GEOPHYSICS, VOL. 84, NO. 6 (NOVEMBER-DECEMBER 2019); P. M15–M24, 10 FIGS., 1 TABLE.
10.1190/GEO2019-0222.1

Uncertainty quantification in Bayesian inverse problems with model and data dimension reduction

Dario Grana[1], Leandro Passos de Figueiredo[2], and Leonardo Azevedo[3]

ABSTRACT

The prediction of rock properties in the subsurface from geophysical data generally requires the solution of a mathematical inverse problem. Because of the large size of geophysical (seismic) data sets and subsurface models, it is common to reduce the dimension of the problem by applying dimension reduction methods and considering a reparameterization of the model and/or the data. Especially for high-dimensional nonlinear inverse problems, in which the analytical solution of the problem is not available in a closed form and iterative sampling or optimization methods must be applied to approximate the solution, model and/or data reduction reduce the computational cost of the inversion. However, part of the information in the data or in the model can be lost by working in the reduced model and/or data space. We have focused on the uncertainty quantification in the solution of the inverse problem with data and/or model order reduction. We operate in a Bayesian setting for the inversion and uncertainty quantification and validate the proposed approach in the linear case, in which the posterior distribution of the model variables can be analytically written and the uncertainty of the model predictions can be exactly assessed. To quantify the changes in the uncertainty in the inverse problem in the reduced space, we compare the uncertainty in the solution with and without data and/or model reduction. We then extend the approach to nonlinear inverse problems in which the solution is computed using an ensemble-based method. Examples of applications to linearized acoustic and nonlinear elastic inversion allow quantifying the impact of the application of reduction methods to model and data vectors on the uncertainty of inverse problem solutions. Examples of applications to linearized acoustic and nonlinear elastic inversion are shown.

INTRODUCTION

Most of the modeling problems in geophysics can be written as mathematical inverse problems in which the values of the properties of interest are estimated from a set of indirect geophysical measurements. Examples of inverse problems in geophysics include seismic inversion for the prediction of elastic properties (Aki and Richards, 1980; Sheriff and Geldart, 1995; Sen and Stoffa, 2013); rock physics inversion for the prediction of reservoir properties (Doyen, 2007; Bosch et al., 2010; Azevedo and Soares, 2017); controlled-source electromagnetic inversion for the estimation of electrical resistivity (Chen et al., 2007; Tompkins et al., 2011; MacGregor, 2012); and inversion of tomographic, time-lapse seismic, and gravity data. Reviews of inverse methods for geophysical problems can be found in Tarantola (2005), Aster et al. (2011), and Sen and Stoffa (2013).

One of the main challenges in geophysical inverse problems for the prediction of rock properties in the subsurface based on large, noisy, low-resolution geophysical data sets is the nonuniqueness of the solution. Therefore, the quantification of the uncertainty in the solution of the inverse problem (i.e., the uncertainty of the predicted values of the model variables) should be as important as the prediction of the most likely model. In geophysical inverse problems, probabilistic methods are often used to quantify the uncertainty in the model predictions through the probability distributions of the model variables given the geophysical data (Tarantola, 2005). The Bayesian approach is the most common probabilistic method used for geophysical inverse problems (Scales and Tenorio, 2001; Ulrych et al., 2001; Buland and Omre, 2003; Tarantola, 2005). Several applications of the Bayesian approach to geophysical inverse prob-

Manuscript received by the Editor 8 April 2019; revised manuscript received 14 May 2019; published ahead of production 22 July 2019; published online 09 October 2019.
[1]University of Wyoming, Department of Geology and Geophysics, Laramie, Wyoming, USA. E-mail: dgrana@uwyo.edu (corresponding author).
[2]LTrace Geophysical Solutions, Florianópolis, Brazil and Federal University of Santa Catarina, Informatics and Statistics Department, Florianópolis, Brazil. E-mail: leandrop.fgr@gmail.com.
[3]Lisbon University, CERENA, DECivil, Instituto Superior Técnico, Lisbon, Portugal. E-mail: leonardo.azevedo@tecnico.ulisboa.pt.

lems, under different statistical assumptions and with different geophysical models, have been proposed (Sen and Stoffa, 1996; Buland and Omre, 2003; Eidsvik et al., 2004; Hansen et al., 2006; Larsen et al., 2006; Chen et al., 2007; Gunning and Glinsky, 2007; Bosch et al., 2009; Ulvmoen and Omre, 2010; Rimstad et al., 2012; Grana 2016; Jullum and Kolbjørnsen, 2016; Grana et al., 2017; de Figueiredo et al. 2018). Monte Carlo and Markov chain Monte Carlo methods have also been proposed to iteratively approximate the posterior distribution (Hastings, 1970; Mosegaard and Tarantola, 1995; Sambridge and Mosegaard, 2002; de Figueiredo et al., 2019). Stochastic sampling methods such as the ensemble Kalman filter, probability perturbation method, gradual deformation, and ensemble smoother (Evensen, 2007; Oliver et al., 2008; Emerick and Reynolds, 2013; Chen and Oliver, 2017) are more commonly used in data assimilation problems for reservoir engineering rather than geophysics.

However, the uncertainty inferred from posterior distributions of the model variables only accounts for the uncertainty in the data and the physical relations between data and model variables. Generally, the uncertainty quantification does not account for errors in the data processing or reparameterizations of the model parameters. A discussion about the uncertainty in inverse problems with biased data can be found in Oliver and Alfonzo (2018). Model and data reduction methods (Hastie et al., 2002) are commonly used to reduce the computational complexity of the inverse problem. Several dimension reduction methods have been proposed in the literature to reduce the dimension of the model and/or data spaces. Examples of dimension reduction methods include, among the others, principal component analysis (PCA) (Jolliffe, 2011), multidimensional scaling (MDS) (Cox and Cox, 2000), wavelet transforms (Rao, 2002), and recent machine-learning algorithms (Bishop, 2006). Examples of applications in geosciences include mostly studies in the reservoir engineering field (Satija and Caers, 2015; Scheidt et al., 2015, 2018; Sun et al., 2017; He et al., 2018; Jeong et al., 2018; Lima et al., 2019) as well as in geophysical inverse problems (Dejtrakulwong et al., 2012; Azevedo et al., 2013; Liu and Grana, 2018). The reduction of the dimension of the data and model vectors generally improves the computational efficiency of the physical operators. However, the loss of information due to the lower dimension of the problem leads to underestimation or overestimation of the uncertainty in the solution of the problem. Studies on uncertainty assessment for inverse problems and the impact of data and/or model reduction on uncertainty quantification are still missing in geophysics.

We investigate the uncertainty of the solution of geophysical inverse problems with and without data and/or model dimension reduction methods. We propose a complete assessment of the uncertainty using numerical examples in two different settings: in a Bayesian setting for linear inverse problems with analytical solutions and for nonlinear inverse problems, in which the solution is obtained using a stochastic sampling method, namely, the ensemble smoother with multiple data assimilation (ES-MDA) (Emerick and Reynolds, 2013).

METHODOLOGY

The goal of this study is to investigate the uncertainty in the predictions of model variables $\mathbf{m}$ given a set of measured data $\mathbf{d}$ in geophysical inverse problems with and without data and/or model reparameterization. We assume that the set of physical relations $\mathbf{g}$, i.e., the operator that models the data response for known model variables, is known. The forward problem can be written in the general form

$$\mathbf{d} = \mathbf{g}(\mathbf{m}) + \mathbf{e}, \tag{1}$$

where $\mathbf{e}$ represents the measurement errors. For linear operators, if we indicate with $\mathbf{G}$ the matrix associated with the operator $\mathbf{g}$, the forward problem can then be rewritten as

$$\mathbf{d} = \mathbf{G}\mathbf{m} + \mathbf{e}. \tag{2}$$

Deterministic and probabilistic methods for inverse problems have been widely studied in the field of geophysics (Tarantola, 2005; Aster et al., 2011; Sen and Stoffa, 2013). Peculiar issues related to geophysical inverse problems include the low signal-to-noise ratio of the data, band-limited nature of the data, spatial correlations of model properties, and accuracy and approximations of the geophysical operators (such as seismic wave propagation and rock physics modeling).

Because the interest is on uncertainty quantification, we focus on probabilistic methods, specifically on Bayesian inversion methods, with the goal of computing the posterior distribution $p(\mathbf{m}|\mathbf{d})$ of the model variables $\mathbf{m}$ given the measured data $\mathbf{d}$. In this approach, the uncertainty is interpreted as the variance of the posterior distribution. For a general inverse problem as shown in equation 1, the posterior distribution $p(\mathbf{m}|\mathbf{d})$ is the product of the prior distribution $p(\mathbf{m})$ of the model variables $\mathbf{m}$ and the conditional probability $p(\mathbf{d}|\mathbf{m})$ of the data $\mathbf{d}$ given the model variables $\mathbf{m}$ divided the probability $p(\mathbf{d})$ of the data:

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d})}. \tag{3}$$

The probability $p(\mathbf{m})$ is the probability distribution of the model properties before integrating the information obtained from the data, the probability $p(\mathbf{d}|\mathbf{m})$ is the likelihood of observing the data $\mathbf{d}$ for a given value of the model variables $\mathbf{m}$, and the $p(\mathbf{d})$ is a normalizing constant to make the posterior distribution $p(\mathbf{m}|\mathbf{d})$ a valid probability density function (i.e., $\int p(\mathbf{m}|\mathbf{d})d\mathbf{m} = 1$).

For linear inverse problems (equation 2), under specific statistical assumptions, the posterior distribution $p(\mathbf{m}|\mathbf{d})$ can be analytically written in a closed form. In the linear case, if the errors are independent of the model $\mathbf{m}$ and Gaussian distributed with $\mathbf{0}$-mean and known covariance matrix $\mathbf{\Sigma_d}$, $p(\mathbf{e}) = N(\mathbf{e}; 0, \mathbf{\Sigma_d})$, then the likelihood $p(\mathbf{d}|\mathbf{m})$ is a Gaussian distribution. If the prior probability distribution is Gaussian $p(\mathbf{m}) = N(\mathbf{m}; \boldsymbol{\mu}_\mathbf{m}, \mathbf{\Sigma_m})$, then the posterior $p(\mathbf{m}|\mathbf{d})$ is also a Gaussian distribution:

$$p(\mathbf{m}|\mathbf{d}) = N(\mathbf{m}; \boldsymbol{\mu}_{\mathbf{m}|\mathbf{d}}, \mathbf{\Sigma}_{\mathbf{m}|\mathbf{d}}), \tag{4}$$

with analytical expressions for the conditional mean $\boldsymbol{\mu}_{\mathbf{m}|\mathbf{d}}$ and conditional covariance matrix $\mathbf{\Sigma}_{\mathbf{m}|\mathbf{d}}$:

$$\boldsymbol{\mu}_{\mathbf{m}|\mathbf{d}} = \boldsymbol{\mu}_\mathbf{m} + \mathbf{\Sigma_m}\mathbf{G}^T(\mathbf{G}\mathbf{\Sigma_m}\mathbf{G}^T + \mathbf{\Sigma_d})^{-1}(\mathbf{d} - \mathbf{G}\boldsymbol{\mu}_\mathbf{m}), \tag{5}$$

$$\mathbf{\Sigma}_{\mathbf{m}|\mathbf{d}} = \mathbf{\Sigma}_m - \mathbf{\Sigma_m}\mathbf{G}^T(\mathbf{G}\mathbf{\Sigma_m}\mathbf{G}^T + \mathbf{\Sigma_d})^{-1}\mathbf{G}\mathbf{\Sigma_m}. \tag{6}$$

Several dimension reduction methods can be applied to reduce the dimension of the model and/or data vectors. Because of the simplicity of the implementation, we focus on linear dimension

reduction and adopt PCA, but any other method could be used. PCA computes a set of uncorrelated variables from observed correlated variables with the full observed variance. The dimension of an inverse problem with large data and/or model spaces can be reduced by applying PCA to the covariance matrix $\mathbf{\Sigma}$, using the eigenvalues/eigenvectors decomposition or truncated singular value decomposition (SVD) (for large problems) and selecting a subset of eigenvectors (principal components) corresponding to the larger singular values. By applying the eigenvalues/eigenvectors decomposition theorem (Golub and Van Loan, 2013), the covariance matrix $\mathbf{\Sigma}$ of any probability distribution can be written as $\mathbf{\Sigma} = \mathbf{V\Lambda V}^T$, where $\mathbf{V}$ is the matrix of orthogonal eigenvectors and $\mathbf{\Lambda}$ is the diagonal matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_n)$ with the corresponding eigenvalues $\lambda_{i=1,\ldots,n}$. Given a set of n spatially correlated random variables $\mathbf{x} = (x_1, \ldots, x_n)^T$ representing realizations of the property of interest at different locations, for example, data observations, with mean $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n)^{\mathbf{T}}$ and covariance matrix $\mathbf{\Sigma} = \mathbf{V\Lambda V}^T$, the vector $\mathbf{x}$ can be approximated as a linear combination of a subset of $\tilde{n} \leq n$ eigenvectors:

$$\mathbf{x} \approx \boldsymbol{\mu} + \tilde{\mathbf{V}}\,\tilde{\mathbf{x}}, \tag{7}$$

where $\tilde{\mathbf{V}}$ is the truncated matrix, of size $n \times \tilde{n}$, including the $\tilde{n}$ principal eigenvectors (matrix columns) and $\tilde{\mathbf{x}}$ is a vector of length $\tilde{n}$, representing the reduced sample vector. The reduced vector $\tilde{\mathbf{x}}$ can be then computed as

$$\tilde{\mathbf{x}} \approx \tilde{\mathbf{V}}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \approx \tilde{\mathbf{V}}^T(\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T)^{-1}(\mathbf{x} - \boldsymbol{\mu}), \tag{8}$$

where $\tilde{\mathbf{V}}^{-1}$ is the generalized inverse matrix of the truncated matrix $\tilde{\mathbf{V}}$. Equation 8 is exact only if $n = \tilde{n}$; otherwise, it is an approximation due to the truncation. The number of samples $\tilde{n}$ is generally chosen by arranging the eigenvalues in decreasing order and choosing a subset of the first $\tilde{n}$ eigenvectors associated with the $\tilde{n}$ eigenvalues that represents a fraction f_σ of the total variance:

$$\sum_{i=1}^{\tilde{n}} \lambda_i \geq f_\sigma \sum_{j=1}^{n} \lambda_j. \tag{9}$$

Data dimension reduction

We first apply PCA to the data vector for data reduction. The data vector $\mathbf{d}$ might represent, for example, the seismogram at a given location. By applying the PCA-based data reduction, we obtain

$$\tilde{\mathbf{d}} = \tilde{\mathbf{V}}_{\mathbf{d}}^{-1}(\mathbf{d} - \boldsymbol{\mu}_{\mathbf{d}}), \tag{10}$$

where $\tilde{\mathbf{V}}_{\mathbf{d}}$ is the truncated matrix of the principal eigenvectors and $\tilde{\mathbf{V}}_{\mathbf{d}}^{-1}$ is its generalized inverse. We then reformulate the inverse problem in equation 2, using the reduced data vector in equation 10:

$$\tilde{\mathbf{d}} + \tilde{\mathbf{V}}_{\mathbf{d}}^{-1}\boldsymbol{\mu}_{\mathbf{d}} = \tilde{\mathbf{V}}_{\mathbf{d}}^{-1}\mathbf{G}\mathbf{m} + \tilde{\mathbf{V}}_{\mathbf{d}}^{-1}\mathbf{e} \equiv \mathbf{H}\mathbf{m} + \mathbf{e}_{\tilde{\mathbf{d}}}, \tag{11}$$

where $\mathbf{e}_{\tilde{\mathbf{d}}} \equiv \tilde{\mathbf{V}}_{\mathbf{d}}^{-1}\mathbf{e}$ is the reduced noise with $\mathbf{0}$-mean and covariance matrix $\mathbf{\Sigma}_{\tilde{\mathbf{d}}} = \tilde{\mathbf{V}}_{\mathbf{d}}\mathbf{\Sigma}_{\mathbf{d}}\tilde{\mathbf{V}}_{\mathbf{d}}^T$. The operators $\mathbf{G}$ and $\tilde{\mathbf{V}}_{\mathbf{d}}^{-1}$ are linear; hence, the operator $\mathbf{H} \equiv \tilde{\mathbf{V}}_{\mathbf{d}}^{-1}\mathbf{G}$ is also linear. If the prior probability distribution of the model $\mathbf{m}$ is Gaussian, $p(\mathbf{m}) = N(\mathbf{m}; \boldsymbol{\mu}_{\mathbf{m}}, \mathbf{\Sigma}_{\mathbf{m}})$, then, the posterior $p(\mathbf{m}|\tilde{\mathbf{d}})$ is also a Gaussian distribution:

$$p(\mathbf{m}|\tilde{\mathbf{d}}) = N(\mathbf{m}; \boldsymbol{\mu}_{\mathbf{m}|\tilde{\mathbf{d}}}, \mathbf{\Sigma}_{\mathbf{m}|\tilde{\mathbf{d}}}) \tag{12}$$

with conditional mean $\boldsymbol{\mu}_{\mathbf{m}|\tilde{\mathbf{d}}}$ and conditional covariance matrix $\mathbf{\Sigma}_{\mathbf{m}|\tilde{\mathbf{d}}}$:

$$\boldsymbol{\mu}_{\mathbf{m}|\tilde{\mathbf{d}}} = \boldsymbol{\mu}_{\mathbf{m}} + \mathbf{\Sigma}_{\mathbf{m}}\mathbf{H}^T(\mathbf{H}\mathbf{\Sigma}_{\mathbf{m}}\mathbf{H}^T + \mathbf{\Sigma}_{\tilde{\mathbf{d}}})^{-1}(\tilde{\mathbf{d}} + \tilde{\mathbf{V}}_{\mathbf{d}}^{-1}\boldsymbol{\mu}_{\mathbf{d}} - \mathbf{H}\boldsymbol{\mu}_{\mathbf{m}}) \tag{13}$$

$$\mathbf{\Sigma}_{\mathbf{m}|\tilde{\mathbf{d}}} = \mathbf{\Sigma}_{\mathbf{m}} - \mathbf{\Sigma}_{\mathbf{m}}\mathbf{H}^T(\mathbf{H}\mathbf{\Sigma}_{\mathbf{m}}\mathbf{H}^T + \mathbf{\Sigma}_{\tilde{\mathbf{d}}})^{-1}\mathbf{H}\mathbf{\Sigma}_{\mathbf{m}}, \tag{14}$$

respectively. The multivariate Gaussian distribution $p(\mathbf{m}|\tilde{\mathbf{d}})$ in equations 12–14 is the solution of the inverse problem in equation 11 with reduced data $\tilde{\mathbf{d}}$.

Model dimension reduction

Similarly, we can apply PCA to the model vector for model reduction and derive the solution of the inverse problem in equation 2 with a reduced model vector $\tilde{\mathbf{m}}$:

$$\tilde{\mathbf{m}} = \tilde{\mathbf{V}}_{\mathbf{m}}^{-1}(\mathbf{m} - \boldsymbol{\mu}_{\mathbf{m}}), \tag{15}$$

where $\tilde{\mathbf{V}}_{\mathbf{m}}$ is the truncated matrix of the principal eigenvectors. We then reformulate the inverse problem in equation 2, using the reduced model vector in equation 15:

$$\mathbf{d} - \mathbf{G}\boldsymbol{\mu}_m = \mathbf{G}\tilde{\mathbf{V}}_{\mathbf{m}}\tilde{\mathbf{m}} + \mathbf{e} \equiv \mathbf{F}\tilde{\mathbf{m}} + \mathbf{e}. \tag{16}$$

If the prior probability distribution of the model $\mathbf{m}$ is Gaussian, $p(\mathbf{m}) = N(\mathbf{m}; \boldsymbol{\mu}_{\mathbf{m}}, \mathbf{\Sigma}_m)$, then the prior distribution $p(\tilde{\mathbf{m}})$ of the reduced model $\tilde{\mathbf{m}}$ is a Gaussian distribution with $\boldsymbol{\mu}_{\tilde{\mathbf{m}}}$ and covariance matrix $\mathbf{\Sigma}_{\tilde{\mathbf{m}}}$, $p(\tilde{\mathbf{m}}) = \mathbf{N}(\tilde{\mathbf{m}}; \boldsymbol{\mu}_{\tilde{\mathbf{m}}}; \mathbf{\Sigma}_{\tilde{\mathbf{m}}})$. By construction, the mean $\boldsymbol{\mu}_{\tilde{\mathbf{m}}}$ of the reduced model $\tilde{\mathbf{m}}$ is $\mathbf{0}$, and the covariance matrix $\mathbf{\Sigma}_{\tilde{\mathbf{m}}}$ contains the truncated eigenvalues $\{\lambda_{\mathbf{m}i}\}_{i=1,\ldots,\tilde{n}}$ of the prior covariance matrix $\mathbf{\Sigma}_{\mathbf{m}} = \mathbf{V}_{\mathbf{m}}\mathbf{\Lambda}_{\mathbf{m}}\mathbf{V}_{\mathbf{m}}^T$, with $\mathbf{\Lambda}_{\mathbf{m}} = \text{diag}(\lambda_{\mathbf{m}1}, \ldots, \lambda_{\mathbf{m}n})$.

The operators $\mathbf{G}$ and $\tilde{\mathbf{V}}_{\mathbf{m}}$ are linear; hence, the operator $\mathbf{F} \equiv \mathbf{G}\tilde{\mathbf{V}}_{\mathbf{m}}$ is also linear. Then, the posterior distribution $p(\tilde{\mathbf{m}}|\mathbf{d})$ of the reduced model given the data is also a Gaussian distribution

$$p(\tilde{\mathbf{m}}|\mathbf{d}) = N(\tilde{\mathbf{m}}; \boldsymbol{\mu}_{\tilde{\mathbf{m}}|\mathbf{d}}, \mathbf{\Sigma}_{\tilde{\mathbf{m}}|\mathbf{d}}) \tag{17}$$

with conditional mean $\boldsymbol{\mu}_{\tilde{\mathbf{m}}|\mathbf{d}}$ and conditional covariance matrix $\mathbf{\Sigma}_{\tilde{\mathbf{m}}|\mathbf{d}}$:

$$\boldsymbol{\mu}_{\tilde{\mathbf{m}}|\mathbf{d}} = \mathbf{\Sigma}_{\tilde{\mathbf{m}}}\mathbf{F}^T(\mathbf{F}\mathbf{\Sigma}_{\tilde{m}}\mathbf{F}^T + \mathbf{\Sigma}_{\mathbf{d}})^{-1}(\mathbf{d} - \mathbf{G}\boldsymbol{\mu}_m) \tag{18}$$

$$\mathbf{\Sigma}_{\tilde{\mathbf{m}}|\mathbf{d}} = \mathbf{\Sigma}_{\tilde{\mathbf{m}}} - \mathbf{\Sigma}_{\tilde{\mathbf{m}}}\mathbf{F}^T(\mathbf{F}\mathbf{\Sigma}_{\tilde{m}}\mathbf{F}^T + \mathbf{\Sigma}_{\mathbf{d}})^{-1}\mathbf{F}\mathbf{\Sigma}_{\tilde{m}}. \tag{19}$$

The solution of the inverse problem in equation 16 is the posterior distribution $p(\mathbf{m}|\mathbf{d}, \tilde{\mathbf{m}})$ of the model $\mathbf{m}$ given the data $\mathbf{d}$ and the reduced model $\tilde{\mathbf{m}}$. Because the transformation in equation 15 is linear, the posterior distribution $p(\mathbf{m}|\mathbf{d}, \tilde{\mathbf{m}})$ can be computed as

$$p(\mathbf{m}|\mathbf{d}, \tilde{\mathbf{m}}) = N(\mathbf{m}; \boldsymbol{\mu}_{\mathbf{m}|\mathbf{d},\tilde{\mathbf{m}}}, \mathbf{\Sigma}_{\mathbf{m}|\mathbf{d},\tilde{\mathbf{m}}}) \tag{20}$$

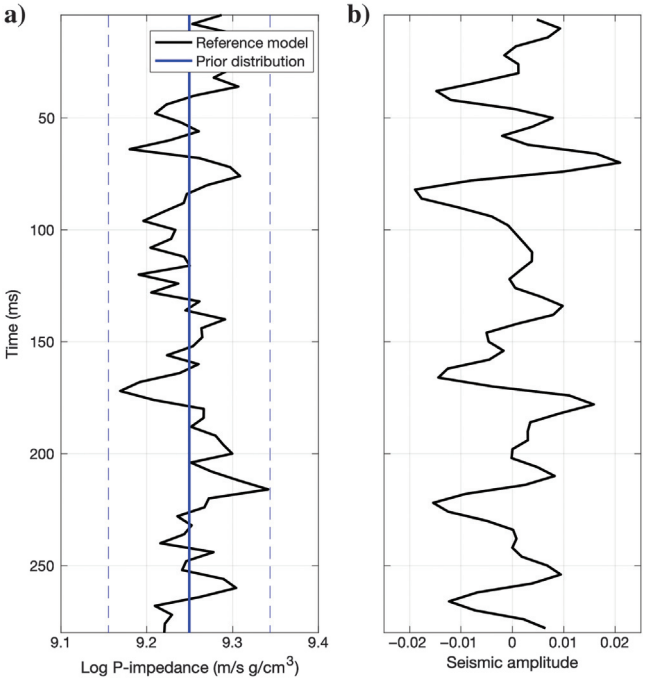

Figure 1. Synthetic example of linear seismic inversion. (a) Logarithm of acoustic impedance (the solid black line represents the well log, the solid blue line represents the mean of the prior distribution, the dashed blue lines represent the 95% confidence interval); (b) measured seismic data (zero-offset).

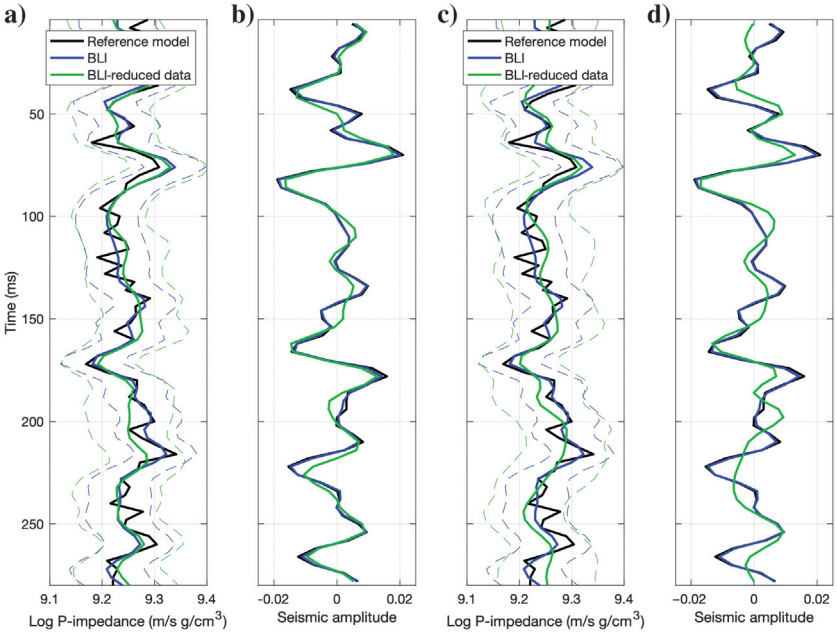

Figure 2. Bayesian linear seismic inversion for the logarithm of P-impedance with and without data reduction. Plots (a and b) show the comparison between BLI and BLI with 90% variance data reduction; plots (c and d) show the comparison between BLI and BLI with 75% variance data reduction. The blue lines represent the posterior distribution and the predicted data from BLI, the green lines represent the BLI with reduced data, and the solid black lines represent the reference model and data. The dashed lines represent the 95% confidence interval.

with conditional mean $\boldsymbol{\mu}_{\mathbf{m}|\mathbf{d},\tilde{\mathbf{m}}}$ and conditional covariance matrix $\boldsymbol{\Sigma}_{\mathbf{m}|\mathbf{d},\tilde{\mathbf{m}}}$:

$$\boldsymbol{\mu}_{\mathbf{m}|\mathbf{d},\tilde{\mathbf{m}}} = \boldsymbol{\mu}_{\mathbf{m}} + \tilde{\mathbf{V}}_{\mathbf{m}}\boldsymbol{\mu}_{\tilde{\mathbf{m}}|\mathbf{d}} \tag{21}$$

$$\boldsymbol{\Sigma}_{\mathbf{m}|\mathbf{d},\tilde{\mathbf{m}}} = \tilde{\mathbf{V}}_{\mathbf{m}}\boldsymbol{\Sigma}_{\tilde{\mathbf{m}}|\mathbf{d}}\tilde{\mathbf{V}}_{\mathbf{m}}^{T}. \tag{22}$$

The formulation in equations 4–6 of the full inverse problem (equation 2) and the formulations in equations 12–14 and 20–22 of the solutions of the inverse problems with reduced data (equation 11) and reduced model (equation 16), respectively, are used to quantify the changes in the uncertainty with and without order reduction.

Nonlinear inverse problems

We then extend the dimension reduction method to nonlinear inverse problems. Linear and nonlinear dimension reduction methods can be applied. We limit our discussion to linear dimension reduction methods applied to nonlinear inverse problems, but analogous considerations can be done for nonlinear dimension reduction methods as well. Several probabilistic methods have been proposed for nonlinear inverse problems. Ad hoc methods have been developed for geophysical inverse problems to account for the spatial correlation of the model parameters. We adopt here an ensemble-based method, namely, the ES-MDA (Emerick and Reynolds, 2013), but the same approach could be applied using any stochastic inverse method, such as Markov chain Monte Carlo or stochastic optimization algorithms.

In ensemble-based methods, a set of initial realizations of the model property is sampled from a prior Gaussian distribution, for example, using geostatistical simulation methods (Doyen, 2007), and then the realizations are iteratively updated using a Bayesian approach based on the Kalman filter equations (Evensen, 2007). Because the forward operator in the inverse problem is nonlinear, the covariance matrix of the model properties cannot be analytically computed; therefore, it is approximated using the sample covariance matrix of the ensemble of models obtained in the previous iteration of the inverse method. Therefore, the ensemble of models must be large enough to avoid the collapse of the ensemble. If the ensemble size is small, other methods such as covariance localization or covariance inflation could be used (Evensen, 2007). In the ensemble smoother, all the data are assimilated simultaneously, whereas in other ensemble-based methods, such as the ensemble Kalman filter, data are assimilated sequentially (Evensen, 2007). In the ES-MDA, the data are assimilated multiple times with inflation factors of the covariance matrix.

The method can be summarized as follows. We define the number of iterations N and the inflation coefficients $\{\alpha_i\}_{i=1,\ldots,N}$ with $\sum_{i=1}^{N}\alpha_i^{-1} = 1$. We sample an initial set of N_e realizations $\{\mathbf{m}_j^0\}_{j=1,\ldots,N_e}$ of the model properties and compute the forward operator to compute the predicted

data $\{\mathbf{d}_j^0\}_{j=1,\ldots,N_e}$. At each iteration and for each ensemble member, we apply a perturbation to the data vector $\mathbf{d}$ and obtain the perturbed data $\mathbf{d_p} = \mathbf{d} + \sqrt{\alpha_i}\boldsymbol{\Sigma}_\mathbf{d}^{1/2}\mathbf{z_d}$, where $\mathbf{z_d} \sim N(0, \mathbf{I}_n)$, where n is the length of the data. At each iteration i, we update each ensemble member as

$$\begin{aligned}\mathbf{m}_j^i &= \mathbf{m}_j^{i-1} + \boldsymbol{\Sigma}_{\mathbf{m,d}}^{i-1}(\boldsymbol{\Sigma}_{\mathbf{d,d}}^{i-1} + \boldsymbol{\Sigma}_\mathbf{d})^{-1}(\mathbf{d}_j^{i-1} - \mathbf{d}_{\mathbf{p}j}) \\ &\equiv \mathbf{m}_j^{i-1} + \mathbf{K_i}(\mathbf{d}_j^{i-1} - \mathbf{d}_{\mathbf{p}j}) \end{aligned} \tag{23}$$

for $j = 1, \ldots, N_e$, where $\mathbf{K_i} \equiv \boldsymbol{\Sigma}_{\mathbf{m,d}}^{i-1}(\boldsymbol{\Sigma}_{\mathbf{d,d}}^{i-1} + \boldsymbol{\Sigma}_\mathbf{d})^{-1}$ is generally called the Kalman gain, $\boldsymbol{\Sigma}_{\mathbf{m,d}}^{i-1}$ is the cross-covariance matrix between the vector of model parameters $\mathbf{m}^{i-1}$ and the vector of the corresponding predicted data $\mathbf{d}^{i-1}$ at the previous iteration, and $\boldsymbol{\Sigma}_{\mathbf{d,d}}^{i-1}$ is the $n \times n$ covariance matrix of the predicted data at the previous iteration. The covariance matrices are estimated using the ensemble of models and predictions.

Large data sets require a large ensemble of initial models, which increases the computational time for complex nonlinear models. It is a common practice in many application fields to reduce the size of the data set and use a smaller number of elements in the ensemble of models. The data reduction can be performed by spatially subsampling the data or by applying more sophisticated data reduction methods. Similar to the data reduction for linear inverse problems in equation 10, we can compute a reduced data vector $\tilde{\mathbf{d}} = \tilde{\mathbf{V}}_\mathbf{d}^{-1}(\mathbf{d} - \boldsymbol{\mu}_\mathbf{d})$ and update the ensemble according to

$$\begin{aligned}\mathbf{m}_j^i &= \mathbf{m}_j^{i-1} + \boldsymbol{\Sigma}_{\mathbf{m,\tilde{d}}}^{i-1}(\boldsymbol{\Sigma}_{\tilde{\mathbf{d}},\tilde{\mathbf{d}}}^{i-1} + \boldsymbol{\Sigma}_\mathbf{d})^{-1}(\tilde{\mathbf{d}}_j^{i-1} - \tilde{\mathbf{d}}_{\mathbf{p}j}) \\ &\equiv \mathbf{m}_j^{i-1} + \tilde{\mathbf{K}}_\mathbf{i}(\tilde{\mathbf{d}}_j^{i-1} - \tilde{\mathbf{d}}_{\mathbf{p}j}) \end{aligned} \tag{24}$$

where the term $\tilde{\mathbf{K}}_\mathbf{i} \equiv \boldsymbol{\Sigma}_{\mathbf{m,\tilde{d}}}^{i-1}(\boldsymbol{\Sigma}_{\tilde{\mathbf{d}},\tilde{\mathbf{d}}}^{i-1} + \boldsymbol{\Sigma}_\mathbf{d})^{-1}$ is the Kalman gain for the reduced data space, $\tilde{\mathbf{d}}^{i-1}$ represents the predicted data in the reduced data space, and the term $\tilde{\mathbf{d}}_\mathbf{p} = \tilde{\mathbf{d}} + \sqrt{\alpha_i}\boldsymbol{\Sigma}_{\tilde{\mathbf{d}}}^{1/2}\mathbf{z}_{\tilde{\mathbf{d}}}$ (with $\mathbf{z}_{\tilde{\mathbf{d}}} \sim N(0, \mathbf{I}_{\tilde{n}})$, where $\tilde{n}$ is the length of the reduced data) is the perturbation of the observed data in the reduced data space. The parameters of the posterior distribution, i.e., the posterior mean and the posterior covariance matrix, are inferred from the updated ensemble. Additional research results on data assimilation in the reduced data space can be found in Evensen (2007) and Chen and Oliver (2017).

In the next section, we consider a simple 1D inverse problem with the goal of predicting elastic properties from the measured seismic trace and we analyze the uncertainty of the solution with and without data and/or model reduction, for the linear and nonlinear cases.

APPLICATION

We first focus on a simple linear inverse problem for the prediction of acoustic impedance from a zero-offset seismic trace. The model $\mathbf{m}$ is a vector of n_m samples of logarithm of acoustic impedance, and the data $\mathbf{d}$ is a vector of $n_d = n_m - 1$ samples of seismic amplitudes. The forward operator $\mathbf{g}$ is a convolution of a wavelet $\mathbf{w}$ and the linear reflection coefficient expression for an incident angle of 0°:

$$\begin{aligned}\mathbf{d}(t) = \mathbf{w}(t) \times \mathbf{r}(t) + \mathbf{e}(t) &= \int \mathbf{w}(u)\mathbf{r}(t-u)du + \mathbf{e}(t) \\ &= \frac{1}{2}\int \mathbf{w}(u)\frac{d\mathbf{m}(t-u)}{dt}du + \mathbf{e}(t). \end{aligned} \tag{25}$$

In the discrete case, the forward operator can be expressed as a matrix $\mathbf{G} = 1/2\mathbf{WD}$ of size $n_d \times n_m$, where $\mathbf{W}$ is the convolution matrix and $\mathbf{D}$ is the first-order differential matrix (Buland and Omre, 2003). The forward operator is then linear in the model variable, i.e., the logarithm of acoustic impedance.

We introduce here a synthetic example (Figure 1) to illustrate the inversion method. The model vector $\mathbf{m}$ includes $n_m = 70$ samples of logarithm of acoustic impedance, and the data vector $\mathbf{d}$ includes $n_d = 69$ samples of seismic amplitudes (Figure 1). The wavelet in the forward model is a Ricker function with dominant frequency of 30 Hz, and we apply a signal-to-noise ratio of 10. We assume that the model $\mathbf{m}$ is prior distributed according to a Gaussian distribution $N(\mathbf{m}; \boldsymbol{\mu}_\mathbf{m}, \boldsymbol{\Sigma}_\mathbf{m})$, where the prior mean $\boldsymbol{\mu}_\mathbf{m}$ is a $n_\mathrm{m} \times 1$ vector and the prior covariance $\boldsymbol{\Sigma}_\mathbf{m}$ is a $n_m \times n_m$ matrix. The prior model is shown in Figure 1. The prior mean is constant and equal to 9.25 in the log domain (corresponding to an actual value of 10,400 m/sg/cm^3), and the prior covariance matrix is the prior variance (spatially invariant and equal to 0.0023) multiplied by a symmetric spatial correlation matrix obtained from an exponential correlation function with a range equal to 25 samples. In real applications, the prior mean is a vector including a low-frequency (background) model and the covariance matrix includes a spatial correlation model obtained from the empirical spatial correlation function.

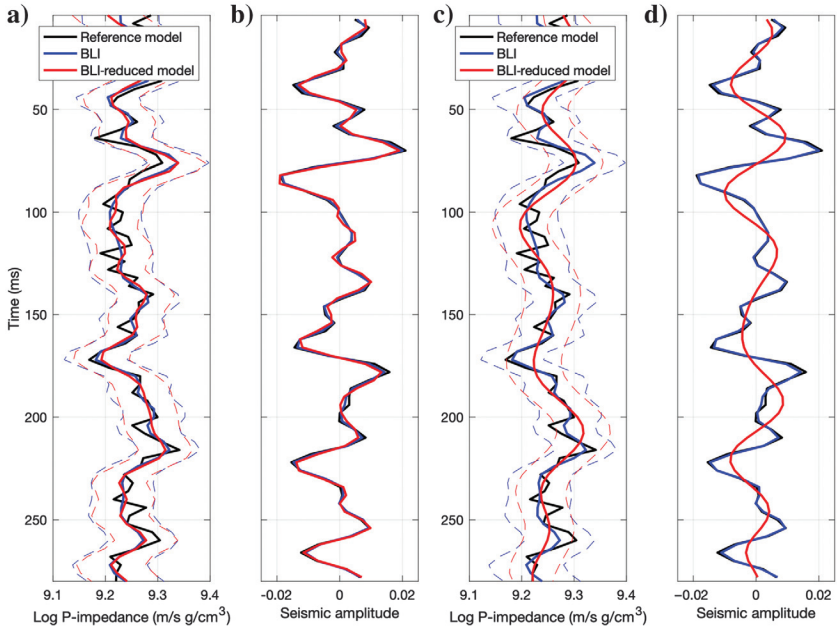

Figure 3. Bayesian linear seismic inversion for the logarithm of P-impedance with and without model reduction. Plots (a and b) show the comparison between BLI and BLI with 90% variance model reduction; plots (c and d) show the comparison between BLI and BLI with 75% variance model reduction. The blue lines represent the posterior distribution and the predicted data from BLI, the red lines represent the BLI with reduced model, and the solid black lines represent the reference model and data. The dashed lines represent the 95% confidence interval.

We apply the Bayesian linear inversion (BLI) (Buland and Omre, 2003). The results are the point-wise posterior distributions of the logarithm of acoustic impedance conditioned by the seismic data. The results are shown in Figure 2. The posterior mean matches the reference model. The posterior variance is constant (equation 6 does not depend on the data) except for at the top and the bottom of the interval due to border effects.

Dimension reduction for linear inverse problems

We first compare the results of BLI with the results of BLI with reduced data. We study two scenarios in which we reduce the data to preserve 90% and 75% of the total variability of the data, respectively. The results of BLI with 90% variance data reduction (i.e., preserving 90% of the data variability, obtained using 22 principal components) show good agreement with BLI. The uncertainty (i.e., the posterior variance) is slightly larger, and the predicted data are slightly smoother; however, the data reduction does not affect the inversion results largely. The quality of the inversion results decreases for the BLI with 75% variance data reduction (i.e., preserving 75% of the data variability, obtained using 16 principal components). The uncertainty in the solution is larger, and the mismatch between predicted model and reference model as well as between predicted data and measured data is worse than the 90% variance data reduction case. The resulting predicted model and data are smoother than the actual model due to the loss of information (i.e., high frequencies) in the reduced data.

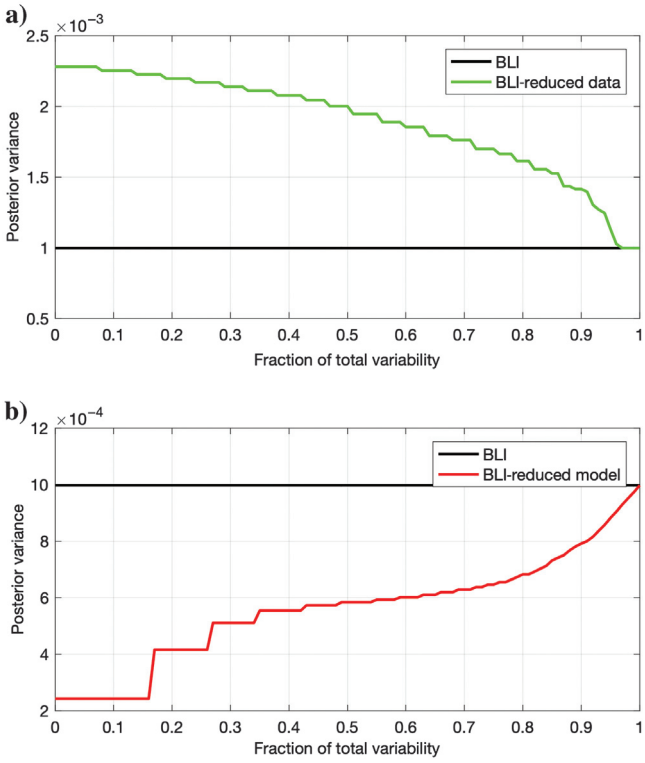

Figure 4. Uncertainty analysis: variances of the posterior distribution as a function of the fraction of total variability of the reduced data versus the full data (a) and the reduced model versus the full model (b). The black line represents the posterior variance of BLI ($\sigma^2 = 10^{-3}$), the green line represents the posterior variance of BLI with data reduction, and the red line represents the posterior variance of BLI with model reduction.

We then compare the results of BLI with the results of BLI with reduced model (90% and 75% of the total variance of the model, corresponding to 30 and 13 principal components, respectively). The inversion results of BLI with 90% variance model reduction show a good agreement with BLI (Figure 3); however, we notice that the uncertainty (i.e., the posterior variance) is smaller than BLI. The uncertainty reduction is more pronounced for the BLI with 75% variance model reduction. Similar to the data reduction case, the mismatch between the predicted model and the reference model as well as between predicted data and measured data is worse for 75% variance model reduction than 90% variance model reduction.

We compute the posterior distribution of the inverse problem with data and model reduction for all the percentages of total variance from 0% to 100%, in which the reduction with 100% of the total variance corresponds to the full inverse problem. Figure 4 shows the posterior variance as a function of the fraction of total variability of reduced data versus full data (Figure 4a) and the posterior variance as a function of the fraction of total variability of reduced model versus full model (Figure 4b). Figure 5 shows the same results as a function of the number of principal components used to reduce the data (Figure 5a) and model (Figure 5b). The reference variance is the posterior variance of the BLI, analytically computed using equation 6 and equal to 10^{-3}. In the data reduction case, the posterior variance always overestimates the BLI posterior variance. The uncertainty overestimation increases as the fraction of the total variability decreases (i.e., with fewer principal components). In the model reduction case, the posterior variance always underestimates the BLI

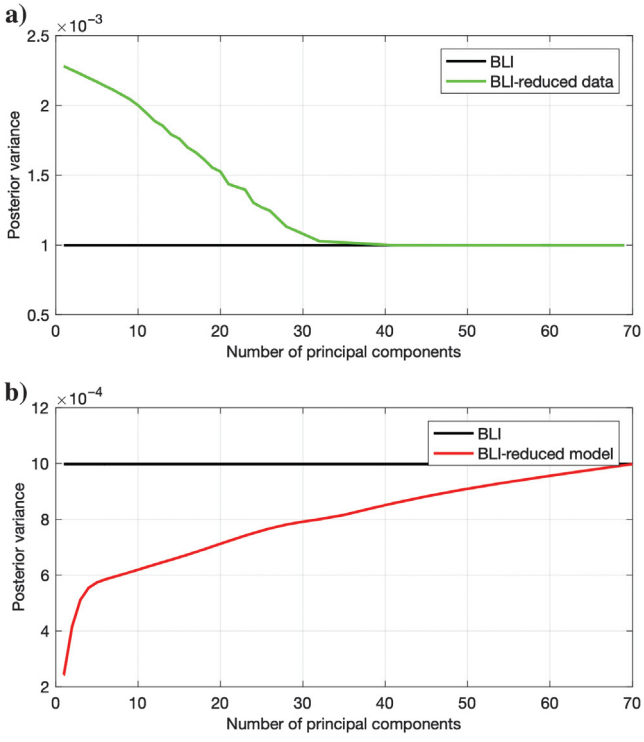

Figure 5. Uncertainty analysis: variances of the posterior distribution as a function of the number of principal components preserved in the reduced data (a) and the reduced model (b). The black line represents the posterior variance of BLI ($\sigma^2 = 10^{-3}$), the green line represents the posterior variance of BLI with data reduction, and the red line represents the posterior variance of BLI with model reduction.

posterior variance. The uncertainty underestimation increases as the fraction of the total variability decreases. The posterior variances of the BLI with data and model reduction converge to the BLI posterior variance when the number of principal components is approximately 40 for data reduction and approximately 68 for model reduction. These results can be explained using equations 14 and 22. In the data reduction case, the posterior variance of the model conditioned by the reduced data (equation 14) is always greater than or equal to the posterior variance conditioned by the full data set. Indeed, it can be numerically shown that

$$(\boldsymbol{\Sigma}_{\mathbf{m}}(\tilde{\mathbf{V}}_{\mathbf{d}}^{-1}\mathbf{G})^T(\tilde{\mathbf{V}}_{\mathbf{d}}^{-1}\mathbf{G}\boldsymbol{\Sigma}_m(\tilde{\mathbf{V}}_{\mathbf{d}}^{-1}\mathbf{G})^T+\tilde{\mathbf{V}}_{\mathbf{d}}\boldsymbol{\Sigma}_{\mathbf{d}}\tilde{\mathbf{V}}_{\mathbf{d}}^T)^{-1}\tilde{\mathbf{V}}_{\mathbf{d}}^{-1}\mathbf{G}\boldsymbol{\Sigma}_{\mathbf{m}})_{i,i} \leq (\boldsymbol{\Sigma}_{\mathbf{m}}\mathbf{G}^T(\mathbf{G}\boldsymbol{\Sigma}_{\mathbf{m}}\mathbf{G}^T+\boldsymbol{\Sigma}_{\mathbf{d}})^{-1}\mathbf{G}\boldsymbol{\Sigma}_{\mathbf{m}})_{i,i} \quad (26)$$

for $i = 1, \ldots n$. Similarly, in the model reduction case, the posterior variance of the model conditioned by the data and reduced model (equation 22) is always less than or equal to the posterior variance of the full model because it can be numerically shown that

$$(\tilde{\mathbf{V}}_{\mathbf{m}}(\boldsymbol{\Sigma}_{\tilde{\mathbf{m}}}-\boldsymbol{\Sigma}_{\tilde{\mathbf{m}}}(\mathbf{G}\tilde{\mathbf{V}}_{\mathbf{m}})^T(\mathbf{G}\tilde{\mathbf{V}}_{\mathbf{m}}\boldsymbol{\Sigma}_{\tilde{\mathbf{m}}}(\mathbf{G}\tilde{\mathbf{V}}_{\mathbf{m}})^T+\boldsymbol{\Sigma}_{\mathbf{d}})^{-1}\mathbf{G}\tilde{\mathbf{V}}_{\mathbf{m}}\boldsymbol{\Sigma}_{\tilde{\mathbf{m}}})\tilde{\mathbf{V}}_{\mathbf{m}}^T)_{i,i} \leq (\boldsymbol{\Sigma}_{\mathbf{m}|\mathbf{d}})_{i,i} \quad (27)$$

for $i = 1, \ldots n$.

In Figure 6, we compare the results of BLI with the results of ES-MDA. The two methods provide similar results in terms of model predictions. The root-mean-square error for BLI is 0.1539, whereas the root-mean-square error for ES-MDA ranges from 0.10 to 0.20 (depending on the initial realizations). However, the ES-MDA underestimates the variance of the posterior distribution ($\sigma^2 = 0.0003$ for an ensemble of 1000 model) compared to BLI ($\sigma^2 = 0.0010$), despite a large number of models in the ensemble. This effect could be mitigated by applying a covariance localization (Evensen, 2007).

The uncertainty quantification results are summarized in Table 1 for all the inversion cases presented above. The variances of the data reduction cases are higher than the BLI variance, whereas the variances of the model reduction cases are smaller than the BLI variance. The 95% coverage ratio (i.e., the percentage of samples of the true model falling in the 95% confidence interval) also shows the overestimation of the uncertainty in the data reduction case and the underestimation in the model reduction case. If the uncertainty is correctly assessed, the 95% coverage ratio should be 0.95. The 95% coverage ratio for BLI is equal to 0.98, possibly due to the limited number of samples in the model vector. The 95% coverage ratios for the data reduction cases are equal to 1, meaning that the uncertainty is overestimated; the 95% coverage ratios for the model reduction cases are less than or equal to 0.95, meaning that the uncertainty is underestimated. The Table 1 also includes the inversion results obtained using the ES-MDA using 100, 1000, and 10,000 models and shows the underestimation of the variance for all of these cases.

Dimension reduction for nonlinear inverse problems

We then focus on a nonlinear problem for the prediction of elastic properties from three partial angles stacks of seismic data. The model $\mathbf{m}$ is a vector of $3n_m$ samples of P-wave velocity, S-wave velocity, and density, and the data $\mathbf{d}$ is a vector of $3n_d = 3n_m - 3$ samples of near-, mid-, and far-angle seismic data. The forward operator $\mathbf{g}$ is a convolution of the wavelets and the Zoeppritz equation for the reflection coefficients of the near, mid, and far angles.

The synthetic example for the nonlinear case is shown in Figure 7. We assume that the model $\mathbf{m}$ is prior distributed according to a trivariate Gaussian distribution $N(\mathbf{m}; \boldsymbol{\mu}_{\mathbf{m}}, \boldsymbol{\Sigma}_{\mathbf{m}})$, where the prior mean

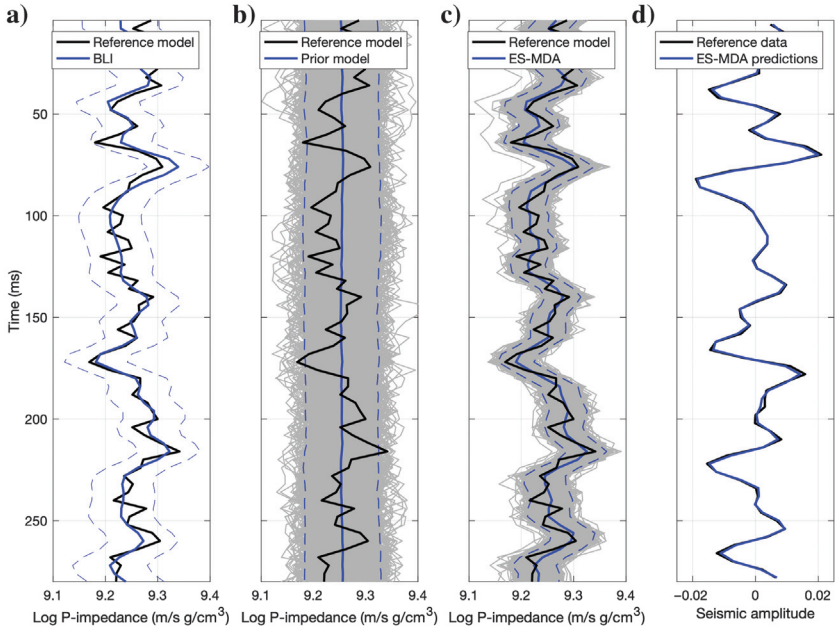

Figure 6. Comparison between BLI and ES-MDA for seismic linearized inversion: (a) posterior distribution of the logarithm of P-impedance from BLI (in blue), (b) prior distribution of the logarithm of P-impedance for ES-MDA (in blue) (gray lines represent the prior realizations), (c) posterior distribution of the logarithm of P-impedance from ES-MDA (in blue) (gray lines represent the posterior realizations), and (d) predicted seismic amplitudes are indicated in blue. The black line represents the reference model and data. The dashed lines represent the 95% confidence interval.

Table 1. Uncertainty quantification for the linear inverse problem using BLI (variances and 95% coverage ratios).

	Variance (Ln I_P)	95% Coverage ratio (Ln I_P)
BLI	0.0010	0.98
BLI — data reduction 90%	0.0016	1.00
BLI — data reduction 75%	0.0018	1.00
BLI — model reduction 90%	0.0008	0.95
BLI — model reduction 75%	0.0006	0.94
ES-MDA (100 models)	0.0002	0.85
ES-MDA (1000 models)	0.0003	0.97
ES-MDA (10,000 models)	0.0003	0.97

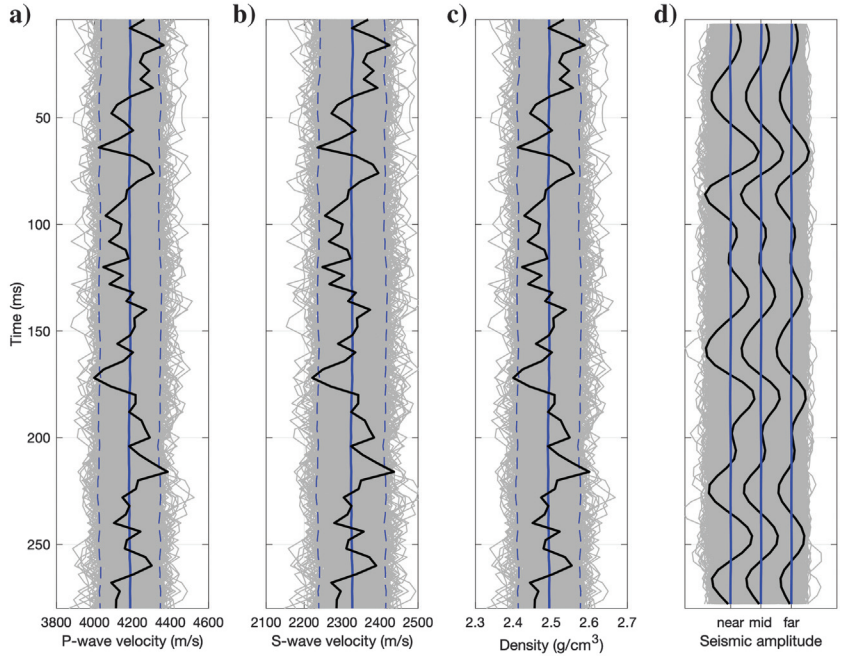

Figure 7. Synthetic example of nonlinear seismic inversion. From left to right: (a) P-wave velocity, (b) S-wave velocity, (c) density, and (d) seismic data (near, mid, and far angle). The black lines represent the reference model, the blue lines represent the prior distributions of the model variables (the solid line represents the mean, and the dashed lines represent the 95% confidence interval), and the gray lines represent 1000 prior realizations and the corresponding predicted data.

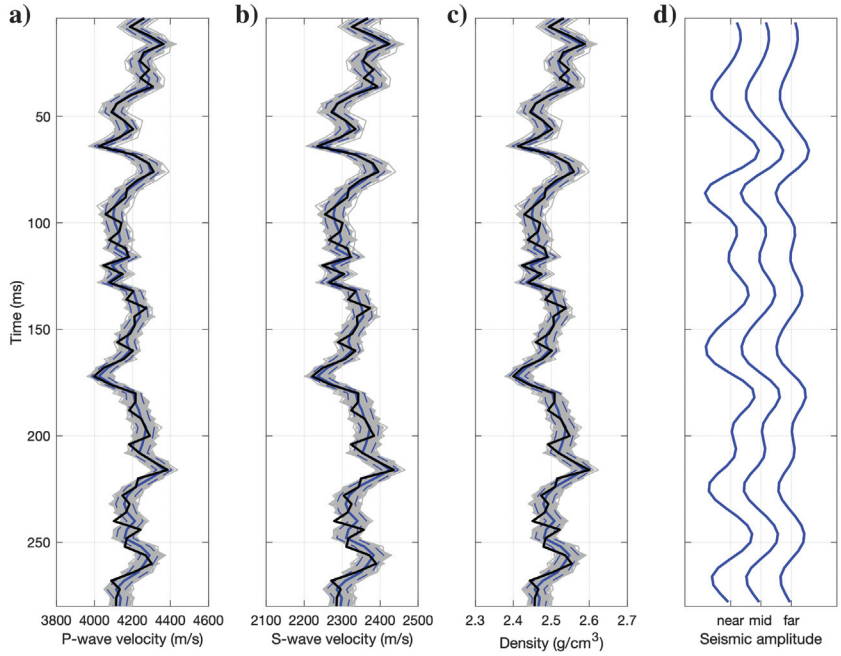

Figure 8. Nonlinear inversion results obtained from the ES-MDA. From left to right: (a) P-wave velocity, (b) S-wave velocity, (c) density, and (d) seismic data (near, mid, and far angle). The black lines represent the reference model, the blue lines represent the posterior distributions of the model variables (the solid line represents the mean, and the dashed lines represent the 95% confidence interval), and the gray lines represent 1000 posterior realizations and the corresponding predicted data (the predicted data overlap with the measured data).

$\boldsymbol{\mu}_{\mathbf{m}}$ is a $3n_m \times 1$ vector and the prior covariance $\boldsymbol{\Sigma}_{\mathbf{m}}$ is a $3n_m \times 3n_m$ matrix, obtained as the Kronecker product of the spatially independent covariance matrix (estimated from the reference model) and a spatial correlation matrix (computed from an exponential correlation function with range equal to 25 samples). The prior model is shown in Figure 7. A set of $N_e = 1000$ geostatistical realizations is computed and used as the ensemble of prior models (Figure 7). We adopt the ES-MDA method to update the ensemble of models and compute the approximate posterior distribution from the posterior ensemble. A sensitivity analysis on the number of models in the ensemble was performed; at least 1000 models are required to obtain stable statistics of the posterior distribution; with less than 100 models the ensemble collapses after the first iteration. The inversion results for the problem with the full data and model are shown in Figure 8. The posterior mean matches the reference model; however, the posterior variance shows that the uncertainty is fairly small, as expected for stochastic inverse methods. We then apply the same inverse method with a reduced data vector obtained by applying PCA. Figure 9 shows the results of the inversion with a reduced vector of $3n_{\tilde{d}} = 66$, preserving 90% of the total variance of the data. Figure 10 shows the results of the inversion with a reduced vector of $3n_{\tilde{d}} = 27$, preserving 75% of the total variance of the data. The posterior variance increases as the fraction of the total variability decreases. The mismatch between predicted model and reference model and between predicted data and measured data is larger for the 75% variance data reduction case than for the 90% variance data reduction case. Overall, the variances are larger for the data reduction cases than the full data case. It seems that the data reduction compensates the uncertainty reduction generally obtained using ES-MDA method with a limited number of models; however, the accuracy of the posterior mean and the data prediction decreases with the greater data reduction.

DISCUSSION

The linear inverse problem allows assessing the changes in uncertainty quantification when data or model reduction is applied. The inverse problem is solved in a Bayesian setting under Gaussian assumptions, in which the mean and the variance of the posterior distribution can be analytically assessed. The variance of the posterior distribution of the model is compared to the posterior variance of the solution of the full problem. The posterior variance of the reduced data problem overestimates the posterior variance of the full data problem, and it increases when the fraction of the variance of reduced data versus the full data

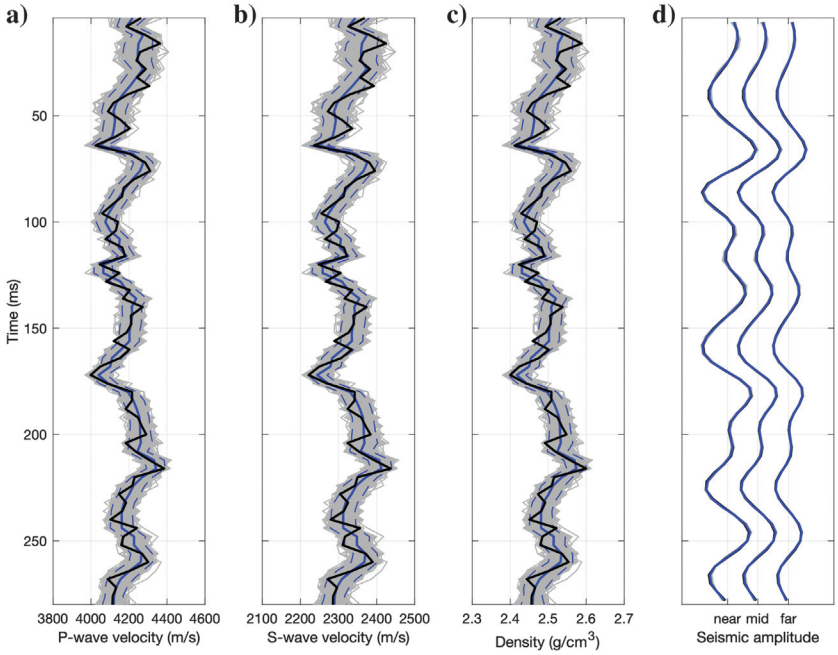

Figure 9. Nonlinear inversion results obtained from the ES-MDA with 90% variance data reduction. From left to right: (a) P-wave velocity, (b) S-wave velocity, (c) density, and (d) seismic data (near, mid, and far angle). The black lines represent the reference model, the blue lines represent the posterior distributions of the model variables (the solid line represents the mean, and the dashed lines represent the 95% confidence interval), and the gray lines represent 1000 posterior realizations and the corresponding predicted data (the predicted data overlap with the measured data).

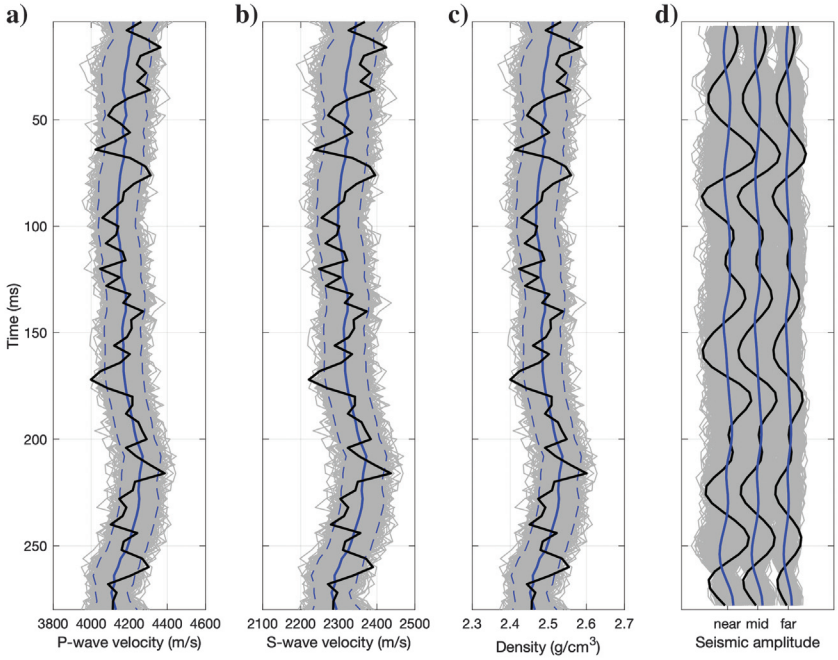

Figure 10. Nonlinear inversion results obtained from the ES-MDA with 75% variance data reduction. From left to right: (a) P-wave velocity, (b) S-wave velocity, (c) density, and (d) seismic data (near, mid, and far angle). The black lines represent the reference model, the blue lines represent the posterior distributions of the model variables (the solid line represents the mean, and the dashed lines represent the 95% confidence interval), and the gray lines represent 1000 posterior realizations and the corresponding predicted data (the predicted data overlap with the measured data).

decreases. The posterior variance of the reduced model problem underestimates the posterior variance of the full model problem, and it decreases when the fraction of the variance of reduced data versus the full data decreases, according to the analytical formulation and the numerical evidence. Indeed, reducing the data makes the inverse problem underdetermined (more model variables than measured data), whereas reducing the model makes the inverse problem overdetermined (more measured data than model variables).

Other linear dimension reduction methods can also be applied. Data could be reduced by subsampling the data vector, for example, by sampling the seismic data every other crossline or inline. Nonlinear reduction methods, such as MDS of the data vector or wavelet transform of the model vector, could also be applied, but analytical solutions are not available.

Analytical solutions for the posterior distribution for nonlinear problems are generally not available. In this work, we proposed the ES-MDA, but any other stochastic inverse method (Markov chain Monte Carlo or stochastic optimization algorithms) could be used. Similar conclusions to the linear case can be drawn for the nonlinear case. However, it is important to point out that the accuracy and precision of the solution of ES-MDA depend on the number of models in the initial ensemble. A limited number of models in the ensemble leads to an underestimation of the uncertainty. The presented methodology can be extended to 2D and 3D problems using a trace-by-trace approach. Advanced dimension reduction methods based on machine learning algorithms, such as the convolutional autoencoder, can also be applied to capture the spatial features of the data and/or model distribution (Liu and Grana, 2018).

CONCLUSION

We formalized the use of linear reduction of data and model in the Bayesian linear workflow and compared the results with traditional BLI. The assessment of the uncertainty for Gaussian-linear inverse problem can be performed analytically using closed forms of the parameters of the posterior distribution of the model given the data. The application to linearized acoustic and nonlinear elastic inversion shows the impact of reduction methods on the posterior uncertainty of the predicted models. We proved with the analytical formulation and the numerical experiments that data reduction leads to the overestimation of the uncertainty and model reduction leads to the underestimation of the uncertainty. We also extended the approach to nonlinear inverse problems by using an ensemble-based method. Further studies should be conducted for nonlinear reduction

methods, but an assessment of the posterior uncertainty is strongly recommended for any geophysical inverse problem with a nonunique solution.

DATA AND MATERIALS AVAILABILITY

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

REFERENCES

Aki, K., and P. G. Richards, 1980, Quantitative seismology: W. H. Freeman and Co.

Aster, R. C., B. Borchers, and C. H. Thurber, 2011, Parameter estimation and inverse problems: Elsevier.

Azevedo, L., R. Nunes, P. Correia, A. Soares, L. Guerreiro, and G. S. Neto, 2013, Multidimensional scaling for the evaluation of a geostatistical seismic elastic inversion methodology: Geophysics, **79**, no. 5, M1–M10, doi: 10.1190/geo2013-0037.1.

Azevedo, L., and A. Soares, 2017, Geostatistical methods for reservoir geophysics: Springer.

Bishop, C. M., 2006, Pattern recognition and machine learning: Springer.

Bosch, M., C. Carvajal, J. Rodrigues, A. Torres, M. Aldana, and J. Sierra, 2009, Petrophysical seismic inversion conditioned to well-log data: Methods and application to a gas reservoir: Geophysics, **74**, no. 2, O1–O15, doi: 10.1190/1.3043796.

Bosch, M., T. Mukerji, and E. Gonzalez, 2010, Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review: Geophysics, **75**, no. 5, 75A165–75A176, doi: 10.1190/1.3478209.

Buland, A., and H. Omre, 2003, Bayesian linearized AVO inversion: Geophysics, **68**, 185–198, doi: 10.1190/1.1543206.

Chen, J., M. Hoversten, D. W. Vasco, Y. Rubin, and Z. Hou, 2007, Joint inversion of seismic AVO and EM data for gas saturation estimation using a sampling based stochastic model: Geophysics, **72**, no. 2, WA85–WA95, doi: 10.1190/1.2435082.

Chen, Y., and D. S. Oliver, 2017, Localization and regularization for iterative ensemble smoothers: Computational Geosciences, **21**, 13–30, doi: 10.1007/s10596-016-9599-7.

Cox, T. F., and M. A. Cox, 2000, Multidimensional scaling: Chapman and Hall/CRC.

de Figueiredo, L. P., D. Grana, F. L. Bordignon, M. Santos, M. Roisenberg, and B. B. Rodrigues, 2018, Joint Bayesian inversion based on rock-physics prior modeling for the estimation of spatially correlated reservoir properties: Geophysics, **83**, no. 5, M49–M61, doi: 10.1190/geo2017-0463.1.

de Figueiredo, L. P., D. Grana, M. Roisenberg, and B. B. Rodrigues, 2019, Gaussian mixture Markov chain Monte Carlo method for linear seismic inversion: Geophysics, **84**, no. 3, R463–R476, doi: 10.1190/geo2018-0529.1.

Dejtrakulwong, P., T. Mukerji, and G. Mavko, 2012, Using kernel principal component analysis to interpret seismic signatures of thin shaly-sand reservoirs: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-1013.1.

Doyen, P., 2007, Seismic reservoir characterization: EAGE.

Eidsvik, J., P. Avseth, H. Omre, T. Mukerji, and G. Mavko, 2004, Stochastic reservoir characterization using prestack seismic data: Geophysics, **69**, 978–993, doi: 10.1190/1.1778241.

Emerick, A. A., and A. C. Reynolds, 2013, Ensemble smoother with multiple data assimilation: Computers and Geosciences, **55**, 3–15, doi: 10.1016/j.cageo.2012.03.011.

Evensen, G., 2007, Data assimilation: The ensemble Kalman filter: Springer.

Golub, G. H., and F. Van Loan, 2013, Matrix computations: JHU Press.

Grana, D., 2016, Bayesian linearized rock-physics inversion: Geophysics, **81**, no. 6, D625–D641, doi: 10.1190/geo2016-0161.1.

Grana, D., T. Fjeldstad, and H. Omre, 2017, Bayesian Gaussian mixture linear inversion for geophysical inverse problems: Mathematical Geosciences, **49**, 493–515, doi: 10.1007/s11004-016-9671-9.

Gunning, J., and M. E. Glinsky, 2007, Detection of reservoir quality using Bayesian seismic inversion: Geophysics, **72**, no. 3, R37–R49, doi: 10.1190/1.2713043.

Hansen, T. M., A. G. Journel, A. Tarantola, and K. Mosegaard, 2006, Linear inverse Gaussian theory and geostatistics: Geophysics, **71**, no. 6, R101–R111, doi: 10.1190/1.2345195.

Hastie, T., R. Tibshirani, and J. Friedman, 2002, The elements of statistical learning: Springer.

Hastings, W. K., 1970, Monte Carlo sampling methods using Markov chains and their applications: Biometrika, **57**, 97–109, doi: 10.1093/biomet/57.1.97.

He, J., P. Sarma, E. Bhark, S. Tanaka, B. Chen, X.-H. Wen, and J. Kamath, 2018, Quantifying expected uncertainty reduction and value of information using ensemble-variance analysis: SPE Journal, **23**, 428–448, doi: 10.2118/182609-PA.

Jeong, H., A. Y. Sun, J. Lee, and B. Min, 2018, A learning-based data-driven forecast approach for predicting future reservoir performance: Advances in Water Resources, **118**, 95–109, doi: 10.1016/j.advwatres.2018.05.015.

Jolliffe, I., 2011, Principal component analysis, *in* M. Lovric, ed., International encyclopedia of statistical science: Springer.

Jullum, M., and O. Kolbjørnsen, 2016, A Gaussian based framework for Bayesian inversion of geophysical data to rock properties: Geophysics, **81**, no. 3, R75–R87, doi: 10.1190/geo2015-0314.1.

Larsen, A. L., M. Ulvmoen, H. Omre, and A. Buland, 2006, Bayesian lithology/fluid prediction and simulation on the basis of a Markov-chain prior model: Geophysics, **71**, no. 5, R69–R78, doi: 10.1190/1.2245469.

Lima, M. M., A. E. Emerick, and C. E. P. Ortiz, 2019, Data-space inversion with ensemble smoother: arXiv:submit/2623697.

Liu, M., and D. Grana, 2018, Ensemble-based seismic history matching with data re-parameterization using convolutional autoencoder: 88th Annual International Meeting, SEG, Expanded Abstracts, 3156–3160, doi: 10.1190/segam2018-2997988.1.

MacGregor, L., 2012, Integrating seismic, CSEM, and well-log data for reservoir characterization: The Leading Edge, **31**, 268–277.

Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, **100**, 12431–12447, doi: 10.1029/94JB03097.

Oliver, D. S., and M. Alfonzo, 2018, Calibration of imperfect models to biased observations: Computational Geosciences, **22**, 145–161, doi: 10.1007/s10596-017-9678-4.

Oliver, D. S., A. C. Reynolds, and N. Liu, 2008, Inverse theory for petroleum reservoir characterization and history matching: Cambridge University Press.

Rao, R., 2002, Wavelet transforms: Encyclopedia of imaging science and technology: John Wiley & Sons Inc.

Rimstad, K., P. Avseth, and H. Omre, 2012, Hierarchical Bayesian lithology/fluid prediction: A North Sea case study: Geophysics, **77**, no. 2, B69–B85, doi: 10.1190/geo2011-0202.1.

Sambridge, M., and K. Mosegaard, 2002, Monte Carlo methods in geophysical inverse problems: Reviews of Geophysics, **40**, 1–29, doi: 10.1029/2000RG000089.

Satija, A., and J. Caers, 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, doi: 10.1016/j.advwatres.2015.01.002.

Scales, J. A., and L. Tenorio, 2001, Prior information and uncertainty in inverse problems: Geophysics, **66**, 389–397, doi: 10.1190/1.1444930.

Scheidt, C., L. Li, and J. Caers, 2018, Bayesian evidential learning, Quantifying uncertainty in subsurface systems: John Wiley & Sons Inc.

Scheidt, C., P. Renard, and J. Caers, 2015, Prediction-focused subsurface modeling: Investigating the need for accuracy in flow-based inverse modeling: Mathematical Geosciences, **47**, 173–191, doi: 10.1007/s11004-014-9521-6.

Sen, M. K., and P. L. Stoffa, 1996, Bayesian inference, Gibbs sampler and uncertainty estimation in geophysical inversion: Geophysical Prospecting, **44**, 313–350, doi: 10.1111/j.1365-2478.1996.tb00152.x.

Sen, M. K., and P. L. Stoffa, 2013, Global optimization methods in geophysical inversion: Cambridge University Press.

Sheriff, R. E., and L. P. Geldart, 1995, Exploration seismology: Cambridge University Press.

Sun, W., M.-H. Hui, and L. J. Durlofsky, 2017, Production forecasting and uncertainty quantification for naturally fractured reservoirs using a new data-space inversion procedure: Computational Geosciences, **21**, 1443–1458.

Tarantola, A., 2005, Inverse problem theory: SIAM.

Tompkins, M. J., J. L. Fernández Martínez, L. Alumbaugh, and T. Mukerji, 2011, Scalable uncertainty estimation for nonlinear inverse problems using parameter reduction, constraint mapping, and geometric sampling: Marine controlled-source electromagnetic examples: Geophysics, **76**, no. 4, F263–F281, doi: 10.1190/1.3581355.

Ulrych, T. J., M. D. Sacchi, and A. Woodbury, 2001, A Bayes tour of inversion: A tutorial: Geophysics, **66**, 55–69, doi: 10.1190/1.1444923.

Ulvmoen, M., and H. Omre, 2010, Improved resolution in Bayesian lithology/fluid inversion from prestack seismic data and well observations: Part 1 — Methodology: Geophysics, **75**, no. 2, R21–R35, doi: 10.1190/1.3294570.