Bulletin of Geophysics and Oceanography Vol. 67, n. 2, pp. 381-404; June 2026
DOI 10.4430/bgo00515

Probabilistic amplitude-versus-angle inversion via annealed Stein variational gradient descent, automatic differentiation, and model-space compression

S. Berti, F. Rincón and M. Aleardi

Earth Sciences Department, University of Pisa, Pisa, Italy

(Received: 16 July 2025; accepted: 19 November 2025; published online: 20 February 2026)

ABSTRACT Amplitude-versus-angle (AVA) inversion is a widely used geophysical technique for estimating subsurface elastic properties from seismic reflection data. However, its ill-posed nature necessitates robust and efficient methods for uncertainty quantification. This work presents a novel Bayesian AVA inversion framework that integrates annealed Stein variational gradient descent (A-SVGD) with discrete cosine transform (DCT) compression of the model space. A-SVGD introduces an annealing schedule to enhance posterior exploration, while DCT significantly reduces problem dimensionality, improving stability and convergence. The inversion employs full Zoeppritz equations as the forward model and leverages automatic differentiation for efficient and accurate gradient computation. Synthetic tests demonstrate that A-SVGD outperforms standard SVGD in convergence speed and uncertainty quantification. The DCT-based model-space compression achieves performance comparable to full-domain inversion, but at a fraction of the computational cost. Further tests under realistic conditions (e.g. where errors are introduced in both the estimated source wavelet and noise statistics) highlight the robustness of the proposed approach. Additionally, replacing the Jacobian with a DCT-projected analytical approximation shows promise for accelerating computation, though its applicability is model-dependent. Overall, this study demonstrates that A-SVGD in a DCT-compressed domain provides a powerful and efficient solution for probabilistic AVA inversion and achieves final predictions comparable with those obtained with a gradient-based Markov chain Monte Carlo method.

Key words: AVA inversion, probabilistic inversion, automatic differentiation.

1. Introduction

Amplitude-versus-angle (AVA) inversion is a well-established geophysical technique for estimating the elastic properties of the subsurface [such as P-wave velocity (*Vp*), S-wave velocity (*Vs*), and density] from pre-stack seismic reflection data. Due to the nonlinear and ill-posed nature of AVA inverse problems, solutions are highly sensitive to measurement noise, limited data coverage, and inaccuracies in the forward model (Aleardi, 2015; Aleardi and Tognarelli, 2016; Moghanloo *et al*., 2018; Dai *et al*., 2022). In this context, an accurate uncertainty estimation on the inversion results is crucial. However, an analytical and fast uncertainty assessment is possible only in very specific conditions: Gaussian assumptions for model parameters, data noise and linear forward operator. Consequently, extensive research has focused on formulating nonlinear

AVA inversion within robust, efficient, and accurate probabilistic frameworks. The Bayesian approach addresses this challenge by rigorously treating subsurface parameters as random variables and expressing the solution in terms of posterior probability distribution (Tarantola, 2005). This enables not only point estimates but also the quantification of uncertainty and model plausibility. Seminal work by Buland and Omre (2003) demonstrated the feasibility of linearised Bayesian AVA inversion under Gaussian assumptions, leading to closed-form posterior distributions and analytical credibility intervals. Further developments have extended the Bayesian AVA framework to include more complex rock physics models, petrophysical facies, and non-Gaussian priors (Grana and Mukerji, 2015; Azevedo *et al*., 2019; Guo *et al*., 2023).

As previously noted, nonlinear and high-dimensional inverse problems result in posterior distributions that are no longer analytically tractable and must instead be explored through numerical methods. Markov chain Monte Carlo (MCMC) algorithms, such as the Metropolis-Hastings, are widely used for this purpose as they enable asymptotically exact sampling from the posterior. However, their convergence can be prohibitively slow, particularly in high-dimensional settings or when each forward model evaluation is computationally expensive (Fernandez-Martinez and Fernandez-Muniz, 2020).

To improve efficiency, a class of gradient-based MCMC (GB-MCMC) methods has emerged. These algorithms leverage the gradient of the posterior to propose informed transitions that explore the parameter space more effectively (Zhao and Sen, 2021; Berti *et al*., 2024a, 2024b). Hamiltonian Monte Carlo (Fichtner *et al*., 2019) introduces auxiliary momentum variables and simulates Hamiltonian dynamics to generate distant, low-rejection proposals with high acceptance rates. Stochastic gradient Langevin dynamics (Brosse *et al*., 2018) and its variants approximate Langevin diffusion using stochastic mini-batches and noise injection, making them scalable to large datasets. These methods greatly benefit from the accurate and efficient computation of posterior gradients, which are crucial for maintaining sampler efficiency and stability.

In parallel, variational inference (VI) (Blei *et al*., 2017) approaches have been developed as an alternative to MCMC for large-scale Bayesian inference problems. Among these, Stein variational gradient descent (SVGD) (Liu and Wang, 2016) and its annealed extension [annealed Stein variational gradient descent (A-SVGD): D'Angelo and Fortuin (2021)] offer a powerful non-parametric framework for approximating the posterior. These methods evolve a set of particles along trajectories guided by posterior gradients, with A-SVGD employing a temperature schedule to improve exploration of multimodal distributions (Berti *et al*., 2024c, 2025; Corrales *et al*., 2025). However, like GB-MCMC, the applicability of SVGD methods critically depend on accurate and computationally efficient gradients evaluations.

Posterior-gradient information can be obtained in several ways. The most direct, but computationally expensive, approach is to approximate the Jacobian matrix using finite differences, requiring at least one forward simulation parameter for mode with a large number of unknowns, this results in a prohibitive computational cost. To address this challenge, we adopt automatic differentiation [AD: Beda *et al*. (1959); Corliss *et al*. (2002)], a technique widely used in machine learning frameworks such as PyTorch, TensorFlow, and JAX which enables the exact computation of derivatives through complex computational graphs at a limited computational cost, much lower than that associated with finite-difference (FD) computation. AD enables efficient and accurate gradient evaluation for the full posterior, including the contributions from the forward reflectivity model, regularisation/prior terms, and any model reparameterisation, without requiring manual and awkward derivation of adjoint operators (Sambridge *et al*., 2007; Liu *et al*., 2025).

To further reduce computational complexity in Bayesian nonlinear inversion, various model parameterisation techniques can be employed, including Legendre polynomials (Aleardi, 2019), wavelet transforms (Fu *et al*., 2024), convolutional autoencoders (Qian *et al*., 2018), and implicit neural representations (Sitzmann *et al*., 2020; Sun *et al*., 2023). In this work we incorporate model-space compression via the discrete cosine transform [DCT: Aleardi (2020)]. This reparameterisation enables a sparse and efficient representation of spatially smooth parameter fields, reducing the number of unknowns and improving stability (Berti *et al*., 2024a; Rincón *et al*., 2025). The reduced number of parameters not only accelerates the inversion process but also decreases the number of particles required for robust uncertainty quantification.

In this study, we propose a novel Bayesian AVA inversion framework that integrates:

1. a compressed parameter space using one dimensional (1D) DCT to reduce dimensionality and mitigate ill-posedness; this reduction enables us to easily meet the theoretical requirements of SVGD approaches in which the number of particles equals the number of model parameters (Ba *et al*., 2022). In this context, the DCT inherently reduces the number of forward evaluations with respect to the inversion working in the full model space;
2. the A-SVGD algorithm for efficient posterior approximation in high-dimensional spaces;
3. the use of AD to efficiently compute exact posterior gradients, in place of costly finite-difference Jacobians.

We validated our approach on synthetic AVA inversion problems constructed from a realistic subsurface model and also compared the proposed method with the predictions of a GB-MCMC sampling. The results demonstrate the capability of the proposed method to recover elastic parameters with high accuracy, quantify uncertainty in a robust manner, and mitigate the computational burden typically associated with nonlinear Bayesian AVA inversion.

2. Methods

2.1. Discrete cosine transform

DCT is a powerful tool for signal representation and compression, as it transforms spatial domain data into frequency domain, concentrating most of the signal's energy into a small number of coefficients. Of the available DCT variants, we adopt type-II DCT (DCT-II), the version most commonly used in geophysics and signal processing (Lochbuhler *et al*., 2014). From this point forward, we refer to DCT-II simply as DCT. In this work we apply the 1D DCT to each elastic property profile associated to a specific common mid-point (CMP) location.

The 1D DCT transformation of a 1D signal x of length P, can be written as follows:

$$y(k) = \sqrt{\frac{2}{P}} \sum_{n=1}^{P} x(n) \frac{1}{\sqrt{1+\delta_{k1}}} \cos\left(\frac{\pi}{2N}(2n-1)(k-1)\right) \quad (1)$$

where δ_{k1} represents the Kronecker delta, y is the set of P coefficients of the transformation that describe original signal x in the compressed space, and k represents the order of each DCT coefficient. In matrix form, Eq. (1) becomes:

$$\mathbf{y} = \mathbf{B}\mathbf{x} \quad (2)$$

where vectors **y** and **x** represent the transformed and the original signal, respectively, while **B** is a *P*-by-*P* matrix expressing the basis DCT functions. The DCT concentrates most of the energy of original signal **x** in the lower order DCT coefficients. This means that an approximation of signal **x** can be obtained by considering only the first *q* DCT base functions (*q*<*P*):

$$\tilde{\mathbf{x}} = \mathbf{B}_q^T \mathbf{y}_q \tag{3}$$

where $\tilde{x}$ is the approximated signal, B_q^T is a partition of matrix *B*, with *P* rows and *q* columns representing the first *q* DCT base functions, whereas vector y_q contains the first *q* coefficients multiplying the cosine functions. These coefficients are the unknowns to be determined in an inverse problem running in a DCT compressed model space.

2.2. Annealed Stein variational gradient descent

Bayesian inversion relies on Bayes' theorem to formally characterise uncertainty in model parameters through posterior probability distribution:

$$p(\mathbf{m}|\mathbf{d}_{obs}) = \frac{p(\mathbf{d}_{obs}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d}_{obs})}. \tag{4}$$

In this expression, $p(\mathbf{m})$ denotes the prior distribution, which reflects existing knowledge about the model parameters. Data likelihood $p(\mathbf{d}_{obs} \mid \mathbf{m})$ measures how well model *m* explains observed data $\mathbf{d}_{obs}$, and denominator $p(\mathbf{d}_{obs})$, called the evidence, acts as a normalising constant. When direct computation of the posterior becomes intractable due to the complexity of the required integrals, VI serves as a practical alternative. VI attempts to approximate the posterior minimising the Kullback-Leibler divergence (Kullback and Leibler, 1951), which measures the differences between a candidate distribution and the target posterior. Among others, SVGD stands out for its flexibility, trying to approximate the posterior distribution by iteratively updating a set of particles, $\{m_i\}_{i=1}^{Npar}$, without requiring a parametric form for the candidate surrogate distribution. Considering a finite set of particles, the updating rule for the *i*-th particle can be written as follows:

$$\mathbf{m}_i^{l+1} = \mathbf{m}_i^l + \epsilon_l \phi^*\left(\mathbf{m}_i^l\right) \tag{5}$$

with

$$\phi^*\left(\mathbf{m}_i^l\right) = \frac{1}{N_{par}} \sum_{j=1}^{N_{par}} \left[k(\mathbf{m}_j^l, \mathbf{m}_i^l) \nabla_{m_j^l} \log p\left(\mathbf{m}_j^l | \mathbf{d}_{obs}\right) + \nabla_{m_j^l} k(\mathbf{m}_j^l, \mathbf{m}_i^l)\right]$$

where ε_l is the step size and *l* represents the iteration. In optimal update direction ϕ^* we can distinguish between the first term, known as the attractive component, which directs particles towards high-posterior density regions, and the second repulsive term which preserves particle diversity.

Despite its effectiveness, SVGD can suffer from two major issues, especially when the number of unknowns exceeds the number of particles (Zhuo *et al.*, 2018): mode collapse, where particles concentrate in few regions, and variance collapse, where the posterior uncertainty is

underestimated.

To mitigate these drawbacks, we adopt the A-SVGD technique, a modified version of SVGD that introduces a temperature-like coefficient, $\alpha(l)$, to balance attraction and repulsion during optimisation, enabling a better exploration of the model space in the initial iterations:

$$\phi^*(\mathbf{m}) = \frac{1}{N_{par}} \sum_{j=1}^{N_{par}} [\alpha(l) k(\mathbf{m}_j, \mathbf{m}) \nabla_{m_j} \log p(\mathbf{m}_j | \mathbf{d}_{obs}) + \nabla_{m_j} k(\mathbf{m}_j, \mathbf{m})] \,. \tag{6}$$

This scheduling enables two distinct phases:

- exploration: when $\alpha(l)$ is close to zero, the repulsive force dominates, promoting wider particle spread;
- exploitation: as $\alpha(l)$ approaches 1, the method concentrates particles in high-probability regions.

To guarantee convergence to the true posterior, $\alpha(l)$ must asymptotically approach 1. Following D'Angelo and Fortuin (2021), we use a hyperbolic tangent to define the annealing schedule:

$$\alpha(l) = \tanh[(1.3 \frac{l}{K})^c] \tag{7}$$

where K is the total number of iterations and c is a tuneable hyperparameter controlling the transition between 0 and 1. In our experiments, we found c values ranging from 1.5 to 3 to be reasonable.

The posterior distribution is modelled under the Gaussian assumption about data and model parameter distribution:

$$\begin{aligned} p(\mathbf{m}|\mathbf{d}_{obs}) \propto \exp\,[-0.5 \times (\mathbf{d}_{obs} - g(\mathbf{m})]^T \mathbf{C}_d^{-1} [\mathbf{d}_{obs} - g(\mathbf{m})] \\ \times \exp\left[-0.5 \times \left(\mathbf{m} - \mathbf{m}_{prior}\right)^T \mathbf{C}_m^{-1} \left(\mathbf{m} - \mathbf{m}_{prior}\right)\right] \end{aligned} \tag{8}$$

where g represents the Zoeppritz equations (Aki and Richards, 1980), whereas $\mathbf{C}_m$ and $\mathbf{C}_d$ are the model and data covariance matrices, respectively. Since forward operator g is nonlinear, the posterior distribution is non-Gaussian, even when both the prior and data likelihood are Gaussian. Hence, the posterior cannot be evaluated analytically and must be approximated, for example, via numerical approaches. In our application kernel function k of Eq. (6) also incorporates information about the prior's curvature (as coded by matrix $\mathbf{C}_m$) and is defined using a radial basis function:

$$k(\mathbf{m}, \mathbf{m}') = \exp\left(-\frac{(\mathbf{m}-\mathbf{m}')^T \mathbf{C}_m^{-1} (\mathbf{m}-\mathbf{m}')}{2h^2}\right) \tag{9}$$

where h is a scale factor determining the interaction intensity among different particles. On the line of Liu and Wang (2016), h is determined as $h=\tilde{d}^2/\log N_{par}$, where $\tilde{d}$ is the median of pairwise distances between all particles.

To enhance convergence speed, step length ϵ is automatically adjusted at each iteration using the adaptive subgradient method [AdaGrad: Duchi *et al*. (2011)]. A distinctive aspect of AdaGrad is its adaptive learning rate, which automatically adjusts for each model parameter based on the cumulative gradients observed for that parameter over iterations.

The inversion runs in a DCT compressed model space where the 1D-DCT is separately applied to each elastic property associated with each CMP gather location (see Aleardi, 2020). Assuming that each CMP gather is associated to L time samples, this means that L×3 unknowns must be estimated by AVA inversion running in the full model space. Instead with the DCT, only q×3 parameters need to be inferred, with $q<<L$ denoting the number of DCT coefficients used to compress each elastic property.

The major computational requirement of the method is related to the computation of the posterior gradient for each particle at each iteration. Because our AVA inversion relies on gradients to drive A-SVGD particle updates, the gradient-estimation scheme directly impacts efficiency, convergence behaviour, and implementation complexity. Efficient and accurate evaluation of the posterior gradient with respect to model parameters is essential for modern Bayesian inversion schemes, particularly those relying on gradient-informed sampling. Depending on the approach adopted, Jacobian matrix J can be used to evaluate the gradient. A classic way to approximate the Jacobian is using forward FDs. The j-th column of the Jacobian around generic model m is approximated as:

$$\mathbf{J}_j \approx \frac{g(\mathbf{m})-g(\mathbf{m}+pe_j)}{p} \tag{10}$$

where e_j is the unit vector in the j-th direction, and p is a small perturbation.

Computing gradients via forward FDs requires several forward evaluations equal to the number of model parameters. While conceptually straightforward, this approach has two major drawbacks: 1) it yields only approximate derivatives, whose accuracy heavily depends on the choice of perturbation size, and 2) its computational cost scales linearly with the number of parameters, making it inefficient for high-dimensional problems. By contrast, AD (Baydin *et al.*, 2018) computes exact derivatives by systematically applying the chain rule through the computational graph. Unlike symbolic differentiation, AD does not require analytical expressions, and unlike FDs, it avoids truncation errors and noise amplification. Derivatives are accumulated with machine precision and, while not strictly independent of the number of parameters, reverse-mode AD enables efficient gradient computation for scalar-valued functions, with a computational cost that is typically comparable to a single model evaluation (Griewank and Walther, 2008). Modern libraries, such as PyTorch and JAX, implement reverse-mode AD efficiently, making it particularly advantageous for large-scale inverse problems where gradient evaluations are required at every iteration. This enables scalable, gradient-based inversion frameworks, such as those encountered in VI and GB-MCMC.

In this work, we adopted AD to efficiently compute the gradient of the target posterior distribution. To enable this, we reformulated the forward model by implementing the Zoeppritz equations directly in PyTorch using tensor operations. This approach preserved the computational graph throughout the forward modelling process, enabling automatic and efficient backpropagation of gradients with respect to the model parameters. Although this required an initial restructuring of the code to ensure AD compatibility, the resulting flexibility and computational efficiency were substantial. Additionally, we also present results in which the Jacobian, typically used to evaluate the gradient around each particle when the FD approach is adopted, is replaced by a DCT projection of the analytical, linear approximation of the Zoeppritz equations, as formulated by Aki and Richards (1980).

3. Results

The method is applied to a synthetic inversion in which the true model mimics a turbidite sequence hosting gas saturated reservoir sand. The observed data are computed according to a convolutional model based on the Zoeppritz equations, using a 35 Hz Ricker wavelet and three incidence angles of 0°, 20°, and 40°. Uncorrelated Gaussian noise is added to the observed data, with a standard deviation equal to 20% of that of the noise-free observations, resulting in a signal-to-noise ratio of 12 dB. The total number of CMP locations is 70, whereas 50 is the number of time samples with a time sampling interval of 0.004 s. This results in 50×3 = 150 elastic properties to be inferred for each inverted seismic gather. The prior mean model has been derived from a heavy smoothed version of the true elastic model, while the prior covariance matrix has been derived from two columns extracted from the true model (located in correspondence of horizontal coordinates 1000 and 2000 m), thus simulating the availability of borehole information. This prior covariance also codes an exponential variogram model to constrain the vertical variability of the retrieved elastic parameters (Fig. 1). Given the linearity of the DCT compression, such Gaussian prior can be analytically projected onto the compressed domain. As an example, Fig. 1d shows the prior model covariance in the reduced model space.

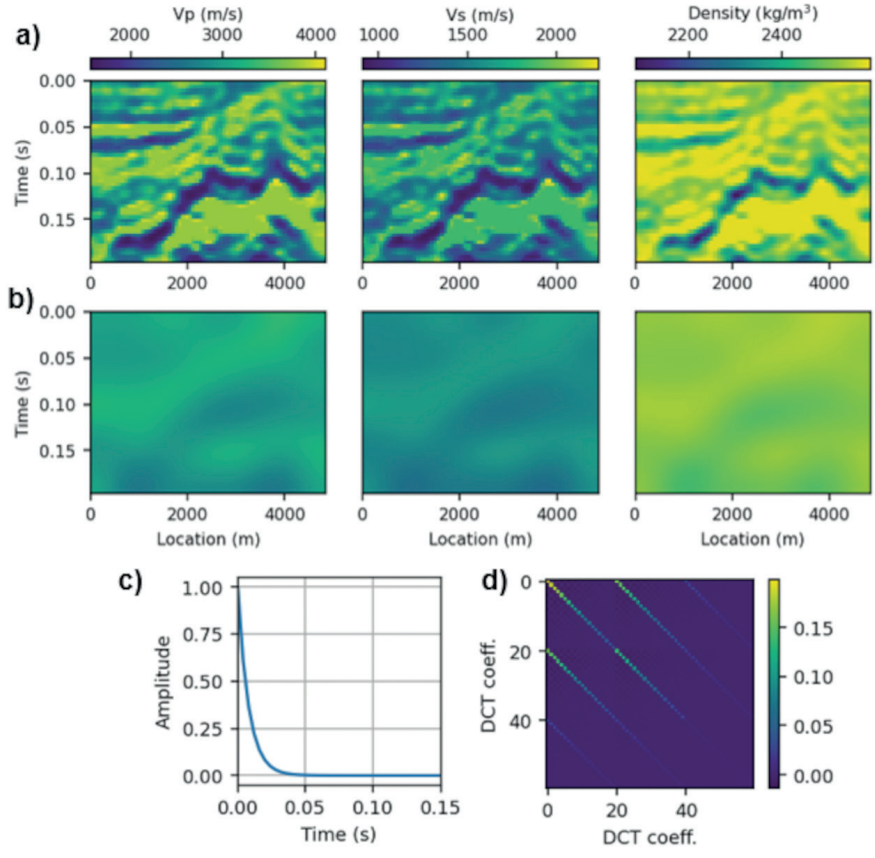

Fig. 1 - a) True model of *Vp*, *Vs*, and density. b) Prior mean model obtained by heavily smoothing the true model. c) Vertical correlation structure evaluated from borehole profiles at 1000 m and 2000 m and used to construct the prior covariance matrix with vertical variogram constraints. d) The prior model covariance matrix projected in the DCT domain showing the variances and covariances associated with the 60 unknown parameters.

The simulated borehole data also serve to determine the optimal number of DCT coefficients to retain for each elastic property in the inversion phase. This number has been determined by measuring how the explained variability, computed as the ratio of the standard deviation of the approximated and original model (Aleardi, 2020), changes when the number of DCT coefficients vary. We found that just 20 coefficients are enough to approximately explain the total variability of each logged elastic property (Fig. 2). This means that the full 150D parameter space can be conveniently compressed in a 20×3 = 60D domain.

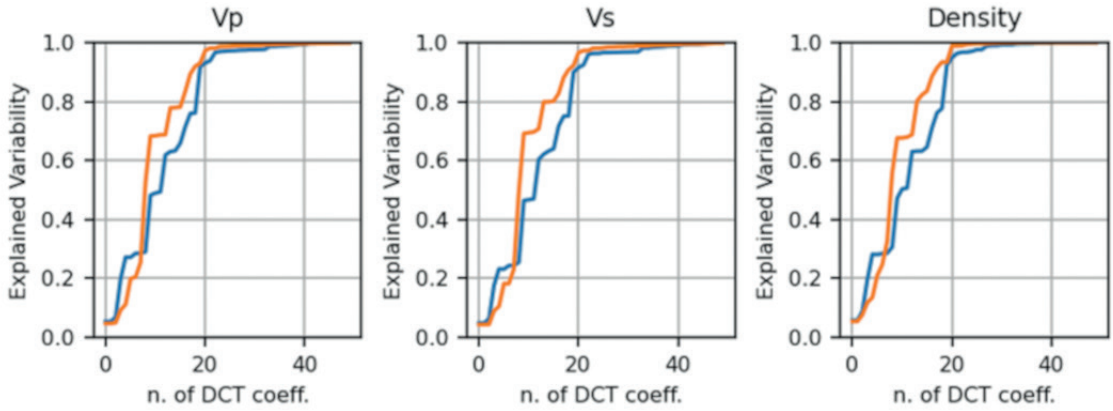

Fig. 2 - Explained variability versus the number of retained coefficients for the *Vp*, *Vs*, and density values extracted from two columns located at 1000 m (blue line) and 2000 m (orange line) along the true model. We observe that just 20 coefficients are enough in both cases to retain almost the entire variability.

We first compare the gradient evaluations in the DCT domain obtained using a forward FD approach and our implemented AD framework, computed over 100 random prior realisations. Fig. 3 shows that the two gradient computation methods yield consistent results, confirming the reliability of the implemented AD strategy. For what concerns the computing time for evaluating the gradient for all the 100 particles, the AD approach takes 1 s approximately, while the FD more than 11 seconds. This results in a reduction of computational cost by more than one order of magnitude. It is also worth noting that, in the FD approach, each column of the Jacobian is computed in parallel by distributing the forward evaluations across multiple cores. A serial implementation would require significantly more computing time. These computing times refer to the following hardware equipment: Intel Core i7-14650HX@2200Mhz with 32 GB RAM.

We first discuss three different inversion experiments, in which the gradient is computed via the AD strategy:

Test 1. An A-SVGD inversion running in the DCT domain;

Test 2. A standard SVGD inversion (i.e. without the introduction of the annealing schedule) still running in the compressed parameter space;

Test 3. An A-SVGD inversion working in the full model space (without DCT compression).

In all cases, we assume the same prior model (shown in Fig. 1) and that both the source wavelet and the noise statistics are perfectly known. To meet the theoretical requirement of the VI inversion, we employ a total number of particles equal to the number of unknowns, with the initial ensemble of particles randomly drawn from the Gaussian prior model. This means that we use 60 particles in tests 1 and 2 and 150 in Test 3. The total number of iterations is set to 50.

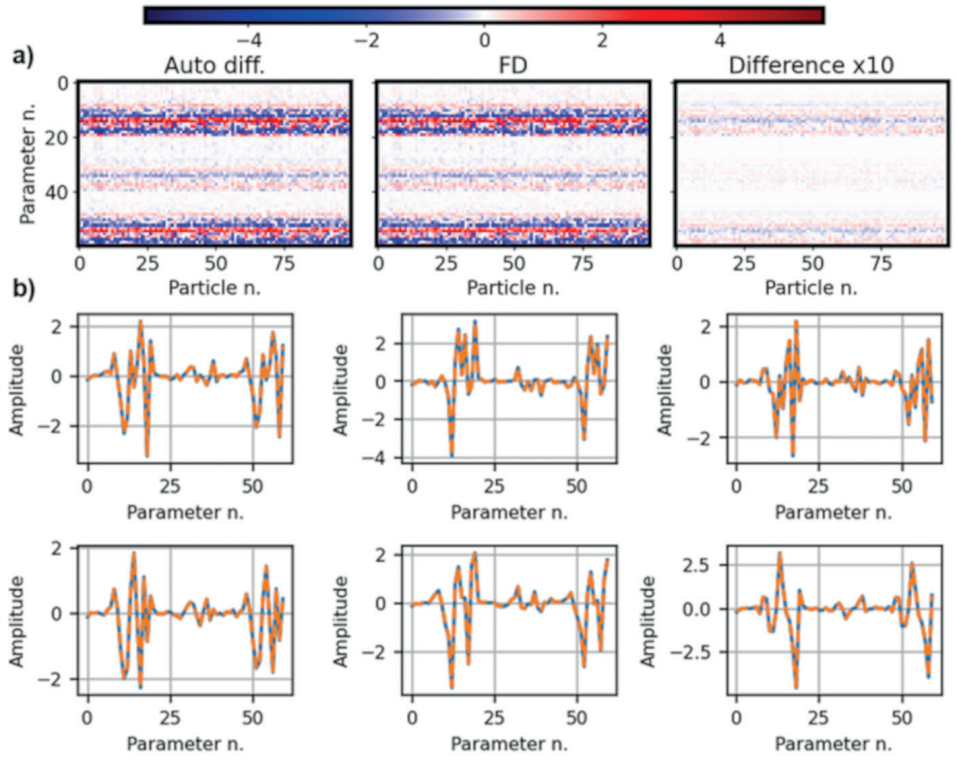

Fig. 3 - a) Comparison between the gradient computed with AD and FDs in DCT domain considering 20 coefficients per elastic property (represented along the vertical axis) and 100 particles representing elastic profiles randomly drawn from the prior (represented along the horizontal axis). In addition, the magnified difference between the two gradient computations is also shown (rightmost plot). b) Some close-ups representing the AD (blue) and FD (dashed orange) gradients computed on six out of the 100 elastic models shown in panel a.

Typically, this number must be chosen to ensure that all particles ultimately converge to a similar data misfit value.

In Fig. 4, the evolution of the L2 norm data misfit value computed for all the particles and for a representative CMP gather inversion are compared for tests 1, 2, and 3. We observe that when compared to the standard SVGD, the annealed version not only ensures faster convergence but also guarantees that all particles achieve the same final data misfit value. This highlights

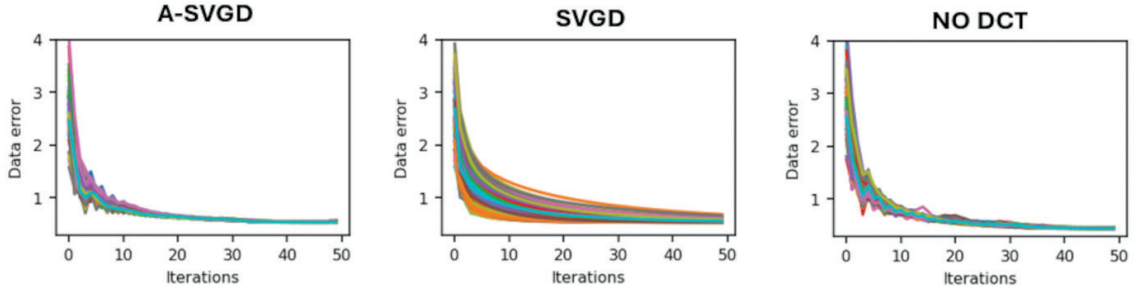

Fig. 4 - Data misfit evolution for one representative CMP gather over 50 iterations. A-SVGD shows fastest convergence. Both A-SVGD inversions (with and without DCT) reach similar misfits.

the superior ability of the A-SVGD to efficiently explore the parameter space compared to the standard SVGD. In addition, we can also appreciate that both tests 1 and 3 reach the same final data misfit values with very similar convergence speeds. This means that, despite model compression, we preserve the ability for the compressed parameterisation to achieve good data predictions.

Fig. 5 compares the true model with the estimated posterior means provided by the different tests. In all cases we observe a good alignment between the true and the estimated elastic

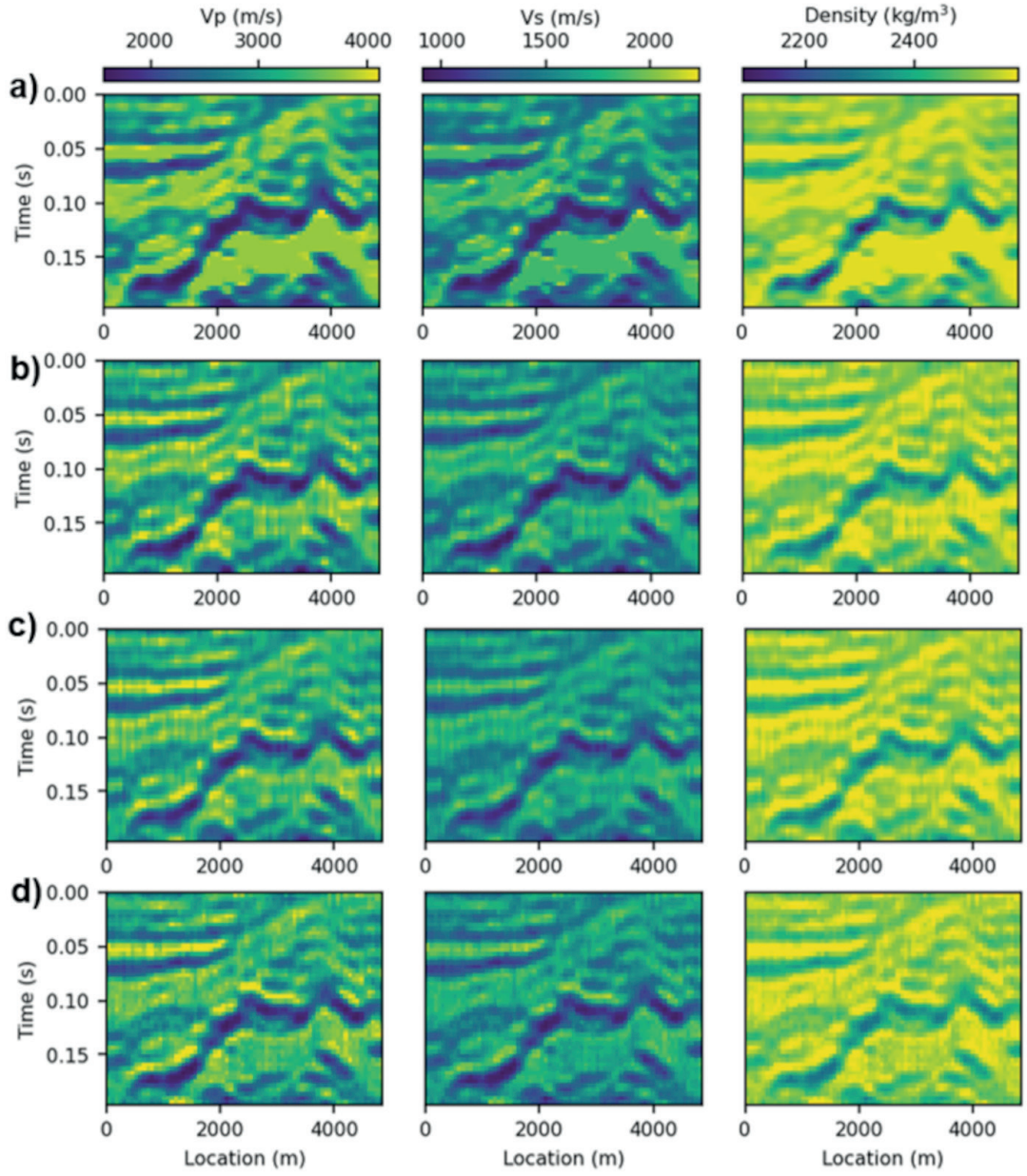

Fig. 5 - Posterior mean estimates of elastic parameters obtained from tests 1, 2, and 3 (panel b, c, and d, respectively). The true model is shown in panel a.

properties. The SVGD and A-SVGD give almost the same results, while we observe slightly more scattered predictions for the A-SVGD running in the full elastic space due to the higher ill-conditioning of the problem resulting from the increased model space dimensionality. Fig. 6 represents the posterior standard deviations estimated in the three tests. In all cases we observe higher uncertainties in correspondence of the highest elastic property values. Test 2 tends to overestimate the posterior uncertainty compared to Test 1. This behaviour is expected due to the lack of convergence to the same data misfit region for all the particles within the maximum number of iterations allowed. As again expected, when the inversion is run without the DCT, the higher dimensionality of the problem results in an increased posterior uncertainty.

As an example of data matching, Fig. 7 compares the three observed and predicted angle stacks for Test 1 along with the associated differences. A good data prediction is achieved and very similar plots would have been obtained for the other two inversions experiments. This illustrates that the possibility to compress the model space does not degrade the final achieved data misfit. We now examine in greater detail the marginal posterior distributions obtained from tests 1 and 2 for three DCT coefficients associated with each elastic property and derived from the CMP located at the 2500-metre horizontal coordinate (Fig. 8). In all cases, we observe that A-SVGD achieves higher accuracy (peak of the marginal posterior closer to the true property value) and precision (narrower posterior uncertainty) with respect to SVGD. When considering the full model space, we obtain the marginal posteriors represented in Fig. 9, which are associated with the horizontal coordinates of 1500 and 2500 m.

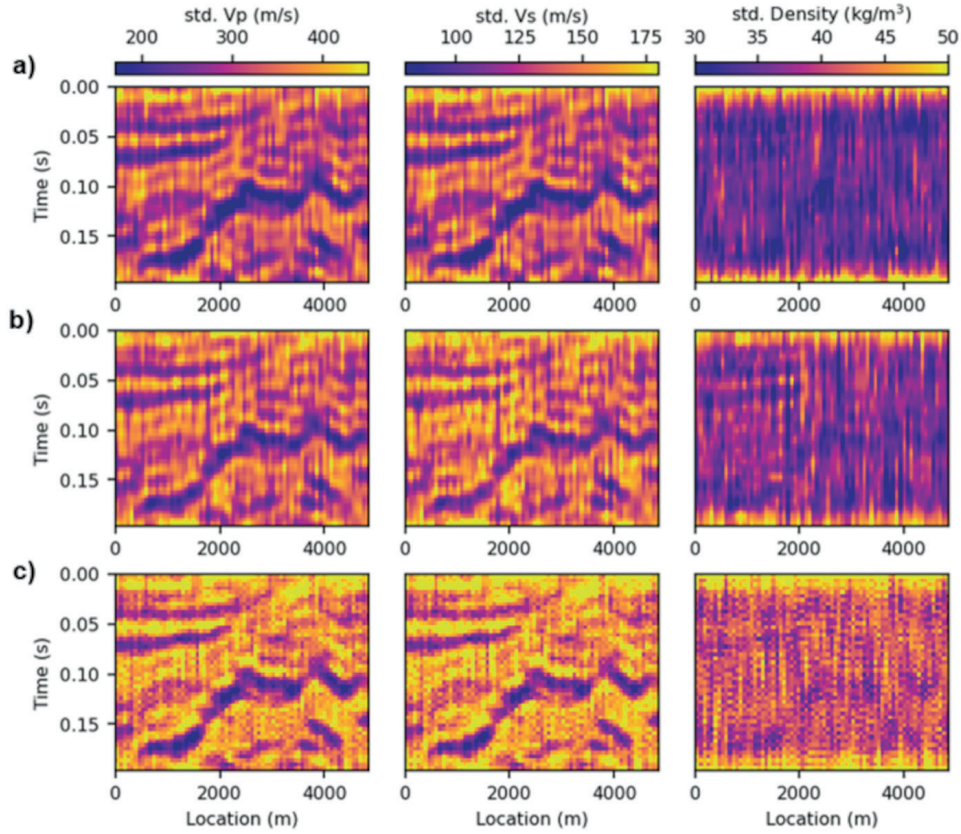

Fig. 6 - Posterior standard deviation of *Vp*, *Vs*, and density obtained in tests 1, 2, and 3 (panels a, b, and c, respectively).

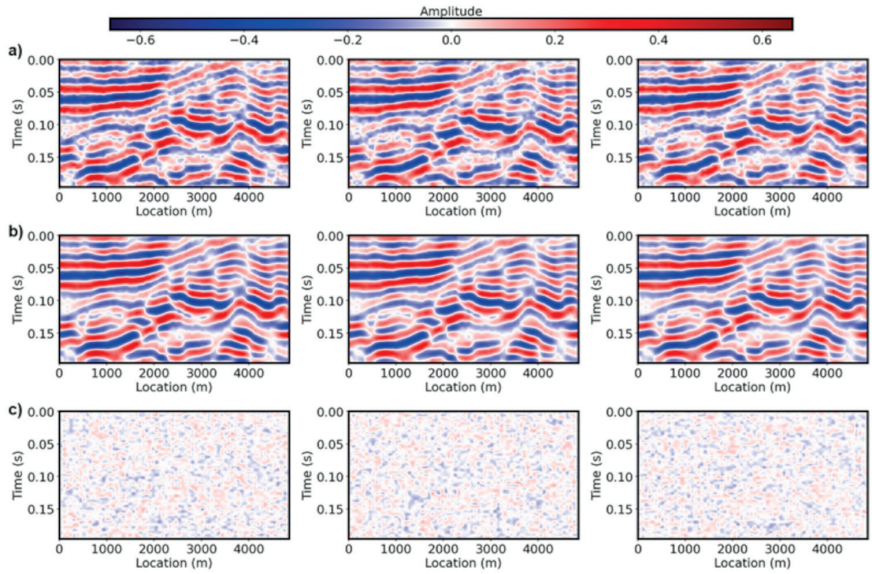

Fig. 7 - Observed (a) and predicted (b) seismic angle stacks (0°, 20°, and 40°, left, centre, and right, respectively) for Test 1 (A-SVGD with DCT). The residual plots in panel c show minimal differences, confirming good data matching.

All three tests seem to give satisfactory predictions, with the true values usually enclosed within the confidence interval. As expected, the inversion running in the full model space provides final predictions with slightly higher vertical resolution at the expense of an increased posterior uncertainty. For a final more quantitative assessment of the inversion results, Table 1 presents the 90% posterior coverage ratio (CR), the correlation coefficient (CC) between true model and

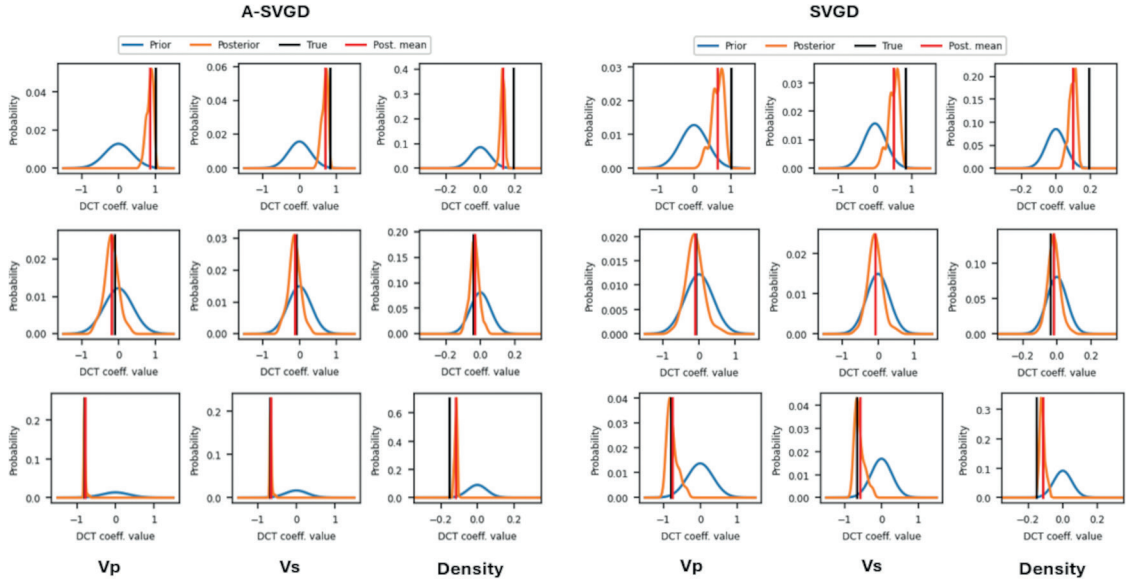

Fig. 8 - Marginal posterior distributions for three selected DCT coefficients (displayed from top to bottom) and for the three properties: *Vp*, *Vs*, and density. This example refers to the CMP location at 2500 m. The comparison between Test 1 (A-SVGD) and Test 2 (SVGD) illustrates that A-SVGD usually achieves higher accuracy and precision.

posterior mean solutions, and the CC between observed data and the data predicted on the posterior mean models previously shown. We observe that the A-SVGD usually provides better uncertainty estimation and model and data predictions than the standard SVGD. The results of Test 1 are consistent with those obtained when A-SVGD runs in the full model space, yet they are achieved with nearly one-third of the forward evaluations used in Test 3. This implies that the total computational cost of Test 1 (20 minutes with the hardware equipment previously presented) is less than 50% of that required by Test 3.

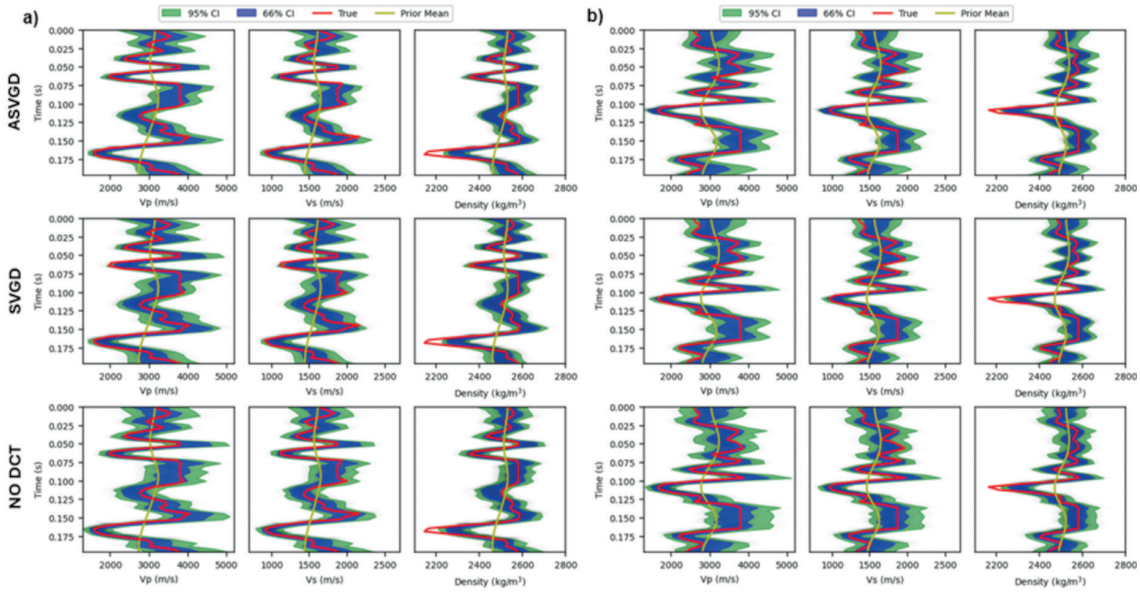

Fig. 9 - Marginal posterior distributions in the full model spaces for selected CMPs at a) 1500 m and b) 2500 m. The 66% and 90% credible intervals (CI) are also shown.

Table 1 - Quantitative assessment of inversion performance for the ideal case (accurate wavelet and noise assumptions). The table reports the 90% posterior CR and CC between true and estimated elastic parameters (*Vp*, *Vs*, and density), and between observed and predicted seismic data.

	A-SVGD		SVGD		NO DCT	
	90% CR	**CC**	**90% CR**	**CC**	**90% CR**	**CC**
Vp	0.97	0.93	0.94	0.91	0.97	0.92
Vs	0.95	0.91	0.93	0.90	0.95	0.91
Density	0.95	0.93	0.93	0.91	0.96	0.93
Data	-	0.97	-	0.95	-	0.97

4. Comparison with a gradient-based Markov chain Monte Carlo method

In this section we show the results obtained for the previous inversion tests when the Stochastic-Newton MCMC is employed. This is a GB-MCMC method that uses the Hessian and gradient of the posterior to define the Markov chain proposal, thereby accelerating posterior

sampling (Martin *et al*., 2012; Berti *et al*., 2024a, 2024b). As for the previous examples, the inversion runs in the DCT compressed model space where 20 coefficients are retained for each elastic property. The observed data and the prior assumptions are the same as in the previous examples. This means that the statistic of the data error and the seismic wavelet are perfectly known.

Unlike the earlier SVGD approach, we cannot fully leverage AD to accelerate the inversion, because the MCMC method used requires not only the local gradient of the posterior but also the Hessian at each iteration, which depends on the Jacobian matrix. Currently, we compute this matrix using a forward FD approach, which requires a number of forward modelling runs equal to the number of unknowns (60 in our case) for each matrix evaluation. Consequently, the number of forward evaluations, and, thus, the total computing time, required by the MCMC method to converge to a stable posterior is significantly higher than that needed by the SVGD methods described above. Therefore, a direct comparison of computational costs between the two approaches is not possible.

To sample the target posterior, we use five independent Markov chains, starting from different random prior realisations and evolving for 500 iterations, with the first 50 iterations considered as the burn-in period. The convergence of the MCMC sampling was monitored through the calculation of the potential scale reduction factor [PSRF: Brooks and Gelman (1998)] and the total number of iterations was chosen so as to achieve a PSRF below 1.1 for most of the unknowns, which indicates convergence to a stable posterior distribution. Fig. 10 shows the mean and standard deviation of the posterior distribution of the three elastic properties estimated using the GB-MCMC sampling. The mean is very similar to that obtained from the

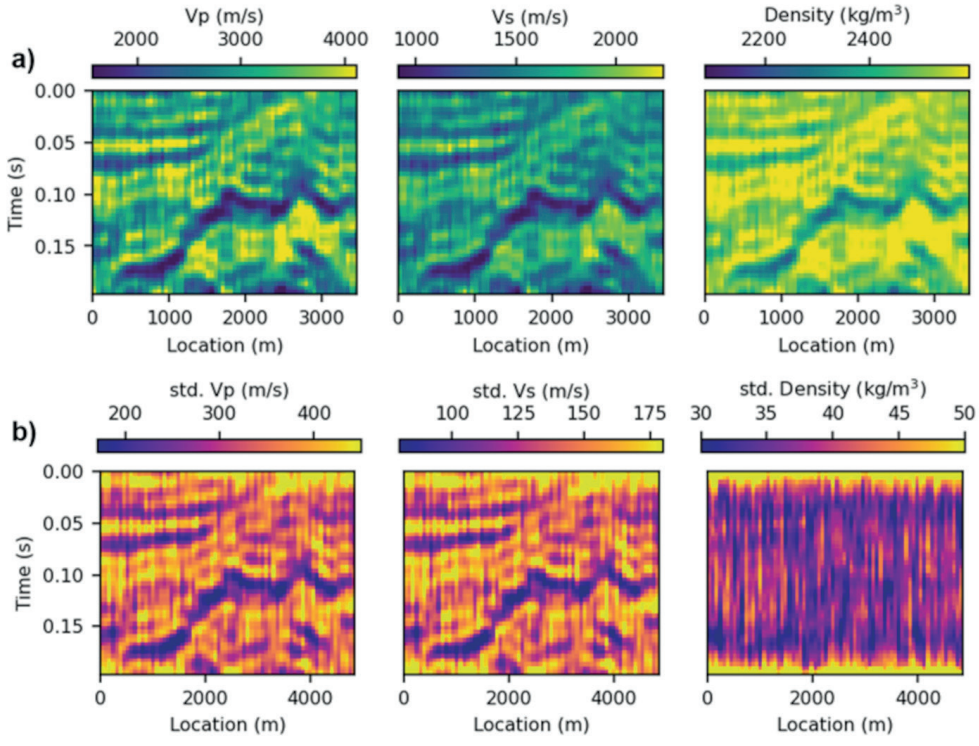

Fig. 10 - Posterior mean for *Vp*, *Vs*, and density estimated by the MCMC inversion (a) and posterior standard deviation for the three elastic properties provided by the MCMC sampling (b).

SVGD inversions described earlier (see Fig. 5). Compared to the previous examples, only a slight increase is noted in the posterior uncertainty . This small difference is, nevertheless, acceptable considering the nonlinearity of the problem and the difficulty of sampling a 60-dimensional parameter space. Fig. 11a illustrates the evolution of the data error for the different chains over 500 iterations for a single CMP inversion. It can be observed that, after an initial burn-in phase of approximately 40–50 iterations, all chains converge to a similar data misfit value, consistent with those previously obtained from the SVGD inversions (see Fig. 4). Fig. 11b shows, for the same CMP gather inversion displayed in Fig. 11a, the evolution of the PSRF for all 60 unknowns. The results indicate that the commonly adopted convergence threshold (PSRF = 1.1) is reached by all parameters after about 450–500 iterations.

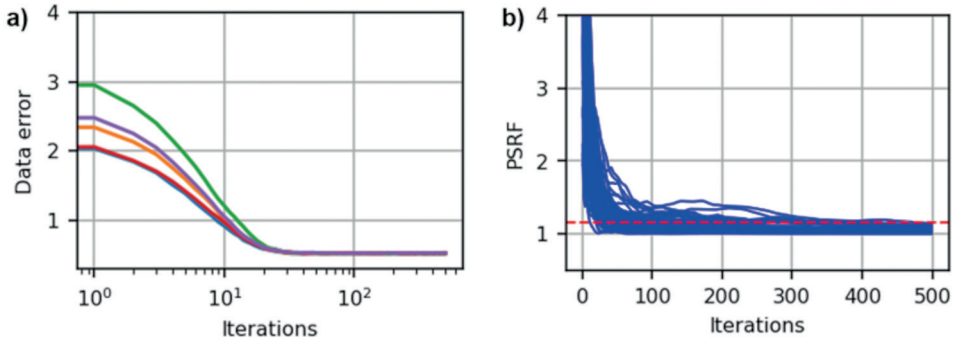

Fig. 11 - a) Data misfit evolution for the five chains and for one representative CMP gather over the 500 GB-MCMC iterations. b) PSRF evolution for all the model parameters and for one representative CMP gather (blue line). The red dashed line represents the threshold of convergence.

Fig. 12 illustrates the marginal posterior distributions projected onto the full model space at 1500- and 2500-metre horizontal coordinates. Once again, no substantial differences are observed compared to the corresponding distributions from the SVGD inversions shown in Fig. 9. Finally, quantitative indicators CR and CC reported in Table 2 demonstrate that the MCMC and A-SVGD inversions yield comparable and consistent results in terms of final predictions. A slight increase in the CR values for various elastic properties is observed in the MCMC results, attributed to the slightly higher posterior standard deviation estimated by this method. In its current implementation, the main drawback of the GB-MCMC inversion is the need to evaluate the Jacobian using the FD approach for each sampled model at every iteration. This significantly reduces its computational efficiency compared to the SVGD inversion. Therefore, a fair comparison of the computational costs between GB-MCMC and A-SVGD would require extending the use of AD to derive the Jacobian matrix needed for the Hessian computation.

5. Inversion with a wrong source wavelet and overestimated noise contamination

We now demonstrate the advantages of the DCT compression over full-domain inversion, in particular the improved predictions obtained when simulating errors in both the estimated noise statistics and source wavelet. The wavelet used to compute the predicted data in the inversion is still a Ricker wavelet but with a 3-Herz error in the estimation of its peak frequency (38 Hz instead

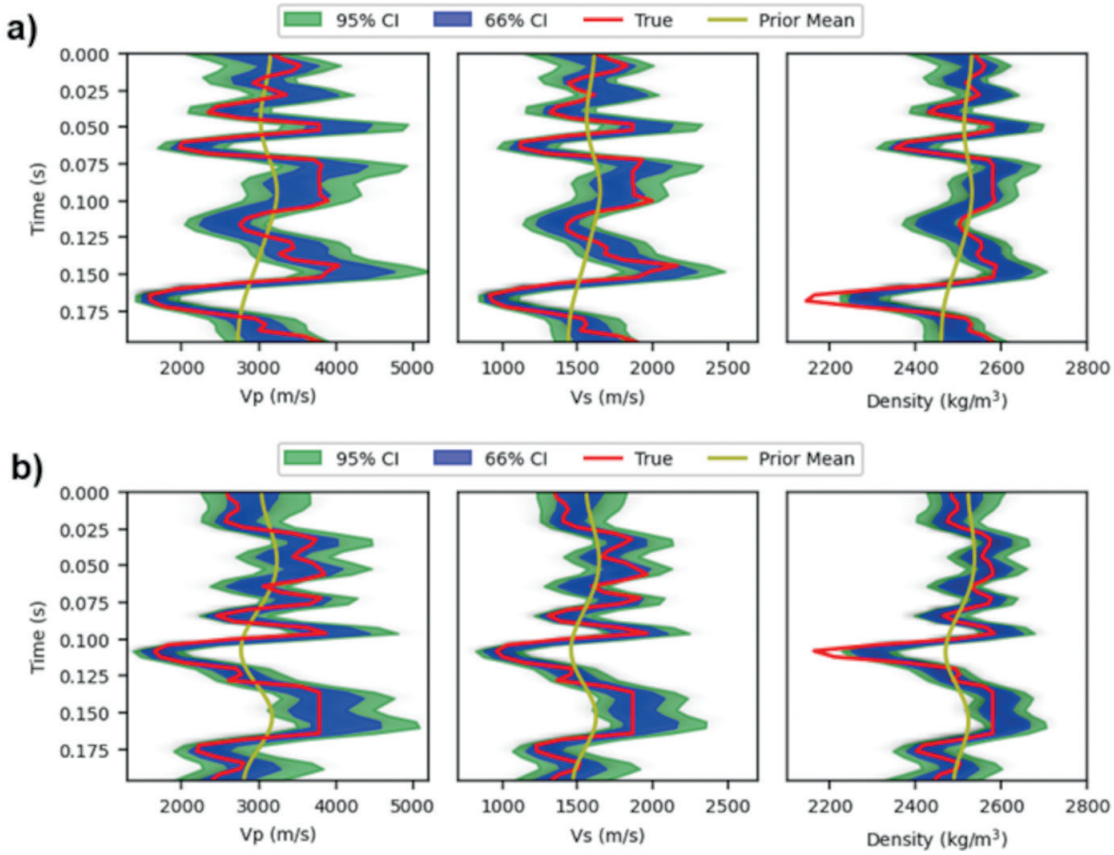

Fig. 12 - Marginal posterior distributions in the full model spaces for selected CMPs at 1500 m (a) and 2500 m (b) as estimated by the GB-MCMC sampling. The 66% and 90% credible intervals (CI) are also shown.

Table 2 - Quantitative assessment of MCMC inversion performance for the ideal case (accurate wavelet and noise assumptions).

	GB-MCMC	
	90% CR	**CC**
Vp	0.98	0.93
Vs	0.96	0.92
Density	0.95	0.93
Data	-	0.97

of the 35 Hz used in the computation of the observed data). We also introduce a 20° phase rotation and apply a 20% error in the estimation of the peak amplitude (Fig. 13). In addition, a 50% overestimation of the standard deviation of the Gaussian noise contaminating the observed data is also simulated. These conditions intend to simulate usual errors made in field data applications in which both the seismic wavelet and the amount of noise affecting the observations are not perfectly known but just estimated and/or inferred from the available data. We repeated the same three SVGD tests previously run, and the posterior mean predictions obtained are illustrated in Fig. 14. All methods yield a reasonably accurate and consistent estimate of the elastic parameters;

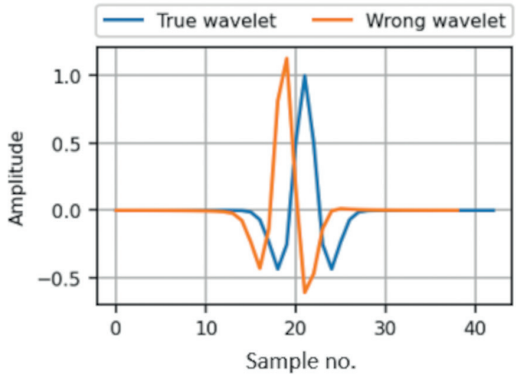

Fig. 13 - Comparison between true and wrong source wavelets.

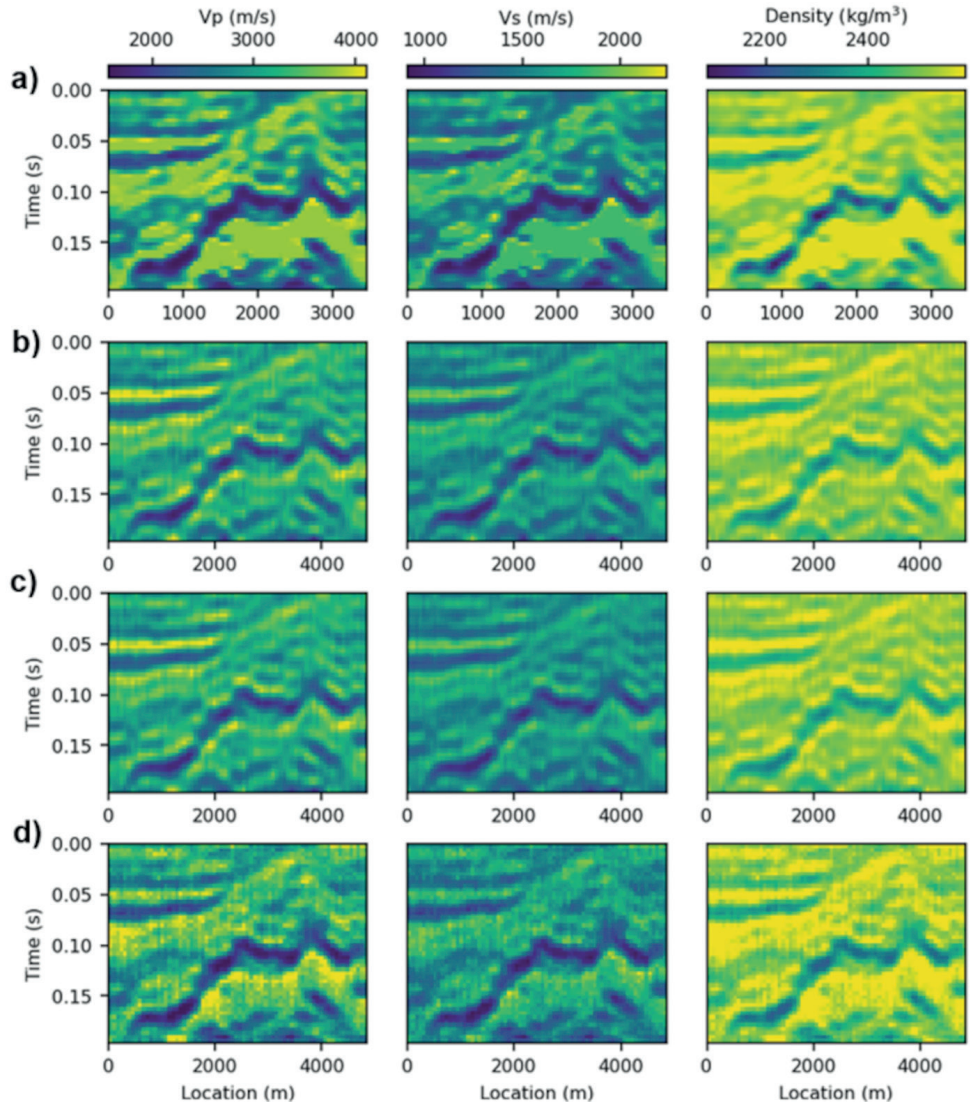

Fig. 14 - Posterior mean estimates of the elastic parameters obtained in tests 1, 2, and 3 (panels b, c, and d, respectively), under incorrect wavelet and noise assumptions. The true model is shown in panel a.

however, it is also evident that the inversion without DCT results in a predicted mean model that is more affected by artefacts and scattering compared to the other two implementations. As expected, and similarly to the previous tests, the inversion without DCT compression tends to overestimate model uncertainty when compared to inversions running in compressed domains (Fig. 15). If we now inspect the evolution of the data error for a single CMP inversion (Fig. 16), we note that the A-SVGD with DCT reaches the lowest final misfit values with the fastest convergence rate; differently from the previous experiments, the inversion without the DCT now shows the

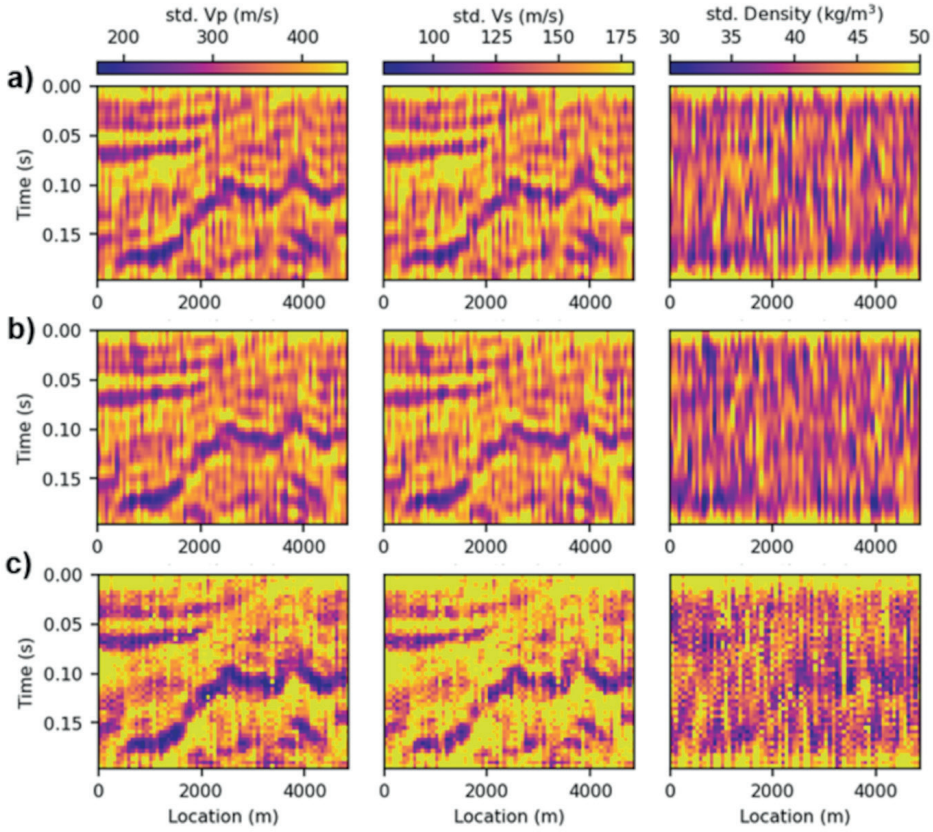

Fig. 15 - Posterior standard deviations under wavelet and noise errors. Panels a, b, and c refer to tests 1, 2, and 3, respectively.

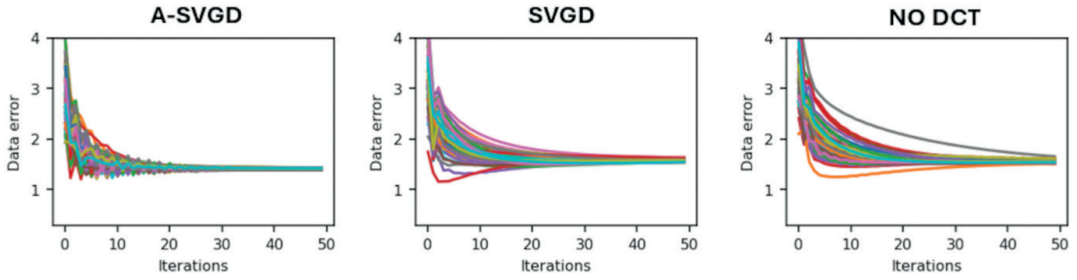

Fig. 16 - Data misfit convergence for the three inversions in the example with erroneous source wavelet and estimated noise statistics.

slowest convergence rate and reaches final misfit values higher than those achieved by the A-SVGD. This seems to highlight the advantages of working in a reduced-dimensional model space when dealing with more realistic scenarios, in which some of the assumptions made during the inversion are not fully satisfied. As an example, Fig. 17 shows a comparison between observed and predicted data associated with the posterior mean model estimated by the A-SVGD. Unlike the previous examples, now some coherent signal is still present in the difference plots, due to the error made in the estimation of the source wavelet. For a more quantitative assessment of the results, Table 3 shows the CR and CC evaluated on the predicted posterior means and the associated predicted seismic data. The inversion working in the full model space achieves higher CR values given the overestimation of the posterior uncertainty; however, this inversion is also characterised by the lowest model accuracy and poorest data fitting as evidenced by the lowest CC between predicted and true model and observed and predicted data.

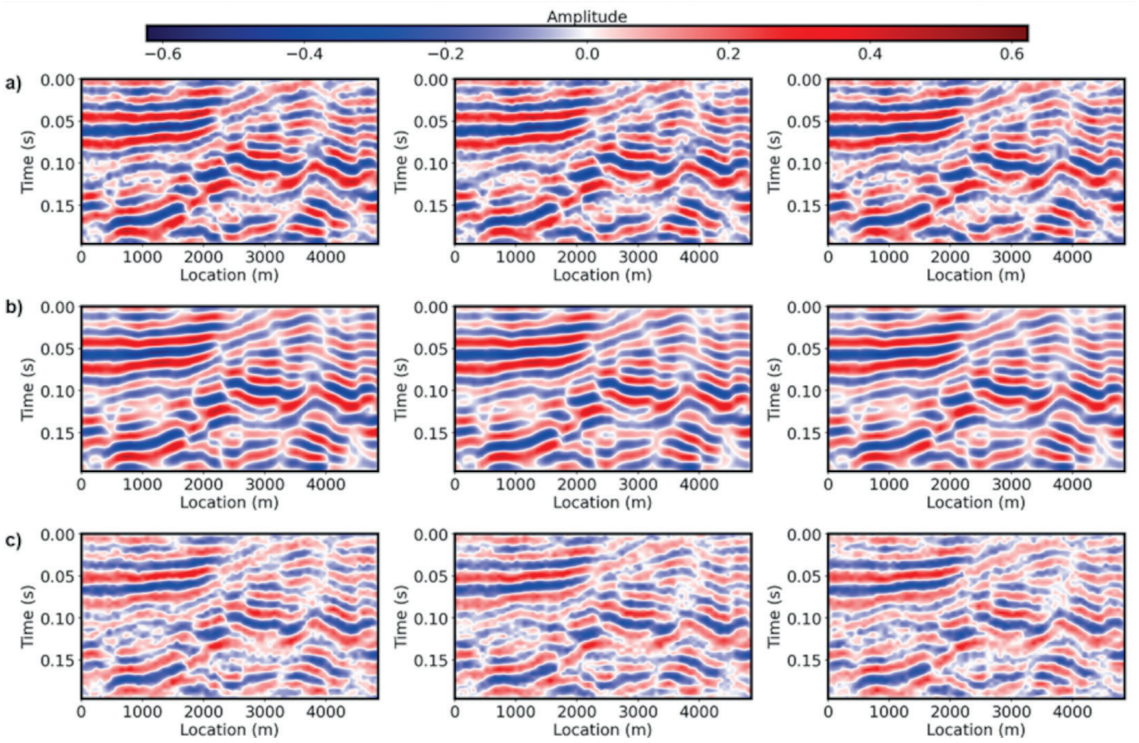

Fig. 17 - Observed (a) and predicted (b) angle stacks (0°, 20°, and 40°, left, centre, and right, respectively) for Test 1 (A-SVGD with DCT). c) Residuals show structured differences resulting from the source wavelet error.

Table 3 - Inversion performance under wavelet and noise estimation errors.

	A-SVGD		SVGD		NO DCT	
	90% CR	**CC**	**90% CR**	**CC**	**90% CR**	**CC**
Vp	0.87	0.83	0.85	0.83	0.89	0.79
Vs	0.86	0.84	0.85	0.83	0.88	0.79
Density	0.88	0.86	0.86	0.85	0.91	0.81
Data	-	0.74	-	0.74	-	0.63

6. Replacing the Jacobian with a linearisation of the Zoeppritz equations

For the A-SVGD with DCT, the results obtained when the Jacobian used for the gradient computation is replaced by a DCT projection of the analytical, linear approximation of the Zoeppritz equations as formulated by Aki and Richards (1980), are also presented. From a theoretical point of view this means that when updating the particles, a constant curvature of the posterior across the entire model space is assumed. From a more practical perspective this approximation can be reliably applied in case of weak elastic contrasts at the reflecting interface and a narrow incidence angle range (i.e. far from the critical reflection and usually not beyond 30°–35°). However, even if these conditions are met, the practical applicability of this approach must be assessed case by case. As in the first examples, the noise and wavelet should be perfectly known. Fig. 18a shows the mean predicted model yielded by the A-SVGD inversion in this scenario. Some artefacts appear in the estimated posterior mean, which are related to the fact that the assumptions underlying the linear approximation are not satisfied in the considered application. Indeed, we include in the inversion incidence angle up to 40° and the observed data are generated from a subsurface model characterised by significant property contrasts at the reflecting interfaces. Fig. 18b points out the increased posterior uncertainty with respect to the test in which the full Zoeppritz equations are used to compute the gradient (Fig. 6a). In the posterior marginals shown in Figs. 18c and 18d, we observe that the true property values are not always contained within the estimated confidence intervals, indicating that the uncertainty quantification is no longer reliable. Table 4 summarises the quantitative evaluation of the inversion results in this test. It is clear that the overall performances in terms of model and data predictions are significantly deteriorated with respect to the previous examples. Although this linear approximation significantly accelerates the inversion (with the inversion process taking less than 5 minutes, as the gradient of the posterior can be computed analytically through matrix-vector multiplication), its applicability is highly case-dependent. In other terms, the choice of employing any linearisation of the nonlinear Zoeppritz equations must be considered case by case and should constitute a compromise between the inversion computational efficiency and the accuracy of the results. Similar conclusions were drawn for tests 2 and 3 and for this reason are not shown here.

Table 4 - A-SVGD inversion performance when the linear approximation is used for the Jacobian evaluation.

	A-SVGD	
	90% CR	**CC**
Vp	0.81	0.33
Vs	0.80	0.33
Density	0.81	0.58
Data	-	0.73

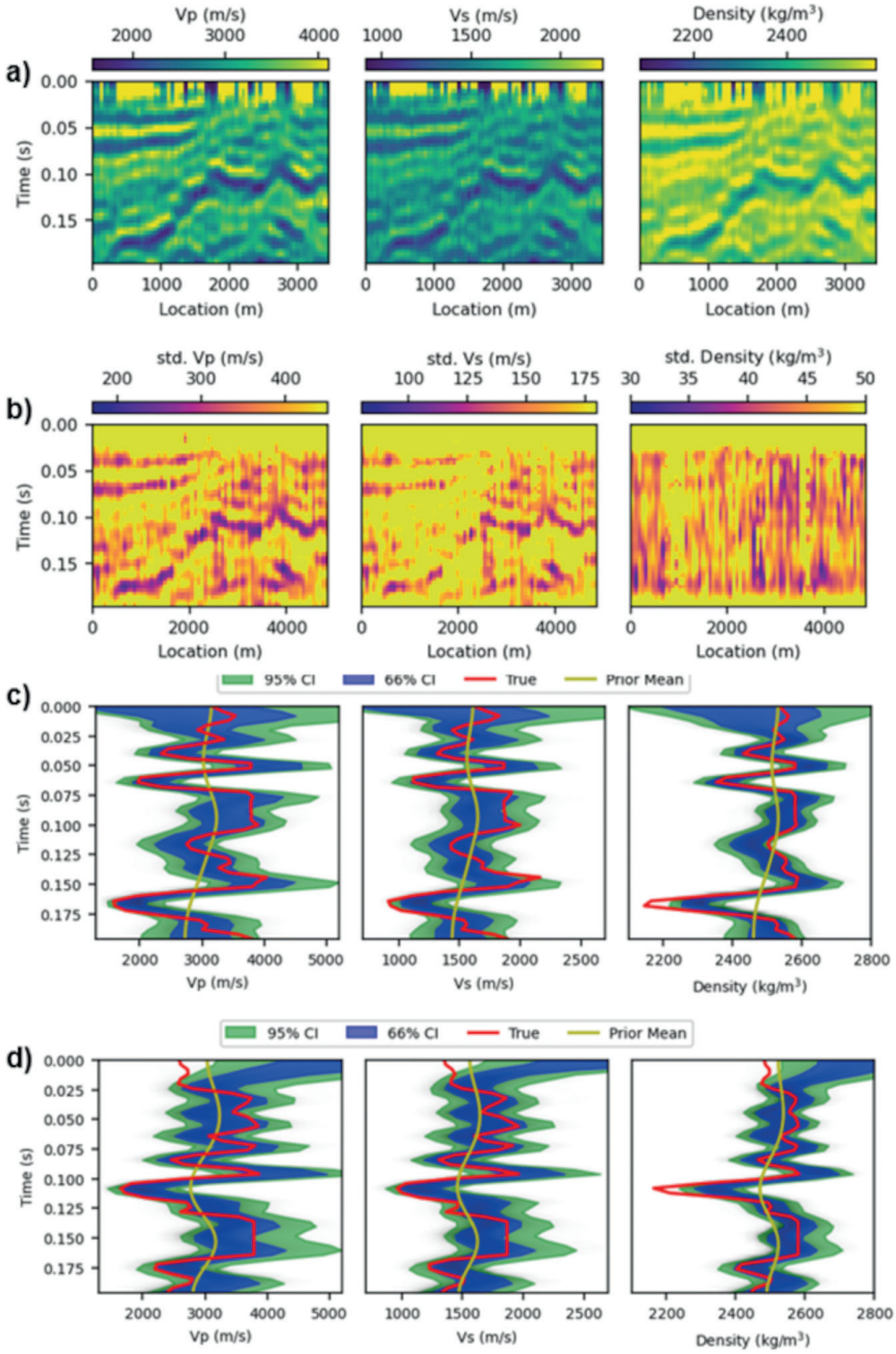

Fig. 18 - a) Posterior mean predictions for Test 1 using the DCT-projected Aki and Richards (1980) linearised Jacobian. Artefacts appear due to the mismatch between the linear approximation and the full Zoeppritz equations. b) Posterior standard deviations for *Vp*, *Vs*, and density. c) and d) Marginal posterior distributions in the full model spaces for the CMPs at 1500- and 2500-metre horizontal locations.

7. Conclusions

This study presents a Bayesian AVA inversion that couples A-SVGD, DCT for model space compression, and AD for fast gradient evaluation. The A-SVGD algorithm enables efficient sampling of high-dimensional posterior distributions, while the DCT transformation reduces model space dimensionality, alleviating computational burdens without compromising inversion accuracy. Synthetic inversion experiments confirm the advantages of A-SVGD over standard SVGD in terms of convergence speed and particle consistency. DCT compression proves to be particularly effective in preserving inversion quality while drastically reducing the number of forward model evaluations, achieving comparable results to full-space inversion at less than half the computational cost.

The proposed approach also produced final predictions that are comparable and consistent with those obtained using a well-established GB-MCMC sampling algorithm. When tested under more realistic conditions, thus including errors in the source wavelet's frequency, phase, and amplitude, A-SVGD combined with DCT compression continues to yield stable and accurate results, with reduced artefacts and better convergence compared to full-space inversion. Furthermore, we explore the use of an analytical approximation of the Jacobian [via Aki and Richards (1980)] projected into the DCT space. While this approach significantly reduces computational demands, its validity is highly dependent on model characteristics, such as elastic contrasts at the reflecting interfaces and incidence angles, and, therefore, must be evaluated on a case-by-case basis. Overall, our results underscore the value of integrating A-SVGD, DCT-based model compression, and AD in AVA inversion workflows. This integrated approach offers significant computational efficiency and improved uncertainty quantification, making it well suited for modern seismic inversion tasks where both robustness and scalability are critical.

REFERENCES

Aki K. and Richards P.G.; 1980: *Quantitative seismology: theory and methods*. W.H. Freeman, New York, NY, USA, 743 pp.

Aleardi M.; 2015: *The importance of the Vp/Vs ratio in determining the error propagation, the stability and the resolution of linear AVA inversion: a theoretical demonstration*. Boll. Geof. Teor. Appl., 56, 357-366.

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, 331-348.

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

Aleardi M. and Tognarelli A.; 2016: *The limits of narrow and wide-angle AVA inversions for high Vp/Vs ratios: an application to elastic seabed characterization*. J. Appl. Geophys., 131, 54-68.

Azevedo L., Grana D. and Amaro C.; 2019: *Geostatistical rock physics AVA inversion*. Geophys. J. Int., 216, 1728-1739.

Ba J., Erdogdu M.A., Ghassemi M., Sun S., Suzuki T., Wu D. and Zhang T.; 2022: *Understanding the variance collapse of SVGD in high dimensions*. In: Tenth International Conference on Learning Representations (Virtual event).

Baydin A.G., Pearlmutter B.A., Radul A.A. and Siskind J.M.; 2018: *Automatic differentiation in machine learning: a survey*. J. Mach. Learn. Res., 18, 1-43.

Beda L.M., Korolev L.N., Sukkikh N.V. and Frolova T.S.; 1959: *Programs for automatic differentiation for the machine BESM*. Institute for Precise Mechanics and Computation Techniques, Academy of Science, Moscow, USSR, Technical Report (in Russian).

Berti S., Aleardi M. and Stucchi E.; 2024a: *A Bayesian approach to elastic full-waveform inversion: application to two synthetic near surface models*. Bull. Geoph. Ocean., 65, 291-308.

Berti S., Aleardi M. and Stucchi E.; 2024b: *A probabilistic full waveform inversion of surface waves*. Geophys. Prospect., 72, 3448-3473.

Berti S., Ravasi M., Aleardi M. and Stucchi E.; 2024c: *Full waveform inversion of surface waves with annealed stein variational gradient descent*. In: Proc. Conference NSG 2024 30th European Meeting of Environmental and Engineering Geophysics, European Association of Geoscientists and Engineers, Helsinki, Finland, pp. 1-5, doi: 10.3997/2214-4609.202420030.

Berti S., Ravasi M., Aleardi M. and Stucchi E.; 2025: *Bayesian full waveform inversion of surface waves with annealed stein variational gradient descent*. Geophys. J. Int., 241, 641-657.

Blei D.M., Kucukelbir A. and McAuliffe J.D.; 2017: *Variational inference: a review for statisticians.* J. Am. Stat. Ass., 112, 859-877.

Brooks S.P. and Gelman A.; 1998: *General methods for monitoring convergence of iterative simulations*. J. Comput. Graph. Stat., 7, 434-455.

Brosse N., Moulines E. and Durmus A.; 2018: *The promises and pitfalls of stochastic gradient Langevin dynamics*. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18), Montréal, Canada, Curran Associates Inc., Red Hook, NY, USA, pp. 8278-8288, doi: 10.48550/arXiv.1811.10072.

Buland A. and Omre H.; 2003: *Bayesian linearized AVO inversion*. Geophys., 68, 185-198.

Corliss G., Faure C., Griewank A., Hascoet L. and Naumann U., eds; 2002: *Automatic differentiation of algorithms*: *from simulation to optimization. 1st ed*. Springer Science and Business Media, Berlin, Germany, 459 pp., doi: 10.1007/978-1-4613-0075-5.

Corrales M., Berti S., Denel B., Williamson P., Aleardi M. and Ravasi M.; 2025: *Annealed stein variational gradient descent for improved uncertainty estimation in full-waveform inversion*. Geophys. J. Int., 241, 1088-1113.

D'Angelo F. and Fortuin V.; 2021: *Annealed stein variational gradient descent*. In: Proc. 3rd Symposium on Advanced in Approximate Bayesian Inference (Virtual event), doi: 10.48550/arXiv.2101.09815.

Dai R., Yin C. and Peng D.; 2022: *AVA inversion of pre-stack seismic data with cross-gradient constraints*. J. Appl. Geophys., 199, 104594.

Duchi J., Hazan E. and Singer Y.; 2011: *Adaptive subgradient methods for online learning and stochastic optimization*. J. Mach. Learn. Res., 12, 2121-2159.

Fernandez-Martinez J.L. and Fernandez-Muniz Z.; 2020: *The curse of dimensionality in inverse problems*. J. Comput. Appl. Math., 369, 112571.

Fichtner A., Zunino A. and Gebraad L.; 2019: *Hamiltonian Monte Carlo solution of tomographic inverse problems*. Geophys. J. Int., 216, 1344-1363.

Fu Z., Yin C., Chen T., Ji Y. and Liao J.; 2024: *Using the wavelet transform for seismic wave impedance inversion*. Geophys., 89, R387-R397.

Grana D. and Mukerji T.; 2015: *Bayesian inversion of time-lapse seismic data for the estimation of static reservoir properties and dynamic property changes*. Geophys. Prospect., 63, 637-655.

Griewank A. and Walther A.; 2008: *Evaluating derivatives: principles and techniques of algorithmic differentiation. 2nd ed*. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 459 pp., doi: 10.1137/1.9780898717761.

Guo Q., Luo C. and Grana D.; 2023: *Bayesian linearized rock-physics amplitude-variation-with-offset inversion for petrophysical and pore-geometry parameters in carbonate reservoirs*. Geophys., 88, MR273-MR287.

Kullback S. and Leibler R.A.; 1951: *On information and sufficiency*. Ann. Math. Stat., 22, 79-86, doi: 10.1214/aoms/1177729694.

Liu F., Li H., Zou G. and Li J.; 2025: *Automatic differentiation-based full waveform inversion with flexible workflows*. J. Geophys. Res.: Mach. Learn. Comput., 2, e2024JH000542.

Liu Q. and Wang D.; 2016*: Stein variational gradient descent: a general purpose Bayesian inference algorithm*. Adv. Neural Inf. Process. Syst., 2378-2386.

Lochbuhler T., Breen S.J., Detwiler R.L., Vrugt J.A. and Linde N.; 2014: *Probabilistic electrical resistivity tomography of a CO2 sequestration analog*. J. Appl. Geophys., 107, 80-92, doi: 10.1016/j.jappgeo.2014.05.013.

Martin J., Wilcox L.C., Burstedde C. and 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.

Moghanloo G.H., Riahi M.A. and Bagheri M.; 2018: *Application of simultaneous prestack inversion in reservoir facies identification*. J. Geophys. Eng., 15, 1376-1388.

Qian F., Yin M., Liu X.Y., Wang Y.J., Lu C. and Hu G.M.; 2018: *Unsupervised seismic facies analysis via deep convolutional autoencoders*. Geophys., 83, A39-A43.

Rincón F., Berti S., Aleardi M., Tognarelli A. and Stucchi E.; 2025: *Integrating deep learning and discrete cosine transform for surface waves full-waveform inversion*. Geophys. J. Int., 240, 805-828.

Sambridge M., Rickwood P., Rawlinson N. and Sommacal S.; 2007: *Automatic differentiation in geophysical inverse problems*. Geophys. J. Int., 170, 1-8, doi: 10.1111/j.1365-246X.2007.03400.x.

Sitzmann V., Martel J.N.P., Bergman A.W., Lindell D.B. and Wetzstein G.; 2020: *Implicit neural representations with periodic activation functions*. Adv. Neural Inf. Process. Syst., 33, 7462-7473.

Sun J., Innanen K., Zhang T. and Trad D.; 2023: *Implicit seismic full waveform inversion with deep neural representation*. J. Geophys. Res.: Solid Earth, 128, e2022JB025964.

Tarantola A.; 2005: *Inverse problem theory and methods for model parameter estimation*. SIAM Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 354 pp., doi: 10.1137/1.9780898717921.

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.

Zhuo J., Liu C., Shi J., Zhu J., Chen N. and Zhang B.; 2018: *Message passing stein variational gradient descent*. In: Proc. 35th International Conference on Machine Learning, Stockholm, Sweden, pp. 6018-6027.

Corresponding author: Sean Berti
Dip. Scienze della Terra, Facoltà di Scienze Naturali, Matematica e Fisica, Università degli Studi
Via Santa Maria 53, 56126 Pisa, Italy
Phone: +39 334 2690 517; e-mail: sean.berti@unifi.it