoustic FWI, recent studies have demonstrated that A-SVGD effectively mitigates mode and variance collapse issues compared to standard SVGD (Corrales *et al.* 2024).

In this paper, we instead employ the A-SVGD method to solve the elastic FWI of surface waves problem for two case studies. First, we apply it to a synthetic near-surface model characterized by significant lateral and vertical velocity variations. Subsequently, we test the A-SVGD method on a real data set acquired in Grenoble, France, in the context of the InterPACIFIC project. In this case, our inversion results can also be validated by the available well-log data measured along the acquisition line. We also compare the predictions of A-SVGD with those of standard SVGD. This work marks the first successful application of SVGD methods to the elastic FWI of field data.

METHOD

Variational inference

The Bayesian approach uses the Bayes' theorem to quantify the model uncertainties expressed by the PPD function in the model space:

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

In eq. (1), $p(\mathbf{m})$ represents the prior distribution of the model parameters \mathbf{m} , encapsulating our knowledge about \mathbf{m} that is independent from the acquired geophysical data. The conditional probability $p(\mathbf{d}_{\text{obs}}|\mathbf{m})$ denotes the likelihood of observing the data \mathbf{d}_{obs} given an Earth model \mathbf{m} . The denominator $p(\mathbf{d}_{\text{obs}})$ acts as a normalization constant known as the evidence. The combination of these terms yields the posterior distribution $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$, describing the conditional probability of all possible models given the observed data.

Variational inference (VI) aims to select an optimal distribution $q^*(\mathbf{m})$ from a family of known distributions, denoted as the variational family $\mathcal{Q}(\mathbf{m}) = \{q(\mathbf{m})\}$, that best approximates this target PPD. This selection is achieved by identifying the distribution $q(\mathbf{m})$ that minimizes the distance (difference) between the variational and posterior distributions, typically measured by the Kullback–Leibler

(KL) divergence:

$$\text{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{\text{obs}})] = E_q[\log q(\mathbf{m}) - \log p(\mathbf{m}|\mathbf{d}_{\text{obs}})]. \quad (2)$$

The KL divergence quantifies the disparity between $q(\mathbf{m})$ and $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$, with $\text{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{\text{obs}})] \geq 0$, achieving equality when $q(\mathbf{m}) = p(\mathbf{m}|\mathbf{d}_{\text{obs}})$.

If we expand the posterior pdf $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$ using Bayes' theorem, the evidence $\log p(\mathbf{d}_{\text{obs}})$ appears in the second term of eq. (2). This poses significant computational challenges since the marginal pdf over \mathbf{d}_{obs} of the joint distribution $p(\mathbf{m}, \mathbf{d}_{\text{obs}})$ requires integration of the forward function over the full prior $p(\mathbf{m})$.

By rearranging terms of the eq. (2), we can derive the evidence lower bound (ELBO):

$$\begin{aligned} \text{ELBO}[q] &= \log p(\mathbf{d}_{\text{obs}}) - \text{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{\text{obs}})] \\ &= E_q[\log p(\mathbf{d}_{\text{obs}}|\mathbf{m}) + \log p(\mathbf{m})] - E_q[\log q(\mathbf{m})]. \end{aligned} \quad (3)$$

The right-hand side of eq. (3) does not involve the intractable evidence term and can be practically computed using analytical or numerical methods. Minimizing the KL divergence is equivalent to maximizing the ELBO with respect to $q(\mathbf{m})$, providing a fully probabilistic solution to Bayesian inverse problems through the use of optimization techniques.

Stein variational gradient descent

We now focus our attention on a specific VI approach, the SVGD, and in particular its variant, the annealed SVGD (A-SVGD) that is used in this work. SVGD operates as a deterministic sampling algorithm that iteratively minimizes the KL divergence between the chosen approximating distribution and the target density (Zhang & Curtis 2020b). This ensures that the final set of particles adheres closely to the desired PPD.

Initially, we define a set of particles as $\{\mathbf{m}_i\}_{i=1}^N$, which are updated using a smooth transform $T(\mathbf{m}_i) = \mathbf{m}_i + \epsilon\phi(\mathbf{m}_i)$, where $\phi = \{\phi_1, \dots, \phi_k\}$ is a smooth vector function describing the perturbation direction and ϵ is the perturbation magnitude. Assuming T is invertible and defining $q_T(\mathbf{m})$ as the transformed probability distribution of pdf $q(\mathbf{m})$, we can compute the gradient of the KL divergence between q_T and the posterior pdf p with respect to ϵ as:

$$\nabla_{\epsilon} \text{KL}[q_T||p]|_{\epsilon=0} = -E_q[\text{trace}(A_p\phi(\mathbf{m}))], \quad (4)$$

where A_p is the Stein operator defined by $A_p\phi(\mathbf{m}) = \nabla_{\mathbf{m}} \log p(\mathbf{m})\phi(\mathbf{m})^T + \nabla_{\mathbf{m}}\phi(\mathbf{m})$.

The optimal ϕ^* that maximizes the expectation in eq. (4) above can be found using kernel functions. If we consider $x, y \in X$ and define a mapping φ from X to a space where an inner product $\langle \cdot, \cdot \rangle$ is defined (i.e. a Hilbert space), a *kernel* is a function that satisfies the following property: $k(x, y) = \langle \varphi(x), \varphi(y) \rangle$. We assume a kernel $k(\mathbf{m}', \mathbf{m})$ and ϕ^* can be expressed as:

$$\phi^* \propto E_{\{\mathbf{m}' \sim q\}}[A_p k(\mathbf{m}', \mathbf{m})]. \quad (5)$$

Since we are using particles $\{\mathbf{m}_i\}_{i=1}^N$ to represent q in SVGD, the expectation can be approximated using the sample mean over the particles. The KL divergence can thus be iteratively minimized as follows:

$$\begin{aligned} \phi_l^*(\mathbf{m}) &= \frac{1}{N} \sum_{i=1}^N [k(\mathbf{m}'_i, \mathbf{m}) \nabla_{\mathbf{m}'_i} \log p(\mathbf{m}'_i|\mathbf{d}_{\text{obs}}) + \nabla_{\mathbf{m}'_i} k(\mathbf{m}'_i, \mathbf{m})] \\ \mathbf{m}_i^{l+1} &= \mathbf{m}_i^l + \epsilon^l \phi_l^*(\mathbf{m}'_i) \end{aligned} \quad (6)$$

Here, l denotes the current iteration, N is the number of particles and ϵ^l is the step size. Assuming the step size to be sufficiently

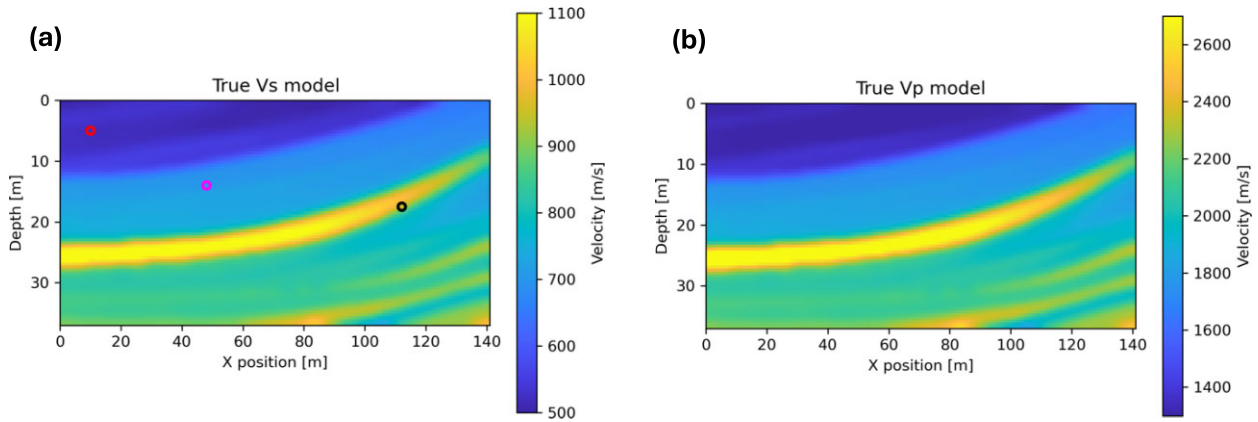


Figure 1. (a, b) True V_S and V_P models used for our synthetic test, respectively. The coloured circles in the left image highlight the locations of three cells used for the analysis of marginal distributions.

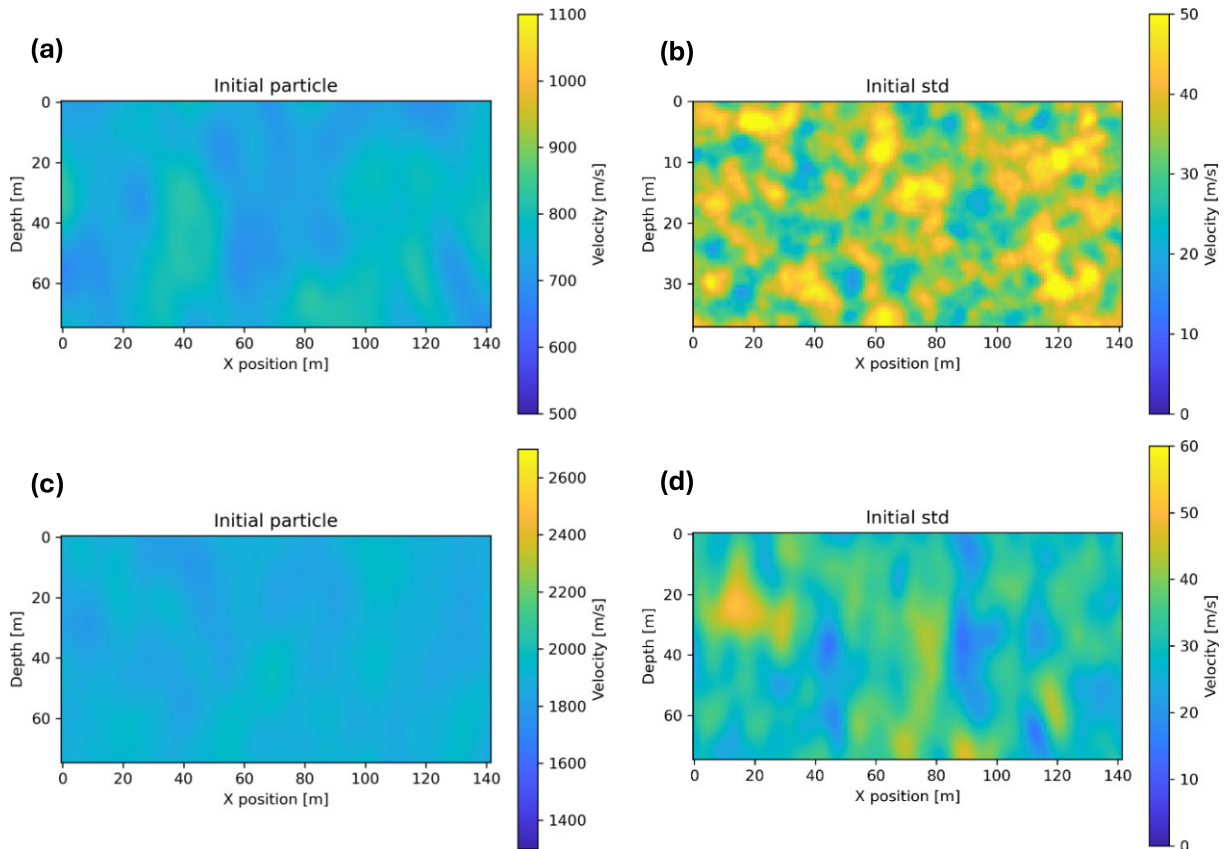


Figure 2. (a, c) Example of an initial particle for the V_S and V_P models, obtained applying GRF perturbations to a homogeneous model with a constant velocity value of 750 and 1800 m s^{-1} , respectively; (b, d) Initial standard deviations computed considering all the 100 initial particles for both the V_S and V_P models.

small, the process asymptotically converges to the target posterior as the number of particles tends to infinity.

Various kernel functions have been proposed over the last few years (Liu & Wang 2016; Gorham & Mackey 2017; Wang *et al.* 2019). However, in our study, we adopt the radial basis function (RBF) defined as:

$$k(\mathbf{m}', \mathbf{m}) = \exp\left[-\frac{\|\mathbf{m} - \mathbf{m}'\|^2}{2h^2}\right], \quad (7)$$

where h is a scale factor that intuitively regulates the strength of interaction between different particles based on their distances. As

suggested in several studies (Liu & Wang 2016; Zhang & Curtis 2020b), we set h to $\tilde{d}/\sqrt{2 \log N}$ where \tilde{d} represents the median of pairwise distances between all particles and N is the total number of particles. For the RBF kernel, the second term of ϕ^* in the equation above becomes $\sum_i \frac{m - m_i}{h^2} k(\mathbf{m}_i, \mathbf{m})$, which pushes the particle \mathbf{m} away from its neighbouring particles when the kernel assumes high values.

In eq. (6), two distinct terms can be identified. The first term, containing the gradient of $\log p(\cdot)$ represents a weighted average steepest descent direction of the log-target density, guiding the particles towards high probability regions of $p(\cdot)$. The second term,

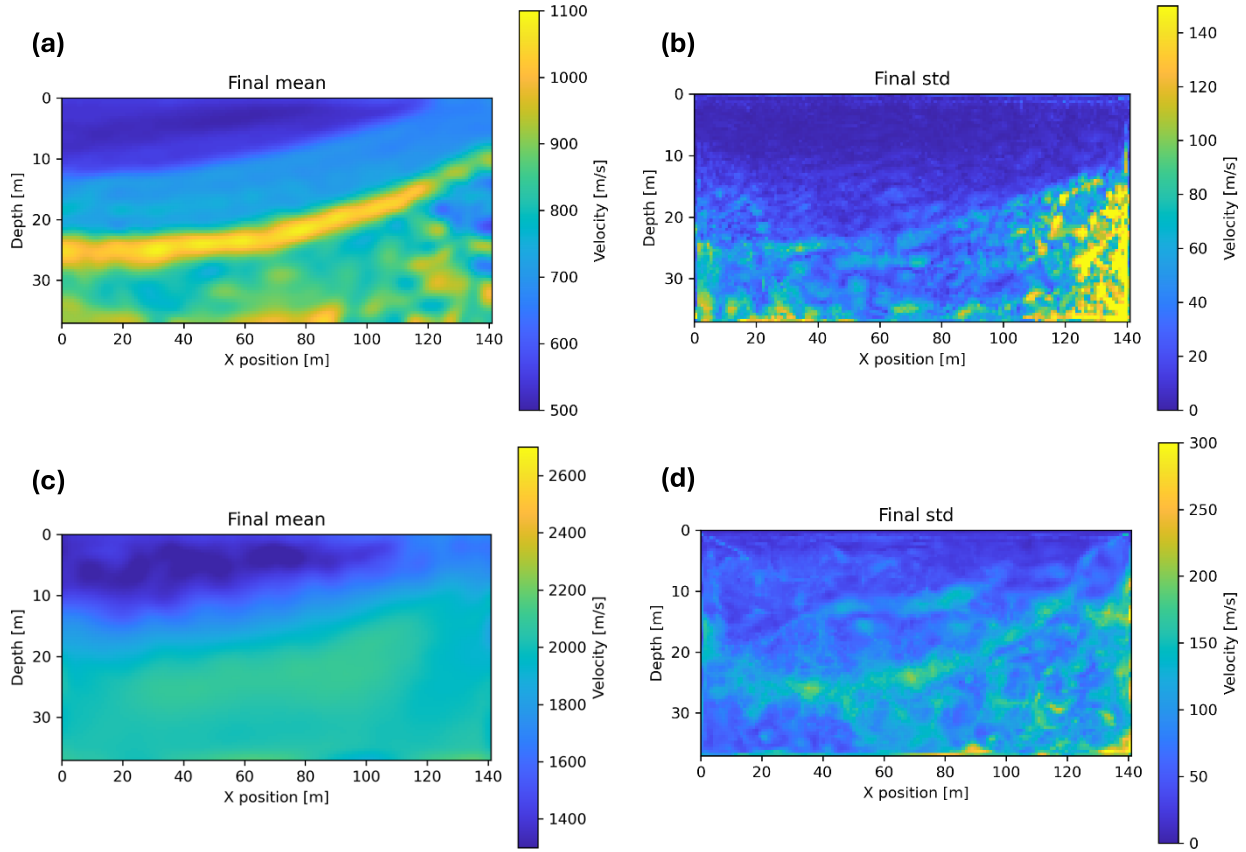


Figure 3. (a) Final mean V_S model computed using all 100 particles; (b) Final standard deviation map for the V_S model; (c) Final mean V_P model computed using all 100 particles; (d) Final standard deviation map for the V_P model.

involving $\nabla_{\mathbf{m}} k$, acts as a repulsive force, preventing particles from collapsing together by spreading them apart.

In their work, Ba *et al.* (2022) discovered that the variance collapse phenomenon, which typically affects standard SVGD algorithms as the problem dimensionality increases, is related to the bias introduced by deterministic updates present in the ‘driving force’ of the SVGD update. They empirically demonstrated that removing such bias leads to more accurate variance estimations, resulting in the damped SVGD approach.

Annealed SVGD

The Annealed SVGD (D’Angelo & Fortuin 2021) represents a variant of the traditional SVGD approach similar to the approach proposed by Ba *et al.* (2022). It incorporates an annealing schedule to enhance exploration and improve mode coverage, while maintaining the deterministic nature of SVGD. This approach aims to mitigate initialization constraints and limitations imposed by a finite number of particles. An annealing parameter $\alpha(l) \in [0, 1]$, which varies with the current iteration l , regulates the updating rule:

$$\phi_l^*(\mathbf{m}) = \frac{1}{N} \sum_{i=1}^N \left[\underbrace{\alpha(l) k(\mathbf{m}'_i, \mathbf{m}) \nabla_{\mathbf{m}'_i} \log p(\mathbf{m}'_i | \mathbf{d}_{\text{obs}})}_{\text{driving force}} + \underbrace{\nabla_{\mathbf{m}'_i} k(\mathbf{m}'_i, \mathbf{m})}_{\text{repulsive force}} \right]. \quad (8)$$

Varying the α parameter within the $[0, 1]$ range induces two distinct phases. The exploratory phase is dominated by a strong repulsive force that disperses particles from their initial positions, facilitating broad coverage of the target distribution. The exploitative phase, where the driving force predominates, concentrates particles distribution around different modes.

The modification to the classic SVGD framework introduces a temperature parameter $A(l) = 1/\alpha(l)$, which adapts the target distribution over the evolution of the approximating density. Careful selection of the annealing schedule is crucial to maintain SVGD’s convergence properties, ensuring that the final iterations operate effectively on the true target density, that is, $\lim_{l \rightarrow \infty} \alpha(l) = 1$. The transition between the two inference phases is sharp, emphasizing significant evolution in one phase rather than a gradual transition. In this study, we employ the hyperbolic tangent as an annealing schedule, with $\alpha(l) = \tanh[(1.3 \frac{l}{M})^p]$, where p is a user-defined parameter controlling the rate of transition between the two phases (usually an integer between 1 and 5), l is the current iteration and M is the total number of iterations.

RESULTS

Synthetic inversion test

To evaluate the effectiveness of the proposed methodology, we first tested it on a synthetic velocity model featuring significant lateral and vertical velocity variations that pose challenges for standard surface-wave analysis methods like MASW. In this case, we assumed a V_P/V_S ratio of 2.5 (Fig. 1). The model parameters to be estimated are the V_P and V_S values, with density assumed to be perfectly known and fixed at 1.6 g cm^{-3} .

The seismic data were generated on a rectangular grid with dimensions of $n_z = 75$ and $n_x = 142$, representing the number of

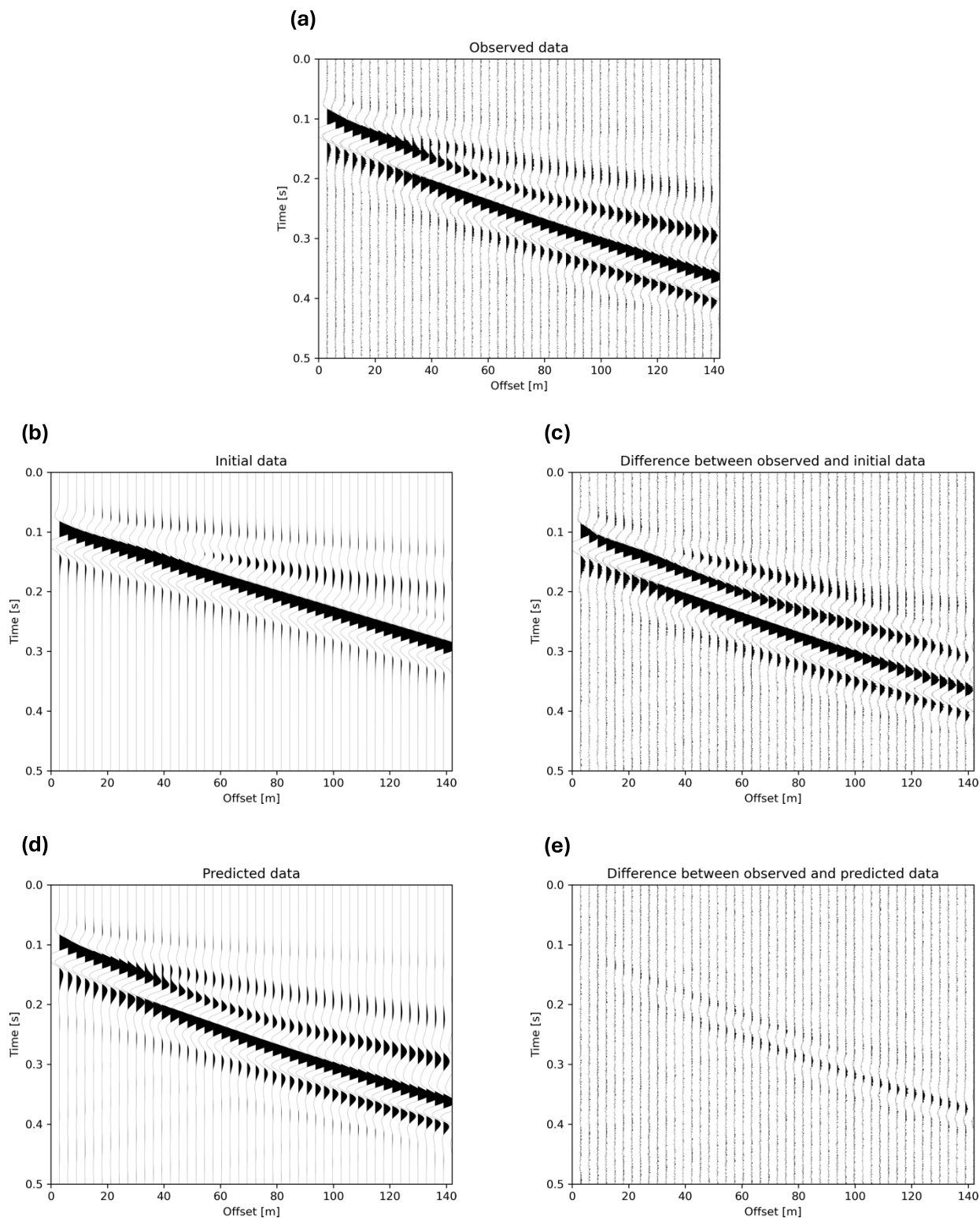


Figure 4. (a) The first shot of the noisy observed data; (b) the same shot generated using the initial model of Figs 2(a) and (c) their sample-by-sample difference; (d) the first shot of the predicted data computed using the final mean model of Figs 3(a) and (e) their sample-by-sample difference.

grid points along the vertical and horizontal directions, respectively. The grid spacing was set to 1 m in the horizontal direction and 0.5 m in the vertical direction, deliberately chosen to minimize numerical dispersion in the finite difference modelling. Since we are dealing with surface waves, the minimum number of

points per wavelength was set to 20, following the work of Pierini *et al.* (2020). The total number of model parameters for this test is 21 300.

The seismic source was modelled using a Ricker wavelet with a peak frequency of 15 Hz, using five shots evenly spaced along the

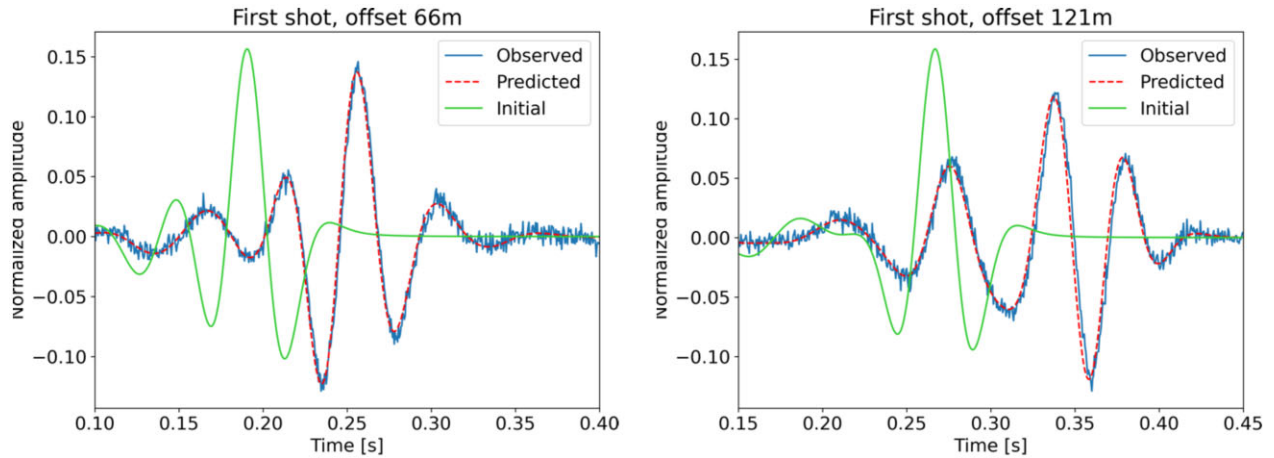


Figure 5. Comparison of mid and far offset traces of the first shot using the true model (observed), the final mean model estimated by the A-SVGD method (predicted) and the initial particle of Fig. 2(a) (initial).

horizontal axis and recorded by 48 receivers placed at 3-m intervals. The recording time was set to 0.5 s and the sample interval used for the forward modelling was 0.5 ms, chosen to ensure stability in the finite difference modelling by adhering to the Courant–Friedrichs–Lewy (CFL) condition. Uncorrelated Gaussian noise was added to the observed data, resulting in a signal-to-noise ratio (SNR) of 11 dB.

Elastic forward modelling was performed using Deepwave (Richardson 2022), a PYTORCH-based framework (Paszke *et al.* 2019) that implements wave propagators as modules, enabling efficient gradient computations via backpropagation on both CPUs and GPUs.

Following the approach of Izzatullah *et al.* (2024), we initialized the A-SVGD FWI with particles generated by perturbing a homogeneous model with a constant velocity value of 750 m s^{-1} using Gaussian random field (GRF) perturbations, with varying variances for both V_P and V_S (Fig. 2).

For the hyperparameters employed in this study, we opted for the Gaussian RBF kernel, using the median trick for the bandwidth selection (eq. 7). The particles were updated at each iteration using the Adam optimizer. We set the number of particles to 100 and the number of iterations to 300. We employed the hyperbolic tangent as the annealing schedule setting p to 3 (see eq. 8) and keeping the α fixed to 1 for the final 100 iterations, following the work of Corrales *et al.* (2024) who applied the A-SVGD algorithm to a FWI problem.

The choice of the number of particles was made to achieve an optimal balance between computational efficiency and quality of the inversion results. Employing a NVIDIA A100 GPU, the entire inversion process took just 3.5 hr.

Fig. 3 presents the inversion results for both the V_S and V_P models, displaying the mean and standard deviation computed from the final ensemble of particles after 300 iterations. The mean V_S model (Fig. 3a) successfully captures all the main velocity variations of the true model (Fig. 1a), in particular the high-velocity dipping layer between 10 and 30 m of depth. The standard deviation map shown in Fig. 3(b) reveals a low uncertainty zone, located above this high velocity dipping layer, indicating good illumination and data coverage. In contrast, a region of high uncertainty is observed below this layer, with standard deviation values above 100 m s^{-1} ,

particularly at the bottom edges of the model, where we have lower illumination.

For the V_P model instead, the final mean (Fig. 3c) is characterized by a low velocity zone in the first 10 m of depth, consistent with the true model (Fig. 1b). However, below this depth, the inversion struggles to accurately replicate the variations of the true model, only showing a slight increase in velocity in the central part of the model.

The standard deviation map associated with the V_P model shown in Fig. 3(d), similarly to the one associated with the V_S , distinguishes two regions: a shallow, low-uncertainty area above the high-velocity dipping layer, and a deeper, high-uncertainty area, with standard deviation values exceeding 200 m s^{-1} , especially at the bottom edges of the model.

These findings underscore the higher sensitivity of Rayleigh waves to variations of the V_S parameter compared to the V_P .

Fig. 4(a) shows the first shot (with the source located at the left edge of the model) of the noisy observed data, providing a baseline for evaluating our inversion results. Figs 4(b) and (c) display, respectively, the same shot simulated using one of the initial particles employed for the inversion (the one depicted in Fig. 2a) and the corresponding sample-by-sample difference with the observed data. This comparison underscores the challenges faced by the inversion process due to significant discrepancies between initial and observed data.

Finally, Figs 4(d) and (e) present the predicted data computed using the final mean model shown in Fig. 3(a), along with its residual difference compared to the observed data. The data matching shows substantial improvement when transitioning from the starting model to the estimated mean, with just some minor differences at larger offsets.

To further illustrate the effectiveness of our proposed method, Fig. 5 presents a comparison of two traces (a mid and a far offset) extracted from the first shot of the observed noisy data, the initial data shown in Fig. 4(b) and the predicted data of Fig. 4(d). In these close-ups, we can clearly appreciate the significant cycle skipping between the observed and the initial data, that makes this inversion test challenging. However, these cycle-skips disappear when considering the predicted data, where the predicted traces are almost perfectly aligned with those of the observed data.

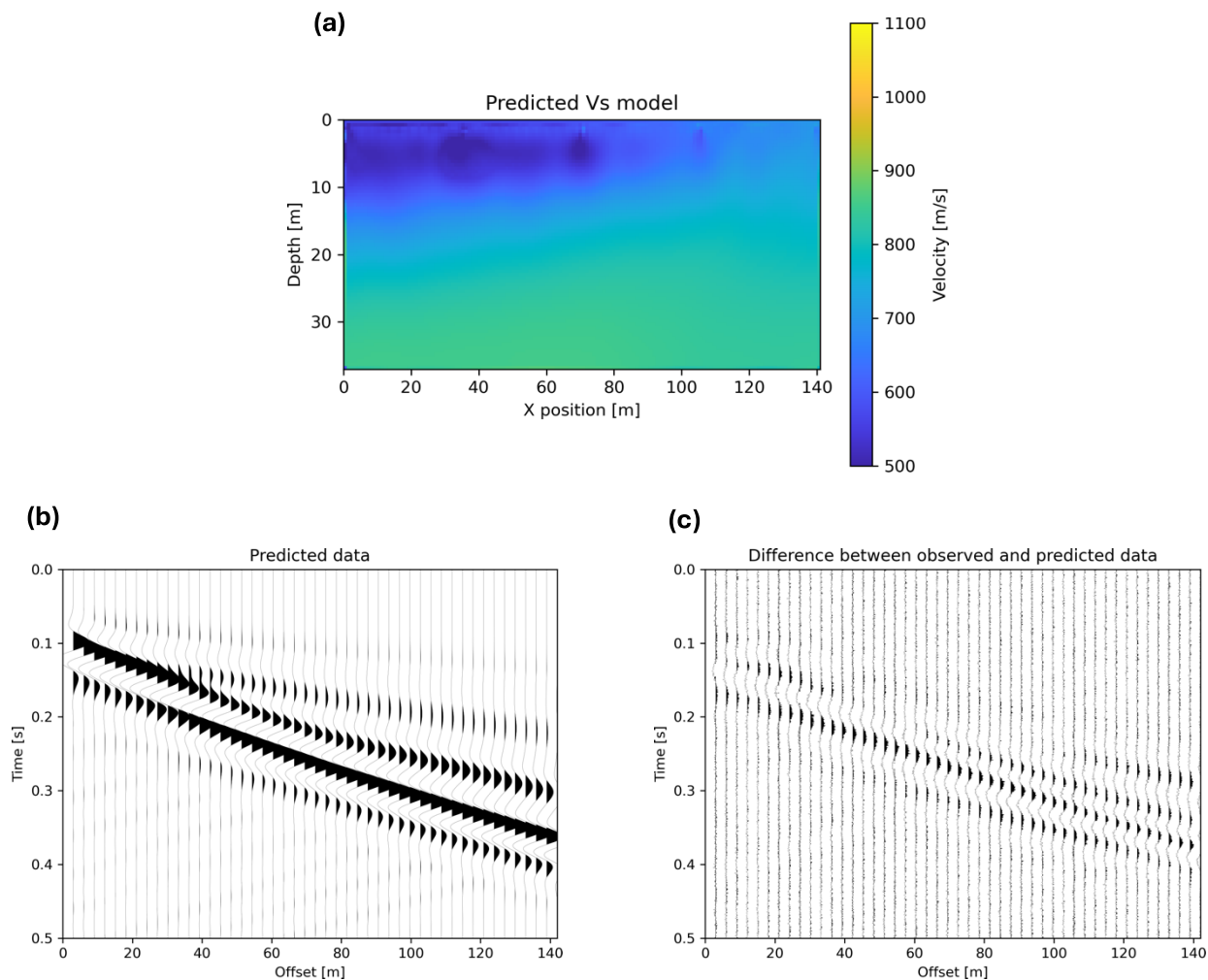


Figure 6. (a) Predicted V_S model after 100 iterations using the L-BFGS algorithm; (b, c) First shot calculated using the L-BFGS predicted model and its sample-by-sample difference with the observed noisy data of Fig. 4(a), respectively.

Table 1. Comparison of the results obtained with A-SVGD and L-BFGS, in terms of SNR with respect to the true V_S model, RPE with respect to the observed data and 99 per cent coverage ratio.

	A-SVGD	L-BFGS
SNR	25.7	19.4
RPE per cent	19.6	27.5
Coverage ratio per cent	86.5	/

Comparison with local inversion

In this section, we illustrate the comparison between the results obtained with the proposed approach, with those resulting from an advanced deterministic inversion method: the L-BFGS algorithm (Liu & Nocedal 1989).

The L-BFGS is a quasi-Newton method that employs an approximation of the Hessian matrix to update the model parameters at each iteration. The full Hessian is not stored or computed explicitly, making it suitable for high-dimensional problems like FWI. It is generally faster to converge compared to other deterministic methods (like steepest descent), thanks to the second-order information coming from the Hessian approximation.

The forward modelling is again performed with Deepwave, using the initial particle depicted in Fig. 2(a) as the starting model,

and maintaining the same acquisition geometry as before. We have considered a total of 100 iterations, with a maximum of 3 iterations for the line-search. The inversion took 35 min on the same NVIDIA A100 GPU employed for the A-SVGD test.

In Fig. 6(a) we present the predicted V_S model, where we can discriminate the first low-velocity dipping layer. However, the L-BFGS inversion struggles to accurately capture features of the true model below a depth of 15 m, unlike the A-SVGD algorithm, which yields more accurate results, as shown in Fig. 3(a).

Figs 6(b) and (c) show the first shot of the predicted data and its sample-by-sample difference with the same shot of the observed noisy data (Fig. 4a), respectively. While the difference has been significantly reduced compared to the initial data misfit shown in Fig. 4(c), it remains larger than the corresponding difference for the predicted data calculated using the final mean model obtained with the A-SVGD inversion (Fig. 4e). This comparison highlights that the A-SVGD method achieves superior results, in terms of both quality of the predicted model and data misfit, compared to advanced local inversion techniques, like L-BFGS. Furthermore, the A-SVGD algorithm offers the additional benefit of providing reliable uncertainty estimations using a limited number of particles.

For a more quantitative evaluation, Table 1 illustrates the key differences between the results obtained using the A-SVGD and

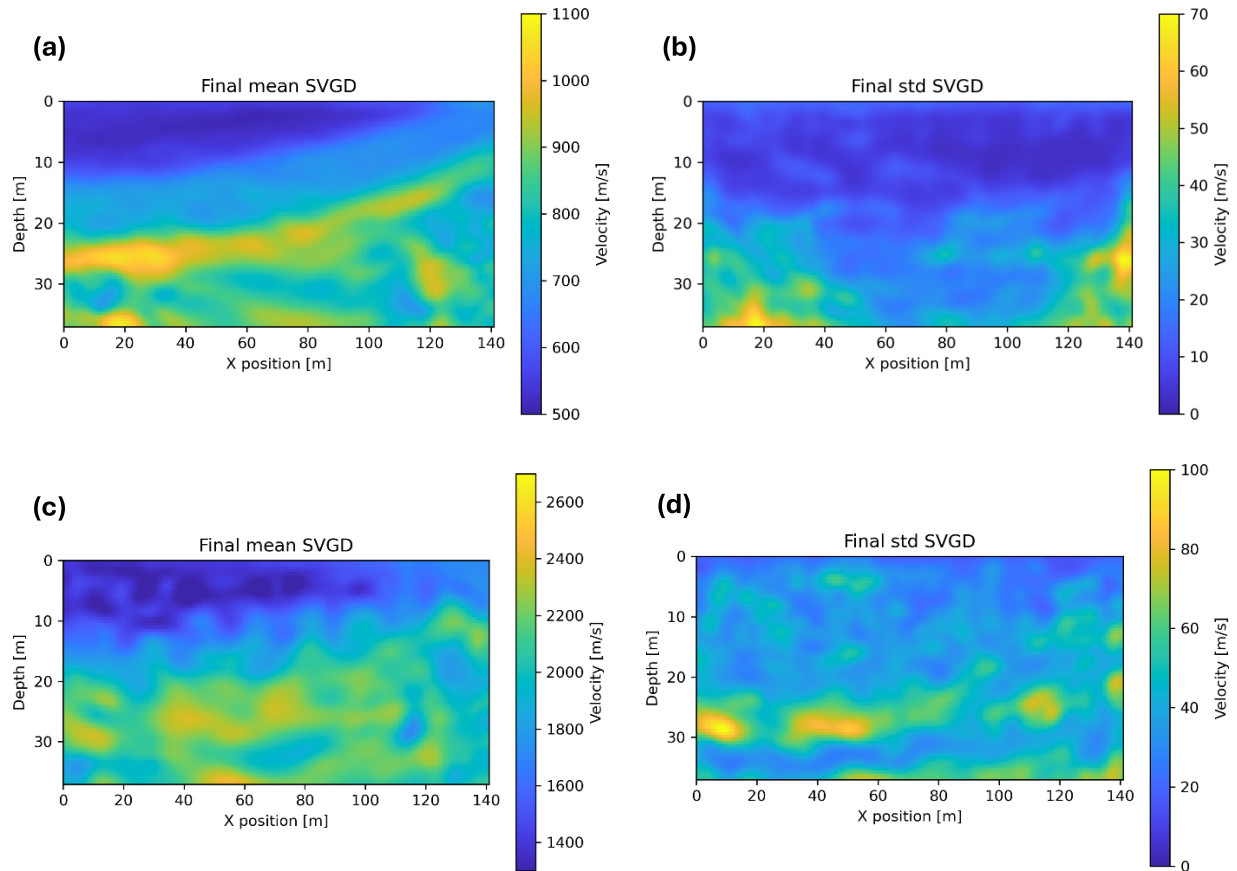


Figure 7. Results obtained with the classic SVGD method: (a) Final mean V_S model computed using all 100 particles; (b) Final standard deviation map for the V_S model; (c) Final mean V_P model; (d) Final standard deviation map for the V_P model.

L-BFGS methods, in terms of SNR of the V_S model prediction with respect to the true V_S model and the relative percentage error (RPE) of the predicted data with respect to the observed noisy data, proving again the superior performance of our proposed A-SVGD method.

Table 1 also includes the 99 per cent coverage ratio obtained with the A-SVGD algorithm, quantifying the percentage of V_S values in the true model that falls within the 99 per cent confidence interval estimated by our proposed method. The very high value (above 85 per cent) we obtain confirms the robustness and reliability of the estimated solutions.

Comparison with standard SVGD

In this section we compare the results obtained using the standard SVGD algorithm with those achieved by the A-SVGD inversion. To ensure consistency and enable a fair evaluation, the SVGD inversion employs the same initial set of particles as the A-SVGD inversion.

The final mean V_S model obtained with the standard SVGD algorithm (Fig. 7a) reproduces the main features of the true model but slightly underestimates the velocities of the dipping layer between 10 and 30 m of depth. The standard deviation map (Fig. 7b) reveals similar patterns to those observed in the A-SVGD inversion result (Fig. 3b), with the highest uncertainties below the high-velocity dipping layer and at the bottom edges of the model. However, the standard deviation values estimated by the SVGD are significantly lower, with maximum values of approximately 70 m s^{-1} . This is

expected and is attributed to the comparatively more limited capacity of SVGD, relative to A-SVGD, to efficiently explore the model space.

For what concerns the V_P , the final mean model (Fig. 7c) captures the low-velocity zone in the first 10 m of depth and the increasing velocity trend below this depth. However, it fails to accurately reconstruct the high-velocity dipping layer visible in the true model (Fig. 1b). The standard deviation map for the V_P (Fig. 7d) shows a slight increase in uncertainty towards the deeper parts of the model, with maximum values reaching 100 m s^{-1} . Again, these uncertainties are notably lower than those observed in the A-SVGD results (Fig. 3d).

Fig. 8(a) compares the marginal distributions of V_S obtained from the A-SVGD and SVGD inversions, calculated for the cells highlighted by the coloured circles in Fig. 1(a). In all cases, the initial distribution is significantly distant from the true model. The leftmost histograms correspond to the shallowest cell, located approximately 5 m deep and 5 m horizontally. The maximum-*a-posteriori* solutions for both algorithms are very close to the true value. The central plot represents a cell located at roughly 15 m depth and 50 m horizontally. In this case, the A-SVGD posterior distribution is nearly perfectly centered on the true V_S , whereas the SVGD tends to underestimate the actual velocity. Notably, only the A-SVGD posterior includes the true model.

Finally, the rightmost plot corresponds to the deepest cell, located at approximately 20 m depth and 110 m horizontally. The A-SVGD posterior distribution is once again almost perfectly centred around

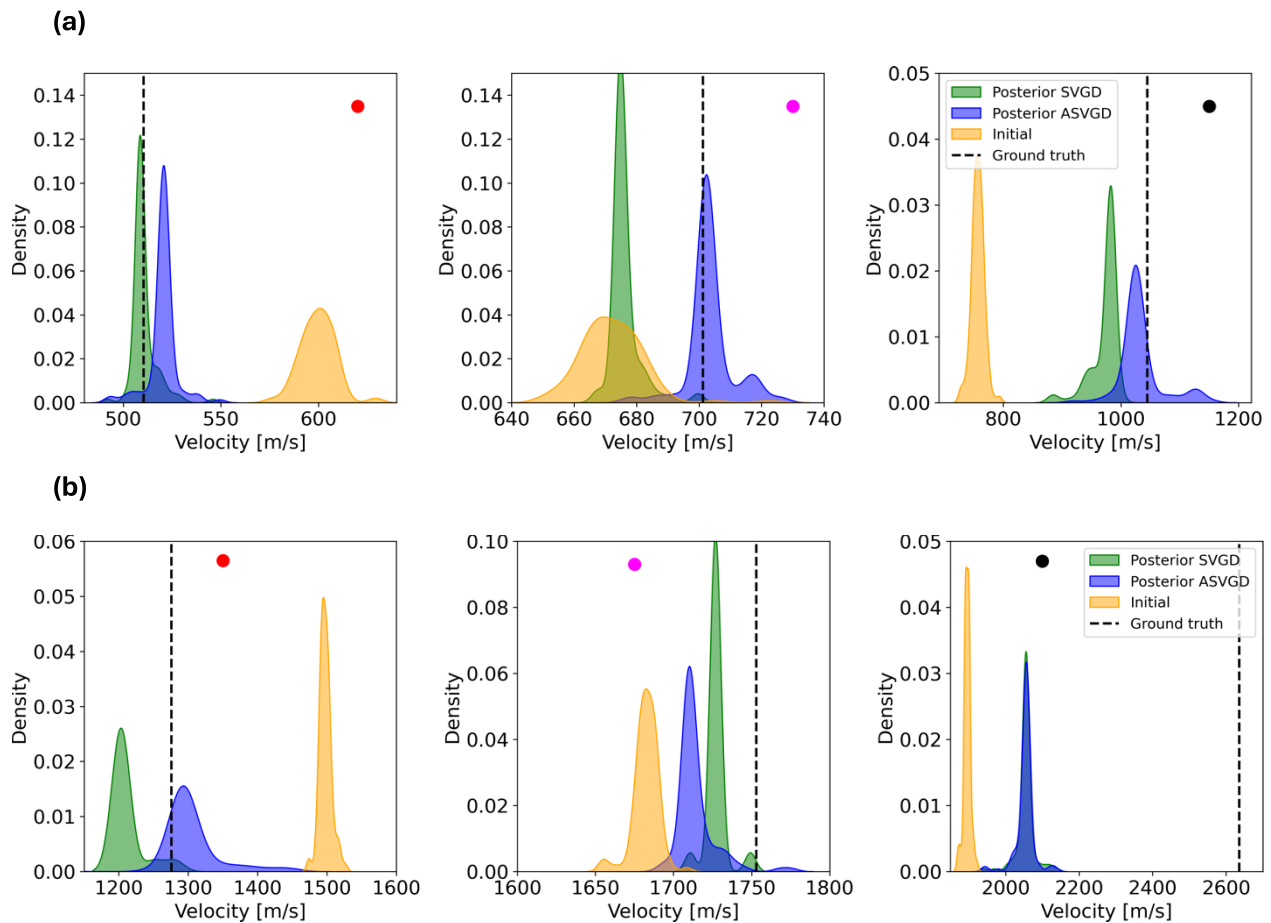


Figure 8. From left to right, marginal distributions associated with three different cells highlighted respectively by the red, magenta and black circles in Fig. 1(a). (a) for the V_S and (b) for the V_P . Green and blue represent the marginal posterior distributions for the SVGD and A-SVGd algorithms, respectively; orange indicates the marginal initial distribution, whereas the black vertical dashed line represents the true velocity value.

Table 2. Comparison of the results obtained with A-SVGd and SVGd, in terms of SNR with respect to the true V_S model, RPE with respect to the observed data and 99 per cent coverage ratio.

	A-SVGd	SVGd
SNR	25.7	21.2
RPE per cent	19.6	25.1
Coverage ratio per cent	86.5	71.4

the true value, while the SVGd significantly underestimates the true V_S .

In Fig. 8(b) we compare the marginal distributions associated with the same three cells but related to the V_P parameter. The initial distribution is again quite far from the true velocity values. For the first two cells the two posterior pdfs include the true V_P . In contrast, for the deepest cell, we observe a severe underestimation of the true value. This can be attributed to two factors: the initial distribution being far from the ground truth and the fact that V_P is much less constrained by the data compared to V_S . As a result, the inversion struggles to effectively update the V_P model below a certain depth.

As previously done for the local inversion, for a more quantitative comparison Table 2 illustrates the differences between the results obtained using the standard SVGd and its annealed variant, in terms of SNR of the V_S model prediction with respect to the true V_S model and the relative percentage error (RPE) of the predicted data

with respect to the observed noisy data. These metrics again prove the superior performance of our proposed A-SVGd method with respect to the standard algorithm. Finally, the higher coverage ratio value obtained with the A-SVGd algorithm confirms the ability of the proposed method to mitigate the variance collapse issue.

Real data set inversion

After validating our proposed approach on synthetic data, we now demonstrate its effectiveness in inverting field data.

This section presents the results of applying the A-SVGd algorithm to a field data set acquired in Grenoble, France, in the framework of the InterPACIFIC project (Garofalo *et al.* 2016). This data set was chosen primarily due to the availability of some borehole information located at 10 m along the acquisition line, allowing for direct validation and comparison with our V_S predictions.

The data acquisition was carried out using an 8 kg sledgehammer as the active source to energize the subsurface. The acquisition surface is flat, featuring a linear array of 48 vertical geophones with a natural frequency of 4.5 Hz and a spacing of 1 m. The data set consists of three shot gathers: one split-spread and two off-ends, with maximum offsets of 23.5 and 51 m, respectively.

Unlike the synthetic case, where the source wavelet was known and the observed data was contaminated with Gaussian noise and

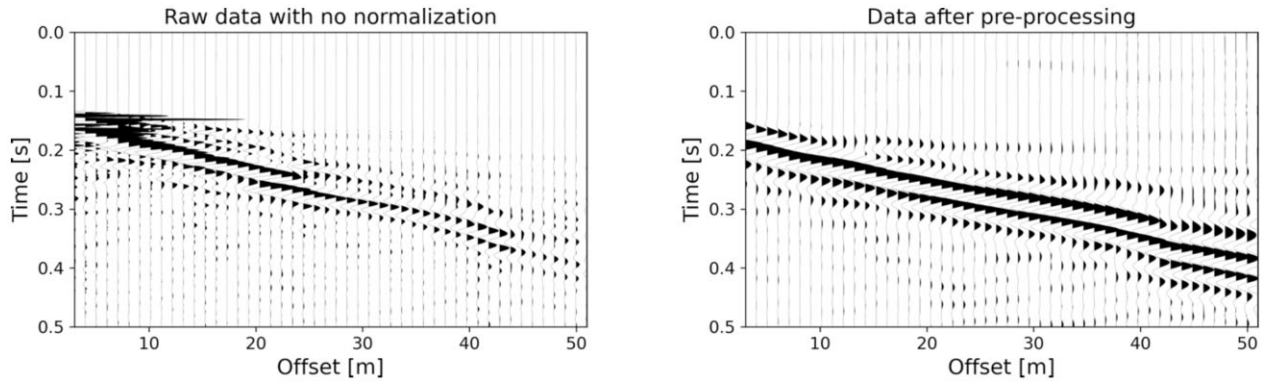


Figure 9. Comparison between the first shot of the raw observed data, on the left, and the same shot after the application of the pre-processing sequence, on the right.

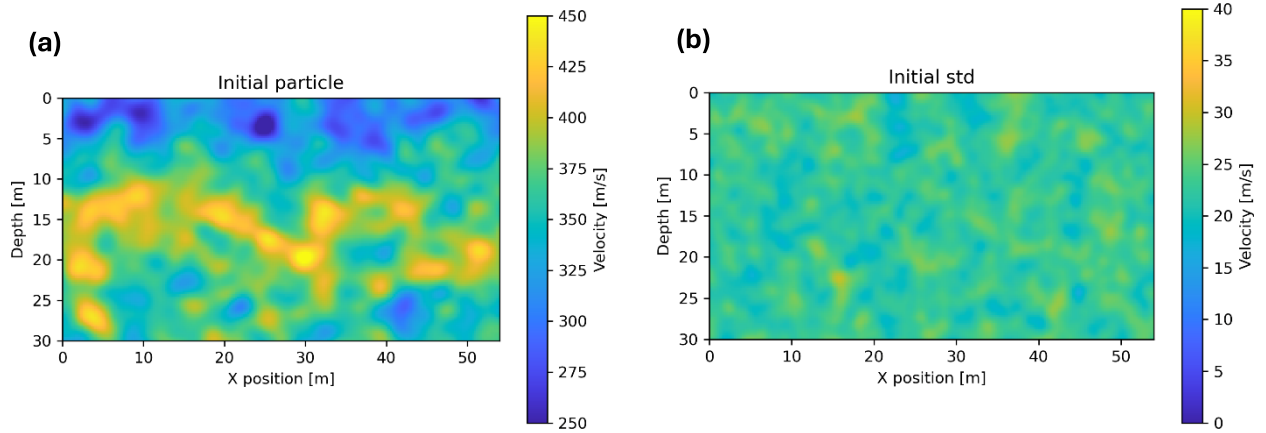


Figure 10. (a) Example of an initial particle for the V_S model, obtained applying GRF perturbations to a highly smoothed version of the available borehole information; (b) Standard deviation computed considering all the 250 initial particles.

computed with the same modelling engine as the simulated data, the field data required some initial pre-processing steps. We applied a trace-by-trace amplitude normalization to our seismic data and a zero-phase band-pass filter (3–30 Hz) to reduce background noise and constrain the maximum frequency for the inversion. Additionally, a 3-D to 2-D correction (Forbriger *et al.* 2014) was necessary to account for the geometrical spreading difference between the real point source and the 2-D forward modelling, which implicitly assumes line source for the simulations. Fig. 9 shows a comparison between the same shot gather before and after applying the processing sequence. Notably, no mute was applied to either the observed or modelled seismograms. This same processing sequence was applied to both the observed and simulated data.

A critical step when dealing with real data is the estimation of the source wavelet from the observed data, which can be carried out using several approaches. For instance, the average or least-squares fit of the reflected waveforms across multiple traces can be performed or we can apply the singular-value-decomposition (SVD) to a selected data matrix (Kirilin & Done 1999). In this work, we employed the SVD approach due to its computationally efficiency and its proven effectiveness in previous studies using the same data set (Lamuraglia *et al.* 2022; Berti *et al.* 2024c). This procedure works as follows: we defined sliding temporal-offset estimation windows, and each window contains time samples and traces that form a matrix containing reflected and interfering events, plus random noise. The SVD was then applied to each of these

matrices, and the reflected waveform was separated from interfering events and random noise by extracting the first eigenvalue (Biondi *et al.* 2014). Finally, an average across the columns is computed to obtain the estimated wavelet. As in the synthetic case, the model parameters to be estimated are the V_P and V_S values, with density assumed to be fixed at 2 g cm^{-3} .

For the simulated data, we employed a grid consisting of 216 grid points along the horizontal direction and 120 along the vertical one, with a uniform spacing of 0.25 m. Sources and receivers were both placed at the free surface. The total number of model parameters in this case is 51 840.

Elastic forward modelling was again performed using Deepwave (Richardson 2022), with a time sampling rate of 0.1 ms and a registration time of 0.5 s (following again the dispersion and stability conditions). Both the predicted and observed data were subsequently resampled to a 2 ms time interval.

The initial ensemble of particles (Figure 10) for the A-SVGD FWI was generated by applying GRF perturbations (as we did in the synthetic inversion test) to a highly smoothed version of the available borehole information for the V_S (for the V_P we used a scaled version of the V_S).

We again employed the Gaussian RBF kernel with the median trick for the bandwidth selection (eq. 7) and used the Adam optimizer for updating the particles at each iteration. The number of particles was fixed to 250 (given the higher number of unknowns, in this case we increased the number of particles with respect to

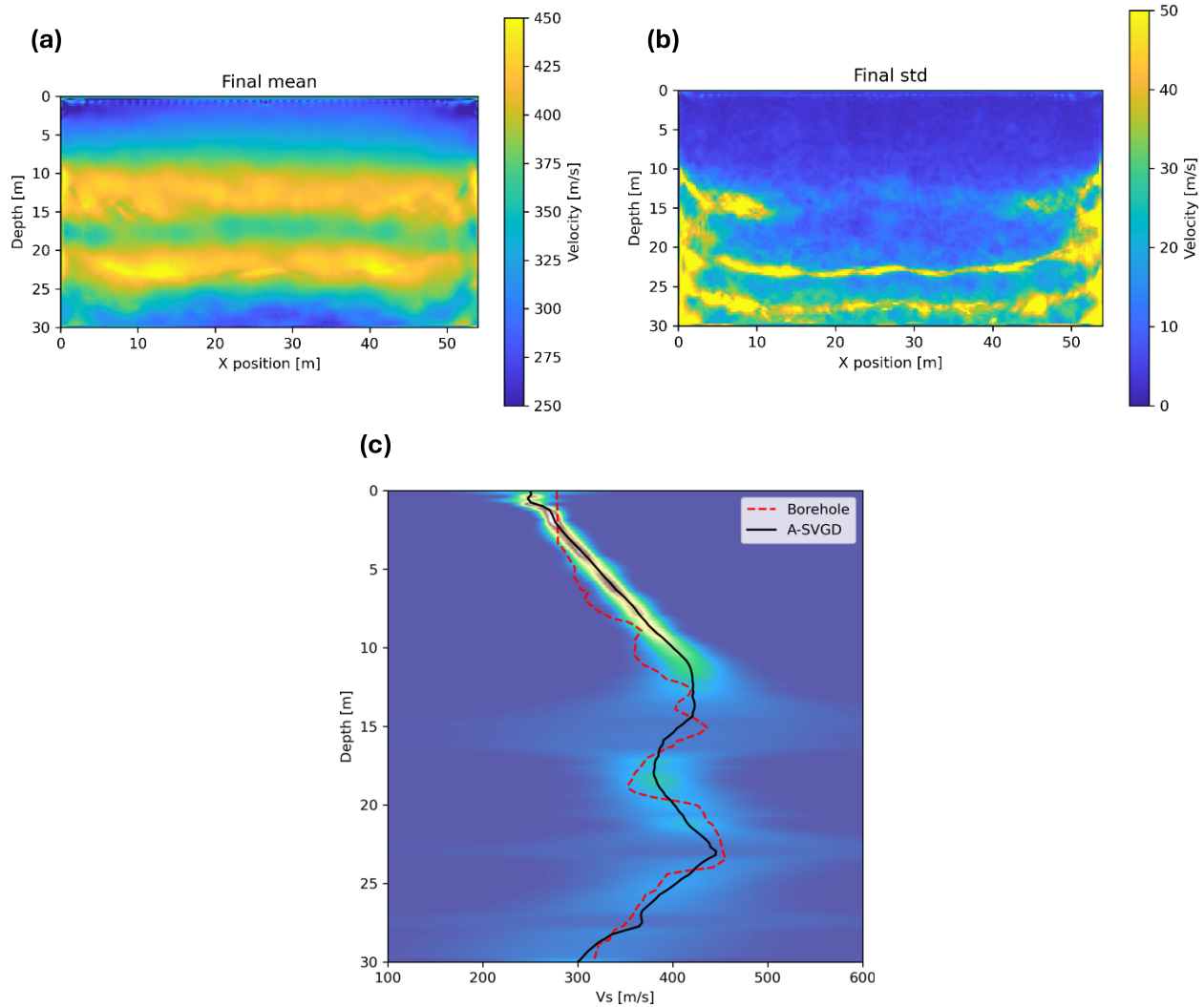


Figure 11. (a) Final mean V_S model computed using all 250 particles; (b) Final standard deviation map for the V_S model; (c) Marginal distribution at the spatial position of 10 m (cold colors correspond to zero probability while hot colours correspond to high probability) along with the mean of the available borehole data. The 1-D velocity profile extracted from the final mean V_S model is displayed in black.

the synthetic test) and the number of iterations at 500, with the inversion process taking 19 hr of computing time using an NVIDIA A100 GPU. The hyperbolic tangent was again employed as annealing schedule (eq. 8), fixing $\alpha = 1$ for the last 20 per cent of iterations.

Figs 11(a) and (b) show the final mean V_S model calculated using the final ensemble of particles and the corresponding standard deviation map, respectively.

From the mean V_S model, we can distinguish a shallow layer characterized by velocity values around $300\text{--}350\text{ m s}^{-1}$, two high-velocity layers (above 400 m s^{-1}) between $10\text{--}15$ and $20\text{--}25$ m of depth, another low-velocity layer at the bottom of the model and a velocity inversion between 15 and 20 m of depth.

The standard deviation map reveals a typical uncertainty distribution, with very small values in the first 10 m, where we have a good data coverage and illumination, and higher values (up to 50 m s^{-1}) towards the bottom and the edges of the model. We can also observe higher uncertainty in correspondence of the two high-velocity layers and at the interfaces between the different layers visible in the mean model.

Validation against the available borehole data confirms the accuracy of the velocity profile extracted from the final mean V_S model at the spatial position of 10 m, which corresponds to the location of the borehole. Fig. 11(c) illustrates the accurate detection of the main velocity variations visible in the borehole data, including the two high-velocity layers at depths of $10\text{--}15$ m and $20\text{--}25$ m, with V_S around 420 m s^{-1} . Additionally, a thinner layer (approximately 5 m thick) characterized by a lower V_S around 390 m s^{-1} , is also identified. Notably, the velocity inversion at a depth of 25 m, near the bottom edge of the model, is successfully captured, showing V_S values around 300 m s^{-1} .

Figs 12(a) and (c) compare the leftmost shot of the observed data with the data computed using one particle from the initial ensemble (specifically, the one shown in Fig. 10a) and the predicted data generated from the final mean model of Fig. 11(a). Significant discrepancies are evident between the observed and initial data, with many cycle-skipped events. These differences almost disappear when comparing the observed data with the A-SVGD predicted data, further demonstrating the robustness and reliability of the A-SVGD method.

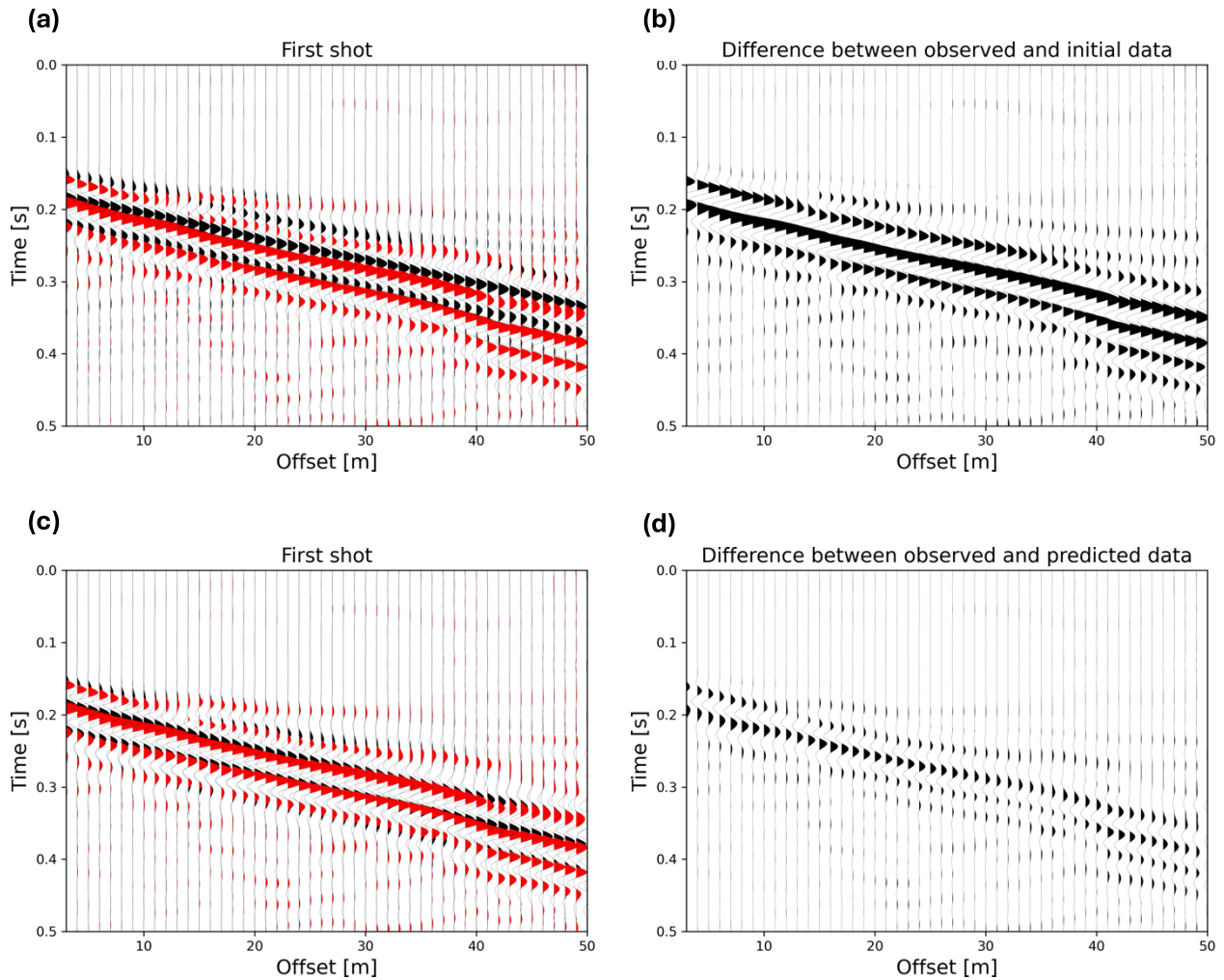


Figure 12. (a, b) Comparison between the leftmost shot of the observed data (in red) and the initial data, calculated using the initial particle of Fig. 10(a) (in black), and their sample-by-sample difference; (c, d) Comparison between the same shot of the observed data (in red) and the predicted data, calculated using the final mean model of Fig. 11(a) (in black) and their sample-by-sample difference.

Comparison with local inversion

In this section, we present the results obtained using the L-BFGS method and compare them with the outcomes achieved through the proposed approach. To ensure a fair comparison, we employed the same forward modelling engine (Deepwave) and used the initial particle shown in Fig. 10(a) as starting model for the deterministic inversion. We performed a total of 150 iterations, allowing a maximum of 5 iterations per optimization step. The inversion took 70 min, using the same NVIDIA A100 GPU employed for the A-SVGD test.

Fig. 13(a) shows the predicted V_S model, which exhibits reasonable velocity magnitudes compared to those obtained with the A-SVGD method. Additionally, a higher velocity region, with values exceeding 400 m s^{-1} , can still be identified between depths of 10 and 25 m. However, the features of the L-BFGS inversion result lack lateral continuity and display numerous low-velocity anomalies distributed across the entire model.

Figs 13(b) and (c) show the comparison between first shot of the predicted data (in black) and the same shot of the observed data (in red), along with their sample-by-sample difference. Although the misfit has been significantly reduced compared to the

one of the initial data shown in Fig. 12(b), it remains larger than the misfit associated with the predicted data calculated using the final mean model obtained from the A-SVGD inversion (Fig. 12d).

For a better quantitative assessment of the inversion results, we calculated the RPE considering all three shots, using the observed data after the pre-processing steps as the reference. We compared the RPE values obtained using the data computed from the initial model of Fig. 10(a), the data calculated using the final mean model of the A-SVGD inversion (Fig. 11a) and the one computed from the L-BFGS predicted model shown in Fig. 13(a). The initial RPE value was 59.6 per cent, which was significantly reduced by both inversion approaches, reaching 36.3 per cent for the L-BFGS inversion and 29.2 per cent for the A-SVGD approach.

This comparison highlights that the A-SVGD method delivers superior results in terms of both the quality of the predicted model and the data misfit, outperforming advanced local inversion techniques such as L-BFGS, even when applied to field data sets. Furthermore, the A-SVGD algorithm provides the added advantage of generating reliable uncertainty estimations while requiring only a limited number of particles.

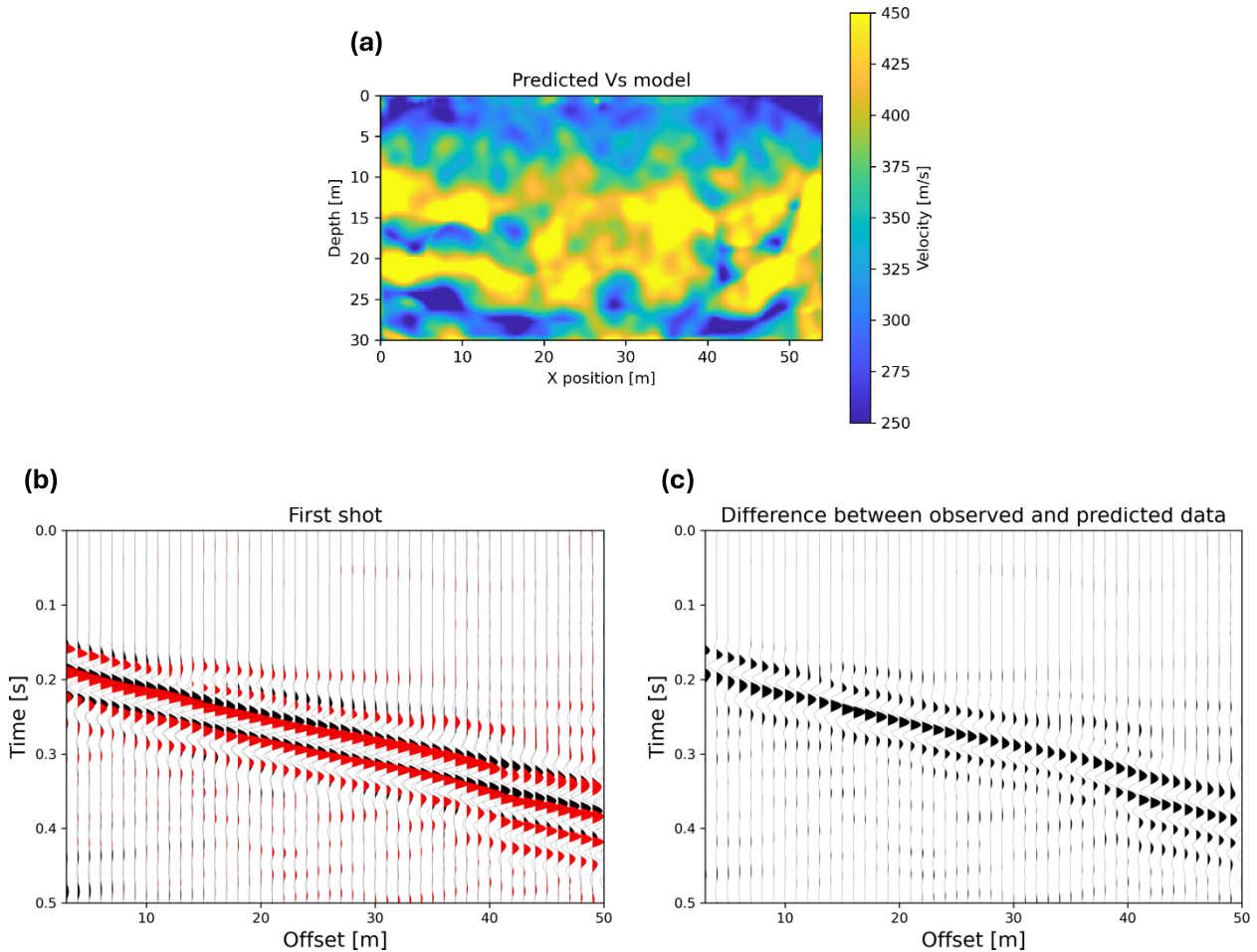


Figure 13. (a) Predicted V_S model after 150 iterations using the L-BFGS algorithm; (b, c) First shot calculated using the L-BFGS predicted model and its sample-by-sample difference with the observed data of Fig. 9, respectively.

Comparison with standard SVGD

In this section, we illustrate the results obtained with the standard SVGD method and compare them with those obtained using the proposed approach. To ensure a fair comparison, we employed the same forward modelling engine and used the same initial ensemble of particles adopted for the A-SVGD inversion.

The posterior mean V_S model, calculated using all particles at the final SVGD iteration, exhibits a structure similar to the result obtained from the A-SVGD inversion (compare Figs 14a and 11a). It reveals two high-velocity layers, a velocity inversion at a depth of 15–20 m, and another low-velocity layer at the bottom of the model. The standard deviation map (Fig. 14b) is comparable to that estimated by the A-SVGD inversion (Fig. 11b), although it exhibits slightly lower magnitudes. This underestimation of uncertainty in the SVGD results can once again be attributed to the variance collapse issue, which is effectively mitigated by the A-SVGD method.

Figs 14(c) and (d) illustrate the observed and predicted data for the first shot, along with their sample-by-sample difference. The SVGD inversion achieves a good data matching, comparable to the one obtained with the A-SVGD.

To quantitatively compare the two inversion results, we again calculated the RPE considering all three shots, using observed data after the pre-processing steps as the reference. We obtained very similar results, reaching a RPE value of 31.1 per cent for the SVGD

inversion, slightly higher than the 29.2 per cent obtained with the A-SVGD approach.

This comparison highlights that the A-SVGD method achieves slightly superior results in terms of the quality of the predicted model, data misfit and uncertainty quantification compared to the standard SVGD algorithm, even in the context of field data set inversion.

DISCUSSION

In this work, we presented the results of a variant of the standard SVGD method applied to elastic FWI of surface waves, for both synthetic and real data sets.

Our proposed approach belongs to the class of VI methods, which approximate the intractable posterior distribution by turning the problem into an optimization task. Within this class of techniques, SVGD stands out for its flexibility, enabling optimization problems to be solved through standard gradient descent algorithms for a certain number of particles, while introducing interparticle communication.

In recent years, SVGD has emerged as a valuable tool for addressing the FWI problem by approximating the posterior distribution using a limited set of particles. However, the optimal condition for applying standard SVGD is that the number of particles must be

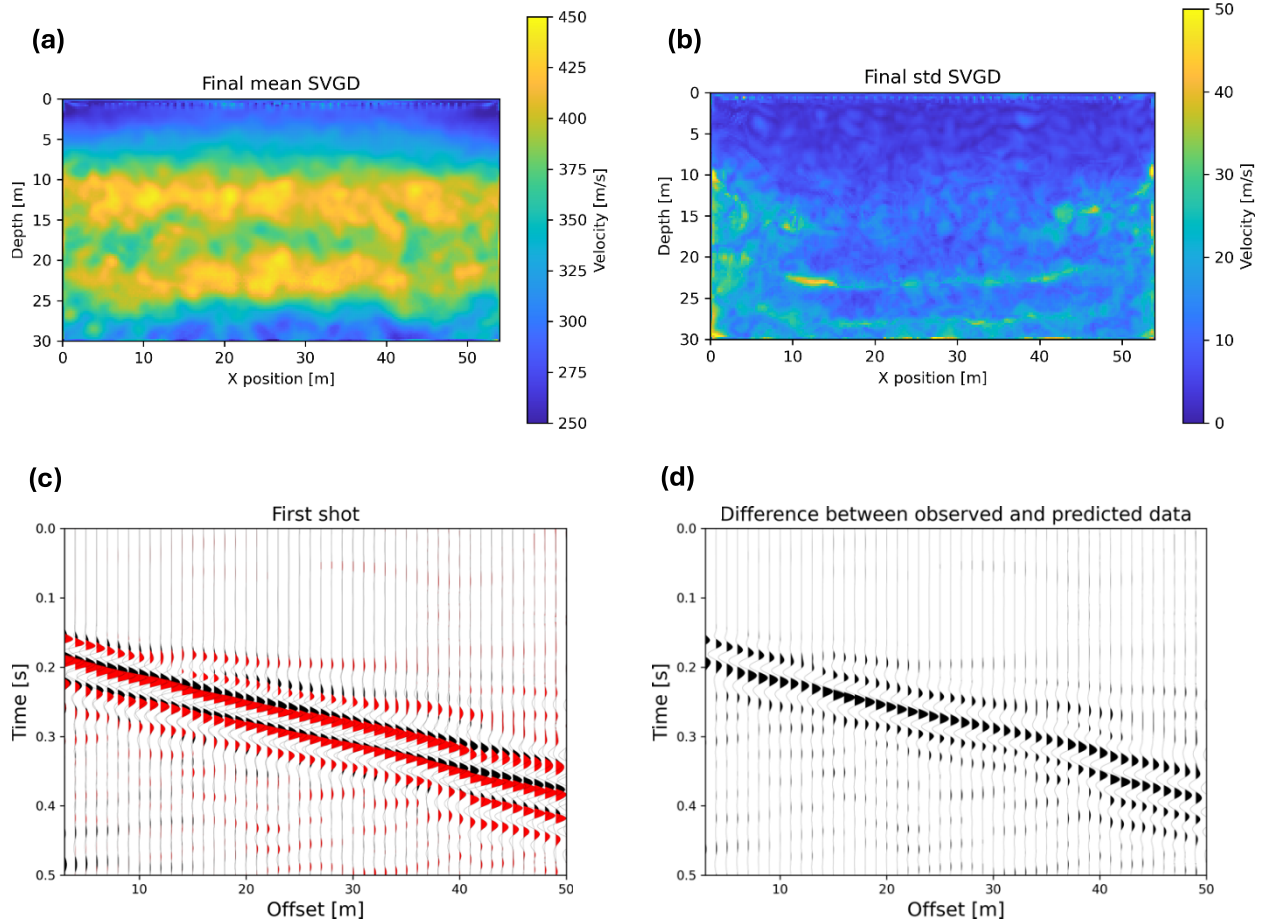


Figure 14. Results obtained using the SVGD approach: (a, b) Final mean V_S model computed using all 250 particles and the final standard deviation map, respectively; (c, d) First shot calculated using the final mean model obtained using the SVGD method (black) superimposed to the observed shot gather (red) and their sample-by-sample difference, respectively.

equal or larger than the number of unknowns, which is unfeasible in large-scale inverse problems like elastic FWI. Under a limited regime of particles, it has been demonstrated that SVGD suffers of two main challenges, specifically the mode and variance collapse issues. Trying to mitigate these issues, we adopted an annealed variant of SVGD, called A-SVGD, that allows a more efficient exploration of the model space by incorporating an annealing schedule in the particle update.

We have demonstrated the efficacy of this method in solving the elastic FWI problem, even when starting from initial particles far from the optimal solution and when dealing with a field data set. Additionally, A-SVGD also provides reasonable uncertainty estimations with a limited computational cost, particularly when compared to MCMC sampling techniques.

Based on the results reported by Berti *et al.* (2024b), who applied a gradient-based MCMC algorithm to the same field data set and achieved comparable outcomes, we observe that their computation time on the same GPU was 31 hr, compared to 19 hr required for the A-SVGD inversion. Furthermore, their MCMC algorithm operated in a DCT-compressed domain, which limited the resolution of their predictions, whereas the A-SVGD method operates in the full space.

There are opportunities to further improve the inversion performance, by employing, for example, different kernel functions (IMQ; Gorham & Mackey 2017) or different annealing schedules (cyclic

schedule; D'Angelo & Fortuin 2021) that can be found in literature, or incorporating some prior information into the inversion framework.

Another promising direction is to apply the SVGD algorithm (or its annealed variant) in a reduced or projected space, which would reduce the number of unknowns in the inverse problem. This approach would align with the ideal conditions for SVGD, where the number of particles equals or exceeds the number of unknowns. It holds the potential to significantly enhance the application of SVGD in FWI, improving both the accuracy of the inversion and the reliability of uncertainty quantification.

CONCLUSIONS

We presented a probabilistic FWI of surface waves using the annealed variant of the standard SVGD algorithm.

The incorporation of the hyperbolic tanh as an annealing schedule in the particle update induces two distinct phases in our inversion process: a first exploratory phase where the repulsive force of the particle update is dominant, resulting in a dispersion of the particles from their starting positions and enhancing the exploration of the model space; a second, exploitative phase, where the driving force predominates, concentrating particles distribution around different modes.

This modification effectively mitigates the mode and variance collapse issues that typically affect standard SVGD algorithms when applied to large-scale inverse problems like elastic FWI.

The results obtained in the synthetic case validate the applicability and reliability of our approach which, unlike conventional deterministic inversion approaches, provides a comprehensive assessment of uncertainties affecting the estimated solution, with a reasonable computational cost. The application to a real data set shows a good data matching between predicted and observed data and an accurate reproduction of the main velocity variations visible in the available borehole data.

The proposed inversion method demonstrates convergence towards reliable and plausible solutions in both synthetic and real-world tests, outperforming advanced deterministic approaches and standard SVGD inversion. Additionally, it provides accurate uncertainty estimations for the recovered velocity model while achieving higher computational efficiency compared to MCMC techniques.

ACKNOWLEDGMENTS

This research was partially conducted at King Abdullah University of Science and Technology (KAUST). This paper has been supported from Italian Ministerial grant PRIN 2022 ‘Joining Exploration and Earthquake Seismology for Velocity Model Building and Seismogenic Process Characterization’, CUP I53C24002650006.

DATA AVAILABILITY

The data presented in this work are available on reasonable request to the corresponding authors. The Deepwave package used in this work is freely available and can be found at the following link: <https://ausargeo.com/deepwave/>.

REFERENCES

- Aleardi, M. 2019. Using orthogonal Legendre polynomials to parameterize global geophysical optimizations: applications to seismic-petrophysical inversion and 1D elastic full-waveform inversion, *Geophys. Prospect.*, **67**(2), 331–348.
- Aleardi, M. 2020. Discrete cosine transform for parameter space reduction in linear and non linear AVA inversions, *J. Appl. Geophys.*, **179**, 104106. doi:10.1016/j.jappgeo.2020.104106
- Aleardi, M. & Salusti, A. 2020. Hamiltonian Monte Carlo algorithms for target- and interval oriented amplitude versus angle inversions, *Geophysics*, **85**(3). doi:10.1190/geo2019-0517.1
- Ba, J., Erdogdu, M.A., Ghassemi, M., Sun, S., Suzuki, T., Wu, D. & Zhang, T. 2022. Understanding the Variance Collapse of SVGD in High Dimensions, *International Conference on Learning Representations*.
- Berti, S., Aleardi, M. & Stucchi, E. 2023. A computationally efficient Bayesian approach to full waveform inversion, *Geophys. Prospect.*, **72**(2), 580–603.
- Berti, S., Aleardi, M. & Stucchi, E. 2024a. A Bayesian approach to elastic full-waveform inversion: application to two synthetic near surface models, *Bull. Geophys. Oceanogr.*, **65**(2), 291–308.
- Berti, S., Aleardi, M. & Stucchi, E. 2024b. A probabilistic full waveform inversion of surface waves, *Geophys. Prospect.*, **72**, 3448–3473.
- Berti, S., Ravasi, M., Aleardi, M. & Stucchi, E. 2024c. Full waveform inversion of surface waves with Annealed Stein Variational Gradient Descent, in *NSG 2024 30th European Meeting of Environmental and Engineering Geophysics*, pp. 1–5, NSG, Helsinki, Finland.
- Biondi, E., Stucchi, E. & Mazzotti, A. 2014. Nonstretch normal move-out through iterative partial corrections and deconvolution, *Geophysics*, **79**(4), V131–V141.
- Blei, D.M., Kucukelbir, A. & McAuliffe, J.D. 2017. Variational inference: a review for statisticians, *J. Am. Stat. Assoc.*, **112**(518), 859–877.
- Boaga, J., Cassiani, G., Strobbia, C. & Vignoli, G. 2013. Mode misidentification in Rayleigh waves: ellipticity as a cause and a cure, *Geophysics*, **78**(4), EN17–EN28.
- 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.
- Bunks, C., Saleck, F.M., Zaleski, S. & Chavent, G. 1995. Multiscale seismic waveform inversion, *Geophysics*, **60**, 1457–1473.
- Corrales, M., Berti, S., Denel, B., Williamson, P., Ravasi, M. & Aleardi, M. 2024. Annealed Stein Variational Gradient Descent for improved uncertainty estimation in full-waveform inversion, *arXiv, eprint:2410.13249*.
- Curtis, A. & Lomax, A. 2001. Prior information, sampling distributions and the curse of dimensionality, *Geophysics*, **66**(2), 372–378.
- D’Angelo, F. & Fortuin, V. 2021. Annealed Stein Variational Gradient Descent, *3rd Symposium on Advances in Approximate Bayesian Inference*, 1–12.
- Datta, D. & Sen, M.K. 2016. Estimating a starting model for full-waveform inversion using a global optimization method, *Geophysics*, **81**(4), R211–R223.
- Fichtner, A. & Simute, S. 2018. Hamiltonian Monte Carlo inversion of seismic sources in complex media, *J. geophys. Res.: Solid Earth*, **123**(4), 2984–2999.
- Forbriger, T., Groos, L. & Schafer, M. 2014. Line-source simulation for shallow-seismic data. Part 1: theoretical background, *Geophys. J. Int.*, **198**(3), 1387–1404.
- Garofalo, F. 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 Dyn. Earthq. Eng.*, **82**, 222–240.
- Gebräud, L., Boehm, C. & Fichtner, A. 2020. Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo, *J. geophys. Res.: Solid Earth*, **125**(3), doi:10.1029/2019JB018428
- Gelis, C., Virieux, J. & Grandjean, G. 2007. Two-dimensional elastic waveform inversion using Born and Rytov formulations in the frequency domain, *Geophys. J. Int.*, **168**, 605–633.
- Gorham, J. & Mackey, L. 2017. Measuring Sample Quality with Kernels, in *Proceedings of the 34th International Conference on Machine Learning*, pp. 1292–1301, eds Precup, D. & Teh, Y. W., ICML, Sydney
- 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.
- Guo, P., Visser, G. & Saygin, E. 2020. Bayesian trans-dimensional full waveform inversion: synthetic and field data application, *Geophys. J. Int.*, **222**(1), 610–627.
- Izzatullah, M., Alali, A., Ravasi, M. & Alkhalifah, T. 2024. Physics-reliable frugal local uncertainty analysis for full waveform inversion, *Geophys. Prospect.*, **72**(2), 2718–2738.
- Izzatullah, M., Ravasi, M. & Alkhalifah, T. 2023. Frugal uncertainty analysis for full waveform inversion, in *84th EAGE Annual Conference and Exhibition*, EAGE, Vienna, Austria, pp. 1–5.
- Kirlin, R.L. & Done, W.J. 1999. *Covariance Analysis for Seismic Signal Processing*, SEG.
- Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A. & Blei, D.M. 2017. Automatic differentiation variational inference, *J. Mach. Learn. Res.*, **18**(1), 430–474.
- Kullback, S. & Leibler, R.A. 1951. On Information and Sufficiency, *Ann. Math. Stat.*, **22**(1), 79–86.
- Kumar, R., Kotsi, M., Siahkoobi, A. & Malcolm, A. 2021. Enabling uncertainty quantification for seismic data preprocessing using normalizing flows (NF)—An interpolation example, *First International Meeting for Applied Geoscience & Energy Expanded Abstracts, SEG Technical Program Expanded Abstracts*, 1515–1519.

- Lailly, P. 1984. *Migration Methods: Partial but Efficient Solutions to the Seismic Inverse Problem: Inverse Problems of Acoustic and Elastic Waves*, Society of Industry and Applied Mathematics.
- 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 Surf. Geophys.*, **21**, 21–38. doi:10.1002/nsg.12243
- Liu, D.C. & Nocedal, J. 1989. On the limited memory BFGS method for large scale optimization, *Math. Program.*, **45**(1), 503–528.
- Liu, Q. & Wang, D. 2016. Stein Variational Gradient Descent: a General Purpose Bayesian Inference Algorithm, in *Advances in Neural Information Processing Systems*, vol. **29**, Curran Associates, Inc.
- Lomas, A., Luo, S., Irakarama, M., Johnston, R., Vyas, M. & Shen, X. 2023. 3D probabilistic full waveform inversion: application to Gulf of Mexico field data, in *84th EAGE Annual Conference and Exhibition*, EAGE, Vienna, Austria, pp. 1–5.
- Malinverno, A. 2002. Parsimonious Bayesian Markov Chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.*, **151**, 675–688.
- Martin, J., Wilcox, L.C., Burstedde, C. & Ghattas, O. 2012. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion, *SIAM J. Sci. Comput.*, **34**, A1460–A1487.
- Mosegaard, K. & Tarantola, A. 2002. Probabilistic approach to inverse problems, *Int. Geophys. Ser.*, **81**, 237–268.
- Nawaz, M.A. & Curtis, A. 2018. Variational Bayesian inversion (VBI) of quasi-localized seismic attributes for the spatial distribution of geological facies, *Geophys. J. Int.*, **214**(2), 845–875.
- Nazarian, S., Stokoe, K.H. & Hudson, W.R. 1983. Use of spectral analysis of surface waves method for determination of moduli and thickness of pavement system, *Transp. Res. Rec.*, **930**, 38–45.
- Pan, Y., Gao, L. & Bohlen, T. 2019. High-resolution characterization of near-surface structures by surface-wave inversions: from dispersion curve to full waveform, *Surv. Geophys.*, **40**, 167–195.
- Park, C.B., Miller, R.D. & Xia, J., 1999. Multichannel analysis of surface waves, *Geophysics*, **64**(3), 800–808.
- Paszke, A. et al. 2019. PyTorch: an Imperative Style, High-Performance Deep Learning Library, *33rd Conference on Neural Information Processing Systems*, Vancouver, Canada.
- Pe' rez Solano, C., Donno, D. & Chauris, H., 2014. Alternative waveform inversion for surface wave analysis in 2-D media, *Geophys. J. Int.*, **198**(3), 1359–1372.
- 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, Vol. 20*, EAGE, pp. 1–4.
- Ray, A., Sekar, A., Hoversten, G.M. & Albertin, U. 2016. Frequency domain full waveform elastic inversion of marine seismic data from the Alba field using a Bayesian trans-dimensional algorithm, *Geophys. J. Int.*, **205**(2), 915–937.
- Rezende, D.J. & Mohamed, S. 2015. Variational Inference with Normalizing Flows, *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France.
- Richardson, A. 2022. Deepwave. Available at: <https://ausargeo.com/deepwave/>.
- Rincón, F., Berti, S., Aleardi, M., Tognarelli, A. & Stucchi, E. 2025. Integrating deep learning and discrete cosine transform for surface waves full-waveform inversion, *Geophys. J. Int.*, **240**(1), 805–828.
- 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., Stucchi, E. & Galuzzi, B. 2014. Comparison of stochastic optimization methods on two analytic objective functions and on a 1D elastic FWI, in *76th EAGE Conference and Exhibition 2014*, EAGE, Amsterdam.
- Sambridge, M. & Mosegaard, K. 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.*, **40**(3), 3–1.
- Sen, M.K. & Stoffa, P.L. 2013. *Global Optimization Methods in Geophysical Inversion*. Cambridge Univ. Press.
- Siahkoochi, A., Rizzuti, G., Louboutin, M., Witte, P.A. & Herrmann, F.J. 2021. Preconditioned training of normalizing flows for variational inference in inverse problems, *3rd Symposium on Advances in Approximate Bayesian Inference*, 1–17.
- Siahkoochi, A., Rizzuti, G., Orozco, R. & Herrmann, F.J. 2022. Reliable amortized variational inference with physics-based latent distribution correction, *arXiv, eprint:2207.11640*.
- Tarantola, A. 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**, 1259–1266.
- 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.
- Wang, D., Tang, Z., Bajaj, C. & Liu, Q. 2019. Stein Variational Gradient Descent with matrix-valued kernels, *33rd Conference on Neural Information Processing Systems*, Vancouver, Canada.
- Wu, R.S., Luo, J. & Wu, B. 2014. Seismic envelope inversion and modulation signal model, *Geophysics*, **79**(3), WA13–WA24.
- Xia, J., Miller, R.D. & Park, C.B., 1999. Estimation of near-surface shear-wave velocity by inversion of Rayleigh waves, *Geophysics*, **64**(3), 691–700.
- 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, C. & Chen, T. 2022. Bayesian slip inversion with automatic differentiation variational inference, *Geophys. J. Int.*, **229**(1), 546–565.
- Zhang, S.X. & Chan, L.S. 2003. Possible effects of misidentified mode number on Rayleigh wave inversion, *J. Appl. Geophys.*, **53**, 17–29.
- Zhang, X. & Curtis, A. 2020a. Seismic tomography using variational inference methods, *J. geophys. Res.: Solid Earth*, **125**, e2019JB018589. doi:10.1029/2019JB018589
- Zhang, X. & Curtis, A. 2020b. Variational full-waveform inversion, *Geophys. J. Int.*, **222**(1), 406–411.
- Zhang, X. & Curtis, A. 2021. Bayesian full-waveform inversion with realistic priors, *Geophysics*, **86**(5), A45–A49.
- Zhang, X., Lomas, A., Zhou, M., Zheng, Y. & Curtis, A. 2023. 3-D Bayesian variational full waveform inversion, *Geophys. J. Int.*, **234**(1), 546–561.
- Zhao, X., Curtis, A. & Zhang, X. 2022. Bayesian seismic tomography using normalizing flows, *Geophys. J. Int.*, **228**(1), 213–239.
- 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.
- Zhuo, J., Liu, C., Shi, J., Zhu, J., Chen, N. & Zhang, B. 2018. Message Passing Stein Variational Gradient Descent, in *Proceedings of the 35th International Conference on Machine Learning*, pp. 6018–6027, eds Dy, J. & Krause, A., PMLR, Stockholm.
- Zunino, A., Ghirrotto, A., Armadillo, E. & Fichtner, A. 2022. Hamiltonian Monte Carlo probabilistic joint inversion of 2D (2.75D) gravity and magnetic data, *Geophys. Res. Lett.*, **49**(20). doi:10.1029/2022GL099789