Discrete cosine transform for parameter space reduction in linear and non-linear AVA inversions

Mattia Aleardi[1]

[1]University of Pisa, Earth Sciences Department, via S. Maria 53, 56126, Pisa, Italy

Corresponding author: Mattia Aleardi, mattia.aleardi@unipi.it

ABSTRACT

Geophysical inversions estimate subsurface physical parameters from the acquired data and because of the large number of model unknowns, it is common practice reparametrizing the parameter space to reduce the dimension of the problem. This strategy could be particularly useful to decrease the computational complexity of non-linear inverse problems solved through an iterative sampling procedure. However, part of the information in the original parameter space is lost in the reduced space and for this reason the model parameterization must always constitute a compromise between model resolution and model uncertainty. In this work, we use the Discrete Cosine Transform (DCT) to reparametrize linear and non-linear elastic amplitude versus angle (AVA) inversions cast into a Bayesian setting. In this framework the unknown parameters become the series of coefficients associated to the DCT base functions. We first run linear AVA inversions to exactly quantify the trade-off between model resolution and posterior uncertainties with and without the model reduction. Then, we employ the DCT to reparametrize non-linear AVA inversions numerically solved through the Differential Evolution Markov Chain and the Hamiltonian Monte Carlo algorithm. To draw general conclusions about the benefits provided by the DCT reparameterization of AVA inversion, we focus the attention on synthetic data examples in which the true models have been derived from actual well log data. The linear inversions demonstrate that the same level of model accuracy, model resolution, and data fitting can be achieved by employing a number of DCT coefficients much lower than the number of model parameters spanning the unreduced space. The non-linear inversions

illustrate that an optimal model compression (a compression that guarantees optimal resolution and accurate uncertainty estimations) guarantees faster convergences toward a stable posterior distribution and reduces the burn-in period and the computational cost of the sampling procedure.

INTRODUCTION

From a mathematical point of view, the estimation of subsurface parameters from the acquired geophysical data is an inverse problem (Zhdanov, 2002; Tarantola, 2005; Menke 2018; Aster et al. 2018). One challenge posed by geophysical inversions is the estimation of several (hundreds, or even thousand) subsurface parameters from noisy, low-resolution, measurements. This lack of information usually results in an ill-conditioned inverse problem in which many models equally fit the observed data. For this reason, it is of crucial importance to quantify the uncertainty affecting the final predictions, and this task is usually accomplished by casting the inverse problem into a Bayesian framework. Some applications of this approach to solve geophysical problems can be found, for example, in Sen and Stoffa (1996), Malinverno (2000), Buland and Omre (2003), Malinverno and Briggs (2004), Bosch et al. (2007), Bodin et al. (2012), Dosso et al. (2012), Rimstad et al. (2012), Zunino et al. (2014), Grana (2016), Sajeva et al. (2017), Ray et al. (2017), Piana Agostinetti and Bodin (2018), Pejic et al. (2018), de Figueiredo et al. (2018), Aleardi and Salusti (2019). The final solution of a Bayesian inversion is the so-called posterior probability distribution (PPD) in the model space, that can be, however, analytically computed only for linear forward modeling operators and under Gaussian assumptions about the model, data, and error distributions. Similarly, an analytical and mathematically exact derivation of the so-called sensitivity analysis tools (e.g. model and data resolution matrices) is only possible for linear inversions.

In case of non-linear problems or/and non-Gaussian assumptions, the analytical solution is not available in a closed form and for this reason it must be numerically assessed using Markov Chain Monte Carlo methods (MCMC; Sambridge and Mosegaard, 2002). These algorithms transform the inverse problem into a sampling problem in which the sampling density is proportional to the PPD.

The first stage of the MCMC sampling is the burn-in period in which the chain moves from a random starting model to a high-probability region. The second stage is often called the sampling stage in which the small fluctuations of the misfit value indicate that the MCMC algorithm has reached the stationary regime.

Although the increasing computational power provided by modern parallel architectures has considerably encouraged the applications of MCMC methods to solve geophysical problems, it is always crucial adopting a specific recipe to guarantee an accurate and computationally efficient sampling of the PPD. For example many MCMC algorithms (e.g. the popular random walk Metropolis) are known to mix slowly between the modes if the target distribution is multimodal with modes separated by low probability regions (Holmes et al. 2017; Scalzo et al. 2019). A simple approach to mitigate this issue makes use of multiple MCMC chains to sample the PPD. This strategy usually offers robust protection against premature convergence because the chains use different trajectories to explore the parameter space. However, this strategy is inefficient in highly dimensional problems where the curse-of-dimensionality makes the target distribution highly localized within each model space dimension. This issue usually increases the probability for the MCMC chains to get trapped in local maxima of the PPD and for this reason a considerable number of sampled models is needed to achieve accurate posterior estimations. There have been many attempts to improve the convergence of MCMC algorithms in case of high-dimensional problems. For examples hybridizing MCMC algorithms with global search methods (e.g. Differential evolution Markov Chain “DEMC”, or Differential evolution adaptive Metropolis; Turner and Sederberg 2012; Vrugt 2016; Aleardi and Mazzotti 2017) or exploiting the Hamiltonian mechanic to include the derivative information of the PPD into the sampling framework (Sen and Biswas, 2017; Fichtner and Simutè, 2018; Fichtner and Zunino, 2019).

The curse of dimensionality can be also mitigated by specific model reparameterizations that reduce the dimensionality of the inverse problem and its computational complexity. Several methods have been proposed using different base functions (e.g. principal component analysis, Chebyshev polynomials, wavelet transforms, Legendre polynomials, Discrete Cosine Transform, machine learning methods). After such reparameterization the unknown parameters become the numerical coefficients that multiply the base functions. Some examples of applications of these methods to geophysical problems can be found in Fernández Martínez et al. (2011), Dejtrakulwong et al. (2012), Lochbühler et al. (2014), Satija and Caers (2015), Azevedo et al. (2016), Fernández Martínez et al. (2017), Aleardi (2019), Szabó and Dobróka (2019), Qin et al. (2019), Grana et al. (2019), Nunes et al. (2019), Azevedo and Demyanov (2019). In the context of Hamiltonian Monte Carlo (HMC) inversions these parameter reduction methods are not only useful to additionally mitigate the curse of dimensionality but also to drastically reduce the computational cost related to the numerical computation of the Jacobian matrix that is needed to estimate the local gradient of the PPD. However, it is well known that the parameterization of an inverse problem must always constitute a compromise between model resolution and model uncertainty (Malinverno, 2000; Menke, 2018). This means that both the model uncertainty and the model resolution decrease as the number of inverted parameters decreases. In other words, the loss of information due to the parameter space reduction leads to underfit the observed data, underestimation of the uncertainty in the final solution, and in a decrease of the model resolution.

In this work, we use the Discrete Cosine Transform (DCT) to reparametrize the amplitude versus angle (AVA) inversion in which the elastic properties of P-, S-wave velocities (*Vp, Vs*, respectively) and density (ρ) are retrieved from partial angle-stacked seismic data. The AVA inversion can be formulated either as a linear or non-linear problem depending on the forward operator employed: the full, non-linear Zoeppritz equations or its linear approximations. The DCT is a linear transformation that projects an N-length signal (vector of model parameters) to an N-length vector containing the coefficients of N different cosine (base) functions. This approach concentrates most of the information of the original signal into the lower-order DCT-coefficient so that only $q<N$ coefficients can be used to accurately approximate the input signal. In the context of geophysical inversion, this means that the numerical values of these q DCT coefficients become the unknowns to be inferred

from the data. Estimating the retained DCT-coefficients reduces the parameter dimensionality and can significantly improve the computational efficiency of the inversion procedure, especially in case of non-linear problems solved through a sampling approach. We use the DCT because its energy compaction efficiency (the ability to concentrate most of the input's signal energy to few DCT coefficients) is greater than any other transformation and it is to date the most widely used transform in image and video compression standards (Wallace, 1991; Le Gall, 1991).

To draw general conclusions about the benefits provided by the DCT reparameterization of AVA inversion, we consider several synthetic inversion tests in which the observed data has been derived from actual borehole logs. First, the effect of this reparameterization on the model resolution and model uncertainty are investigated using analytical AVA inversions for which the sensitivity kernels and the posterior uncertainty can be exactly determined. Then, the DCT is used to reparametrize the DEMC and HMC inversions. In all cases, we compare the outcomes provided by the DCT reformulation and the standard model parameterization. To the best of our knowledge this paper investigates for the first time a DCT reparameterization of linear and non-linear AVA inversions and runs for the first time an HMC inversion in a DCT-reduced space.

METHODS

In this section we present the mathematical framework of the DCT and the reformulation of the Bayesian linear AVA inversion in a DCT-reduced model space. Then, we give a brief overview of the DEMC and HMC algorithms.

Bayesian linearized AVA inversion in the DCT domain

In this work, we use the DCT parameterization because this approach exhibits superior compression power over other compression methods (see Lochbühler et al. 2014). Several variants of DCT exist with slightly modified definitions, but here we employ the so-called DCT-2 formulation that is the most common. Hereafter we simply refer to the DCT-2 transformation as the DCT. The DCT is a

Fourier-related transform that uses only real numbers to express a finite signal in terms of the sum of cosine functions oscillating at different frequencies. The DCT transformation of a 1-D signal *x* of length *N* can be written as follows:

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

where δ_{k1} represents the Kronecker delta, *y* are the *N* coefficients of the transformation that fully describe the original signal *x* in the transformed DCT space, and *k* represents the order of each DCT coefficient. In matrix form, equation 1 becomes:

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

where the vectors **x** and **y** represent the original and the transformed signal, respectively, and **B** is an *N*-by-*N* matrix that contains the cosine functions (base functions) spanning the DCT space. Some examples of DCT base functions of different orders *k* are represented in Figure 1. The DCT is a linear transformation expressed by the orthonormal matrix **B** (so that $\mathbf{B}\mathbf{B}^T = \mathbf{I}$, where **I** is the identity matrix) that concentrates most of the energy of the original signal **x** in the low order DCT coefficients. This means that an approximation of the signal **x** can be obtained by considering only the first *q* DCT base functions:

$$\tilde{\mathbf{x}} = \mathbf{B}_q^T \mathbf{y}_q, \quad (3)$$

where $\tilde{\mathbf{x}}$ is the approximated signal, $\mathbf{B}_q^T$ is a partition of the matrix **B**, with *N* rows and *q* columns representing the first *q* DCT base functions, whereas the vector $\mathbf{y}_q$ contains the first *q* coefficients that multiply the base functions. These coefficients become the unknown parameters to be determined in a DCT reparameterization of the inverse problem. Note that the approximated signal $\tilde{\mathbf{x}}$ tends toward the original signal **x** as the number of considered coefficients *q* tends to *N*. If *q*=*N* the reconstructed signal $\tilde{\mathbf{x}}$ equals the original signal **x**.

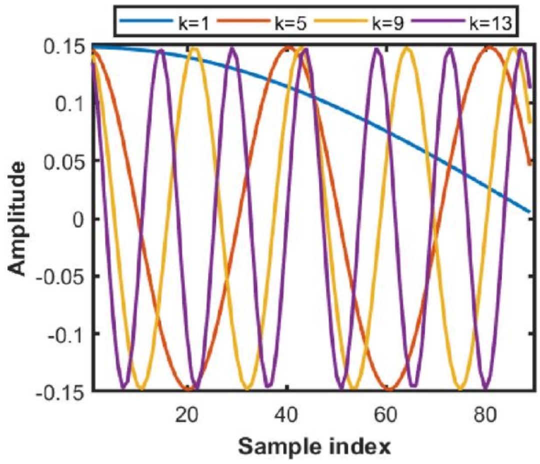

Figure 1: Some examples of DCT base functions of different orders *k*.

For the Bayesian linear AVA inversion, we employ the method proposed by Buland and Omre (2003). In this case, the model vector of $N \times 3$ rows and 1 column, represents the natural logarithm of P-wave velocity (*Vp*), S-wave velocity (*Vs*) and density (ρ) along a 1D vertical profile:

$$\mathbf{m} = \ln[Vp_1, Vp_2, \dots, Vp_N, Vs_1, Vs_2, \dots, Vs_N, \rho_1, \rho_2, \dots, \rho_N]^T. \quad (4)$$

The forward modeling of the linear AVA inversion is given by the time-interval extension of the single-interface Aki and Richards equation (Aki and Richards,1980):

$$Rpp(t,\theta) = \frac{1}{2}(1 + tan^2(\theta))\frac{\partial}{\partial t}\ln Vp(\mathrm{t}) + 4\frac{\overline{Vs}^2(t)}{\overline{Vp}^2(t)}sin^2(\theta)\frac{\partial}{\partial t}\ln Vs(t)$$

$$+\frac{1}{2}\left(1 - 4\frac{\overline{Vs}^2(t)}{\overline{Vp}^2(t)}sin^2(\theta)\right)\frac{\partial}{\partial t}\ln\rho(t)$$

$$= \alpha_{Vp}\frac{\partial}{\partial t}\ln Vp(t) + \alpha_{Vs}\frac{\partial}{\partial t}\ln Vs(t) + \alpha_\rho\frac{\partial}{\partial t}\ln\rho(t), \quad (5)$$

where t is the time, θ is the incidence angle, Rpp is the P-wave reflection coefficient, whereas $\frac{\overline{Vs}^2}{\overline{Vp}^2}$ indicates the average *Vs*/*Vp* ratio at the reflecting interfaces that can be derived, for example, from the so-called low-frequency (LF) elastic back-ground model usually estimated from well log data interpolation. If we consider the convolutional modeling and we adopt the matrix formalism, the seismic gather $\mathbf{d}$ can be computed as:

$$\mathbf{d} = \mathbf{SADm} = \mathbf{Gm} + \mathbf{n}, \quad (6)$$

where $\mathbf{S}$ is the wavelet matrix, $\mathbf{n}$ is the noise vector, $\mathbf{A}$ contains the numerical coefficients α_{Vp}, α_{Vs} and α_ρ of equation 5, $\mathbf{D}$ is the first order numerical derivative operator, and $\mathbf{G}$ is a $M \times (N \times 3)$ forward operator matrix, where M indicates the number of data points. In this context, an approximation of the elastic model can be obtained as:

$$\mathbf{m} \approx \boldsymbol{\mu}_\mathbf{m} + \mathbf{K}_q^T \mathbf{r}, \quad (7)$$

where $\boldsymbol{\mu}_\mathbf{m}$ is the mean elastic model (e.g. the elastic LF model), and $\mathbf{K}_q$ is a block diagonal matrix with $N \times 3$ columns and $q \times 3$ rows given by:

$$\mathbf{K}_q = \begin{bmatrix} \mathbf{B}_q & 0 & 0 \\ 0 & \mathbf{B}_q & 0 \\ 0 & 0 & \mathbf{B}_q \end{bmatrix}, \quad (8)$$

and again $\mathbf{B}_q$ contains the first q DCT base functions, whereas the parameter vector in the reduced $(q \times 3)$-D space is represented by the vector $\mathbf{r}$ in which the first q elements are the coefficients that pertain to *Vp*, the second q elements are associated to *Vs*, and the last q elements pertain to the density. In geophysical inversions this means that the DCT allows for a reduction of the $(N \times 3)$-D full, elastic space to a $(q \times 3)$-D space with $q<N$.

In all the following examples we simply consider a Gaussian elastic prior model $p(\mathbf{m})$, with mean vector $\boldsymbol{\mu}_\mathbf{m}$ and a-priori covariance matrix $\mathbf{C}_\mathbf{m}$:

$$p(\mathbf{m}) = \mathcal{N}(\mathbf{m}; \boldsymbol{\mu}_\mathbf{m}, \mathbf{C}_\mathbf{m}), \quad (9)$$

where $\mathcal{N}$ represents the Gaussian distribution defined over the elastic model space $\mathbf{m}$. Note that to mitigate the ill-conditioning of AVA inversion, the prior covariance matrix $\mathbf{C}_\mathbf{m}$ expresses both the covariance of the elastic properties and their vertical variogram (see Buland and Omre, 2003). Since the DCT is a linear transformation, the prior model in the DCT space is still Gaussian:

$$p(\mathbf{r}) = \mathcal{N}(\mathbf{r}; \boldsymbol{\mu}_\mathbf{r}, \mathbf{C}_\mathbf{r}), \quad (10)$$

with mean vector and covariance matrix equal to:

$$\boldsymbol{\mu}_\mathbf{r} = \mathbf{0}, \quad (11)$$

$$\mathbf{C}_\mathbf{r} = \mathbf{K}_q \mathbf{C}_\mathbf{m} \mathbf{K}_q^T. \quad (12)$$

According to equation 7, note that a null a-priori mean model in the DCT space corresponds to an elastic prior model equal to $\boldsymbol{\mu}_{\mathbf{m}}$. By combining equation 6 with equation 7 we derive the linear forward modeling in the DCT space:

$$\mathbf{d} = \mathbf{G}\left(\mathbf{K}_q^T\mathbf{r} + \boldsymbol{\mu}_{\mathbf{m}}\right) + \mathbf{n} = \mathbf{Pr} + \mathbf{G}\boldsymbol{\mu}_{\mathbf{m}} + \mathbf{n}, \quad (13)$$

in which the matrix $\mathbf{P}$ is a $M \times (q \times 3)$ matrix, computed considering only the first q DCT coefficients for each elastic property, and $\mathbf{n}$ is again the noise vector. As an example, Figure 2 compares the forward modeling matrices **G** and **P**, associated with the elastic and the DCT space.

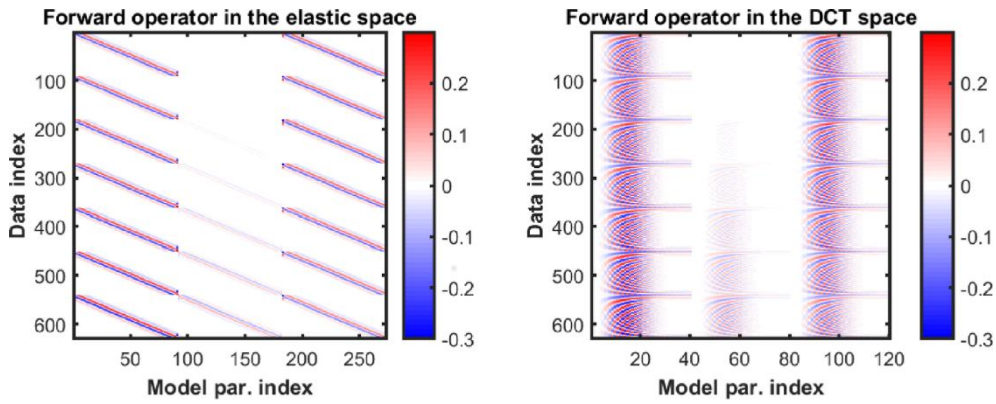

Figure 2: a) Example of forward modeling matrix in the elastic space for 630 data points and 273 model parameters. b) The projection onto the DCT space of the forward modeling matrix shown in a) if only the first 40 DCT coefficients per elastic property are considered (q=40).

If the forward operator is linear, the posterior model $p(\mathbf{r}|\mathbf{d})$ in the DCT space is still Gaussian with posterior mean and covariance given by:

$$\boldsymbol{\mu}_{\mathbf{r|d}} = \boldsymbol{\mu}_{\mathbf{r}} + (\mathbf{P}^T\mathbf{C}_{\mathbf{d}}^{-1}\mathbf{P} + \mathbf{C}_{\mathbf{r}}^{-1})^{-1}\mathbf{P}^T\mathbf{C}_{\mathbf{d}}^{-1}(\mathbf{d} - \mathbf{P}\boldsymbol{\mu}_{\mathbf{r}}), \qquad (14)$$

$$\mathbf{C}_{\mathbf{r|d}} = (\mathbf{P}^T\mathbf{C}_{\mathbf{d}}^{-1}\mathbf{P} + \mathbf{C}_{\mathbf{r}}^{-1})^{-1}. \quad (15)$$

In equation 14 note that we directly invert for the reflectivity contrasts because the low-frequency elastic model is simply added when we project the results back onto the elastic space. Indeed, to be of practical utility, the final solution of the AVA-DCT inversion must express the posterior distribution of elastic properties **m** conditioned upon the observed data **d** and the reduced model **r**: $p(\mathbf{m}|\mathbf{d},\mathbf{r}) = \mathcal{N}\left(\mathbf{m}; \boldsymbol{\mu}_{\mathbf{m|d,r}}, \mathbf{C}_{\mathbf{m|d,r}}\right)$. For a linear inversion, the statistical properties of such posterior model can be computed by projecting the mean and covariance of the $p(\mathbf{r}|\mathbf{d})$ distribution onto the elastic space:

$$\boldsymbol{\mu}_{\mathbf{m}|\mathbf{d},\mathbf{r}} = \boldsymbol{\mu}_{\mathbf{m}} + \mathbf{K}_q^T \boldsymbol{\mu}_{\mathbf{r}|\mathbf{d}}, \qquad (16)$$

$$\mathbf{C}_{\mathbf{m}|\mathbf{d},\mathbf{r}} = \mathbf{K}_q^T \mathbf{C}_{\mathbf{r}|\mathbf{d}} \mathbf{K}_q. \quad (17)$$

For what concerns the sensitivity analysis kernels, the model resolution matrix in the DCT space is defined as (Menke 2018):

$$\mathbf{R}_{\mathbf{r}} = (\mathbf{P}^T \mathbf{C}_{\mathbf{d}}^{-1} \mathbf{P} + \mathbf{C}_{\mathbf{r}}^{-1})^{-1} \mathbf{P}^T \mathbf{C}_{\mathbf{d}}^{-1} \mathbf{P}. \quad (18)$$

The corresponding model resolution matrix in the elastic space given the reduced model $\mathbf{r}$, can be obtained by a projection of the $\mathbf{R}_{\mathbf{r}}$ matrix (Menke, 2018):

$$\mathbf{R}_{\mathbf{m}|\mathbf{r}} = \mathbf{K}_q^T \mathbf{R}_{\mathbf{r}} \mathbf{K}_q. \quad (19)$$

The DEMC and HMC inversions

For the sake of coherency, the numerical inversions consider the same Gaussian prior model (i.e. log Gaussian distributed elastic properties) used by the linear inversion. However, one outstanding benefit of Monte Carlo methods is that they can also manage non-parametric prior models. Both the DEMC and HMC inversion employ a convolutional forward modeling based on the exact Zoeppritz equations. These two approaches make use of the well-known Metropolis-Hasting rule to define the probability of moving from the current state of the chain $\mathbf{e}$ to the proposed state $\mathbf{e}'$. For example, in the implemented DEMC this rule can be written as follows:

$$\alpha = p(\mathbf{e}'|\mathbf{e}) = \min\left[1, \frac{p(\mathbf{e}')}{p(\mathbf{e})} \times \frac{p(\mathbf{d}|\mathbf{e}')}{p(\mathbf{d}|\mathbf{e})}\right], \quad (20)$$

where $\frac{p(\mathbf{e}')}{p(\mathbf{e})}$ and $\frac{p(\mathbf{d}|\mathbf{e}')}{p(\mathbf{d}|\mathbf{e})}$ are the so-called prior and proposal ratios, respectively, whereas in the standard inversion approach $\mathbf{e}$ represents the elastic model $\mathbf{m}$, while in the DCT inversion $\mathbf{e}$ represents the vector $\mathbf{r}$. In both the DEMC and HMC inversions the ensemble of models sampled after the burn-in period is used to numerically compute the statistical properties (e.g. mean, mode, standard deviations, marginal densities) of the PPD. For the inversions running in the DCT space, the a-priori mean vector and covariance matrix are analytically derived from the statistical properties of the elastic prior model through equations 11 and 12. For these inversions, the sampled models are projected back onto the elastic space (see equation 7) just before the forward modeling phase (see equation 13) that gives the predicted data needed to compute the likelihood value. The elastic posterior model can be numerically derived from the ensemble of DCT models collected during the sampling stage, after projection onto the *Vp*-*Vs*-density space.

The DEMC is an advanced MCMC algorithm that uses a population of different chains that are evolved using differential evolution principles (Ter Braak, 2006). In more detail, such differential evolution principles are used to generate multivariate proposals for each DEMC chain: let the d-vector $\mathbf{s}$ represent the state of a single chain, then at each iteration t-1, the Q chains define a population $\mathbf{S} = \{\mathbf{s}_{t-1}^1, \mathbf{s}_{t-1}^2, \dots, \mathbf{s}_{t-1}^Q\}$ which corresponds to an $Q \times d$ matrix. Multivariate proposals $\mathbf{s}_p$ are defined as:

$$\mathbf{s}_p^i = \mathbf{s}_{t-1}^i + \gamma\left(\mathbf{s}_{t-1}^a - \mathbf{s}_{t-1}^b\right) + \epsilon, \quad a \neq b \neq i \quad (21)$$

where i is the index of the current chain, γ denotes the jump rate, a and b are integer values drawn without replacement from {1,…,i-1,i+1,…, Q}, and ϵ represents a small random perturbation drawn from a normal distribution with a small standard deviation σ tailored to the problem at hand: $\epsilon = \mathcal{N}(0, \sigma)$. Each proposal is accepted with Metropolis probability (see equation 20). If the proposal is accepted $\mathbf{s}_t^i = \mathbf{s}_p^i$, otherwise $\mathbf{s}_t^i = \mathbf{s}_{t-1}^i$. The optimal γ parameter depends on the model dimensionality and is usually set to $\gamma = 2.38/2\mathrm{d}$. Besides, with a 10% probability the value of $\gamma = 1$ allows for mode-jumping which is a significant strength of DEMC compared with more conventional MCMC methods (i.e. random walk Metropolis or adaptive Metropolis). Additional and more detailed theoretical insights into the DEMC, together with a Matlab implementation can be found in Vrugt (2016).

Finally, HMC considers a model as a particle that moves from its current position to a new position along a given trajectory that is uniquely determined by the mass matrix (**M**), the kinetic energy (*K*), and the potential energy (*U*). In particular, the potential energy is equal to the negative natural logarithm of the posterior distribution and is interpreted as the misfit function. For a *d*-dimensional parameter space, HMC determines the kinetic energy by introducing an auxiliary variable (momentum variable) $\mathbf{p}$ that is defined over a *d*-dimensional space:

$$K(\mathbf{p}) = \frac{1}{2}\, \mathbf{p}^{T}\, \mathbf{M}^{-1}\mathbf{p}. \quad (22)$$

After defining the kinetic and potential energies, the model **e** moves across the 2×*d*-phase space according to Hamilton's equations:

$$\frac{d\mathbf{e}_i}{d\tau} = \frac{\partial K}{d\mathbf{p}_i}, \qquad \text{with } i = 1, 2, \dots, d, \quad (23)$$

$$\frac{d\mathbf{p}_i}{d\tau} = -\frac{\partial U}{d\mathbf{e}_i}, \quad \text{with } i = 1, 2, \dots, d, \quad (24)$$

where τ indicates the artificially introduced time variable, whereas the vectors **e** and **p** define the so-called phase space. Note that the kinetic energy and the mass matrix are artificially introduced as auxiliary quantities and allow for the inclusion of the derivative information of the misfit function into the sampling framework. Indeed, the right term of equation 24 contains the partial derivative of the potential energy (i.e. the misfit function) with respect to the considered model **e**. For each HMC iteration, the momentum vector is drawn from the normal distribution $p(\mathbf{p}) = \mathcal{N}(\mathbf{p}; 0, \mathbf{M})$, then the proposed model is found by numerically solving the Hamilton's equations starting from the current state of the chain.

In this work the potential energy is defined as (Fichtner et al. 2019):

$$U(\mathbf{e}) = \frac{1}{2}\big(\mathbf{d} - G(\mathbf{e})\big)^{T}\mathbf{C}_{d}^{-1}\big(\mathbf{d} - G(\mathbf{e})\big) + \frac{1}{2}\big(\mathbf{e} - \mathbf{e}_{prior}\big)^{T}\mathbf{C}_{\mathrm{e}}^{-1}\big(\mathbf{e} - \mathbf{e}_{prior}\big), \quad (25)$$

where *G* represents the non-linear, exact Zoeppritz equations, $\mathbf{C}_d$ is the data covariance matrix, **d** is the observed data (partial angle stacks at different incidence angles), $\mathbf{e}_{prior}$ is the prior model (either in the elastic or in the DCT space) with prior covariance matrix given by $\mathbf{C}_\mathrm{e}$. Note that the potential energy $U(\mathbf{e})$ in equation 25 equals the negative natural logarithm of the posterior probability of Bayesian inversions (Tarantola, 2005). The mass matrix and the number of time integration steps (L) in the phase space are two crucial hyperparameters that must be accurately set to ensure the convergence of the HMC algorithm. For setting the L parameter we follow the approach of Mackenze (1989) that in each iteration randomly draws the number of time integration steps from a previously defined uniform distribution, that in our case is $U(5,10)$. On the other hand, following Fichtner et al. (2019) we compute the mass matrix as a local approximation (around the currently evaluated model) of the inverse of the posterior covariance matrix:

$$\mathbf{M} = \mathbf{J}^T \mathbf{C}_d^{-1} \mathbf{J} + \mathbf{C}_\mathrm{e}^{-1}, \quad (26)$$

where $\mathbf{J}$ is the numerically computed Jacobian matrix that expresses the partial derivative of the data with respect to model parameters. We use a forward finite difference scheme to compute the Jacobian and in this context the model reparameterization provided by the DCT is particularly useful to reduce the number of forward evaluations needed for the Jacobian computation. Indeed, the HMC algorithm was developed for problems in which the derivative of the target probability density can be computed quickly. For example, if L is the number of integration steps for solving the Hamilton's equations and for a d-dimensional model space, the number of forward evaluations per iteration needed to numerically compute the Jacobian with a forward finite difference scheme is equal to $(d+1)\times L$. In this context, the computational cost of the HMC sampling exponentially increases with the number of unknowns.

Additional theoretical details about the HMC method can be found in Neal (2011) and Betancourt (2017), while Sen and Biswas (2017), Fichtner et al. (2019), Fichtner and Zunino (2019), Aleardi and Salusti al. (2020), and Aleardi et al. (2020) presented some applications of this method to solve geophysical inverse problems.

APPLICATIONS AND RESULTS

Linear AVA inversions

We now discuss the results provided by linear Bayesian AVA inversions running in the full and the reduced DCT space. We consider two different examples in which the true models have been derived from logged elastic properties recorded along different wells (hereafter called well A and well B) drilled in the same area and through similar geological formations. The covariance of the elastic prior model is the same in all the following examples and has been derived from borehole information extracted from three other wells drilled in the same zone. The variance of the DCT prior model has been analytically computed by projecting the prior variance in the elastic space onto the DCT space (see equation 12). The prior covariance is assumed to be stationary along the entire inverted vertical profile. The a-priori mean model in the elastic space is equal to a heavily low-pass filtered version of the true model. This prior corresponds to a null mean vector in the DCT space (see equation 11). The observed data are computed through equation 6 and considering a sampling interval of 1 ms and a 55-Hz Ricker wavelet as the source signature. The observed data vector is contaminated with random Gaussian noise with a standard deviation of 0.03.

For well A the true model is formed by 91 time samples of *Vp*, *Vs*, and density for a total number of $91 \times 3 = 273$ elastic parameters to be determined. For well B the true model includes 140 time samples of *Vp*, *Vs* and density, thus resulting in $140 \times 3 = 420$ elastic parameters. Figure 3 shows the explained variability of the logged *Vp*, *Vs*, and density values along well A and well B as the number of the considered DCT coefficients increases. For both wells, we note that only 25 coefficients per elastic property (q=25) explain more than the 90% of the variability and that the 95%, approximately, of the total variability, is explained by only 50 coefficients. Therefore, the same number of DCT base functions can conveniently be used to compress signals with different lengths, but with similar statistical properties (i.e. vertical variability, variance).

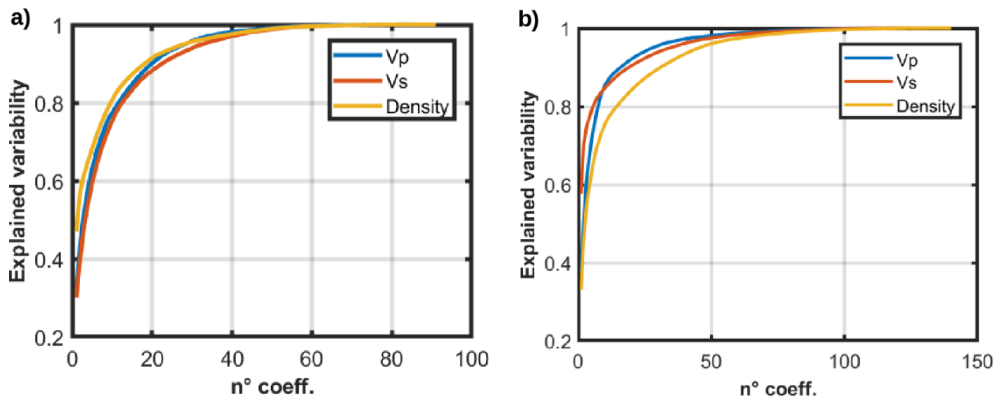

Figure 3: Explained variability of the true *Vp*, *Vs,* and, density profiles along well A (a), and B (b) as the number of considered DCT coefficients increases.

We start by describing the inversion tests on well A. Figure 4 shows the results in the DCT space if 20 and 40 coefficients are considered. In both cases the inversion reliably predicts the coefficient values associated with the true elastic model. Moreover, for the low-order coefficients, which describe a major part of the original variability of the elastic profile, the true model usually lies in the 95% posterior confidence intervals. In Figure 4b we observe that the match between predicted and actual coefficients decreases as the coefficient order increases, and that the predictions of the high-order coefficients (i.e. higher than 27) are primarily guided by the prior information (i.e. the prior and posterior mean and variance are very similar). Indeed, high order coefficients explain minimal, high-frequency variations of the elastic profile that are not constrained by the seismic data. In other terms, these parameters are associated with the lowest singular values of the inversion kernel and span the null-space of solutions. We can also observe that the posterior uncertainties (evidenced by the 95% confidence intervals) increase as we move from the inversion with q=20 to the inversion with q=40 (Figure 4c).

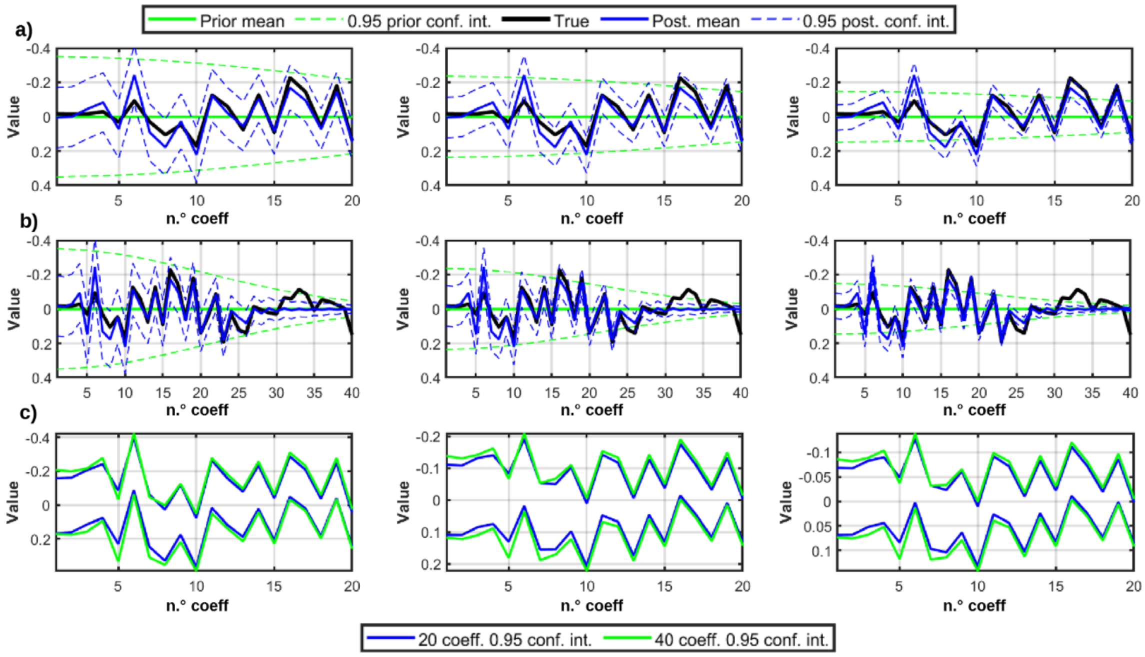

Figure 4: Comparison of the true model, prior model, and posterior model for linear AVA inversions in the DCT domain. a) 20 DCT coefficients per elastic property are considered (q=20). b) 40 DCT coefficients per elastic property are considered (q=40). c) Comparison between the confidence intervals estimated by the inversions shown in a) and b).

Figure 5 compares the outcomes of a standard Bayesian (SB) inversion (Buland and Omre, 2003) and of two DCT inversions in which 15 and 40 coefficients per elastic property are considered. In these cases, we are reducing the full 273-D elastic space to a 45-D and a 120-D parameter space, respectively. For q=15 the AVA inversion is not able to reliably reproduce the actual elastic property contrasts and for this reason the observed data is not properly matched. Differently, 40 coefficients per elastic property provide final estimates (in terms of posterior mean and variance) close to the actual elastic property profiles, and congruent with the predictions of the SB inversion running in the full, elastic space. Figure 6 highlights that the DCT inversion with q=15 underestimates the posterior uncertainties, while the inversion with q=40 yields confidence intervals equal to those provided by the AVA inversion running in the unreduced space.

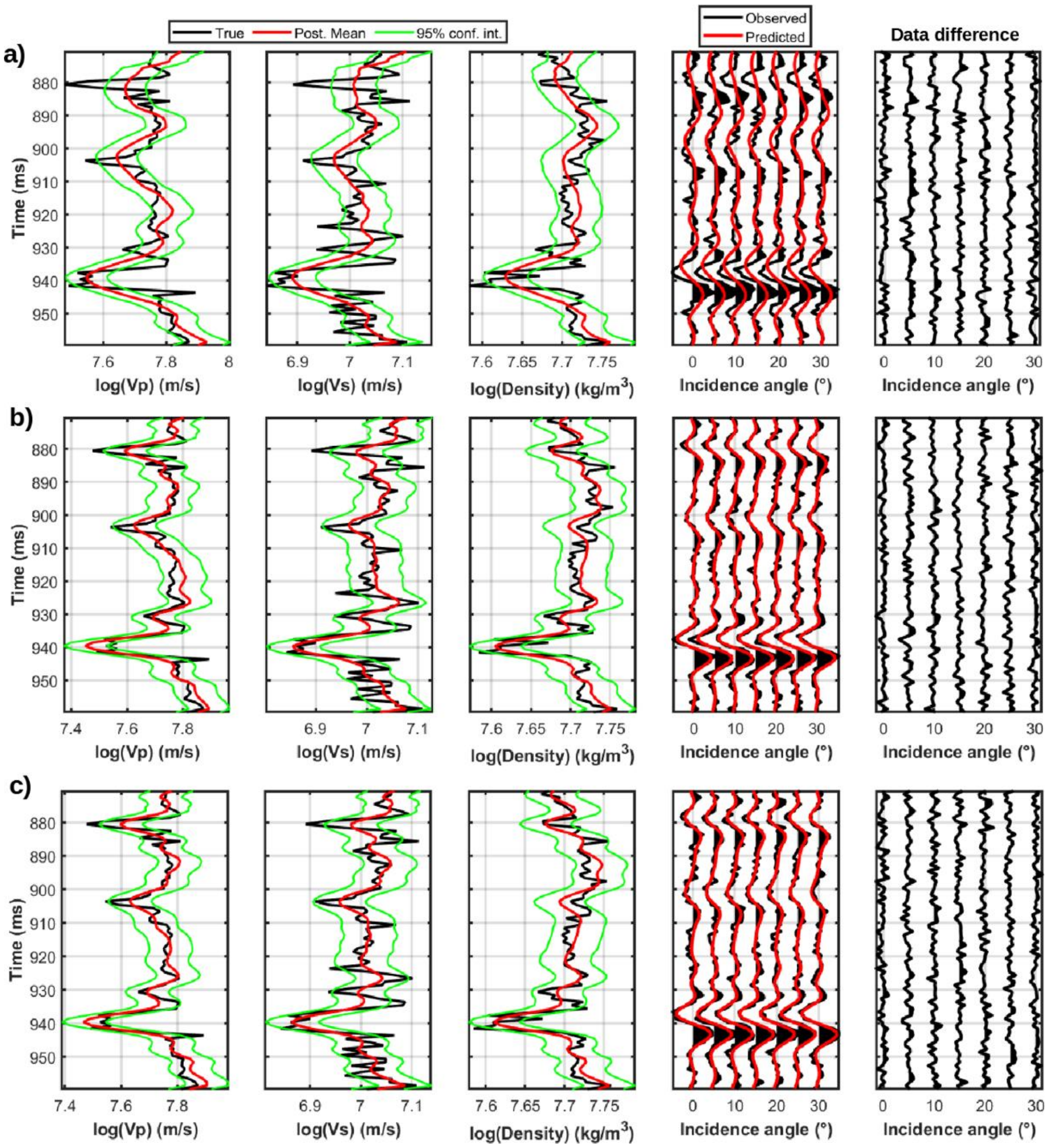

Figure 5: a) DCT inversion results projected onto the elastic space for *q*=15. b) As in a) but for 40 coefficients per elastic property. c) SB inversion results in the full, elastic space. The predicted data correspond to the seismic gathers computed on the a-posteriori mean models. The amplitude scale is the same for the seismic gathers and the data differences.

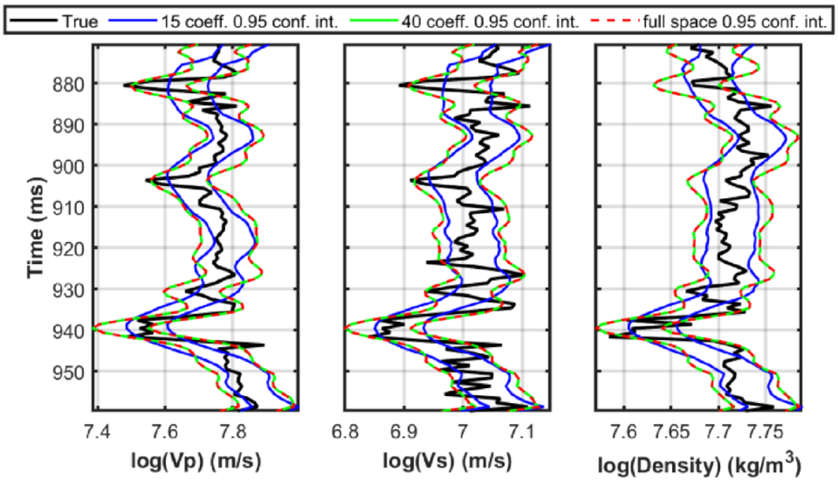

Figure 6: Comparison between the 95% confidence intervals for the DCT inversions running with *q*=15 and *q*=40, and for the AVA inversion running in the unreduced elastic space.

For a more quantitative assessment of the influence played by the number of considered DCT base functions in the posterior model resolution and covariance, we perform a sensitivity analysis of the inversion kernel and we also compute the coverage ratio of the inversion results for different *q* values. Figure 7a shows the ratio between the posterior variances provided by the SB and DCT inversions as the number of DCT coefficients per elastic property (*q*) increases. For the DCT inversion, as expected, we observe that for *q*<38 the posterior variance in the reduced space underestimates the posterior variance of the full model problem, while for *q*>38 the posterior variance of the reduced problem equals the posterior variance of the full model problem. The 0.95 coverage ratios (i.e. the probability that the 95 confidence interval contains the true model parameter values) provided by the SB and DCT inversion confirm that 40 coefficients are more than enough to guarantee the same level of accuracy of a standard inversion running in the full space (Figure 7b). Finally, Figure 7c compares the diagonal entries of the model resolution matrices for the SB and DCT inversions. Again, we note that 40 DCT coefficients per elastic property guarantee the same resolution of a linear AVA inversion running in the unreduced, elastic space. As expected, the model resolution decreases moving from the *Vp* to *Vs* and to density. Indeed, the *Vp* is the parameter that mostly influences the observed seismic amplitudes, while the density is the parameter that exerts the minor influence on the P-wave reflection coefficients. These results confirm that for well A, only 40 DCT coefficients for each elastic property allow for a substantial reduction of the model dimensionality (from the unreduced 237-D elastic space to a 120-D DCT space), while still guaranteeing model resolution and uncertainty estimations similar to those estimated in the full model space.

We now briefly discuss the results we obtain on well B. Figure 8 compares the outcomes of an SB inversion with those yielded by DCT inversions with different compressions of the elastic parameter space. For 15 DCT coefficients we obtain underpredicted posterior uncertainties, a poor match between the true and the predicted properties, and underfit between predicted and observed data. On the contrary, the SB and the DCT inversion running with q=40 provide similar estimates of the posterior mean and posterior variance. In Figure 9 the ratio between the posterior variances estimated in the DCT and in the full space, the 95% coverage ratio, and the model resolution, confirm the conclusions drawn in the previous example on well A: 40 coefficients can successfully recover the vertical variability of the true *Vp*, *Vs* and density profiles and ensure final accuracy and resolution similar to the SB inversion.

Finally, the two examples on wells A and B show that the optimal number of DCT coefficients is independent of the number of samples forming the true elastic property profiles but depends on the actual vertical variability (or in other terms the variance) of the true model. This means that elastic model vectors with different lengths but with similar vertical variability can be conveniently approximated by the same number of DCT coefficients. Therefore, in a DCT inversion the same number of unknowns can be used to infer the elastic property values along vertical intervals of different lengths.

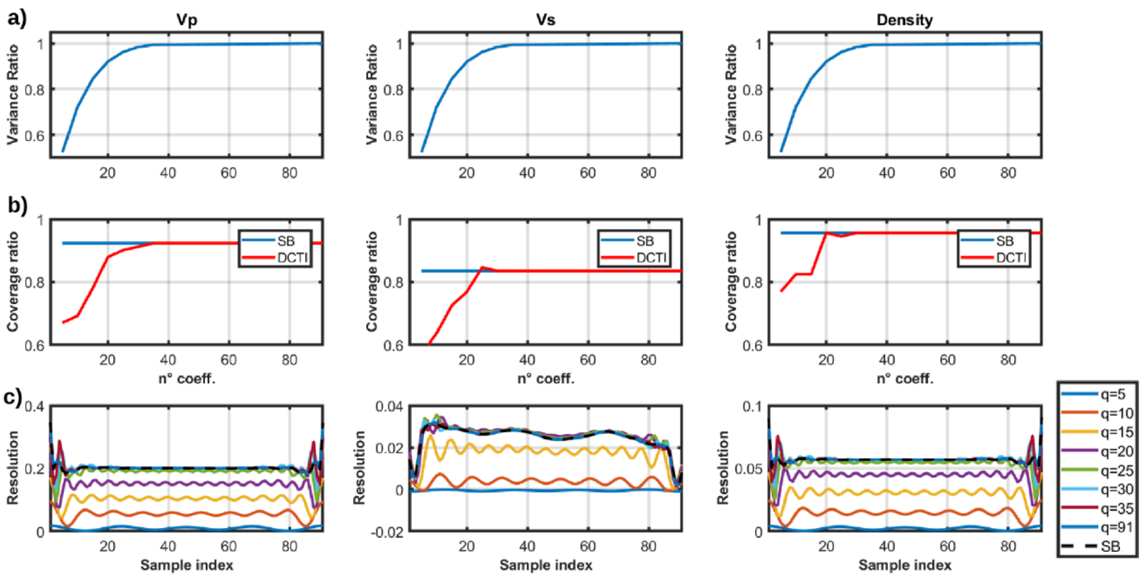

Figure 7: a) Ratio between the posterior variances estimated by DCT inversions for different numbers of base functions and the posterior variance estimated by the SB inversion. b) Comparison of the 0.95 coverage ratios for the DCT and SB inversions. c) Diagonal elements of the model resolution matrices for the SB inversion and DCT inversions with different number of base functions. The acronyms DCTI and SB stand for DCT inversion and standard Bayesian inversion, respectively.

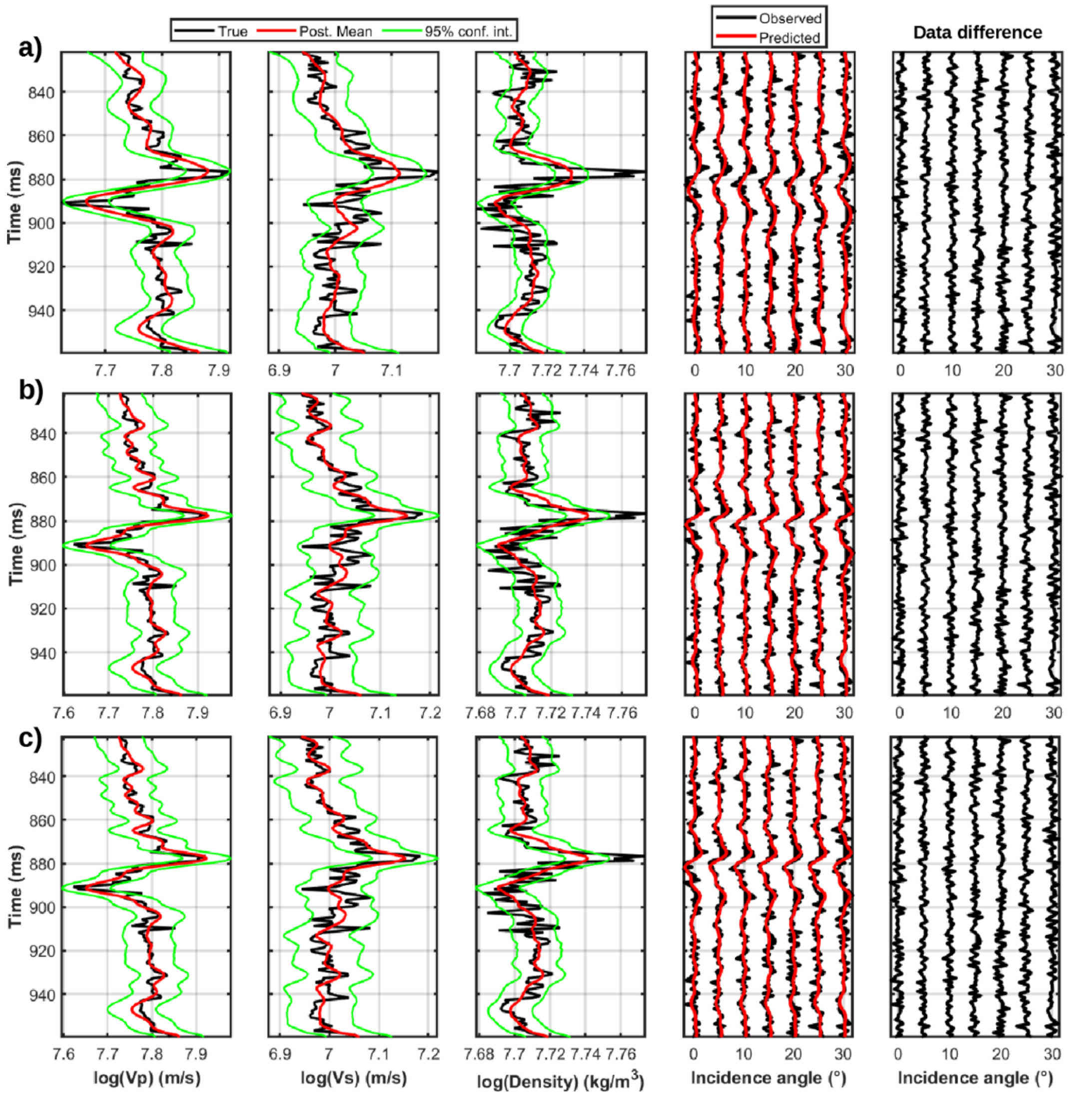

Figure 6: As in Figure 5 but for well B.

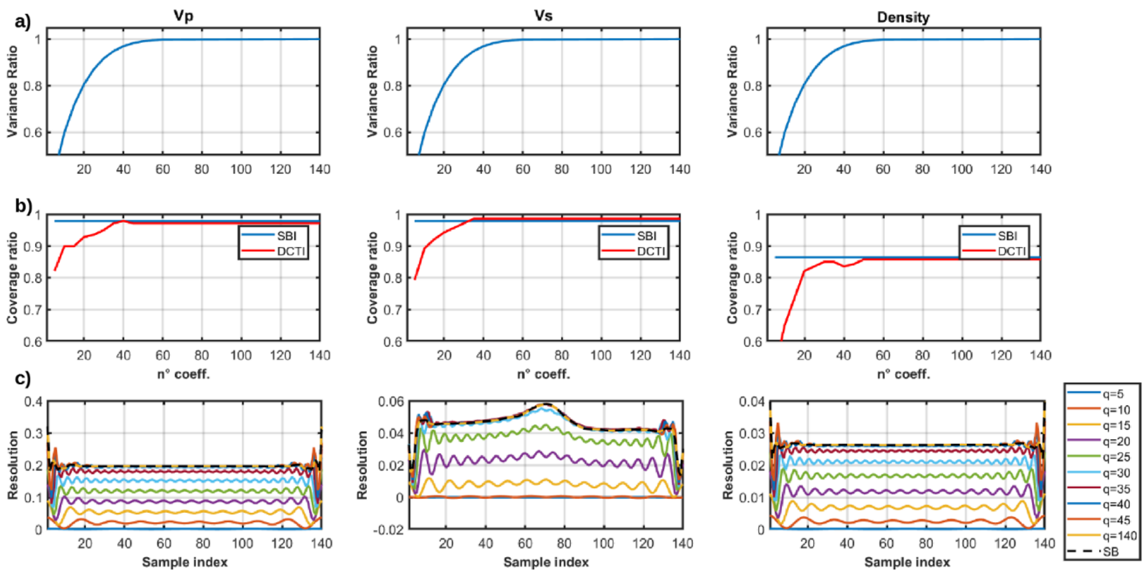

Figure 7: As in Figure 7 but for well B.

DEMC inversion

In all the examples discussed in this section we use the exact, non-linear, Zoeppritz equations as the forward modeling operator for computing both the observed data and the data predicted on each sampled model. To this end, we consider a 55-Hz Ricker wavelet and an angle range between 0-40 degrees. Gaussian random noise with a standard deviation of 0.03 contaminates the observed amplitudes values. The prior models for the SB and DCT inversions are the same previously used in the linear examples. The starting models for each DEMC chain are randomly generated according to the Gaussian prior model.

We start by describing the results on well A. In this case the DEMC inversion employs 20 chains running for 30000 iterations and with a burn-in period of 10000. If only 20 coefficients per elastic property are used (Figure 10a), we get a PPD that underestimates the posterior uncertainty, and a posterior mean with low vertical resolution. Differently, 40 coefficients (Figure 10b) ensure model resolution and uncertainty assessment comparable to that yielded by the DEMC running in the unreduced space (Figure 10c). Figure 11 shows the evolution of the negative log-likelihood values in the first 10000 iterations for the examples depicted in Figure 10. The MCMC chains running in the DCT space always show faster convergence toward the stationary regime than the DEMC inversion running in the full space. In more detail the DEMC-DCT with q=20 shows the fastest convergence rate, although the underparameterization of the model generates underfitting with the observed data (i.e. the chains converge to higher negative log-likelihood values). If q=40 the DEMC-DCT algorithm attains the stationary regime within 2000 iterations approximately, while more than 5000 iterations are needed by the standard DEMC. Note that the DEMC-DCT with q=40 guarantees the same level of data fitting of the standard DEMC, as demonstrated by the same negative log-likelihood values reached by the two inversions. These results show that 40 DCT coefficients constitutes a good compromise between model resolution, data prediction, and accuracy of the final solution.

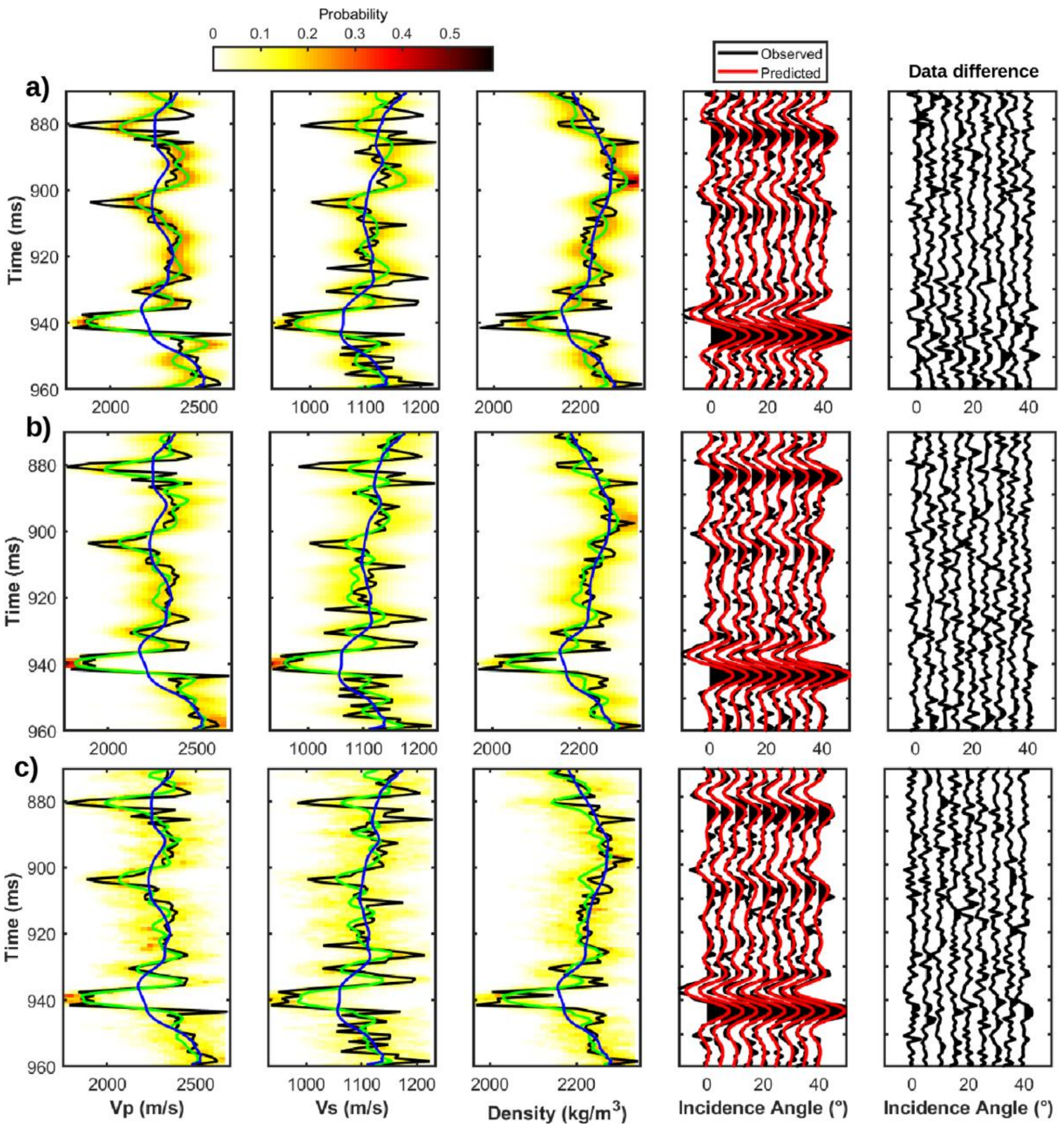

Figure 8: Inversion results on well A. a)-b) DEMC-DCT inversions running with 20 and 40 coefficients for each elastic property, respectively. c) Standard DEMC inversion results. The black, blue, and green lines represent the true, prior, and posterior mean models, respectively, while the colormap codes the posterior probability values. The predicted data correspond to the seismic gathers computed on the a-posteriori mean models. The amplitude scale is the same for the seismic gathers and the data differences. In a) note the underestimation of the posterior uncertainty that is evidenced by the darker colors of the PPD.

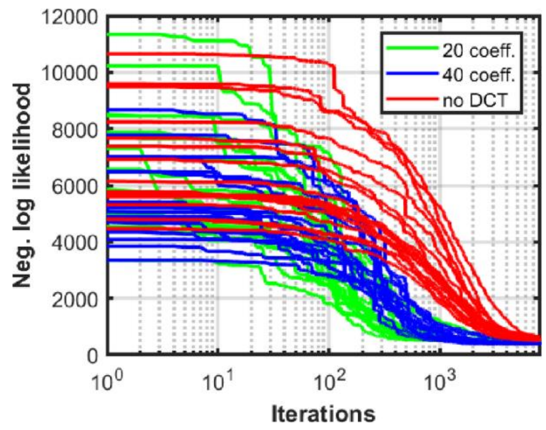

Figure 11: Examples of evolutions of the negative log-likelihood within the first 10000 iterations for the standard DEMC sampling (red lines) and for the DEMC-DCT running with q=20 and q=40.

For a more quantitative demonstration of the faster convergence achieved in the DCT space, we compare for some elastic parameters the evolutions of the potential scale reduction factor (PSRF) computed on the elastic models sampled by the DEMC running with q=40 and by the DEMC running in the full elastic space. We remind that the PSRF quantifies for each model parameter the difference between the "within-walk" and "between-walk" estimated variances (Gelman et al. 1995). The PSRF decreases to 1 as the number of drawn samples tends to infinite. A high PSRF value indicates that for the considered model parameter the variance within the walks is small compared to that between the walks and that additional iterations are needed to converge to a stable distribution. Usually, a PSRF lower than 1.2 for a given unknown proves that convergence has been achieved for that model parameter. Figure 12 shows that the DEMC-DCT algorithm is characterized by faster convergence toward a stable posterior than the standard DEMC. Note that for the considered parameters the standard DEMC never reaches reliable PPD estimations within the first 20000 iterations. Conversely, the DEMC-DCT always converges to a stable posterior model. In practical terms, this means that the DEMC-DCT needs a much lower number of forward modeling evaluations to attain stable posterior estimations than the standard DEMC.

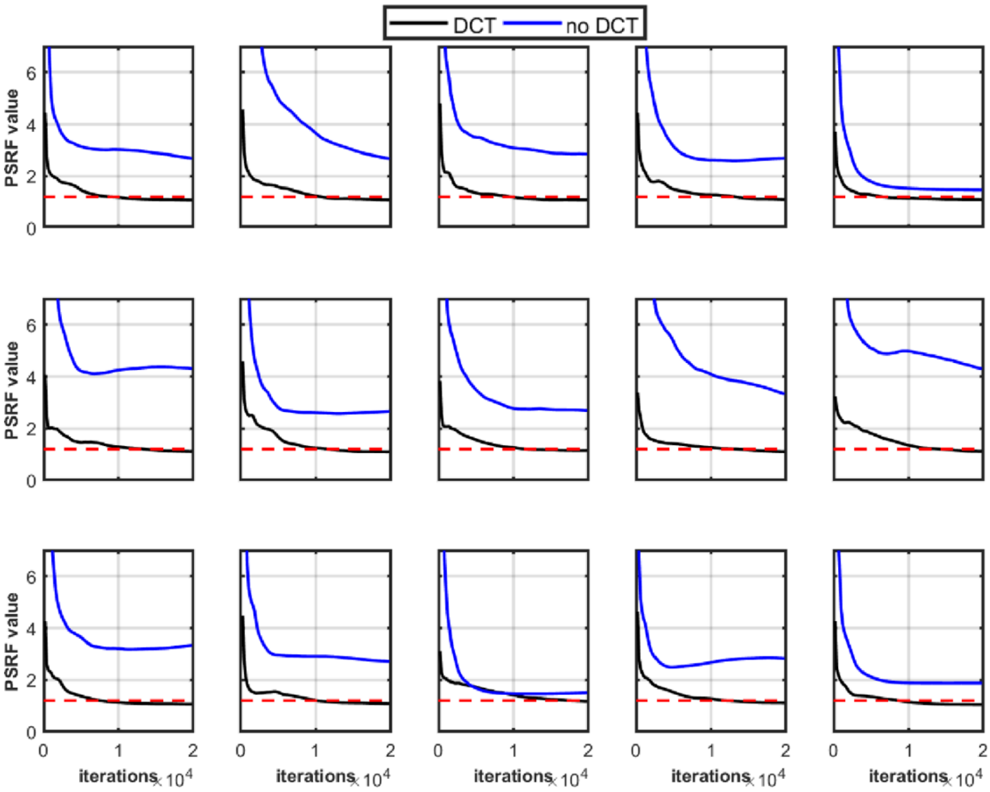

Figure 12: Close-up over the first 20000 iterations that shows examples of PSRF evolution for some model parameters. Top, central, and bottom rows refer to five *Vp*, *Vs*, and density parameters, respectively, evenly extracted along the inverted time interval. The red dashed lines at 1.2 indicate the threshold of convergence.

We now briefly discuss the results on well B. In this case to better highlight the benefit provided by the DCT transformation we decrease the number of MCMC iterations to 20000. The higher dimensionality of the elastic space (now comprising 420 parameters to be estimated) increases the differences between the rates of convergence of the different inversion approaches. The standard DEMC is severely affected by the curse of dimensionality and reaches the stationary regime well beyond the 15000 iterations (Figure 13). Conversely, only 5000 iterations are needed by the DEMC-DCT to reach the stationarity. This means that the algorithm running in the unreduced model space requires a much higher number of sampled models to attain a stable posterior distribution than the algorithm running in the reduced DCT space. In other terms, the standard DEMC required a much higher number of forward evaluations (and then a much higher computational effort) than the DEMC-DCT. The scattering in the final PPD estimated by the standard DEMC (Figure 14) is a direct consequence of the lack of convergence of the algorithm toward the stationary regime and a stable PPD. In Figure 14b the similarity between the prior and the posterior mean models (especially for the density parameter) further demonstrates that the standard DEMC needs more iterations to accurately sample the posterior model. On the contrary, the DEMC-DCT provides accurate posterior assessments and reliable predictions also within the limited number of iterations we employ.

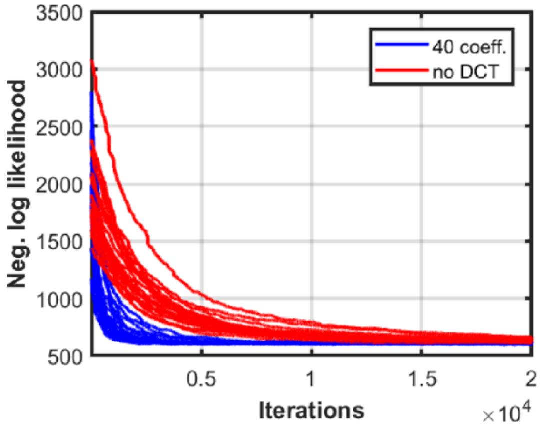

Figure 13: Evolution of the negative log-likelihood values for the different chains in the example on well B.

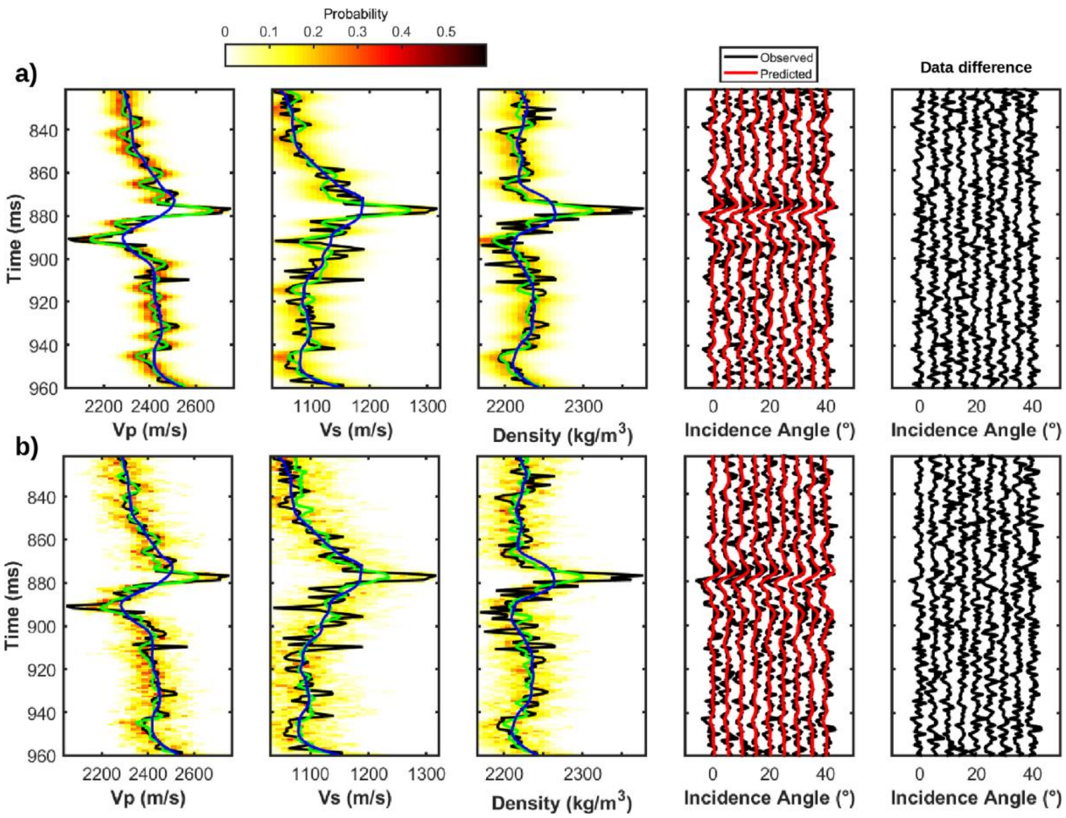

Figure 14: Inversion results on well B. a) DEMC-DCT inversion running with 40 coefficients for each elastic property. b) Standard DEMC inversion results. The black, blue, and green lines depict the true, prior, and posterior mean models, respectively, while the colormap codes the posterior probability values.

HMC inversion

For the sake of brevity here we limit the attention to the well A, but similar conclusions would have been drawn for the well B. Similarly to the previous DEMC experiments, the observed data have been computed from the true model and applying the Zoeppritz equations within an angle range of 0-40 degrees and with a 55-Hz Ricker as the source wavelet. Gaussian random noise with a standard deviation of 0.03 has been added to the synthetic, noise-free, observed seismic gather. In all the inversions that follow we use a single HMC chain running for 10000 iterations with a burn-in of 100. The prior models for the standard and DCT inversions are the same used in the linear examples. The starting points of the HMC sampling correspond to the a-priori mean models.

We start by comparing the predicted elastic models provided by a standard HMC inversion running in the full space with those yielded by HMC-DCT inversions running with different q values (Figure 15). We note that for q=20 (corresponding to a 60-D parameter space) the inversion satisfactory captures the vertical variability of the elastic properties, but the major elastic contrasts are slightly underestimated as are the uncertainty affecting the recovered model (compare Figure 15a with Figures 15b-c). As expected, the underprediction of the elastic contrasts generates underfitting with the observed seismic data. Differently, if we employ 40 DCT coefficients (q=40, thus corresponding to a 120-D parameter space) the HMC-DCT and the standard HMC inversion provide congruent results with similar posterior mean models, posterior uncertainties, and similar matches between observed and predicted seismic amplitudes.

For the HMC inversion with q=40 we represent the results in the DCT space (Figure 16). For the low-order coefficients, which are the parameters better constrained by the data, we observe that the true model usually lies in the posterior 95% confidence interval. Differently, the accuracy of the results decreases as the coefficient order increases, and also moving from the coefficients pertaining to *Vp*, to *Vs*, and to density. The latter characteristic is related to the different influences played by the elastic properties in determining the seismic amplitudes, while the former indicates that the high-order coefficients are not informed by the data. This is also confirmed by Figure 17 that compares the true, prior, and posterior models for 12 DCT coefficients out of 120. We note that the posterior model tends to the prior as the order of the considered DCT coefficient increases. This proves that the estimation of these parameters from the data is a hopelessly ill-conditioned problem, and for this reason the prediction is mainly driven by the a-priori information infused into the inversion.

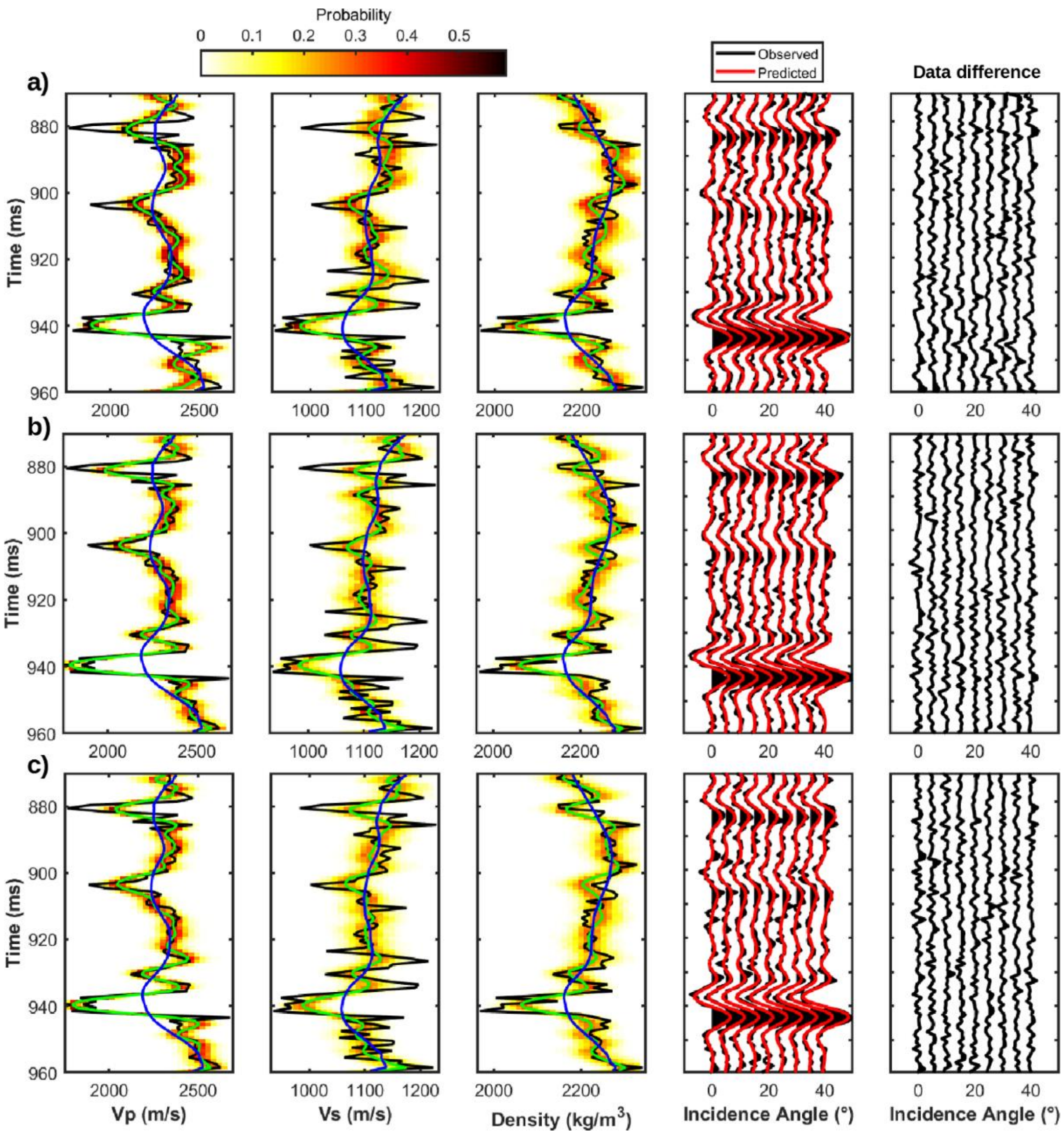

Figure 15: a)-b) HMC-DCT inversions running with 20 and 40 coefficients for each elastic property, respectively. c) Standard HMC inversion results. The black, blue, and green lines represent the true, prior, and posterior mean models, respectively, while the colormap codes the posterior probability values. In a) note the underestimation of the posterior uncertainty evidenced by the darker colors of the PPD.

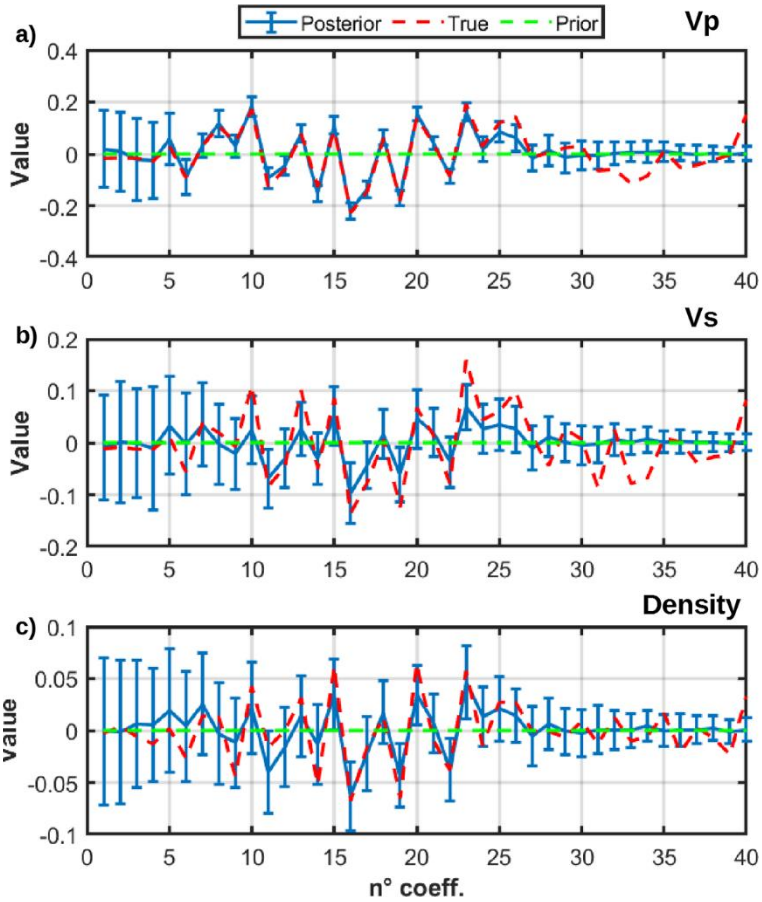

Figure 16: Comparison in the DCT space, between the true model (dashed red lines), the prior mean (green line) and the posterior mean (blue line) estimated by an HMC-DCT inversion running with q=40. a) The 40 DCT coefficients associated with *Vp*. b) The 40 DCT coefficients associated with *Vs*. c) The 40 DCT coefficients associated with the density parameter. The blue bars represent the a-posteriori 95% confidence interval.

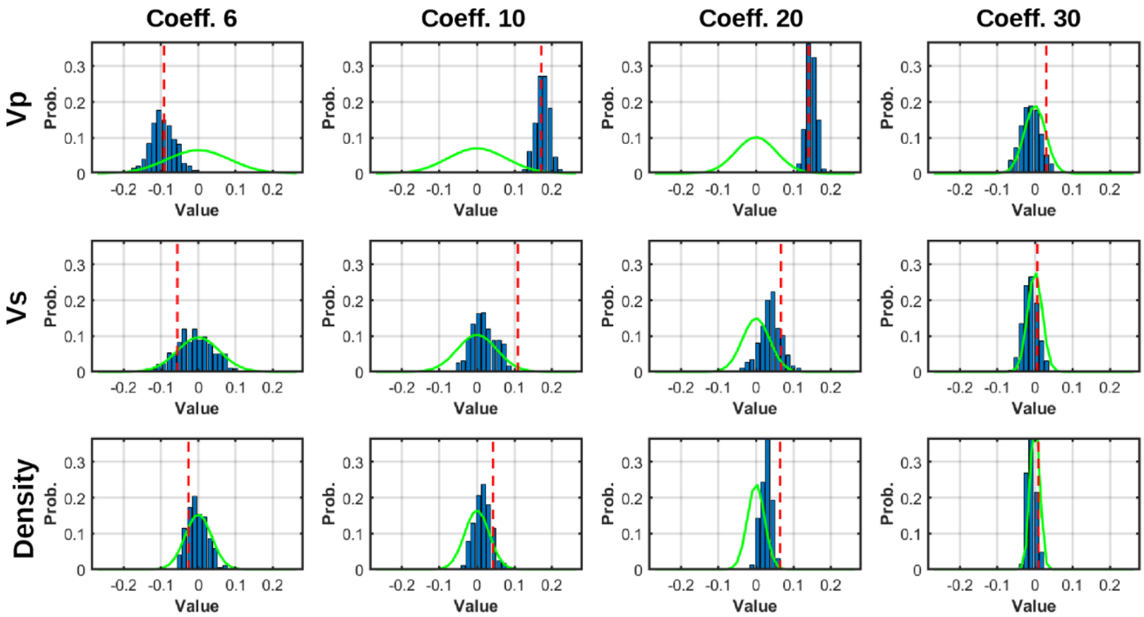

Figure 179: Comparison in the DCT space, between the true model (dashed red lines), the marginal prior (green curves) and the marginal posterior distributions (blue bars) for different DCT coefficients estimated by the HMC-DCT inversion running with q=40.

The evolution of the negative log-likelihood values shows that the length of the burn-in period slightly decreases as the number of considered DCT coefficients decreases (Figure 18). However, if only 20 coefficients are employed, the inversion underfits the observed data, while just 40 coefficients ensure final negative log-likelihood value equal to that achieved by an HMC inversion running in the full space. Note that the reduction of the burn-in period less significant in the DCT-HMC inversion than in the DCT-DEMC. Indeed, the inclusion of the derivative information into the sampling framework allows for a rapid convergence toward the stationary regime also in high-dimensional model spaces. This is a crucial strength of the HMC sampling in comparison with standard MCMC algorithms. Therefore, the major benefit provided by the DCT in the HMC inversion concerns the reduction of the computational cost. In Figure 18b we compare the percentage difference of the computational costs of a standard HMC inversion and HMC-DCT inversions running with different q values. If we consider only 10 DCT coefficients per elastic property, we move from the original 273-D elastic space to a reduced 30-D parameter space. This huge dimensionality reduction saves the 90 % of the total computational cost with respect to a standard inversion, but, as previously discussed, the final results would be characterized by low resolution and underestimated posterior uncertainties. On the other hand, 40 base functions per elastic property guarantee a substantial decrease of the computational cost with respect to a standard HMC inversion (equal to the 58% in our example), but still ensure reliable PPD estimations (see Figure 15).

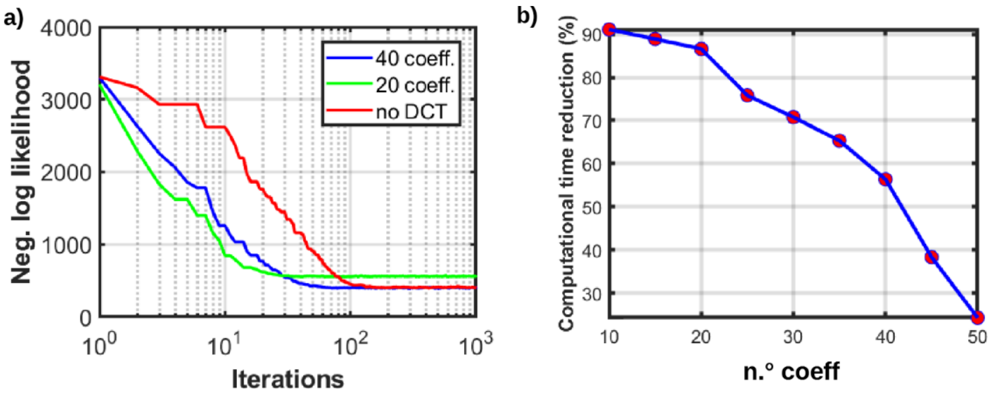

Figure 18: a) Close-up of the evolution of the negative log likelihood values for the three inversions shown in Figure 15. b) Percentage difference between the computational cost of an HMC inversion running in the full space, and HMC-DCT inversions running with different *q* values.

CONCLUSIONS

We used the DCT to reparametrize linear and non-linear 1D Bayesian AVA inversions, the latter solved using both the DEMC and the HMC algorithms. With this parameterization a signal (i.e. expressing the subsurface model vector) is expanded into series of cosine functions oscillating at different frequencies. Usually, most of the variability of the original signal is expressed by the first DCT coefficients (low-order coefficients) and for this reason this mathematical transformation can be used for model compression, which is accomplished by setting the coefficients of the base function terms beyond a certain threshold equal to zero. In this context, the unknown parameters become the numerical values of the retained, non-zero coefficients. The choice of this threshold level constitutes a compromise between the desired resolution, the accuracy of the estimated PPD, and the dimensionality reduction of the parameter space. Indeed, the analytical solutions of linear inversions clearly showed the trade-off between resolution and uncertainty, that is a too strong compression of the parameter space leads to an underestimated posterior variance, decreased model resolution, and underfitting with the observed data. However, our examples showed that for the considered case 40 DCT base functions per elastic property ensured posterior assessments, data predictions, and model resolutions similar to those achieved by Bayesian inversions running in the full, elastic space. This means that the DCT reparameterization allowed for a substantial reduction of the dimensionality and the computational complexity of the AVA inversion. For example, in the inversion experiments on well B, the 420-D elastic model space was conveniently reduced to a 120-D DCT space.

The DEMC examples showed that the length of the burn-in period and the number of iterations needed to attain stable PPD estimations significantly decrease in the reduced space. The DCT compression slightly reduced the length of the burn-in phase of the HMC inversion, but in this context the major benefit provided by this reparameterization was the significant reduction of the computational effort related to the numerical derivation of the Jacobian matrix.

In case of field data applications, the number of DCT coefficients (or in other terms the optimal compromise between the model compression and the accuracy and resolution of the results) can be determined by a DCT decomposition of actual well log data or of elastic property profiles simulated in accordance to the prior model. For example, these simulations can be used to determine the fraction of the entire model variability expressed by different numbers of base functions. Noteworthy, the optimal number of DCT coefficients is independent of the number of model parameters to be estimated and is only related to the variance of the subsurface model. Indeed, the examples on wells A and B showed that model vectors with different dimensionalities but with similar statistical properties (e.g. vertical variability) can be compressed using the same number of DCT coefficients.

The approach presented here can be extended to other model reduction strategies (i.e. using Legendre polynomials, or wavelet transform approaches) and to other geophysical inverse problems. For example, the DCT transformation is also extendible to 2D signals (i.e. images) and for this reason it could be used to reparametrize 2D geophysical inversions (e.g. 2D seismic or electrical resistivity tomography).

Data availability

Data associated with this research are available and can be obtained by contacting the corresponding author

REFERENCES

1. Aleardi M., Mazzotti A. (2017). 1D elastic full-waveform inversion and uncertainty estimation by means of a hybrid genetic algorithm–Gibbs sampler approach. Geophysical Prospecting, 65(1), 64-85.

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

3. Aleardi, M., Salusti, A., and Pierini, S. (2020). Transdimensional and Hamiltonian Monte Carlo inversions of Rayleigh-wave dispersion curves: A comparison on synthetic datasets. Near Surface Geophysics. In print. doi: 10.1002/nsg.12100.

4. Aleardi, M., and Salusti, A. (2019). Markov chain Monte Carlo algorithms for target-oriented and interval-oriented amplitude versus angle inversions with non-parametric priors and non-linear forward modellings. Geophysical Prospecting, 68(3), 735-760.

5. Aleardi, M., and Salusti, A. (2020). Hamiltonian Monte Carlo algorithms for target-oriented and interval-oriented AVA inversions. Geophysics, 85(3), 1-76.

6. Aki, K., and Richards, P. G. (1980). Quantative seismology: Theory and methods. New York, 801.

7. Aster, R. C., Borchers, B., and Thurber, C. H. (2018). Parameter estimation and inverse problems. Elsevier.

8. Azevedo, L., and Demyanov, V. (2019). Multiscale uncertainty assessment in geostatistical seismic inversion. Geophysics, 84(3), R355-R369.

9. Azevedo, L., Tompkins, M. J., and Mukerji, T. (2016). Regularized sparse-grid geometric sampling for uncertainty analysis in non-linear inverse problems. Geophysical Prospecting, 64(2), 320-334.

10. Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. arXiv preprint arXiv:1701.02434.

11. Bodin, T., Sambridge, M., Rawlinson, N., and Arroucau, P. (2012). Transdimensional tomography with unknown data noise. Geophysical Journal International, 189(3), 1536-1556.

12. Bosch, M., Cara, L., Rodrigues, J., Navarro, A., and Díaz, M. (2007). A Monte Carlo approach to the joint estimation of reservoir and elastic parameters from seismic amplitudes. Geophysics, 72(6), O29-O39.

13. Brooks, S. P., Giudici, P., and Roberts, G. O. (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65(1), 3-39.

14. Buland, A., and Omre, H. (2003). Bayesian linearized AVO inversion. Geophysics, 68(1), 185-198.

15. de Figueiredo, L. P., Grana, D., Bordignon, F. L., Santos, M., Roisenberg, M., and Rodrigues, B. B. (2018). Joint Bayesian inversion based on rock-physics prior modeling for the estimation of spatially correlated reservoir properties. Geophysics, 83(5), M49-M61.

16. Dejtrakulwong, P., Mukerji, T., and Mavko, G. (2012). Using kernel principal component analysis to interpret seismic signatures of thin shaly-sand reservoirs. In SEG Technical Program Expanded Abstracts 2012 (pp. 1-5). Society of Exploration Geophysicists.

17. Dosso, S. E., Holland, C. W., and Sambridge, M. (2012). Parallel tempering for strongly nonlinear geoacoustic inversion. The Journal of the Acoustical Society of America, 132(5), 3030-3040.

18. Fernández-Martínez, J. L., Mukerji, T., García-Gonzalo, E., and Fernández-Muñiz, Z. (2011). Uncertainty assessment for inverse problems in high dimensional spaces using particle swarm optimization and model reduction techniques. Mathematical and Computer Modelling, 54(11-12), 2889-2899.

19. Fernández-Martínez, J. L., Fernández-Muñiz, Z., Pallero, J. L. G., and Bonvalot, S. (2017). Linear geophysical inversion via the discrete cosine pseudo-inverse: application to potential fields. Geophysical Prospecting, 65, 94-111.

20. Fichtner, A., and Zunino, A. (2019). Hamiltonian nullspace shuttles. Geophysical research letters, 46(2), 644-651.

21. Fichtner, A., Zunino, A., and Gebraad, L. (2019). Hamiltonian Monte Carlo solution of tomographic inverse problems. Geophysical Journal International, 216(2), 1344-1363.

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

23. Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995). Bayesian Data Analysis, Chapman & Hall, London.

24. Grana, D. (2016). Bayesian linearized rock-physics inversion. Geophysics, 81(6), D625-D641.

25. Grana, D., Passos de Figueiredo, L., and Azevedo, L. (2019). Uncertainty quantification in Bayesian inverse problems with model and data dimension reduction. Geophysics, 84(6), M15-M24.

26. Holmes, C., Krzysztof, L. and Pompe, E. (2017). Adaptive MCMC for multimodal distributions. Technical report. https://pdfs.semanticscholar.org/c75d/f035c23e3c0425409e70d457cd43b174076f.pdf.

27. Le Gall, D., (1991). A video compression standard for Multimedia Applications, Communications of the ACM, 34(4), 46–58.

28. Lochbühler, T., Breen, S. J., Detwiler, R. L., Vrugt, J. A., and Linde, N. (2014). Probabilistic electrical resistivity tomography of a CO2 sequestration analog. Journal of Applied Geophysics, 107, 80-92.

29. Mackenze, P. B. (1989). An improved hybrid Monte Carlo method. Physics Letters B, 226(3-4), 369-371.

30. Malinverno, A. (2000). A Bayesian criterion for simplicity in inverse problem parametrization. Geophysical Journal International, 140(2), 267-285.

31. Malinverno, A., and Briggs, V. A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. Geophysics, 69(4), 1005-1016.

32. Menke, W. (2018). Geophysical data analysis: discrete inverse theory. Academic press.

33. Neal, R.M. (2011). MCMC using Hamiltonian dynamics, in Handbook of Markov Chain Monte Carlo, Brooks, S., Gelman, A., Jones, G. & Meng, X.-Li., Chapman and Hall, 113–162.

34. Nunes, R., Azevedo, L., and Soares, A. (2019). Fast geostatistical seismic inversion coupling machine learning and Fourier decomposition. Computational Geosciences, 23(5), 1161-1172.

35. Pejić, T., Hawkins, R., Sambridge, M., and Tkalčić, H. (2019). Transdimensional Bayesian Attenuation Tomography of the Upper Inner Core. Journal of Geophysical Research: Solid Earth, 124(2), 1929-1943.

36. Piana Agostinetti, N., and Bodin, T. (2018). Flexible coupling in joint inversions: A Bayesian structure decoupling algorithm. Journal of Geophysical Research: Solid Earth, 123(10), 8798-8826.

37. Qin, H., Xie, X., and Tang, Y. (2019). Evaluation of a Straight-Ray Forward Model for Bayesian Inversion of Crosshole Ground Penetrating Radar Data. Electronics, 8(6), 630.

38. Ray, A., Kaplan, S., Washbourne, J., and Albertin, U. (2017). Low frequency full waveform seismic inversion within a tree based Bayesian framework. Geophysical Journal International, 212(1), 522-542.

39. Rimstad, K., Avseth, P., and Omre, H. (2012). Hierarchical Bayesian lithology/fluid prediction: A North Sea case study. Geophysics, 77(2), B69-B85.

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

41. Sajeva, A., Aleardi, M., and Mazzotti, A. (2017). Genetic algorithm full-waveform inversion: uncertainty estimation and validation of the results. Bollettino di Geofisica Teorica ed Applicata, 58(4), 395-414.

42. Satija, A., and Caers, J. (2015). Direct forecasting of subsurface flow response from non-linear dynamic data by linear least-squares in canonical functional principal component space. Advances in Water Resources, 77, 69-81.

43. Scalzo, R., Kohn, D., Olierook, H., Houseman, G., Chandra, R., Girolami, M., and Cripps, S. (2019). Efficiency and robustness in Monte Carlo sampling for 3-D geophysical inversions with Obsidian v0. 1.2: setting up for success. Geoscientific Model Development, 12(7), 2941-2960.

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

45. Sen, M. K., and Stoffa, P. L. (1996). Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion. Geophysical Prospecting, 44(2), 313-350.

46. Szabó, N. P., and Dobróka, M. (2019). Series expansion-based genetic inversion of wireline logging data. Mathematical Geosciences, 51(6), 811-835.

47. Tarantola, A. (2005). Inverse problem theory and methods for model parameter estimation. siam.

48. Ter Braak, C. J. (2006). A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. Statistics and Computing, 16(3), 239-249.

49. Turner, B. M., and Sederberg, P. B. (2012). Approximate Bayesian computation with differential evolution. Journal of Mathematical Psychology, 56(5), 375-385.

50. Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. Environmental Modelling & Software, 75, 273-316.

51. Wallace, G. K., (1991). The JPEG Still Picture Compression Standard, Communications of the ACM, 34(4), 30–40.

52. Zhdanov, M. S. (2002). Geophysical inverse theory and regularization problems. Elsevier.

53. Zunino, A., Mosegaard, K., Lange, K., Melnikova, Y., and Mejer Hansen, T. (2014). Monte Carlo reservoir analysis combining seismic reflection data and informed priors. Geophysics, 80(1), R31-R41.