

Ensemble-based uncertainty estimation in full waveform inversion

J. Thurin,¹ R. Brossier¹ and L. Métivier^{1,2}

¹*Univ. Grenoble Alpes, ISTerre, F-38000 Grenoble, France. E-mail: julien.thurin@univ-grenoble-alpes.fr*

²*Univ. Grenoble Alpes, CNRS, LJK, F-38000 Grenoble, France*

Accepted 2019 August 23. Received 2019 April 12; in original form 2019 July 12

SUMMARY

Uncertainty estimation and quality control are critically missing in most geophysical tomographic applications. The few solutions to cope with that issue are often left out in practical applications when these ones grow in scale and involve complex modelling. We present a joint full waveform inversion and ensemble data assimilation scheme, allowing local Bayesian estimation of the solution that brings uncertainty estimation to the tomographic problem. This original methodology relies on a deterministic square root ensemble Kalman filter commonly used in the data assimilation community: the ensemble transform Kalman filter (ETKF). Combined with a 2-D visco-acoustic frequency domain full waveform inversion scheme, the resulting method allows to access a low-rank approximation of the posterior covariance matrix of the solution. It yields uncertainty information through an ensemble-representation, that can conveniently be mapped across the physical domain for visualization and interpretation. The combination of ETKF with full waveform inversion is discussed along with the scheme design and algorithmic details that lead to our mixed application. Both synthetic and field-data results are presented, along with the biases that are associated with the limited rank ensemble representation. Finally, we review the open questions and developments perspectives linked with data assimilation applications to the tomographic problem.

Key words: Probability distributions; Statistical methods; Tomography; Waveform inversion.

1 INTRODUCTION

Geophysical tomography is a set of fundamental techniques in geophysical exploration, allowing to make sense of physical measurements to characterize subsurfaces properties. Seismic tomography, in particular, aims at estimating said properties from wavefield measurements. As wavefields behaviour and evolution are imposed by the physical properties of their propagating medium, it is possible to determine those physical parameters through inverse problem-solving.

Seismic tomographic applications cover a broad spectrum of scales and targets, as they are commonly used both, for regional to global scale in the academic community (Aki *et al.* 1977; Bedle & Lee 2009; Panning *et al.* 2010; French & Romanowicz 2015) and for crustal-scale exploration industrial applications (crustal scale imaging, reservoir monitoring and civil engineering targets). Amongst the many variants of seismic tomography, full waveform inversion (FWI, Lailly 1983; Tarantola 1984) has been growing in popularity in the past decades. As opposed to most variants of seismic tomography application, FWI aims at taking advantages of the entirety of the recorded data without discarding any valuable phase and amplitude measurements, whereas other techniques tend to focus on specific and small subsets of information (first arrival time, first

arrival phase or amplitude). Despite implying a more complex problem to solve, (as it involves simulating complete wavefield recording instead of portions of data), its superior resolution power makes it a popular imaging technique both in academic geosciences and in the exploration industry.

Academic applications of FWI yield results allowing to better understand complex mechanisms and structures at depth (Fichtner *et al.* 2009; Tape *et al.* 2010; Fichtner *et al.* 2013; Bozdağ *et al.* 2016), that seismic ray tomography resolution may not allow. FWI is routinely applied in many industrial workflows (Plessix 2009; Sirgue *et al.* 2010; Plessix *et al.* 2012; Warner *et al.* 2013; Zhu *et al.* 2015; Operto *et al.* 2015) focused on crustal-scale exploration. In this context, FWI outputs are either used at later processing stage to perform migration of seismic-reflection data or might be used as a model on its own for interpretation (Shen *et al.* 2018). However, one of the main issues of FWI besides cost and complexity is the scarcity of options for quality control and uncertainty estimation. While such an ill-posed, non-linear problem would greatly benefit from it, the state-of-the-art methods have not been massively adopted by the community. Up to now, most of the quality control assessment is either conducted by cross-validation with other geophysical techniques or well-log data (*in situ* comparison) which are either costly or impossible to realize past the shallow crustal scale.

The uncertainty quantification research field in the frame of FWI and seismic tomography is quite recent, as it has been overlooked in favour of a focus on understanding the imaging power and issues of the FWI concept first, improving the formulation to make it more affordable and improve its outcomes. Consequently, the whole topic is still regarded as a challenging problem to tackle (Rawlinson *et al.* 2014). To deal with the uncertainty shortcoming, we can seek a solution in the Bayesian inference framework for general inverse problems, as presented by Tarantola (2005), which would allow expressing uncertainty in a Bayesian and probabilistic formalism, as opposed to FWI deterministic form. Tarantola (2005) states that the posterior covariance of the minimization problem is equivalent to the inverse Hessian operator when the solution is close to the global minimum. Hence, access to the inverse Hessian operator or the posterior covariance might be the solution to achieve uncertainty estimation that FWI is currently lacking. This is especially true in the case of multiparameter FWI where inter-parameters cross-talk are involved (Operto *et al.* 2013). Although recent propositions to evaluate the effects of the Hessian on a vector through second-order adjoint approaches has proven valuable in the context of optimization (Fichtner & Trampert 2011a; Métivier *et al.* 2013, 2014; Matharu & Sacchi 2019), computation of the full Hessian and its inverse is out of reach. In order to access the content of the Hessian operator and estimate the posterior covariance information, the ‘Hessian-based’ uncertainty estimation methodologies, are calling on either dimensionality reduction (making the Hessian size tractable) or evaluating partial information from the operator through various approximations. Although not producing uncertainty estimation, Fichtner & Trampert (2011b) and Fichtner & van Leeuwen (2015) estimate resolution and physical parameters tradeoffs of the solution through probing of the Hessian. Du *et al.* (2012) and later Jordan (2015) are relying on model parameterization with B-Spline functions, that drastically reduce the number of degrees of freedom, to estimate the full Hessian in a reduced problem. Bui-Thanh *et al.* (2013) approximate the Hessian operator at the solution, with the matrix-free Lanczos method to build a low-rank approximation of the Hessian. The pseudo-inverse of this approximation then yield an approximate inverse Hessian. Zhu *et al.* (2016), Eliasson & Romdhane (2017) and Liu & Peter (2019) are relying on the randomized Singular Value Decomposition (SVD) to estimate a truncated Hessian in a tractable way. Finally, using the wavefield reconstruction inversion (WRI) to relax the inverse problem formulation, Fang *et al.* (2018) demonstrate that the WRI cost function is particularly suited for the quadratic approximation that is assumed in all the methods mentioned above and therefore justifies the assumption of Gaussianity of the posterior distribution. To estimate uncertainty, they approximate the Gauss–Newton Hessian, from which the square root makes it possible to sample the posterior covariance once the approximate Gauss–Newton Hessian is computed and stored. However, these methodologies are limited by their computational cost: their Hessian approximation procedures are based on matrix-free Hessian-vector products, that require to solve expensive numerical wavefield simulations. This is also the case for the evaluation of the Gauss–Newton Hessian. These approximation procedure are also sequential by nature and therefore prevent scalability of the aforementioned uncertainty estimation method.

Concurrently, the data assimilation (DA) community has designed, for several decades, methods to solve inverse problems with a large number of degrees of freedom, high degree of complexity and data sparsity, while integrating uncertainty quantification within their inverse problem-solving schemes. Generally, the overall goal

of DA in geophysical applications is to characterize the state of a dynamic system through time, which can be subjected to non-linear dynamics, by combining sparse observations and numerical models. DA has notably been successfully implemented at operational scales in areas such as numerical weather forecasting, oceanography, reservoir characterization and climatology (Rodell *et al.* 2004; Navon 2009; Cosme *et al.* 2010; Lee *et al.* 2016). Most of the dynamic models in those fields of applications consist of solving a forward problem based on prior information on the system state. Unfortunately, the non-linearities inherent to those systems prevent accurate forecast at long timescales, as they tend to diverge if integrated for too long. To deal with this limitation, DA introduces observational knowledge as soon as it is available along the forecast, to correct the state estimates and thus improve the predictions by taking into account model and data biases. The first DA assimilation modern tool based on this idea of sequentially providing a forecast through modelling, and correcting the state estimate with observations, has been introduced by Kalman (1960), through the Kalman filter (KF). The KF, however, is a limited tool only providing solutions to study small-scale, linear systems. Even though the KF can be extended to non-linear problems, it requires manipulations of covariance matrices and large operators preventing to go beyond a few hundred to thousands of parameters. A generalization to large-scale problems has been proposed by Evensen (1994), introducing the ensemble Kalman filter (EnKF). It allows avoiding any explicit covariance matrices manipulations thanks to a low-rank ensemble representation of system states, from which the covariances information can be approximated. The EnKF is currently developed at an operational level and is commonly used in weather on up to 10^9 degrees of freedom (as it is the case for the MGREPS global assimilation system ran at the Met Office (United-Kingdom) or the ICON global domain model ran at the Deutscher Wetterdienst (Germany) as part of their numerical weather prediction routines). As EnKF allows to handle problems non-linearity and is designed to deal with a large number of parameters, as it is the case with the FWI problem, we might be able to take advantage of the EnKF formalism to bring a new look at uncertainty estimation in FWI.

Applying DA or ensemble-based methods to geophysical tomography has already started being investigated. Indeed Jin *et al.* (2008) propose using the EnKF to solve 1-D prestack FWI. Gineste & Eidsvik (2017) and later Gineste *et al.* (2019) proposed to use the Ensemble Kalman Smoother (Evensen & Van Leeuwen 2000) and the Iterative Ensemble Kalman Smoother (Bocquet & Sakov 2014) to inverse 1-D velocity profiles. Liu & Grana (2018) propose to use the Ensemble Kalman Smoother to inverse jointly elastic and petrophysical rock properties in the context of reservoir monitoring. In this study, we wish to suggest an original DA-FWI scheme adapted to take advantage of both worlds, by combining a classical FWI quasi-Newton solver and an ensemble filtering, to characterize the uncertainty of the solution. We expect that including quasi-Newton optimization will speed up the convergence of the filter, while also linearizing the least squares analysis step. It is also worth mentioning that our proposition is also different from other Bayesian methodologies such as Martin *et al.* (2012), Bardsley *et al.* (2014), Biswas & Sen (2017) who propose different approaches based on Markov chain Monte Carlo instead of DA. Other methods directly involving EnKFs as Iglesias *et al.* (2013); Schillings & Stuart (2017) would advocate for solving the FWI with the EnKF as the minimization technique itself, rather than relying on quasi-Newton methods.

In this study, we will develop the theoretical aspects of DA by reviewing the ensemble transform Kalman filter (ETKF) used in our application. Following by a brief review of the FWI problem we

will expose the structure of our mixed ETKF-FWI scheme in details, before presenting applications on both synthetic and field-data. Issues associated with ensemble rank-limited uncertainty estimation will be discussed along with the importance of prior information, advantages and shortcomings of our methodology.

2 INTRODUCING DA TO THE TOMOGRAPHIC PROBLEM

Characterization of systems subject to non-linear dynamics, numerical weather forecasting for example, is significantly different from the general tomographic problem. It generally relies on computationally intensive forecast modelling operators evolving the system state through time, giving access to prediction possibilities. However, natural systems non-linearities will most certainly cause the forecast to diverge at some point, limiting predictions through modelling at brief timescales (Fletcher 2017). The forecast, or modelling task, is also made difficult by the mismatch between the physical world, and its mathematical description; The correct set of equations governing a physical process may not be known, or one might only afford to compute their approximations. Besides, estimating the boundary conditions in limited space simulations, or defining correct and accurate initial conditions, is often challenging (Evensen 2009). Therefore, modelling is not enough to accurately study systems like Earth's atmosphere or oceans dynamics. The role of DA is to make use of the state's measurement and integrate data to correct for the imperfect forecast by connecting modelling and observations into an optimal *analysis* state.

DA can also be viewed as inverse modelling in some fields of research (Fletcher 2017), and is used to perform *retrieval*, which consists of combining prior state statistical insights, with observations. Thus, the DA problem is merely a matter of estimating the model parameters, considered as a set of random variables, from the previous or current state of the system (*prior statistical knowledge*) (van Leeuwen *et al.* 2015). As DA schemes and tools have been developed for non-linear problems at large operational scales, and because of the retrieval capacity of such schemes, we believe DA can be a solution to consider, as for the FWI uncertainty estimation problem. To that extent, we introduce the DA principles that allow us to expand the FWI formulation to an ensemble representation and unlock uncertainty assessment in the next subsections.

2.1 Ensemble Kalman filter

The ETKF (Bishop *et al.* 2001; Ott *et al.* 2004) is a 'square-root' version of the EnKF proposed by Evensen (1994), which addresses several biases that were included in Evensen's original formulation. It allows large-scale dynamic system study, thanks to an ensemble representation. By assuming that model and data errors are Gaussian, the state estimate and uncertainty are represented by first and second order Gaussian moments (mean and variance). From an ensemble of system state vectors, we can compute the relevant metrics involved in the original KF formulation, which makes extension of the KF scheme towards large scale problems affordable.

We adopt the following notations: \mathbf{P} designates a generic covariance matrix while \mathbf{P}^f and \mathbf{P}^a denote the forecast and analysis covariances, respectively. Subscript e in \mathbf{P}_e , \mathbf{P}_e^f and \mathbf{P}_e^a stands for the ensemble representation of said covariance matrices. Furthermore we will denote matrices by bold letters, vectors with a standard font, and use caligraphic letters to define non-linear operators.

Defining an ensemble \mathbf{m} as a $n \times N_e$ matrix whose columns contain N_e state vectors $m^{(i)} \in \mathbb{R}^n$ with n parameters:

$$\mathbf{m} = \{m^{(1)}, m^{(2)}, \dots, m^{(N_e)}\}, \quad (1)$$

the system state estimate (first Gaussian moment) is given by

$$\bar{m} = \frac{1}{N_e} \sum_{i=1}^{N_e} m^{(i)}. \quad (2)$$

The ensemble covariance \mathbf{P}_e provides an approximation of the true covariance \mathbf{P} (second Gaussian moment) provided N_e is sufficiently big. It is computed according to

$$\mathbf{P}_e = \frac{1}{N_e - 1} \mathbf{M} \mathbf{M}^T, \quad (3)$$

where \mathbf{M} is the perturbation matrix defined as $M^{(i)} = (m^{(i)} - \bar{m})$ and the superscript T is the transpose operator. By construction, \mathbf{M} is at best of rank $N_e - 1$ and thus the ensemble covariance matrix is rank limited.

The ETKF cycle is conducted in two successive steps: The modelling is performed on each ensemble members independently (forecast step) until data corresponding to the forecast, $d_{obs} \in \mathbb{R}^d$ with d observables, is available (analysis step). In that case, the analysis corrects the ensemble forecast, based on forecast uncertainties and observations uncertainties. The analysed ensemble then becomes the initial conditions for the next ETKF cycle, and the process can be repeated.

Considering first the *forecast step* in the ensemble formalism, the forecast ensemble from a step k to $k + 1$ is given by

$$m_{k+1}^f{}^{(i)} = \mathcal{F}_k(m_k^{(i)}) + \eta_k \quad i = 1, 2, \dots, N_e, \quad (4)$$

where \mathcal{F} is a non-linear forecast operator generally evolving a dynamic system, η_k is a zero-mean normal noise vector which covariance properties (generally denoted \mathbf{Q}) describe the uncertainties associated to \mathcal{F} . The superscript f denotes the forecast state. In practice, η_k might not be known, as evaluating the statistics of the sources of errors in the modelling is a challenge on its own.

The *analysis step* is given as an approximation of the linear KF analysis (the subscript k is discarded in the following, as all operation are done within step $k + 1$). Following the original formalism and considering for now a linear measurement operator $\mathbf{H} : \mathbb{R}^n \rightarrow \mathbb{R}^d$ projecting the model space into the observation space, the analysis mean \bar{m}^a is given by,

$$\bar{m}^a = \bar{m}^f + \mathbf{K}(d_{obs} - \mathbf{H}\bar{m}^f), \quad (5)$$

and the analysis covariance \mathbf{P}^a by

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f = \frac{1}{N_e - 1} (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{M}^f(\mathbf{M}^f)^T, \quad (6)$$

where d_{obs} are observations, \mathbf{I} the identity operator and \mathbf{K} is the Kalman gain matrix (Kalman 1960):

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1}, \quad (7)$$

where \mathbf{R} is the measurement noise matrix which contains the uncertainty information related to data. The Kalman gain acts as a weighting factor over the residual term $(d_{obs} - \mathbf{H}\bar{m}^f)$, based on a ratio of forecast uncertainty over observation uncertainty.

The analysis equation of the KF can also be written into its variational formulation, where the analysis state is given by the

minimization of

$$\mathcal{J}(m) = \frac{1}{2}(m - \bar{m}^f)^T \mathbf{P}^{f-1} (m - \bar{m}^f) + \frac{1}{2}(d_{obs} - \mathbf{H}\bar{m}^f)^T \mathbf{R}^{-1} (d_{obs} - \mathbf{H}\bar{m}^f). \quad (8)$$

Implementing eqs (6) and (7) would require to compute \mathbf{P}^{f-1} , which is in practice too large to be directly invertible. However, using the ensemble representation of eq. (3) in eqs (7), yields

$$\mathbf{K} = \frac{1}{N_e - 1} \mathbf{M}^f (\mathbf{M}^f)^T \mathbf{H}^T \left[\frac{1}{N_e - 1} \mathbf{H} \mathbf{M}^f (\mathbf{M}^f)^T \mathbf{H}^T + \mathbf{R} \right]^{-1}, \quad (9)$$

where \mathbf{M}^f is the forecast perturbation matrix, allows to express eq. (6) as

$$\mathbf{P}_e^a = \frac{1}{N_e - 1} (\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{M}^f (\mathbf{M}^f)^T = \mathbf{M}^f \left[\frac{1}{N_e - 1} \mathbf{I} + (\mathbf{M}^f)^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{M}^f \right]^{-1} (\mathbf{M}^f)^T. \quad (10)$$

Note that each time the observation operator appears in the final expression of \mathbf{P}_e^a , it is applied to the ensemble perturbation matrix \mathbf{M}^f . In addition, $\mathbf{H}\mathbf{M}^f$ is the first order Taylor approximation of $\mathcal{H}(m^{f(i)}) - \mathcal{H}(\bar{m}^f)$, with \mathcal{H} a non-linear observation operator (Harlim & Hunt 2005). This development can be written as a linearization of the general non-linear case over the ensemble representation. Let us consider any non-linear observation operator \mathcal{H} by introducing the ensemble observation perturbation matrix \mathbf{Y}^f , such that each column is defined as

$$\mathbf{Y}^{f(i)} = \mathbf{H} M^{f(i)} \approx \mathcal{H}(m^{f(i)}) - \mathcal{H}(\bar{m}^f). \quad (11)$$

We can then consider the analysis as an update of the perturbation matrix from \mathbf{M}^f to \mathbf{M}^a such that \mathbf{M}^a satisfies

$$\mathbf{P}_e^a = \frac{1}{N_e - 1} \mathbf{M}^a \mathbf{M}^a{}^T. \quad (12)$$

Given eq. (12), the analysis takes as an input the whole forecast ensemble $\{\mathbf{m}^f = \bar{m}^f + \mathbf{M}^f\}$, and its output is its analysed version, $\{\mathbf{m}^a = \bar{m}^a + \mathbf{M}^a\}$ (Harlim & Hunt 2005), allowing to circumvent any unnecessary covariance manipulation.

To update the ensemble perturbation matrix, Bishop *et al.* (2001) propose to use the effective ensemble uncertainty $\tilde{\mathbf{P}}^a$ spanning the subspace defined by \mathbf{M}^f , satisfying eq. (12), giving

$$\mathbf{P}^a = \frac{1}{N_e - 1} \mathbf{M}^f \tilde{\mathbf{P}}^a \mathbf{M}^f{}^T, \quad (13)$$

with

$$\tilde{\mathbf{P}}^a = \left(\frac{1}{N_e - 1} \mathbf{I} + \mathbf{Y}^{fT} \mathbf{R}^{-1} \mathbf{Y}^f \right)^{-1}. \quad (14)$$

Note that eq. (14) generalizes eq. (10) to the non-linear case. The symmetric matrix $\tilde{\mathbf{P}}^a$ can be expressed as the product of the square-root matrices $\tilde{\mathbf{P}}^a = \mathbf{T}\mathbf{T}^T$, where \mathbf{T} plays the role of a transform operator from which the filter takes its name. The square root of $\tilde{\mathbf{P}}^a$ is then used to perform the transform operation $\mathbf{M}^a = \mathbf{M}^f \mathbf{T}$ and with that, \mathbf{M}^a satisfies eq. (12).

The non-unicity in the choice of the square-root lead to the proposition of several Ensemble Square Root Filters, with various ways of computing \mathbf{T} . Amongst them are schemes referred to as *one-sided* unable to prevent biased state estimate, or having the tendency to produce outliers in the ensemble members (Bishop *et al.* 2001; Wang & Bishop 2003; Evensen 2004; Leeuwenburgh *et al.* 2005). For our methodological development we chose to follow the formalism proposed by Wang *et al.* (2004) and Ott *et al.* (2004), the

spherical simplex ETKF, which mitigates the aforementioned problems. The transform operator is thus defined as

$$\mathbf{T} = \mathbf{C} \Gamma^{-1/2} \mathbf{C}^T, \quad (15)$$

where the columns of \mathbf{C} are the singular vectors of $\tilde{\mathbf{P}}^a$ and Γ is a diagonal matrix containing the singular values of $\tilde{\mathbf{P}}^a$. This allows to update the forecast perturbation matrix following,

$$\mathbf{M}^a = \sqrt{N_e - 1} \mathbf{M}^f \mathbf{C} \Gamma^{-1/2} \mathbf{C}^T. \quad (16)$$

Following eqs (5) and (7), the analysed mean is then given by

$$\bar{m}^a = \bar{m}^f + \mathbf{M}^f \mathbf{C} \Gamma^{-1} \mathbf{C}^T (\mathbf{Y}^f)^T \mathbf{R}^{-1} \left[d_{obs} - \overline{\mathcal{H}(\mathbf{m}^f)} \right], \quad (17)$$

where $\overline{\mathcal{H}(\mathbf{m}^f)}$ is the mean over the forecast observations. Finally, the analysed ensemble is given by

$$m_{k+1}^{a(i)} = \bar{m}_{k+1}^a + M_{k+1}^{a(i)} \quad i = 1, 2, \dots, N_e. \quad (18)$$

In terms of computational cost, the manipulation of large scale covariance matrices is replaced with the SVD of the $N_e \times N_e$ matrix $\tilde{\mathbf{P}}^a$ and operations on $N_e \times n$ matrices through eqs (16), (17) and (18). The ETKF formulation satisfies the hypothesis of the original linear KF, while also providing an unbiased approximation of the KF mean and covariance. Uncertainty estimate is allowed through the evaluation of the ensemble covariance \mathbf{P}_e^a , as a low rank approximation of the true covariance \mathbf{P}^a . While rank reduction allows efficient computation of an approximate of \mathbf{P}^a , let us keep in mind that problems of representativity can arise in the context of ensemble based approximation. In the next subsection, we succinctly recall the nature of FWI, before introducing the ETKF-FWI scheme.

2.2 Frequency domain FWI

FWI can be considered as a constrained optimization problem in which one seeks to generate synthetic wavefield data d_{cal} with a high degree of correspondence with measured data d_{obs} (Virieux *et al.* 2017). In that regard, the object of optimization is the subsurface model, in which synthetic data are generated. Ideally, reducing the misfit between d_{cal} and d_{obs} to a minimum should lead the resulting synthetic model towards the real Earth parameters, provided the initial model allows setting the problem in the correct basin of attraction of the misfit function. Furthermore, to solve the FWI problem, it is required to be able to model the full wavefield in any physical media accurately. We will focus on the acoustic frequency-domain formalism in the following, as it is the chosen approach for the implementation of our proposed ETKF-FWI methodology.

In this framework, it is possible to reduce the wave equation under a compact form (Pratt *et al.* 1998; Virieux *et al.* 2009) as a generalization of the Helmholtz equation. In this section we adopt the matrix representation of the partial-differential operators of the continuous wave equation. The steady-state wave equation can thus be expressed as the following linear system

$$\mathbf{B}(\omega, m)u = s, \quad (19)$$

with \mathbf{B} a complex-valued impedance matrix which values vary according to the frequency ω and the medium properties m (Pratt *et al.* 1998; Operto *et al.* 2006), u is the pressure field (in the acoustic case), and s is the source term. The favoured way of solving the system in (19) is to factorize the \mathbf{B} matrix into a product of lower and upper (LU) triangular matrices, in order to use a direct solver (Virieux & Operto 2009), to efficiently deal with multiple right hand sides.

Solving eq. (19) allows computing the value of the pressure field in the entire medium depending on its physical parameters and a given frequency. The synthetic data d_{cal} are then extracted from the wavefield at receiver locations with a linear observation operator \mathbf{E} . This enables solving the optimization problem by iteratively minimizing the following misfit function

$$\mathcal{C}(m) = \frac{1}{2} \Delta d^\dagger \Delta d, \quad (20)$$

with Δd defined as the misfit vector $\Delta d = d_{obs} - d_{cal}(m)$, and where the superscript \dagger denotes the conjugate transpose. The misfit function in (20) does not contain any prior statistical information such as a model penalty term and data weighting, to keep the formalism simple. In practice though, these types of prior information are improving the solution and stability of the inverse problem by better constraining the initial data and model covariances (Pratt *et al.* 1998).

One must keep in mind that FWI is a local optimization problem implying that the initial parameters model m_0 needs to be in the vicinity of the solution for the global minimum. Thus, to achieve convergence towards the global minimum m_G is to assume that the prior knowledge is already a reasonable estimate of the real physical parameters. One basic way to solve the FWI problem is to apply a gradient descent optimization scheme. Assuming the starting model is adequately defined, the optimization will drive the solution towards the global minimum; The model parameters are iteratively updated according to

$$m_k = m_{k-1} - \alpha_{k-1} \nabla_m \mathcal{C}_{k-1}, \quad (21)$$

where α is the step length and the subscript t denote a FWI iteration number. The parameter α is computed by a line-search strategy.

Let us now introduce compact notation of both the forward and inverse problem embedded in the FWI process, as these will be used in subsequent sections. The forward modelling is only dependent on the parameter model, therefore the forward modelling for a given frequency can be expressed as

$$u = \mathcal{G}(m) = \mathbf{B}^{-1}(m, \omega)s, \quad (22)$$

where \mathcal{G} is a non-linear modelling operator, and m contains the medium physical parameters. Following computation of the wavefield in the full domain the computed pressure field is extracted at receivers location by applying a measurement operator \mathbf{E} , giving

$$y = \mathbf{E}u. \quad (23)$$

Moreover, as the wavefield depends on the model parameters m one can define a non-linear observation operator \mathcal{H} as,

$$y = \mathcal{H}(m) = \mathbf{E}\mathcal{G}(m), \quad (24)$$

such that (24) yields the computed monochromatic wavefield data at the receivers location.

As we conceptually reduce the synthetic data generation to the observation operator \mathcal{H} , we can also define and generalize the complete FWI process as an inversion operator such as

$$m_k = \mathcal{I}_n(m_{k-1}, d_{obs,k}). \quad (25)$$

Here, we consider \mathcal{I}_n as a non-linear operator encapsulating altogether: the forward simulation, computation of the misfit function, its gradient and gradient descent iteration for any given model m . The subscript n denotes the number of non-linear optimization iteration performed. Following the brief exposition of FWI and definition of the FWI operator \mathcal{I}_n , let us review now how the ETKF

can be modified to accommodate a tomographic application into the ETKF-FWI scheme.

2.3 ETKF-FWI Scheme, filter parameters

As geophysical tomography is considered to be a static problem at the considered time scales, applying the ETKF, designed for dynamic systems, is not straightforward. However, with specific parameterization of the ETKF, it is possible to take the specificities of the FWI problem into account. We thus adapt the filter to the FWI requirements and define an ETKF-FWI scheme (Fig. 1) that allows for uncertainty estimation, combining adjoint-base FWI and the ETKF algorithm.

As defined earlier in this section, the ETKF requires a forecast step, based on a projection of the system state ahead of time. In the DA community, it is common to see prediction steps as a forward modelling engine applied to a system state. In our lack of time dependency, we define a proxy for temporal evolution. By adopting a frequency continuation strategy, replacing the timesteps of the KF forecast, by a set of K frequency bands $k = 1, 2, \dots, K$, we can take advantage of the FWI multiscale approach, commonly used to mitigate cycle skipping issues (Bunks *et al.* 1995). Such difficulty is related to the convergence towards local minima when minimizing the FWI misfit functions. These local minima can be associated with the misinterpretation of the time delay in the recorded trace, with a phase ambiguity. Thus, working with broader wavelength data allows minimizing a more convex objective function. Applying FWI from low to high frequency can, therefore, provide a dynamic axis: FWI is treated as a frequency-dependent dynamic process.

This frequency continuation axis, allows defining the forecast operator as our adjoint-based FWI solver. This choice differs from most standard DA application as we replace the forward problem by the resolution of an inverse problem. Consequently, the FWI problem is solved during the forecast by using the operator \mathcal{I}_n defined in eq. (25), on the N_e ensemble members, independently, at a given k from the K frequencies or frequency bands considered. The forecast eq. (4) thus becomes

$$m_k^{f(i)} = \mathcal{I}_n(m_{k-1}^{a(i)}, d_{obs,k}). \quad (26)$$

Typically, n will be set so that \mathcal{I}_n performs 5 to 10 non-linear iterations. Note that as the correct statistics of the process noise is missing, the forecast uncertainty \mathbf{Q} term cannot be included.

Because we seek to evaluate the parameter's estimate uncertainty and our choice of forecast operator, the considered system state are subsurface model parameters. In a standard mono-parameter case, the ensemble \mathbf{m} is thus composed of N_e subsurface velocity models $m^{(i)}$, $i = 1, 2, \dots, N_e$, where we typically have $N_e \ll n$ and $N_e \ll d$ by several orders of magnitude.

The initial ensemble is generated in such a way that it both satisfies the normal distribution hypothesis of the KF and local resolution imposed by the FWI resolution power for a given frequency band (Devaney 1984; Wu & Toksöz 1987). An ensemble member is built by taking an initial model m_0 suited for convergence, to which we add a perturbation. Perturbations are generated by convolution of zero-mean, uniformly distributed random vector u^i (with $i = 1, 2, \dots, N_e$) with a non-stationary Gaussian function \mathcal{G} which correlation length and amplitude are varying according to the local velocity in m_0 , such that

$$m_0^i = m_0 + \mathcal{G}u^i, \quad i = 1, 2, \dots, N_e. \quad (27)$$

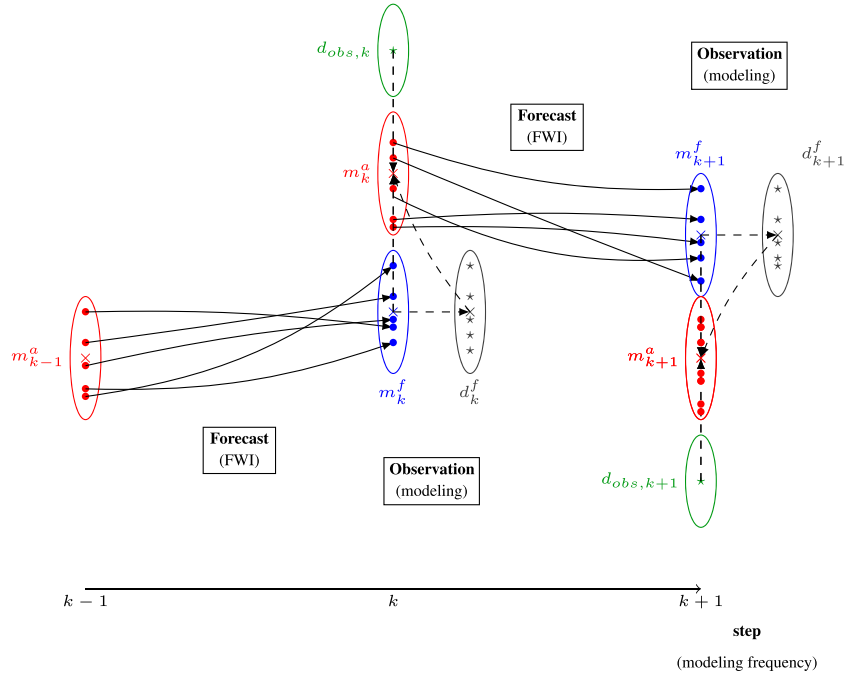


Figure 1. Schematic of the ETKF-FWI scheme. Dots represent state vectors, stars represent the measured state, crosses and ellipses are, respectively, means and covariances. Blue denotes the forecast system state, red the analysis, green the observed data and grey the forecast data. The dynamic follows a modelling frequency continuation axis from low to high frequencies, based on FWI multiscale approach.

$\mathcal{G}u^i$ produces smooth perturbations which wavenumber is half of the wavefields' wavelength, corresponding to the maximum spatial frequency that can be recovered (Wu & Toksöz 1987). The initial ensemble is then inspected with an Eikonal solver, to ensure that the initial population of models will not allow cycle skipping at our starting frequency, that could be provoked by too dramatic initial perturbations. Even though this test only allows assessing the first arrival cycle skipping, it is deemed sufficient as a first-order diagnosis of the initial ensemble quality. To further ensure favourable initial conditions, we verify that the rank of the initial ensemble is equal to N_e .

After applying the forecast operator to the N_e ensemble members, the forecast system state (in blue, Fig. 1), is a set of N_e optimized velocity models with respect to the considered band limited data. After the forecast, synthetic wavefield data are computed in the forecast ensemble to generate the forecast data (grey stars, Fig. 1). To do so, we use the frequency domain forward modelling engine used in the FWI process, to compute the wavefield in the whole velocity model, and extract the pressure values at receiver locations. It corresponds to applying the observation operator \mathcal{H} as defined in eq. (24)

$$d_{cal,k}^{(i)} = \mathcal{H}(m_k^{f(i)}), \quad (28)$$

The forecast data allows computing a misfit between the ensemble and the observed data (green star, Fig. 1), which is required to perform the analysis, calculate the transform operator and then the analysis ensemble (in red, Fig. 1). The cost of the analysis is negligible compared with the numerous forward modelling needed for the forecast step.

After the model-wise adjoint-based inversion of the forecast, the analysis allows performing an additional inversion, ensemble-wise, rearranging the ensemble around the mean solution, and ensuring coherency of the solution. Finally, we justify the compatibility of

the Gaussian assumption of the ETKF, with an application on a FWI problem by the following: provided that all the ensemble members are located along the same minimum of convergence (which is the intent of this methodology), the assumed local convexity of the cost function, necessary to apply a local optimization scheme, is deemed a good first order approximation of Gaussian probability density function.

The scheme has been tested on synthetic and field-data applications, to study the feasibility of the approach and evaluate its shortcomings. The experimental setups and results for both cases are presented and discussed in the next sections.

3 APPLICATION ON 2-D MARMOUSI SYNTHETIC BENCHMARK

We define a framework based on the Marmousi II synthetic model (Martin *et al.* 2006). As our approach is intrinsically based on rank reduction, we decide to evaluate its potential effects on our solution. To that extent, we define three scenarios with increasing ensemble sizes, $N_e = 20, 100, 600$. We consider $K = 15$ ETKF-FWI cycles along with $n_\omega = 15$ mono-frequency complex-valued data from 3 to 10 Hz, with a 0.5 Hz increment between each ETKF-FWI cycle. The considered domain width and depth are, respectively, $x = 16.025 \text{ km}$ and $z = 3.250 \text{ km}$ with vertical and horizontal resolutions of $dx = dz = 25 \text{ m}$, for a total of 83 300 degrees of freedom.

The exact model, the initial ensemble mean and the acquisition footprint, are displayed in Fig. 2. Data are simulated using a fixed spread surface acquisition configuration, with 144 sources and 640 receivers evenly spaced, to mimic realistic marine streamer acquisition, resulting in a data vector d_{obs} with 92160 entries. The total number of discrete data points is equal to $92160 \times n_\omega$ where n_ω is the number of considered frequencies. The modelling/FWI code

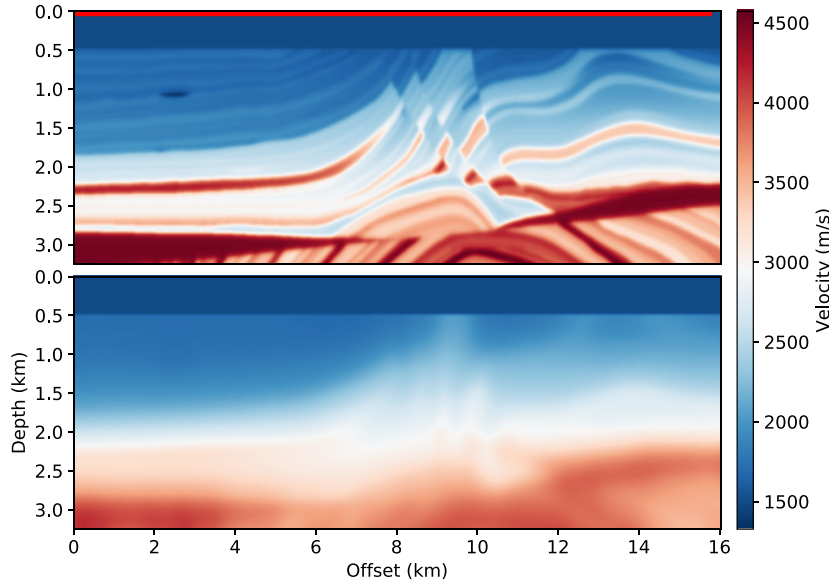


Figure 2. Numerical experiment setting. Top panel: true Marmousi II model. Acquisition is denoted by a red line at the surface. Bottom panel: smoothed version of the true model. It is used to build the starting ensemble, being the initial ensemble mean m_0 .

relies on the open-source TOY2DAC code developed in the SEISCOPE Consortium, coupled with the non-linear optimization toolbox (Métivier & Brossier 2016). The solver used for the forward simulations, relies on an optimized finite-difference discretization strategy with a compact stencil providing accuracy, equivalent to fourth-order methods for the applications considered here (Hustedt *et al.* 2004; Operto *et al.* 2009). The modelling operation (22) is solved using the MUMPS sparse solver (MUMPS team 2017). The forecast operator \mathcal{I}_n is set to perform $n = 10$ minimization iterations with the l-BFGS optimization scheme (Byrd *et al.* 1995; Nocedal & Wright 2006; Métivier & Brossier 2016), on each of the N_e velocity models, with mono-frequency synthetic calculated $d_{cal,k} \in \mathbb{C}^d$ and observed data $d_{obs,k} \in \mathbb{C}^d$ at frequency k . The cost of the methodology is thus linearly linked with the number of ensemble members and non-linear FWI iterations. In this instance, the number of forward modelling is thus $N_e \times 10 \times 2$, as both the incident and adjoint wavefields are computed at each iteration. Once the forecast state is obtained, we compute the forecast data at the frequency k with observation operator \mathcal{H} . The observed data for both the forecast inversions and the analysis are set to be the same $d_{obs,k}$ data at step k . With the same observed data, we are performing two optimizations steps: model-wise first then ensemble-wise. This pragmatical approach deviates from common ETKF scheme where it is assumed to introduce new information during the analysis rather than relying on previous data. In our case, using the same data for both the forecast and the analysis has given us more consistent and stable results in terms of parameters estimate. We attribute this behaviour to the nature of our forecasting operator. Because \mathcal{I}_n is updating the ensemble of models using $d_{obs,k}$, the wavenumber content of these updates is closely tied to the frequency content of $d_{obs,k}$. The ensemble of optimized models obtained from $d_{obs,k}$ are likely to lack the higher wavenumber content to explain or ‘predict’ higher frequency data $d_{obs,k+1}$. By using the same objective data, the analysis’ purpose is to reduce the spread of the ensemble and thus avoid unwanted cycle skipping amongst the ensemble.

To avoid inverse crime, a complex Gaussian random noise was added to the synthetic observed data (Eikrem *et al.* 2019):

$$d_{noisy} = d + \frac{\|d\|}{\sqrt{r * E(\|w\|^2)}} w \quad (29)$$

with d_{noisy} the noisy signal, d the original signal, $\|\cdot\|$ indicating the Euclidean norm and E the expectation. The vector w is defined as

$$w = v_1 + i v_2, \quad (30)$$

where $v_1, v_2 \in \mathbb{R}^d$ are vectors of normally distributed random numbers and r is defined as the signal-to-noise ratio, such that

$$r = \frac{\|d\|^2}{\|w\|^2}. \quad (31)$$

In the following experiments, we set up $r = 8$ as our reference noise value through all the ETKF-FWI cycles. From eq. (29), we can see that the noise added to the data has a variance given by

$$\sigma^2 = \frac{\|d\|^2}{r * E(\|w\|^2)} \quad (32)$$

which allows to define the measurement noise matrix such that

$$\mathbf{R} = \mathbf{I}_d \sigma^2 \quad (33)$$

where \mathbf{I}_d is an identity matrix of size d . This is required as we typically lack information about possible correlated measurement errors in FWI. While the benefits of taking correlated noise structures into account have been highlighted (Stewart *et al.* 2008; Weston *et al.* 2014), they cannot be taken into account in our case. Therefore \mathbf{R} has to be considered diagonal (which is a reasonable assumption). Note also that if we could estimate the off-diagonal terms of \mathbf{R} , the computation of \mathbf{R}^{-1} during the analysis step would become computationally challenging.

We review the various outputs constituting the solution of our ETKF-FWI in the next subsections.

3.1 Parameter estimate

We first present the final ensemble means for $N_e = 20, 100$ and 600 in Fig. 3. All three test cases lead to consistent parameter estimation, as all three results are fairly comparable. Thanks to the careful design of the initial ensemble, which limits the chances of cycle-skipping occurrences, the whole ensemble is in a favourable position to converge close to the global minimum. Moreover, because of the nature of our forecast operator (being an optimization problem), it seems that the ensemble mean is clear of any drifting or divergence effect, commonly encountered in typical dynamic EnKF applications. It is also worth noting that the three results recovered are close to the result of FWI alone given the same experimental setup (initial model, acquisition design, data noise).

3.2 Ensemble variance

Although the mean of the ensemble corresponds to the parameter estimate as a solution of the inverse problem, the posterior ensemble covariance \mathbf{P}_e^a is holding the uncertainty and resolution information we are interested in. The full matrix cannot be computed explicitly because of hardware limitations but computing its diagonal or individual lines are trivial operations when \mathbf{m}^a is stored. Initial and final variance maps plotted over the parameter space are displayed in Figs 4 and 5 and reflect the diversity of solutions for each parameter among the different ensemble members. Final variance values are thus indicative of the convergence quality of our ensemble of solutions.

Contrarily to the state parameter estimate, we observe a substantial lack of result consistency regarding the ensemble size. The various ensemble size tested reveal that the ensemble covariance is strongly affected by what is termed undersampling (Guzzi 2015) in the DA community. Undersampling is an issue arising when a small ensemble fails at being statistically meaningful. In this instance, it translates to variance underestimation. Nonetheless, it is possible to identify consistent features in all three final variance maps. The predominant effect is the link between geometrical spreading and variance, manifested by the net increase towards the depth and lateral limits of the physical domain, where poor illumination is expected. High variance values also tend to focus along sharp velocity contrasts. We might associate high variance at interfaces with the band-limited context of our application: band-limitation is expected to limit the ability of the optimization scheme to recover sharp discontinuities which will tend to smooth the interfaces because of the lack of high-frequency content. Another possible source of variability in interface recovering might be the inherent velocity–depth ambiguity in tomography (Yilmaz 1993). To precisely evaluate how variance changes according to the geological structure, we are interested in the locations of variance maxima in Fig. 6.

To extract those maximum variance peaks, we use a maximum filter of radius 275 m. The maximum filter dilates the variance map, and create local zones of homogeneous values. Peaks (or local maxima) are defined as parameters located where the variance map and the output of the maximum filter are equal. Most of the measured variance peaks are consistently located along interfaces where high-velocity layer are overlaying lower velocity layers. As stated before, band limitation and velocity–depth ambiguity might explain why most of the structural uncertainty is linked to interfaces reconstruction.

As for the quality of the variance estimate, assuming that the test with ensemble $N_e = 600$ is the less affected with variance underestimation, it is possible to make several observations. The

$N_e = 20$ case displayed in Fig. 5 exhibit a severe underestimation of the variance values in most of the physical domain along with non-physical oscillating behaviour in the deeper part of the domain. Those oscillating patterns are a direct result of a small ensemble, and thus a poor covariance approximation. For $N_e = 100$ however, the variance map does not exhibit non-physical oscillations, but is still slightly underestimated. The qualitative aspect of the variance map is at least preserved.

To better understand the results of Fig. 5 and go beyond simple qualitative comparison, we evaluate absolute variance values from a set of ETKF-FWI realizations for $N_e = [20, 600]$. We evaluate the underestimation of variance by computing the mean variance value for every variance maps. We plot the averaged variance against ensemble size in Fig. 7. As it stands, the trend in absolute variance values seems to be consistent with the variance underestimation observed Fig. 5. It is also worth noting that variance estimates behave almost asymptotically, which means we can hope to find a compromise between too small and too big ensembles, even though it seems complicated to estimate this ‘optimal’ ensemble size in advance, nor it is practical to evaluate it by trial and error. To complete this analysis, we have computed the root-mean-square-error (RMSE) values of the ensemble final means, with the true model as reference:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (m_{n,true} - \bar{m}_n^a)^2}, \quad (34)$$

where $m_{n,true}$ is the n th parameter of true velocity model, and \bar{m}_n^a is the n th parameter of the final ensemble mean. Values of RMSE reduction between the initial model m_0 and final ensemble means are displayed in Table 1: RMSE reduction is not affected by the ensemble size and all parameters estimates are nearly identical, which is consistent with the results observed in Fig. 3.

3.3 \mathbf{P}_e^a off-diagonal terms

Storing the ensemble also allows to examine the off-diagonal terms of the covariance matrix, containing complementary information regarding local resolution. However, undersampling makes it difficult to compare the three test cases, as it also impacts the quality of the off-diagonal terms estimates. For the sake of comparison of various results, we propose to work with the dimensionless correlation matrix, instead of the covariance matrix that we estimate. The correlation matrix is a dimensionless operator that contain correlation coefficients from -1 to $+1$ (Feller 2008). When the correlation coefficients tend to $+1$, it reflects a strong positive link between two parameters, implying that they share similar physical properties and are evolving in a similar fashion. Conversely, a negative correlation coefficient of -1 denotes a strong link but expresses an opposite behaviour between parameters. Finally, a correlation coefficient of 0 implies the absence of a physical connection between parameters. To compute the correlation matrix, we first need to define \mathbf{D} as a diagonal matrix containing the variance terms of \mathbf{P}^a . The correlation matrix is then given by,

$$\mathbf{C} = (\mathbf{D})^{-1/2} \mathbf{P}^a (\mathbf{D})^{-1/2}. \quad (35)$$

\mathbf{C} is thus a dimensionless, normalized version of the covariance matrix, which diagonal terms are all equal to 1 (correlation of a parameter with itself). By effectively normalizing the covariance matrices by their variances, we simplify the comparison of our three results.

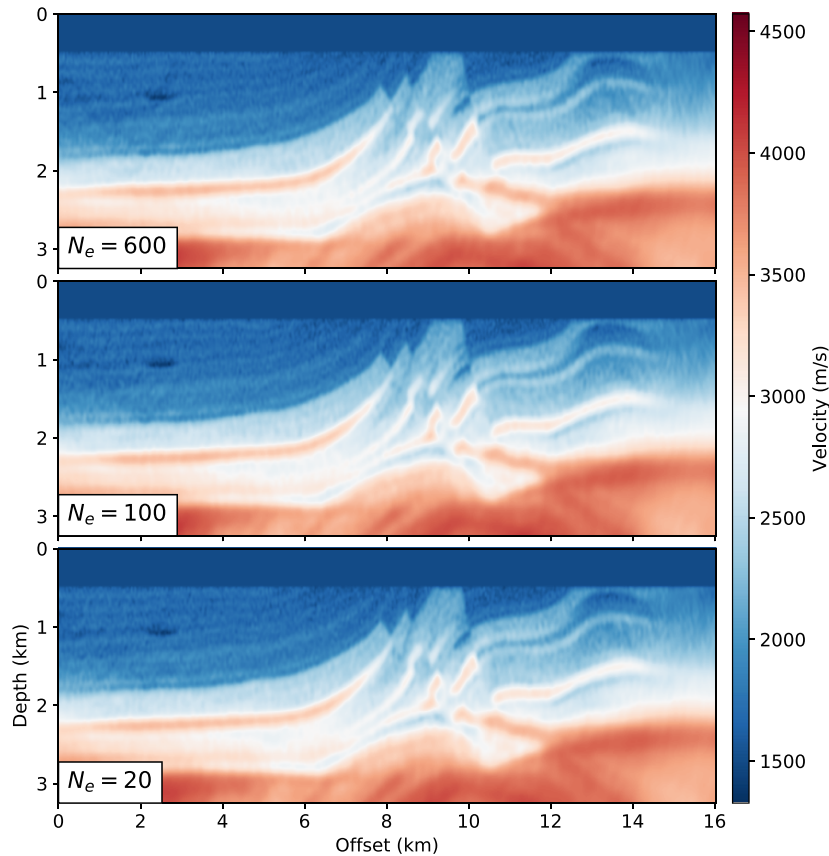


Figure 3. Analysed ensemble means for ensemble sizes $N_e = 600, 100, 20$ after 15 ETKF-FWI cycles from 3 to 10 Hz.

We compute local correlation maps for three arbitrary selected points shown in Fig. 8. This allows to evaluate correlation maps in various conditions: A shallow point on a reflector with relative low ensemble variance (orange), a point extracted from a strong reflector with large relative variance (red) and a point selected at depth, where the structure is only faintly recovered (black). Fig. 8(a) illustrate the three parameters location (coloured crosses) and the subdomains spanned by their local correlation maps (coloured rectangles). We choose to work with local correlation maps rather than the full domain as far-field is affected by spurious correlations and is thus of lesser interest. We focus on local correlation map to evaluate the consistency of local information in the vicinity of the investigated parameters, as we would do with a point-spread function evaluation.

The shallower point (orange, at $z = 1.0$ km; $x = 13.0$ km) located in a low variance area, displays chaotic and incoherent correlation structures for the case $N_e = 20$ (Fig. 8b). Both the layer dip and thickness informations are mostly lost in spurious correlations: even local information is lost. Improvements are visible for $N_e = 100, 600$ (Figs 8c and d) where the dipping structure starts to be visible. A small circular positive correlation zone in the immediate vicinity of the examined parameter denotes the very high local resolution. Despite the presence of spurious correlations, the geometrical information can be inferred.

For the intermediate depth parameter (red, at $z = 1.825$ km; $x = 13.0$ km), we observe on all three cases (Figs 8e, f and g), the polarization of the positive correlation values along the axis of the reflector, with correct thickness information. The limits of this geological feature are well defined on the correlation map. These

results are encouraging, implying that despite the undersampling effect associated with a small ensemble, some coherent dipping information can be retrieved from the final ensemble.

Finally, the local correlation maps for the deeper point (red, at $z = 2.650$ km; $x = 13.0$ km), are characterized by a broad zone of strong positive correlation where the local structure is faintly defined. This broad zone can be interpreted as a low resolution power. It is expected to find such broad positive correlation areas in the deeper domain, as illumination becomes weak. We also observe that despite the significant undersampling hinted by variance underestimation, the correlation maps for $N_e = 20$ still display the essential qualitative information (polarization along interfaces, correlation length radius). The stronger undersampling bias happens in the shallow zone. This leads us to think that undersampling biases might be decreasing with depth, as the problem is less constrained, and the ensemble is less likely to ‘collapse’ towards a unique solution and lose representativity. Those observations are encouraging, as we expect model parameters along the same interface, or in the same geological feature, to be defined with similar physical properties and correlation length to increase with depth.

Overall, this synthetic application showed that the ensemble is sensitive to undersampling when it comes to the covariance estimate, whereas the parameter estimate remains coherent even with a meager number of ensemble members. The next section is dedicated to field-data seismic exploration FWI application to demonstrate the applicability of the ETKF-FWI scheme, in a less favourable case.

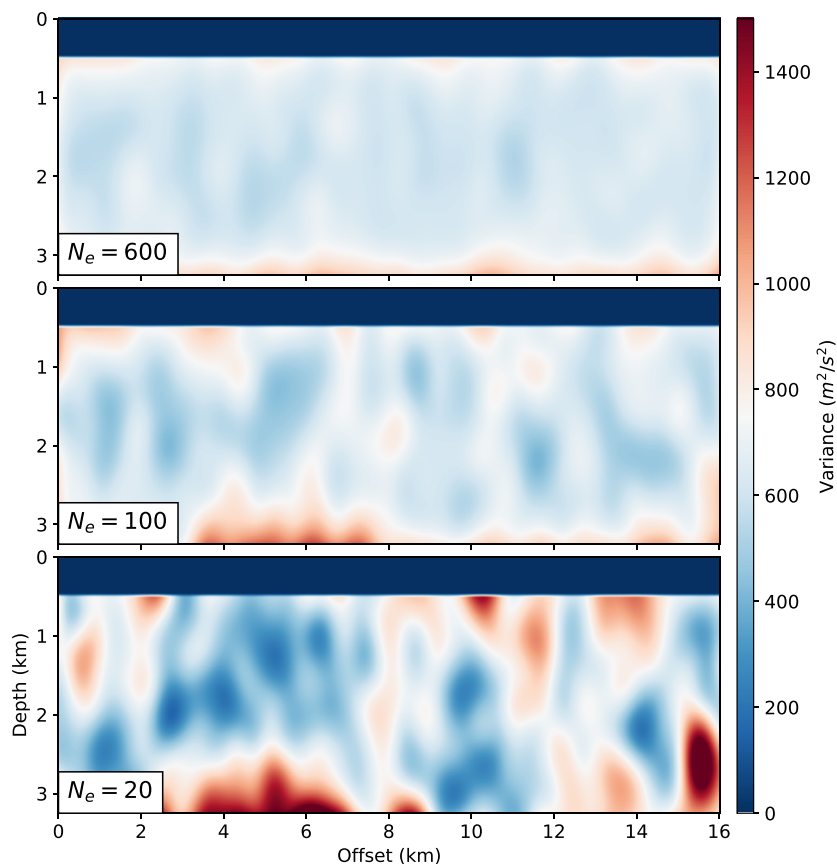


Figure 4. Initial ensemble variance maps for ensemble sizes $N_e = 600, 100, 20$.

4 FIELD-DATA APPLICATION

This field-data application is also based on 2D Frequency Domain FWI, but this time considering a VTI anisotropic medium for the modelling. The data set comes from the Valhall oilfield, located in the Norwegian North sea and is provided by AkerBP. The specificity of the Valhall oil field is its shallow water level (70 m). It features a reservoir with an anticline cap-rock structure, trapping the underlying hydrocarbon resources at a 2.5 km depth. The advantage of this case study is that it is well documented, and FWI has already been applied successfully to this data set (Operto *et al.* 2015). The domain width and depth are, respectively, $x = 16.725 \text{ km}$ and $z = 5.025 \text{ km}$ with vertical and horizontal resolutions of $dx = dz = 25 \text{ m}$, for a total of 134 469 degrees of freedom.

The data set is composed of four component Ocean Bottom Cable (OBC) recordings. From the full acquisition which contains 49 954 shots for 2302 receivers, we extract a 2-D line containing 192 sources and 315 receivers (which makes each frequency data vector composed of 60 480 entries), the same as the one used in (Zhou *et al.* 2018). The total number of discrete data points is equal to $60\,480 \times n_\omega$. OBC receivers are evenly spaced (50 m) and lie fixed on the seabed (70 m depth). The selected sources are also evenly spaced (50 m) at a constant 5 m depth. In this application, we exploit the hydrophone out of the 4 components recording.

The ETKF-FWI scheme is the same as presented in the previous experiment. To ensure the best-case scenario result, we work with an ensemble of $N_e = 600$ members, as the application size is of the same order of magnitude as the synthetic test case. We choose to work with $K = 6$ ordered groups of frequencies ranging from 3.56 to 7.01 Hz. This frequency selection strategy has been suggested

in preliminary work conducted by Zhou (2016) on this data set, and have proven to be adequate for this specific application. Using frequency groups rather than mono-frequency data ensures that each inversion cycle rely on redundant information which mitigate the risks of cycle-skipping. Contrarily to the synthetic test case, we are considering several groups of monochromatic data for both our FWI forecast operator and the analysis, bringing the amount of mono-frequency data pieces to $n_\omega = 15$.

As in the previous application, the measurement noise matrix \mathbf{R} is defined as a scaled Identity matrix. Because this application is based on groups of frequencies, we evaluated the variance for each of the monochromatic data to be included in \mathbf{R} from their signal-to-noise ratio according to eq. (32), which makes \mathbf{R} block diagonal, each block corresponding to a monochromatic data variance as given in equation 32. This time, the optimization scheme used during the forecast inversions is a preconditionned version of the l-BFGS scheme. We review two cases, a monoparameter P -wave velocity test cases, and a multiparameter $\{P$ -wave velocity, density $\}$ inversion, to evaluate the cross-talk between inversion parameters, and the ETKF-FWI behaviour with multiparameter inversion.

4.1 P -wave velocity reconstruction

After the 6 ETKF-FWI cycles, we obtain an ensemble of solutions, as with the synthetic case. Focusing first on the parameter estimate, we compare the initial and final ensemble means in Fig. 9. As expected from the previous experiment, the final mean model provides a net increase in resolution. Layered structures are well defined in

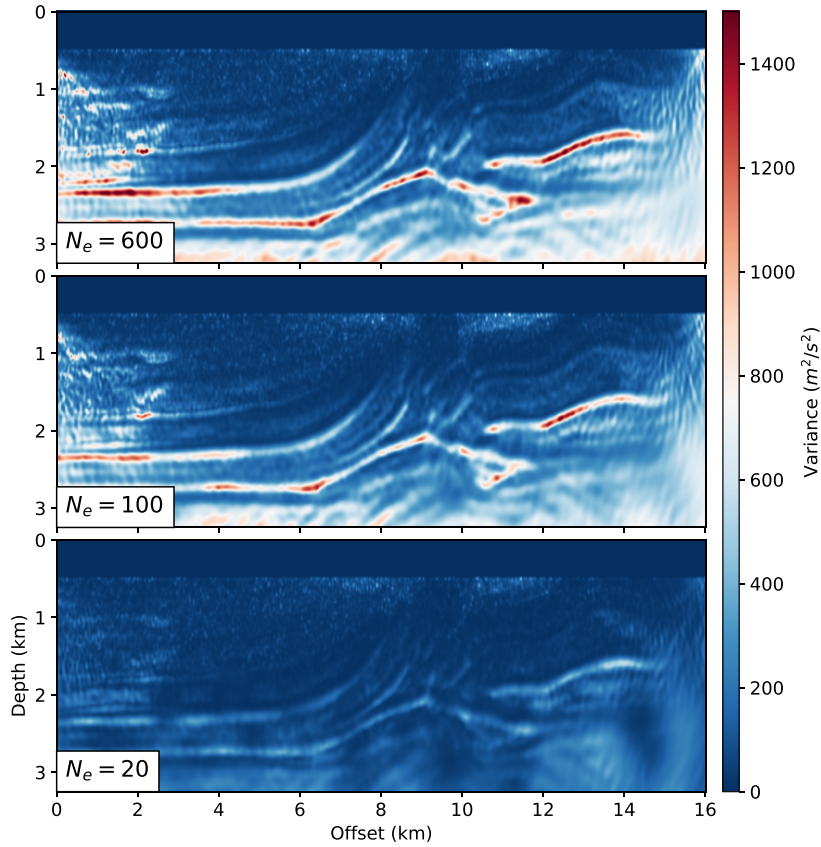


Figure 5. Analysed ensemble variance maps for ensemble sizes $N_e = 600, 100, 20$ after 15 ETKF-FWI cycles from 3 to 10 Hz.

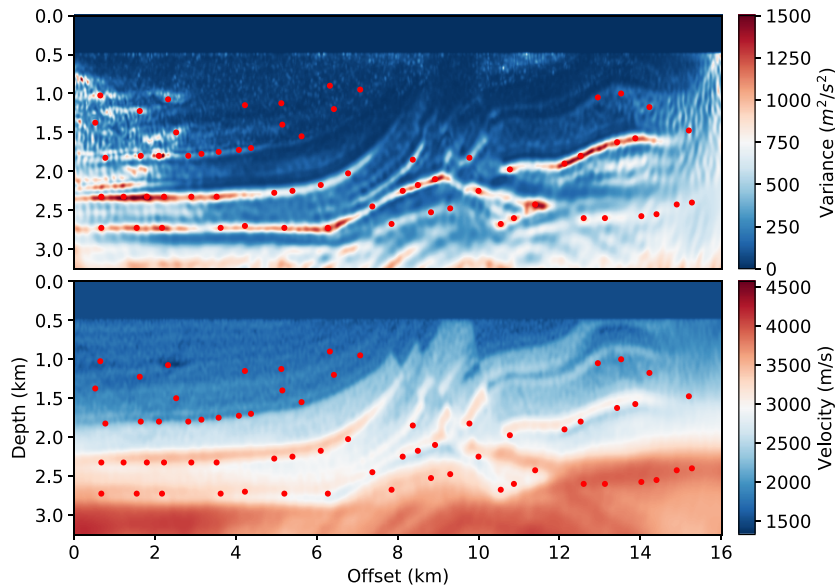


Figure 6. Variance map (top panel) and final mean velocity model (bottom panel). Red points denotes local maximum variance peaks in both maps. Variance peaks are evaluated with a maximum filter defined with a 275 m radius.

the top half of the domain, and from this result, we can identify what can be interpreted as hydrocarbon-charged units overlaying the anticline structure. The deep layered structures are not as sharp as the top section because of the strong impedance contrast between the upper and lower units of the medium. The strong P -wave velocity

contrast between the upper and lower domain is expected to reduce the illumination power in the deeper part of the model, along with the geometrical spreading effect.

We compare initial and final variance maps in Fig. 10. While the initial variance is relatively large in the entire domain (the

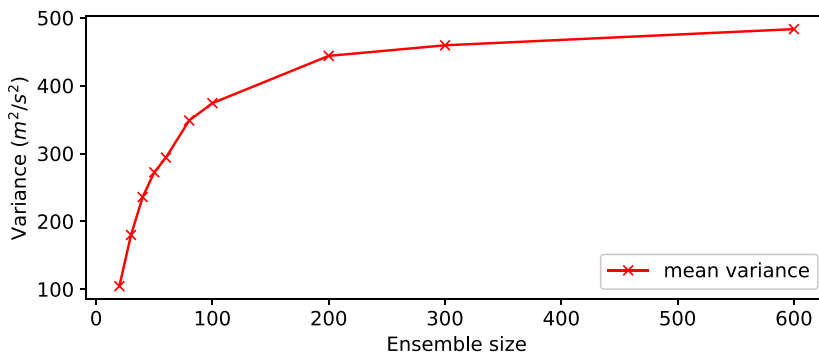


Figure 7. Average variance plotted against ensemble size.

Table 1. Normalized root-mean-square model error (RMSE) reduction with respect to the initial model for various ensemble sizes. The RMSE values are computed between the final ensemble mean and the true model. The amount of RMSE reduction is computed with the initial model RMSE as reference.

N_e	20	30	40	50	60	80	100	200	300	600
RMSE reduction (per cent)	15.4	15.4	15.6	15.0	15.2	15.9	15.7	15.6	15.3	15.6

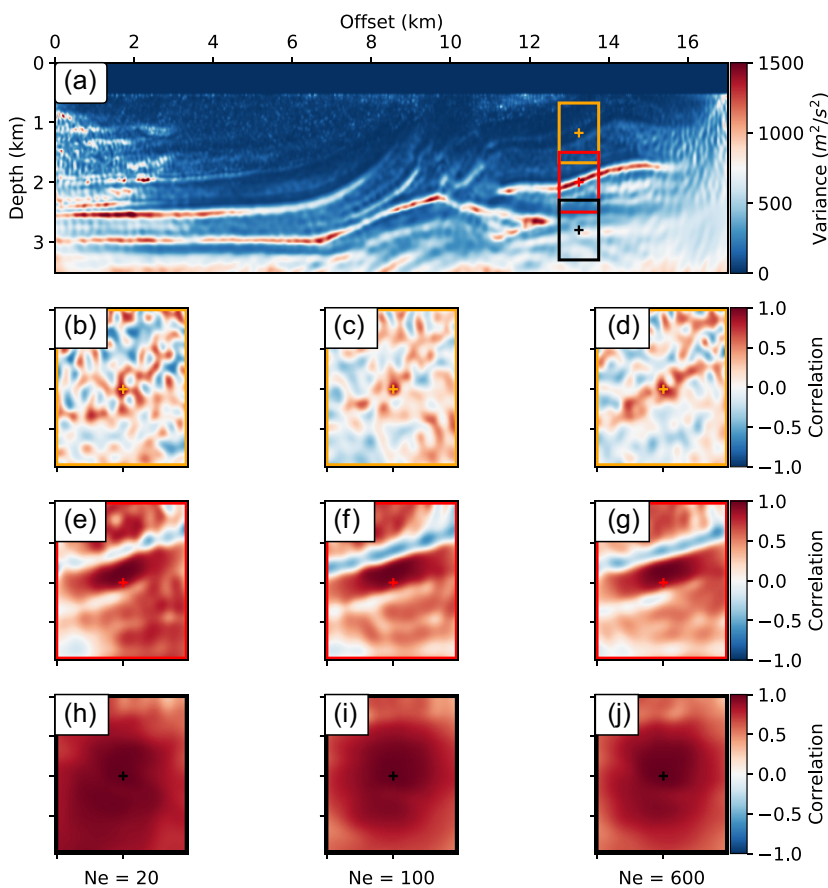


Figure 8. Reference variance map (a) and local correlation maps (b–j) for $N_e = 20$, 100 and 600. Coloured squares on the variance map correspond to the extents of the local correlations for the three considered parameters.

water depth is not perturbed), the final variance displays the same two tendencies as in the synthetic case. The first order uncertainty structure is dominated by the geometrical spreading and the sharp velocity contrast between the upper and lower units at 2.5 km depth. Second to that are the variance values imposed by the velocity structures estimated in the solution. Note that we use

a non-linear colourscale to underline uncertainty associated to the structure.

To repeat the procedure detailed in the synthetic application and evaluate how the variance aligns with the velocity structure, we computed maximum peak locations in the variance map (Fig. 11). The search radius has been reduced to 150 m because of the

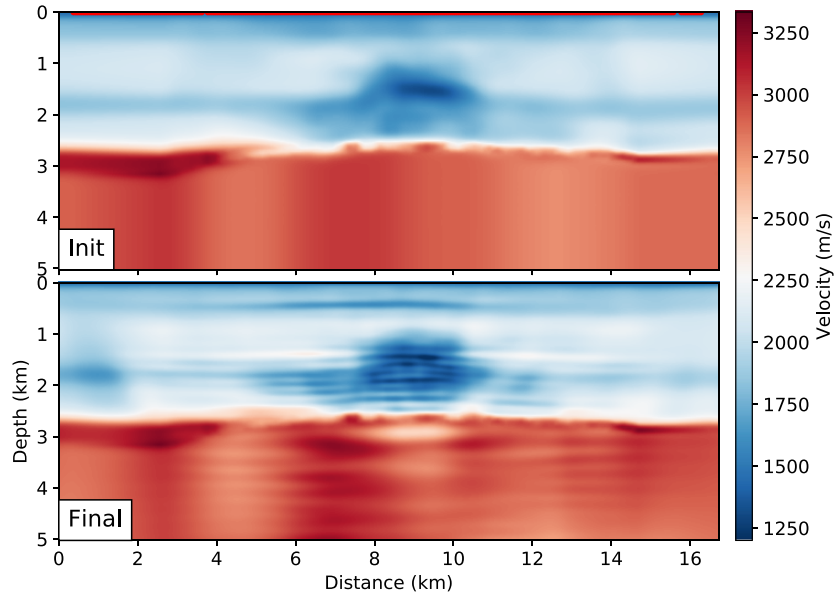


Figure 9. Top panel: initial ensemble mean velocity model m_0 . Acquisition is denoted by a red line at the surface. Bottom panel: final ensemble mean velocity model after 6 ETKF-FWI cycles from from 3.56 to 7.01 Hz with $N_e = 600$.

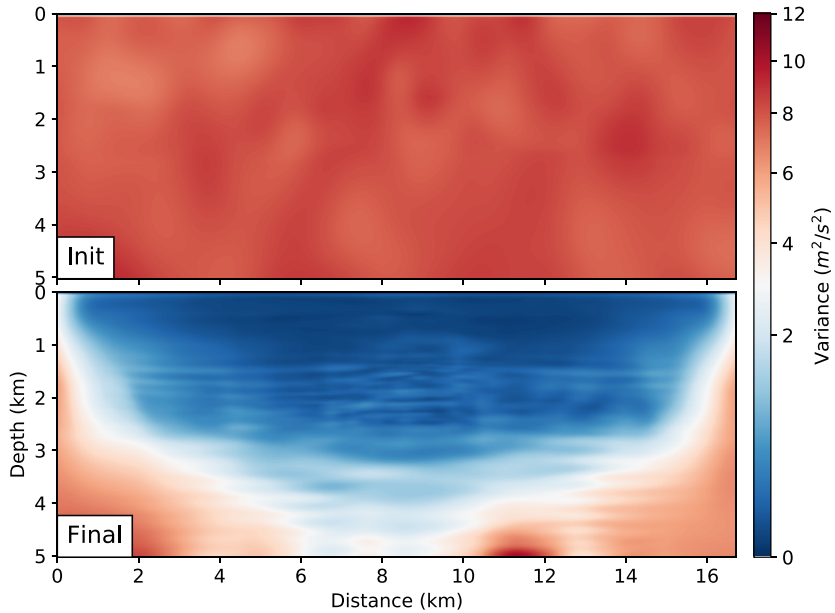


Figure 10. Initial and final variance maps after 6 ETKF-FWI cycles from from 3.56 to 7.01 Hz with $N_e = 600$.

smoothness of the variance map. Despite the map smoothness and the thin layered structure in the final velocity model, we can confirm that local uncertainty maxima are preferentially located along structure discontinuities.

Correlation maps are also computed in the final ensemble, following the same procedure as for the synthetic test, for three parameters denoted as orange, red and black (Fig. 12)

The horizontally layered structures can be observed in all three parameters correlation maps. We also observe the effect of resolution loss between the orange (located at $z = 1.3 \text{ km}$ and $x = 5.0 \text{ km}$ and the black parameter (located at $z = 2.8 \text{ km}$ and $x = 12.6 \text{ km}$), characterized by the increase of the positive correlation radius around the parameter. We also point out the coherency of the correlation maps with the recovered structure as visible for the

black parameter. Indeed, the broad, circular positive correlation zone around the parameter is abruptly stopped by the velocity contrast at 2.5 km depth.

Comparing the velocity model computed through the ensemble method to a classical FWI result (Fig. 13), shows that the ETKF-FWI produces a mean model similar to the corresponding FWI solution. We can, however, notice that the resolution of the mean ETKF-FWI is slightly higher and the contrasts between the layers velocity appear stronger in the ETKF-FWI result, both in the shallow and deep parts of the model. This might be due to the effect of the analysis step, which provides a correction from the estimated covariance matrix. This could have an effect similar to the one of a preconditioner which approximates the inverse Hessian operator. This is further discussed in the following multiparameter application.

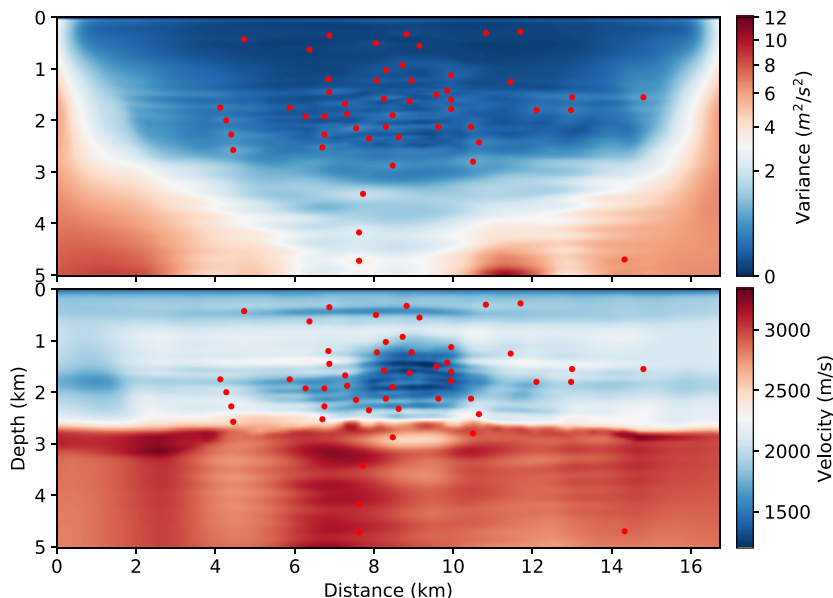


Figure 11. Variance map (top panel) and final mean velocity model (bottom panel). Red points denotes local maximum variance peaks in both maps. Variance peaks are evaluated with a maximum filter defined with a 150 m radius.

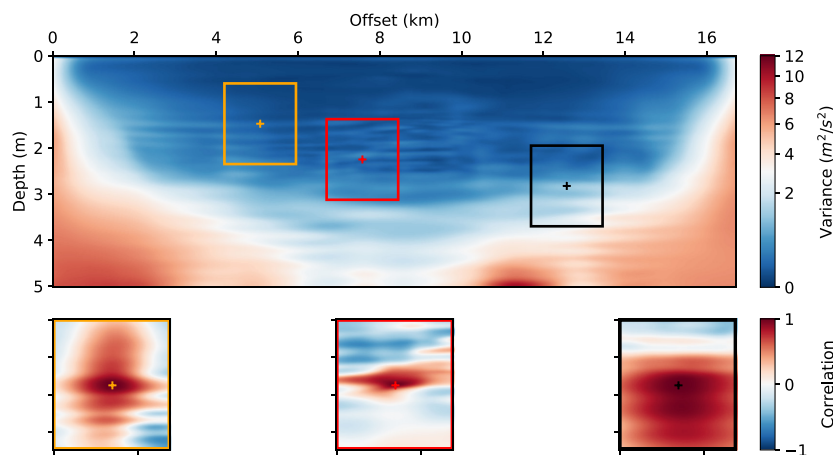


Figure 12. Reference variance map (top panel) and local correlation maps (bottom panel) for $N_e = 600$. Coloured rectangles on the variance map correspond to the extents of the local correlations subdomains for the three considered parameters.

However, the quality of these results is strongly linked to the initial ensemble parameterization. Modifying the initial perturbations correlation length or amplitude will result in a different outcome, or might cause instabilities if incorrectly chosen.

4.2 P-wave velocity and density reconstruction

In the following, we present preliminary multiparameter inversion results to show the potential of the method for uncertainty estimation and parameter estimation. Multiparameter FWI is known as a challenging problem, especially because of the presence of cross-talks between parameters (Operto *et al.* 2013). Recovering information about the uncertainty linked to these cross-talks is thus crucial, and might be an important benefice from strategies such as the ETKF-FWI scheme presented here. We modify the system state vector such that the columns of the ensembles contain both the velocity parameter V_p and the density ρ instead of the velocity

alone

$$m_{V_p, \rho}^{(i)} = \begin{pmatrix} V_p \\ \rho \end{pmatrix}. \quad (36)$$

Considering the joint state $m_{V_p, \rho}^{(i)}$ makes it possible to take the changes of density during the forecast optimization steps into account when the analysis is performed. Note that the extension of the state vector also implies an extension of the state covariance matrix. It is expected that the cross-talk terms between V_p and ρ (off-diagonal blocks of the covariance matrix) will play a role in the Kalman Gain estimate.

The initial density perturbations are derived from the initial perturbed velocity model according to Gardner's empirical relationship (in soil only) (Gardner *et al.* 1974)

$$\rho = 0.31 V_p^{0.25}. \quad (37)$$

This way, initial ensemble members' velocity and density perturbations are physically linked.

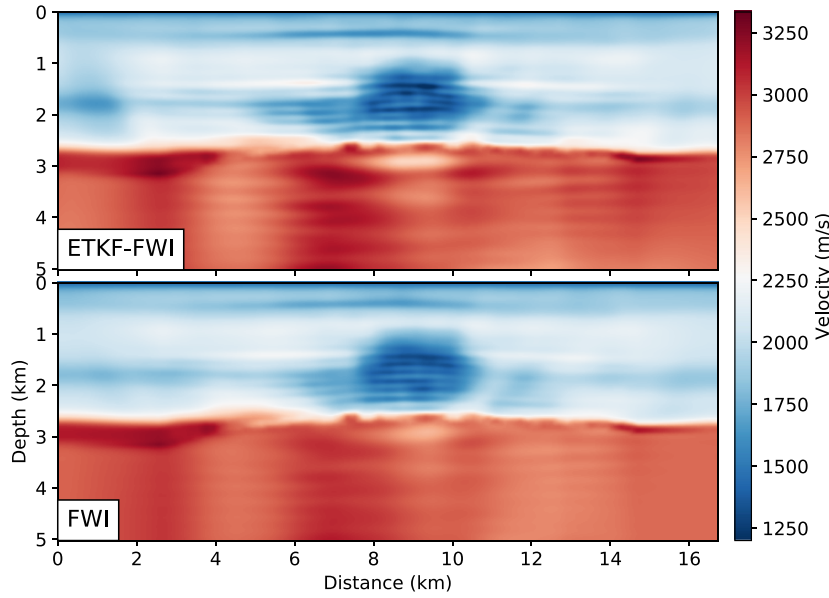


Figure 13. Comparison of monoparameter ETKF-FWI (top panel) and FWI (bottom panel) results with similar inversion setup (inversion parameters, regularization, acquisition geometry and data frequency groups).

The starting ensemble mean velocity and density models are displayed in Fig. 14. The ETKF-FWI scheme is applied following the same setup as detailed for the monoparameter test, except for the forecast that now includes inversion of the density parameter alongside the inverted velocity. The parameter estimation after 6 ETKF-FWI cycles are shown in Fig. 15.

The recovered velocity model is almost identical to the velocity estimate from the monoparameter case. As for the density inversion, the horizontally layered structures observed in the velocity map, are closely matching the density estimate. A lower density is seen in the central area where hydrocarbon charged layers are expected to be located.

The joint covariance matrix for the multiparameter case contains four blocks. Its structure is defined by

$$\mathbf{P}_{[V_p, \rho]} = \begin{bmatrix} \mathbf{P}_{V_p V_p} & \mathbf{P}_{V_p \rho} \\ \mathbf{P}_{\rho V_p} & \mathbf{P}_{\rho \rho} \end{bmatrix}, \quad (38)$$

where $\mathbf{P}_{V_p V_p}$ and $\mathbf{P}_{\rho \rho}$ are the variance matrices of the marginal distribution of V_p and ρ , respectively, and $\mathbf{P}_{V_p \rho}$ and $\mathbf{P}_{\rho V_p}$ are the cross-covariance blocks. Note that since $\mathbf{P}_{V_p, \rho}$ is symmetric, we have $\mathbf{P}_{\rho V_p} = \mathbf{P}_{V_p, \rho}^T$ by definition. The $\mathbf{P}_{V_p V_p}$ block is expected to yield results similar to the covariance matrix in the mono-parameter case, while the $\mathbf{P}_{\rho \rho}$ block is its equivalent for the recovered density. The cross-covariance blocks are instead a measure of the link between the two parameters, and therefore makes it possible to quantify the inversion cross-talk between velocity and density. Starting with the parameter's uncertainty and cross-talk, we extract the four diagonal elements of the block joint-covariance matrix and plot them as variance and cross-covariance maps in Figs 16 and 17.

The initial variance maps are displayed in Fig. 16. The initial velocity variance distribution tends to the monoparameter case starting distribution, while the initial density variance map is very different. This is a result of the use of Gardner's law to produce the initial density models from perturbed velocity. The cross-covariance maps are symmetric and appear to be a combination of both velocity and density variances. The final variance maps are displayed in Fig. 17. As in the previous results, the geometrical spreading effect is the

prevalent source of uncertainty in the velocity reconstruction, while structural uncertainty is the dominant effect in the density variance map. Although the geometrical spreading is not directly visible in the density variance map, the higher variance values are located in the deeper region of the model nonetheless. The cross-covariance maps seem to indicate that the cross-talk between parameters is strongly linked to their respective uncertainties. The differences between the velocity variance map and the density variance map can be linked to wave propagation theory. The prevalence of the geometrical spreading effect can be associated to the higher sensitivity of the body waves to velocity perturbations, while the structural uncertainty in the density map could be explained by the higher sensitivity of reflected-arrivals towards density changes.

Added to the diagonal elements of the block-covariance matrix, individual parameters resolution and cross-talk terms of the block-correlation matrix are evaluated. This is achieved by extracting four corresponding lines out of the different blocks and mapping the correlation coefficients into the physical domain. This procedure is the extension of the correlation maps computation of the previous applications, to the block-diagonal structure. We choose a parameter arbitrarily, located at $z = 2.0$ km; $x = 9.6$ km and plot its initial correlation maps in Fig. 18 followed by their final correlation maps in Fig. 19.

Although the initial correlations are identical in all blocks due to the models' generation, the final correlation patterns are entirely different in the final maps. There is a sharp difference of resolution in velocity and density: velocity correlations are laterally oriented along the structure, while density correlations are oriented along a vertical axis across the domain. The resolution information is coherent with theoretical expectations as stated previously; velocity reconstruction is mostly constrained by diving waves that can explain lateral ambiguity, while density is constrained by short offset reflections arrivals, which can explain the higher vertical uncertainty.

Besides, correlation cross-talk maps allow evaluating the coupling effect between velocity and density across the whole domain. In that case, they are negligible with respect to the parameter

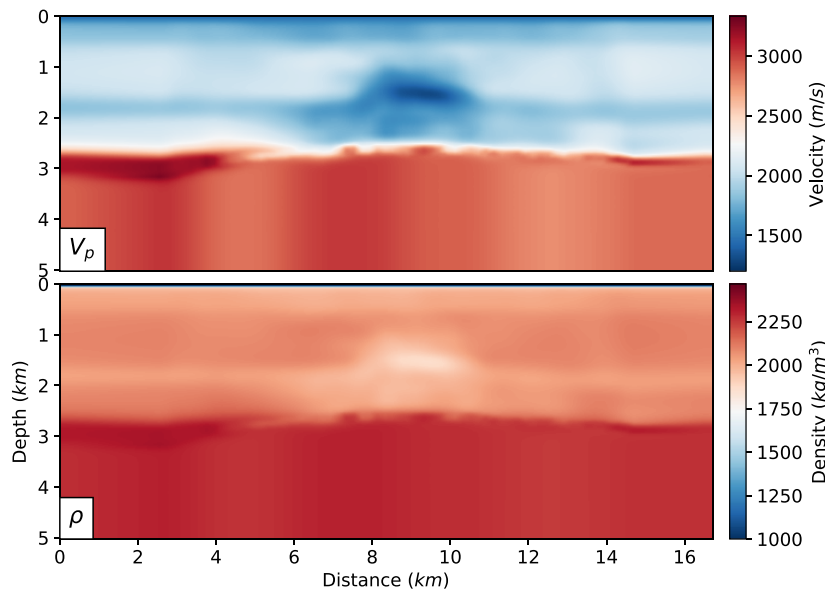


Figure 14. Top panel: initial ensemble mean velocity model m_{0, V_p} . Bottom panel: initial ensemble mean density model $m_{0, \rho}$.

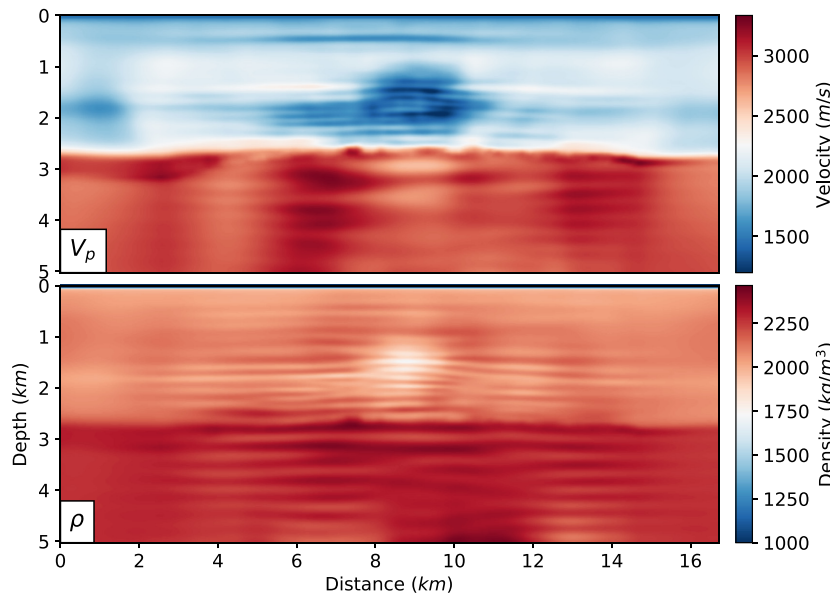


Figure 15. Top panel: final ensemble mean velocity model m_{0, V_p} . Bottom panel: final ensemble mean density model $m_{0, \rho}$.

resolution maps which makes the recovered density map believable (as density reconstruction does not seem to be contaminated by velocity leakage during the inversion).

Finally, we compare the estimated density, with an equivalent multiparameter FWI result, obtained with a similar inversion setup (data selection and processing, number of minimization steps, initial model) in Fig. 20.

Contrarily to the velocity estimation, there are significant discrepancies between the density model recovered by the ETKF-FWI and its FWI equivalent. The density in the hydrocarbon layers is lower in the FWI estimate, while the ETKF-FWI result is characterized by a high wavenumber content and sharper density contrasts. Because of these differences, both density estimates have been evaluated by comparing the synthetic data-fit with the observed common receiver gather data. Time-domain synthetic common-receiver gathers

are plotted in colour over the black-and-white observed data after filtering with a 6–8 Hz band-pass filter in Fig. 21. On this visualization, synthetic blue arrivals should overlap white, observed arrivals, while red should be overlapped by black arrivals (and therefore not be visible). The blue colour is hence indicative of good fit, while visible red is indicative of phases misalignment.

While the FWI result (centre panel) is significantly improving on the initial models (left-hand panel), the ETKF-FWI result (right-hand panel) is exhibiting an overall better data fit. Late arrival diving waves, as well as near offset reflections, are improved (see red ellipses). It seems that the analysis step of the ETKF-FWI acts as a Hessian-like pre-conditioning term, allowing a better convergence, which might enhance parameter disambiguation. While these preliminary results are a call for careful investigations, it seems that the analysis of the joint-space allows for better convergence of the

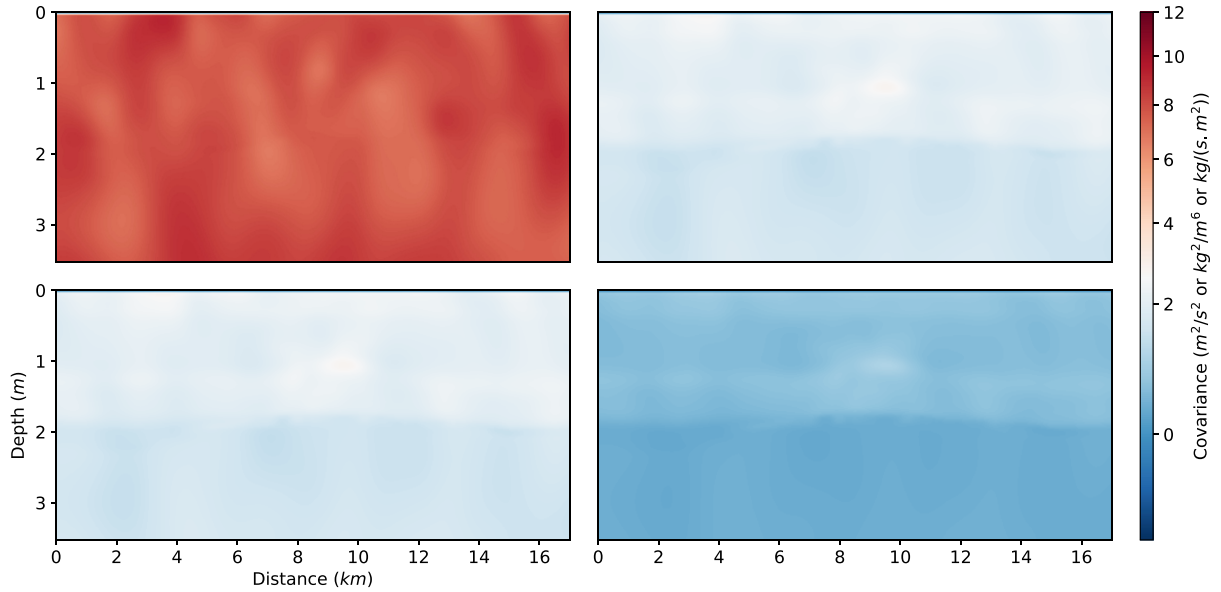


Figure 16. Diagonal elements of the initial joint-covariance matrix, plotted in the physical domain arranged according to their respective position in the block matrix. Top left-hand panel: P -wave velocity variance in $\text{m}^2 \text{s}^{-2}$. Bottom right-hand panel: density variance in $\text{kg}^2 \text{m}^{-6}$. Bottom left and top right-hand panels: V_p, ρ cross-covariance maps in $\text{kg} (\text{s.m}^2)^{-1}$.

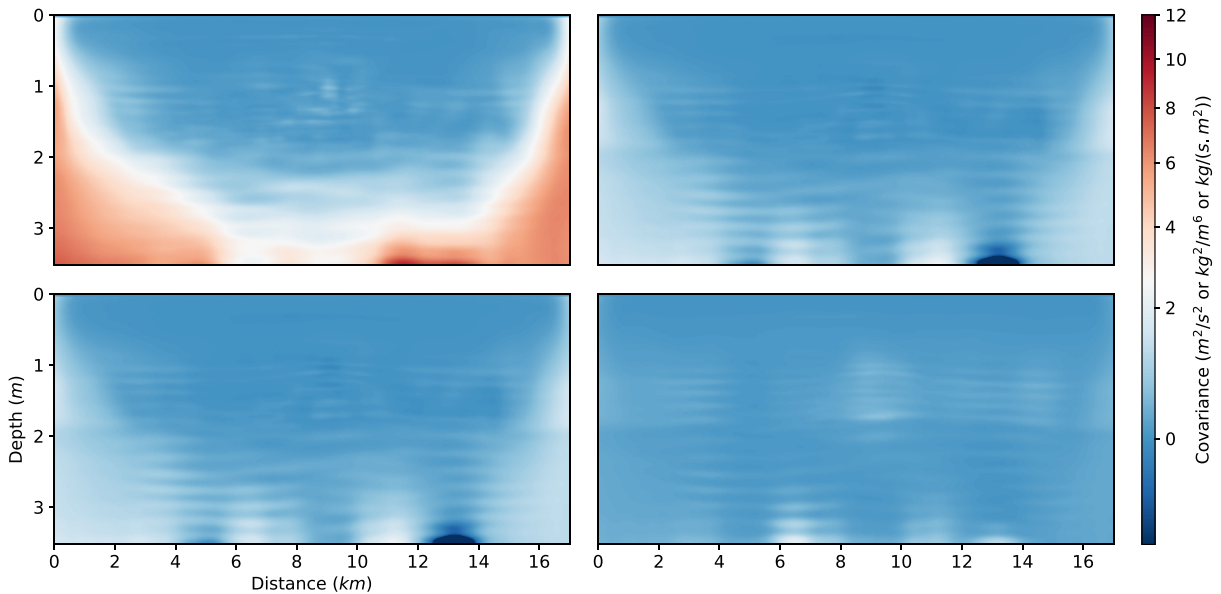


Figure 17. Diagonal elements of the posterior joint-covariance matrix, plotted in the physical domain arranged according to their respective position in the block matrix. Top left-hand panel: P -wave velocity variance in $\text{m}^2 \text{s}^{-2}$. Bottom right-hand panel: density variance in $\text{kg}^2 \text{m}^{-6}$. Bottom left and top right-hand panels: V_p, ρ cross-covariance maps in $\text{kg} (\text{s.m}^2)^{-1}$.

ETKF-FWI scheme, compared to the classical FWI. These results prompt us to investigate the possibilities of extension of the methodology beyond mono-parameter inversion in future studies.

5 DISCUSSION

These ETKF-FWI applications raise several points of discussions and questions that are yet to be answered.

What is the role of the analysis? In our applications, the Analysis step plays a crucial role in limiting the ensemble spread and thus prevents the ensemble from splitting over several local minima. The

role of the analysis is essential to rebalance the ensemble around the optimal mean (in the least squares sense) and lower the variance of the forecast ensemble. We can illustrate this behaviour with the evolution of variance between the forecast and the analysis steps in Fig. 22.

After the forecast, we observe a reduction of variance in shallow areas, but also a significant increase along a sharp velocity contrast (at 1.5 km depth). Because we cannot ensure that all of the N_e models will resolve the interface within the same number of iterations, the variance might increase in this specific area of the model. The analysis is responsible for a decrease of variance in the whole map, but its effect is predominant along with this velocity

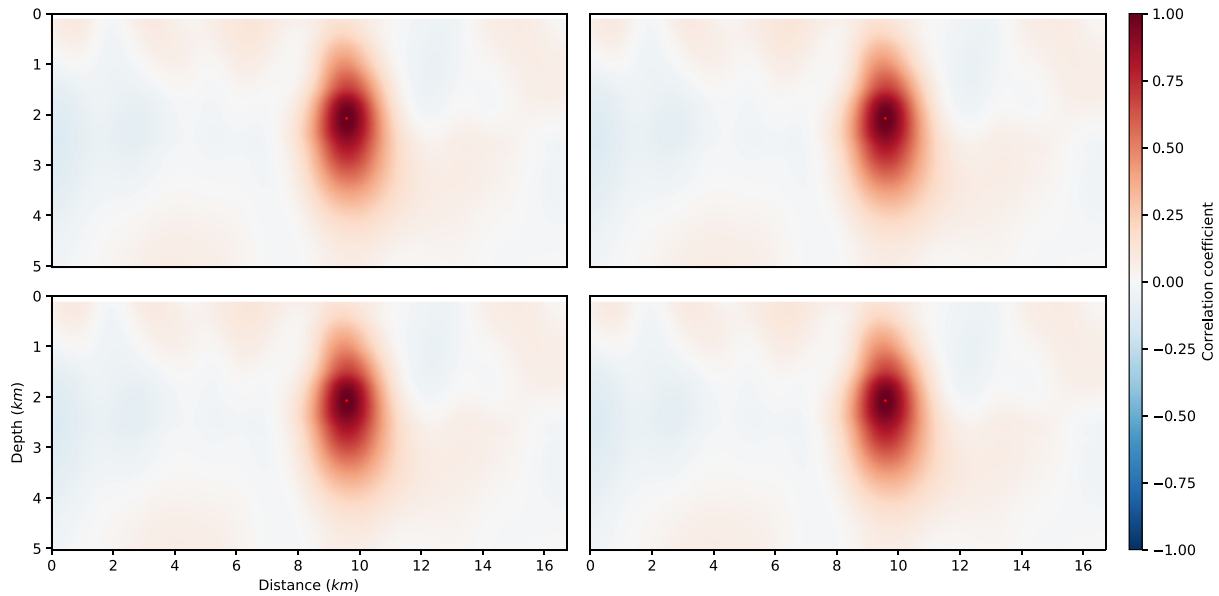


Figure 18. Off-diagonal elements of the initial joint-covariance matrix, plotted in the physical domain. The covariance matrix lines considered are corresponding to the parameter located at $z = 2.1$ km; $x = 9.6$ km. Top left-hand panel: P -wave velocity correlation coefficient. Bottom right-hand panel: density correlation coefficient. Bottom left and top right-hand panels: correlation cross-talk terms.

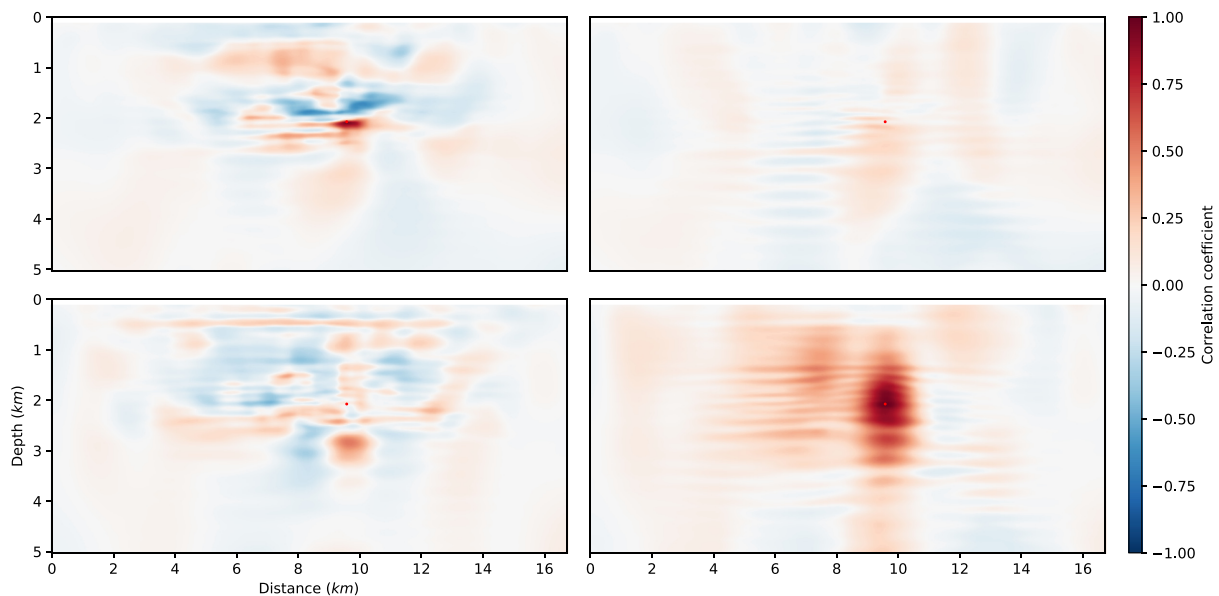


Figure 19. Off-diagonal elements of the posterior joint-covariance matrix, plotted in the physical domain. The covariance matrix lines considered are corresponding to the parameter located at $z = 2.1$ km; $x = 9.6$ km. Top left-hand panel: P -wave velocity correlation coefficient. Bottom right-hand panel: density correlation coefficient. Bottom left and top right-hand panels: correlation cross-talk terms.

contrast. Therefore the analysis is needed to improve the chances to sample a single minimum rather than splitting the ensemble. Note also that the Hessian-like preconditioning effect we have mentioned in the field data application might be related to this phenomenon. Given the forecast generates a high variance in some areas (that appear challenging to recover), most of the Analysis model update will take place in those areas, which should improve their recovery.

How much of a problem is undersampling? Regarding undersampling, its effects seem not too dramatic, as they do not affect the state estimate capabilities of the ETKF-FWI. We attribute this robustness to the inversion scheme that acts as our forecast, which

is not expected to spread-out the ensemble members. The underestimation of variance and spurious correlations might be more of an issue, as they have a direct impact on our ability to use and interpret the quantitative covariance data. Variance underestimation is typically solved through what is referred to as *covariance matrix inflation* in the DA community (Anderson & Anderson 1999). The goal of inflation is to artificially increase the forecast covariance by a factor r to mitigate overconfidence in the forecast. However, due to the necessity of evaluating an appropriate inflation parameter through trials and errors, its implementation in our case is limited. We might find a solution in recent DA schemes that enables automatic inflation settings (Miyoshi 2011), or overcome the inflation

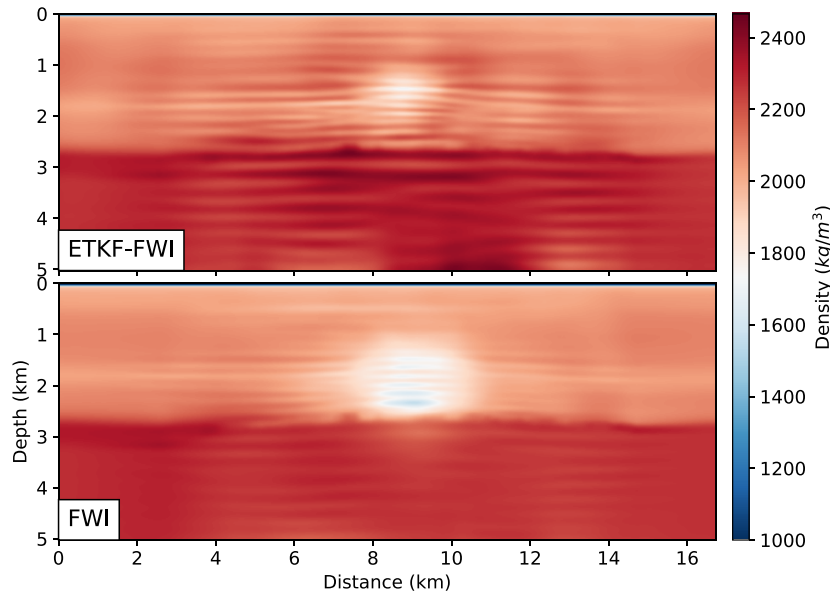


Figure 20. Comparison of ETKF-FWI (top panel) and FWI (bottom panel) density estimate with similar inversion setup (inversion parameters, regularization, acquisition geometry and data frequency groups).

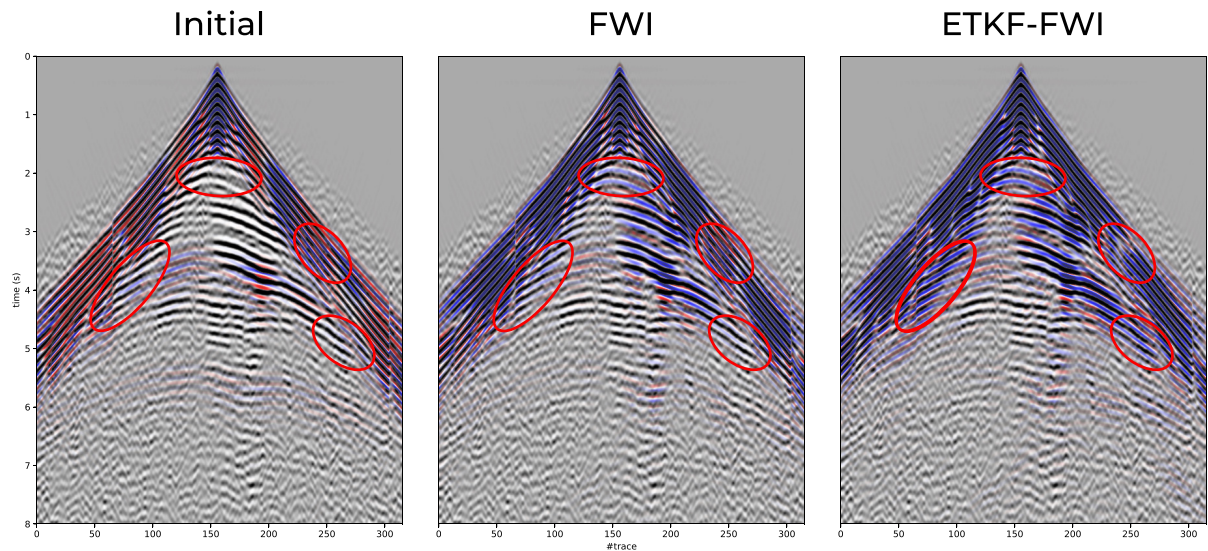


Figure 21. Datafit evaluated on a common receiver gather between (from left to right), the initial models, the FWI outcome and the ETKF-FWI outcome. Blue arrivals denote a good data fit over corresponding white arrivals. Red arrivals overlapping white arrivals are indicative of misaligned phases. Major improvement areas granted by the ETKF-FWI results have been marked with red ellipses in all three common-receiver gathers.

issue altogether like the finite-size EnKF (Myrseth & Omre 2010; Bocquet 2011; Myrseth *et al.* 2013; Bocquet *et al.* 2015). On the other hand, our observation operator is strongly non-local, which prevents us from applying covariance localization, the conventional solution to mitigate spurious correlation terms in the covariance matrix. Thus we propose to rely mostly on local covariance information, that seems to be preserved most of the time (as seen in correlation maps) and appears to be a reliable resolution proxy. Ultimately, the undersampling issues allowed us to address the validity of our low-rank approximation, and evaluate its associated biases. We think this specific point should be investigated in any methodology proposal based on rank reduction or Hessian approximation, which is unfortunately not always discussed in current propositions among the uncertainty estimation literature.

How to characterize prior uncertainty, and define the initial ensemble? Good practices when it comes to initial ensemble building may deserve entire research focus on its own. As it stands, we have adopted a pragmatical approach to generate initial perturbations, but defining ‘optimal’ and how an optimal initial ensemble should be built, is an open question. One might advocate for producing greater variance initial ensembles, to allow further parameter exploration at the cost of stability and convergence. Another option would be to align with the tests we have set up by limiting the spread of the initial ensemble to ensure an optimal parameter estimation. To constrain a strict convergence, one might even choose to add perturbation in limited portions of the model only, to limit the chances of unphysical updates during the analysis. For instance, in our field-data test case, we could remove perturbations in the

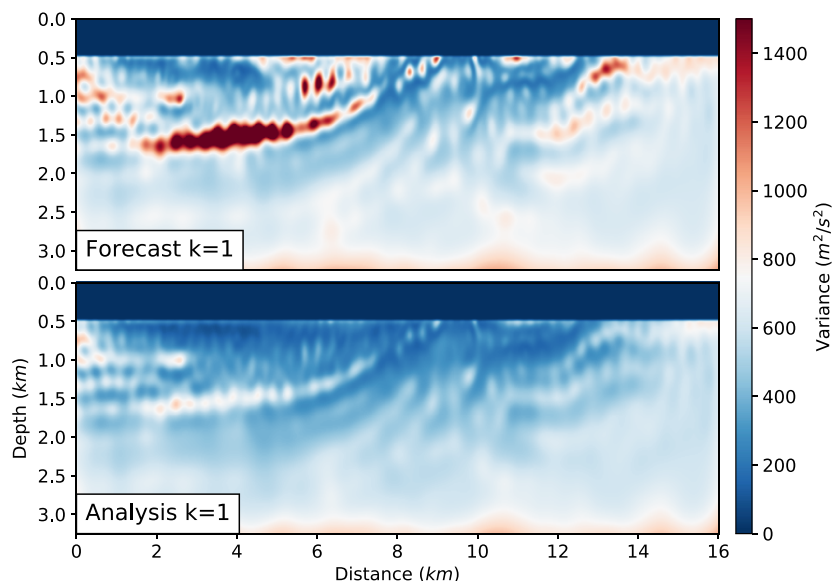


Figure 22. Evolution of variance between the forecast and the analysis step. These variance maps have been extracted from the synthetic application during the first ETKF-FWI cycle. They are, respectively, the first forecast variance (top panel) and the first analysis variance (bottom panel). The forecast is responsible for both increase and decrease of variance, while the analysis only reduces the variance.

lower half of the domain, constrained by a small portion of data. This would prevent any unphysical updates driven by the data term during the analysis. With such questions, we think the initial model building deserves a careful investigation, as the options mentioned above might be logical choices depending on one's goals.

How is the quantitative uncertainty estimate reliable? It has to be reminded that uncertainty estimates are, at best, expressed both in terms of 'local optimization' uncertainty and in the frame of finite-frequency wave propagation. As the wave propagation and the limited coverage act as a filter over the physical domain, it is not possible to link quantitative uncertainty with absolute physical parameter uncertainties. We instead think uncertainty should be expressed in terms of the optimal apparent macromodel as 'seen' by the waves, in similar ways as Capdeville & Métivier (2018) suggestion for downscaling and homogenization problems. Another possibility would be to find a way to express quantitative uncertainty regarding a reference FWI result. Unfortunately, it is unclear if we will be able to move towards real physical parameter uncertainty.

Extension and perspectives. It is worth noting that the technique might be extended to time domain applications to match current industrial standards. Time domain extension, despite requiring a completely different strategy concerning data management (mainly related to the cost of time domain FWI), could allow introducing time-based localization to the ETKF-FWI approach. A more global view of the approach also leads to the question of the variables and observations to consider in the ETKF-FWI. Up to now, only velocity and density have been introduced as variable, but other multiparameters system states could be considered. As an example, including the entire wavefield as an unknown variable of the ETKF, would allow making some links with the WRI proposed by van Leeuwen & Herrmann (2013), as both the physical parameter and the wavefield would be considered as unknowns. Multiparameter FWI fits well into this type of methodology extension, as it grants easy access to the cross-talk terms between inversion parameters, which are currently a challenging issue in multiparameter inversion. The benefits of the joint-inversion, as shown by the improvement on

data-fit in our multiparameter test case, will have to be thoroughly evaluated.

Cost and applicability. The differences and added values of the proposed approach relying on ETKF have to be evaluated to other methods from the literature, along with its practicality when it comes to applications. First, we have not discussed how this methodology compares with global optimization approaches. Global optimization approaches such as Martin *et al.* (2012), Biswas & Sen (2017) and Sajeve *et al.* (2017) are trying to mitigate the non-convexity of the cost function by sampling the entirety of the solution space, rather than sampling the cost function around the solution as we performed in the ETKF-FWI. While these methods seem very appealing, they have to rely on tricks to make this sampling possible and alleviate the *curse of dimensionality* problem they would face otherwise. These approaches are thus either limited to small problems (with a low number of unknown to sample) or rely on clever parameterizations (such as B-spline functions or Voronoi tessellation) to reduce the size of the search space. Nonetheless, most of these methodology will require several thousands of samples [and thus as many partial-differential-equ. (PDE) to solve], which makes them challenging to use as up to now. They also tend to produce very coarse solution to the inverse problem [which nonetheless makes for great potential starting models for local uncertainty estimation, as shown in Sajeve *et al.* (2017)]. The philosophy of local and global approaches differs, as they propose to deal with very different but complementary aspects of uncertainty estimation.

The local approaches that are more akin to our proposition are based on rank-reduction methods. These approximations of the inverse Hessian operator in the vicinity of the solution, make sampling from the posterior covariance matrix affordable. Their low-rank approximation of the inverse Hessian operator, require to solve several forward and adjoint PDEs, typically several hundred to several thousand per frequencies (for example Bui-Thanh *et al.* (2013) is evaluating 1400 PDE to estimate the first 700 eigenvalues of their global FWI application with hundreds of thousands of parameters). Fang *et al.* (2014) requires to solve approximately 6000 forward

modelling problems, with their MCMC sampling to produce an uncertainty estimate (with most of the cost coming from the sampling strategy). Zhu *et al.* (2016) is able to produce an uncertainty estimation along with the solution of the inverse problem at the minimal cost of 144 PDE resolution thanks to the assumption made on the structure of the Hessian operator. Though this cost is indeed reasonably low, it does not include the computational cost of the reverse time migration they are using to precondition their sampling. Finally, the number of PDE solved to sample the posterior covariance in Fang *et al.* (2018) proposition, is the number of sources plus the number of receivers per frequencies (not including the number of PDE to solve the inverse problem). Besides, this method does seem to display challenging memories limitation as it requires to store the optimal wavefields in memory for each frequency bands, which may become challenging for large scale 3-D application. The extension to uncertainty estimation of multiparameter inversion also seems to be non-trivial in this extended domain FWI application, as only recent publications are addressing the multiparameter aspect of WRI (Aghamiry *et al.* 2019). Note also that the low-rank approximation methods of the propositions mentioned above (such as randomized Singular-Value-Decomposition, or Lanczos methods) are sequential by nature which makes these uncertainty methods only as scalable as their PDE solver can get.

In comparison, the cost of ETKF-FWI in our applications ranges from 5000 to 18 000 PDE solve (for the synthetic and field data cases, respectively), which might appear to be a daunting number (although convergence tests have shown we could potentially consider smaller ensemble size). However, unlike the other methods, we are set to solve an embarrassingly parallel problem as all of our ensemble members are evolving independently during the bulk of the computational time (forecast step), which makes our problem not only scalable on the PDE solver but fully scalable on the ensemble size. Thanks to this advantage, and because of the development of hardware capacities towards the exascale and the current trend towards grid computation, we believe that the ETKF-FWI for uncertainty estimation can be a valuable approach even for large-scale FWI problems, as it is currently the case for DA applications.

6 CONCLUSION

We have demonstrated in those applications that the ETKF can be paired with a frequency-domain FWI quasi-Newton solver successfully, and allows for uncertainty estimation of the solution. The results we have obtained so far are encouraging in several regards. The presented method can produce a robust state estimation while allowing to recast our inversion problem in a local Bayesian framework. Variance and correlation maps only require to store the ensemble to be computed. Those maps provide a straightforward way of evaluating the quality of convergence, the correlation links and tradeoffs between parameters. It also allows integrating some form of data weighting terms in the whole tomographic process via the measurement noise matrix \mathbf{R} . If \mathbf{R} is set properly, the resulting uncertainty takes into account the physical properties of assimilated data. The extension perspectives offered by the DA framework and the full scalability of the method makes it a great candidate for uncertainty estimations.

ACKNOWLEDGEMENTS

This study was partially funded by the SEISCOPE consortium (<http://seiscope2.osug.fr>), sponsored by AKERBP, CGG,

CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC and TOTAL. This study was granted access to the HPC resources of the Froggy platform of the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>), which is supported by the Rhône-Alpes region (GRANT CPER07-13 CIRA), the OSUG@2020 labex (reference ANR10 LABX56) and the Equip@Meso project (reference ANR-10-EQPX-29-01) of the programme Investissements d'Avenir supervised by the Agence Nationale pour la Recherche, and the HPC resources of CINES/IDRIS/TGCC under the allocation 046091 made by GENCI. We thank AKERBP ASA and their partner Pandion Energy for providing the data set and permission to present this work, and the help of Ross Milne from AKERBP. We thank Editor . H. Chauris, one anonymous reviewer and Prof A. Fournier for their valuable suggestions which helped improve the paper.

REFERENCES

- Aghamiry, H., Gholami, A. & Operto, S., 2019. ADMM-based multiparameter wavefield reconstruction inversion in VTI acoustic media with TV regularization, *Geophys. J. Int.*, **219**(2), 1316–1333.
- Aki, K., Christofferson, A. & Husebye, E.S., 1977. Determination of the three-dimensional seismic structure of the lithosphere, *J. geophys. Res.*, **82**(2), 277–296.
- Anderson, J.L. & Anderson, S.L., 1999. A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, *Mon. Weather Rev.*, **127**(12), 2741–2758.
- Bardsley, J.M., Solonen, A., Haario, H. & Laine, M., 2014. Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems, *SIAM J. Numer. Anal.*, **36**(4), A1895–A1910.
- Bedle, H. & Lee, S.V.D., 2009. S velocity variations beneath North America, *J. geophys. Res.: Solid Earth*, **114**(B7), doi:10.1029/2008JB005949.
- Bishop, C.H., Etherton, B.J. & Majumdar, S.J., 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects, *Mon. Weather Rev.*, **129**(3), 420–436.
- Biswas, R. & Sen, M., 2017. 2D full-waveform inversion and uncertainty estimation using the reversible jump Hamiltonian Monte Carlo, in *Proceedings of the SEG Technical Program, Expanded Abstracts 2017*, pp. 1280–1285.
- Bocquet, M., 2011. Ensemble Kalman filtering without the intrinsic need for inflation, *Nonlin. Proc. Geophys.*, **18**(5), 735–750.
- Bocquet, M. & Sakov, P., 2014. An iterative Ensemble Kalman smoother, *Quart. J. R. Meteorol. Soc.*, **140**(682), 1521–1535.
- Bocquet, M., Raanes, P.N. & Hannart, A., 2015. Expanding the validity of the Ensemble Kalman filter without the intrinsic need for inflation, *Nonlin. Proc. Geophys.*, **22**(6), 645–662.
- Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N. & Pugmire, D., 2016. Global adjoint tomography: first-generation model, *Geophys. J. Int.*, **207**(3), 1739–1766.
- Bui-Thanh, T., Ghattas, O., Martin, J. & Stadler, G., 2013. A computational framework for infinite-dimensional Bayesian inverse problems Part I: the linearized case with application to global seismic inversion, *SIAM J. Scient. Comput.*, **35**(6), A2494–A2523.
- Bunks, C., Salek, F.M., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics*, **60**(5), 1457–1473.
- Byrd, R.H., Lu, P. & Nocedal, J., 1995. A limited memory algorithm for bound constrained optimization, *SIAM J. Scient. Stat. Comput.*, **16**, 1190–1208.
- Capdeville, Y. & Métivier, L., 2018. Elastic full waveform inversion based on the homogenization method: theoretical framework and 2-D numerical illustrations, *Geophys. J. Int.*, **213**(2), 1093–1112.
- Cosme, E., Brankart, J.-M., Verron, J., Brasseur, P. & Krysta, M., 2010. Implementation of a reduced rank square-root smoother for high resolution ocean data assimilation, *Ocean Modell.*, **33**(1–2), 87–100.
- Devaney, A., 1984. Geophysical diffraction tomography, *Geosci. Remo. Sens., IEEE Trans.*, **GE-22**(1), 3–13.

- Du, Z., Querendez, E. & Jordan, M., 2012. Resolution and uncertainty in 3D stereotomographic inversion, in *Proceedings of the 74th Annual EAGE Meeting Expanded Abstracts*, Copenhagen.
- Eikrem, K.S., Nævdal, G. & Jakobsen, M., 2019. Iterated extended Kalman filter method for time-lapse seismic full-waveform inversion, *Geophys. Prospect.*, **67**(2), 379–394.
- Eliasson, P. & Romdhane, A., 2017. Uncertainty quantification in waveform-based imaging methods—a sleipner CO₂ monitoring study, *Ener. Proc.*, **114**, 3905–3915.
- Evensen, G., 1994. Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. geophys. Res.*, **C5**(99), 143–162.
- Evensen, G., 2004. Sampling strategies and square root analysis schemes for the enfk, *Ocean Dyn.*, **54**(6), 539–560.
- Evensen, G., 2009. *Data Assimilation: The Ensemble Kalman Filter*, Springer.
- Evensen, G. & Van Leeuwen, P.J., 2000. An Ensemble Kalman smoother for nonlinear dynamics, *Mon. Weather Rev.*, **128**(6), 1852–1867.
- Fang, Z., Herrmann, F.J. & Silva, C.D., 2014. Fast uncertainty quantification of 2D full-waveform inversion with randomized source subsampling, in *Proceedings of the 76th Annual EAGE Conference & Exhibition, Expanded Abstracts*, EAGE, Amsterdam.
- Fang, Z., Silva, C.D., Kuske, R. & Herrmann, F.J., 2018. Uncertainty quantification for inverse problems with weak partial-differential-equation constraints, *Geophysics*, **83**(6), R629–R647.
- Feller, W., 2008. *An Introduction to Probability Theory and its Applications*, John Wiley & Sons.
- Fichtner, A. & Trampert, J., 2011a. Hessian kernels of seismic data functionals based upon adjoint techniques, *Geophys. J. Int.*, **185**(2), 775–798.
- Fichtner, A. & Trampert, J., 2011b. Resolution analysis in full waveform inversion, *Geophys. J. Int.*, **187**, 1604–1624.
- Fichtner, A. & van Leeuwen, T., 2015. Resolution analysis by random probing, *J. geophys. Res.: Solid Earth*, **120**(8), 5549–5573.
- Fichtner, A., Kennett, B. L.N., Igel, H. & Bunge, H.P., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.*, **179**(3), 1703–1725.
- Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y. & nor, A.V., 2013. Multiscale full waveform inversion, *Geophys. J. Int.*, **194**, 534–556.
- Fletcher, S., 2017. *Data Assimilation for the Geosciences*, Elsevier.
- French, S.W. & Romanowicz, B.A., 2015. Broad plumes rooted at the base of the earth's mantle beneath major hotspots, *Nature*, **525**(7567), 95.
- Gardner, G.H.F., Gardner, L.W. & Gregory, A.R., 1974. Formation velocity and density—the diagnostic basics for stratigraphic traps, *Geophysics*, **39**, 770–780.
- Gineste, M. & Eidsvik, J., 2017. Seismic waveform inversion using the Ensemble Kalman smoother, in *Proceedings of the 79th EAGE Conference and Exhibition 2017*.
- Gineste, M., Eidsvik, J. & Zheng, Y., 2019. Seismic waveform inversion using an iterative Ensemble Kalman smoother, in *Second EAGE/PESGB Workshop on Velocities*.
- Guzzi, R., 2015. *Data Assimilation: Mathematical Concepts and Instructive Examples*, Springer.
- Harlim, J. & Hunt, B.R., 2005. Local ensemble transform Kalman filter: an efficient scheme for assimilating atmospheric data, preprint.
- Hustedt, B., Operto, S. & Virieux, J., 2004. Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling, *Geophys. J. Int.*, **157**, 1269–1296.
- Iglesias, M.A., Law, K.J. & Stuart, A.M., 2013. Ensemble Kalman methods for inverse problems, *Inverse Problems*, **29**(4), 045001.
- Jin, L., Sen, M.K. & Stoffa, P.L., 2008. One-dimensional prestack seismic waveform inversion using Ensemble Kalman filter, in *Proceedings of the SEG Technical Program, Expanded Abstracts 2008*, pp. 1920–1924, SEG.
- Jordan, M., 2015. Estimation of spatial uncertainties in tomographic images, in *Proceedings of the 77th Annual EAGE Meeting Expanded Abstracts*, Madrid.
- Kalman, R., 1960. A new approach to linear filtering and prediction problems, *J. Basic Eng.*, **82**(1), 35–45.
- Lailly, P., 1983. The seismic problem as a sequence of before-stack migrations, in *Proceedings of the Conference on Inverse Scattering: Theory and Applications*, Philadelphia, SIAM.
- Lee, K., Jung, S. & Choe, J., 2016. Ensemble smoother with clustered covariance for 3D channelized reservoirs with geological uncertainty, *J. Petrol. Sci. Eng.*, **145**, 423–435.
- Leeuwenburgh, O., Evensen, G. & Bertino, L., 2005. The impact of ensemble filter definition on the assimilation of temperature profiles in the tropical pacific, *Quart. J. R. Meteorol. Soc.*, **131**(631), 3291–3300.
- Liu, M. & Grana, D., 2018. Stochastic nonlinear inversion of seismic data for the estimation of petroelastic properties using the ensemble smoother and data reparameterization, *Geophysics*, **83**(3), M25–M39.
- Liu, Q. & Peter, D., 2019. Square-root variable metric based elastic full-waveform inversion-part 2: uncertainty estimation, *Geophys. J. Int.*, **218**(2), 1100–1120.
- Martin, G.S., Wiley, R. & Marfurt, K.J., 2006. Marmousi2: an elastic upgrade for Marmousi, *Leading Edge*, **25**(2), 156–166.
- Martin, J., Wilcox, L., Burstedde, C. & Ghattas, O., 2012. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion, *SIAM J. Scient. Comput.*, **34**(3), A1460–A1487.
- Matharu, G. & Sacchi, M., 2019. A subsampled truncated-newton method for multi-parameter full waveform inversion, *Geophysics*, **84**(3), 1–33.
- Métivier, L. & Brossier, R., 2016. The seiscopie optimization toolbox: a large-scale nonlinear optimization library based on reverse communication, *Geophysics*, **81**(2), F11–F25.
- Métivier, L., Brossier, R., Virieux, J. & Operto, S., 2013. Full Waveform Inversion and the truncated Newton method, *SIAM J. Scient. Comput.*, **35**(2), B401–B437.
- Métivier, L., Bretaudeau, F., Brossier, R., Operto, S. & Virieux, J., 2014. Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures, *Geophys. Prospect.*, **62**, 1353–1375.
- Miyoshi, T., 2011. The gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter, *Mon. Weather Rev.*, **139**(5), 1519–1535.
- MUMPS team, 2017. *MUMPS - Multifrontal Massively Parallel Solver Users' Guide - version 5.1.1 (March 21, 2017)*, ENSEEIHT-ENS Lyon, <http://mumps-solver.org>.
- Myrseth, I. & Omre, H., 2010. Hierarchical ensemble Kalman filter, *SPE J.*, **15**(02), 569–580.
- Myrseth, I., Sætrom, J. & Omre, H., 2013. Resampling the ensemble Kalman filter, *Comput. Geosci.*, **55**, 44–53.
- Navon, I.M., 2009. Data assimilation for numerical weather prediction: a review, in *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, pp. 21–65, Springer.
- Nocedal, J. & Wright, S.J., 2006. *Numerical Optimization*, 2nd edn. Springer.
- Operto, S., Virieux, J., Dessa, J.X. & Pascal, G., 2006. Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: application to the eastern Nankai trough, *J. geophys. Res.*, **111**(B09306), doi:10.1029/2005JB003835.
- Operto, S., Virieux, J., Ribodetti, A. & Anderson, J.E., 2009. Finite-difference frequency-domain modeling of visco-acoustic wave propagation in two-dimensional TTI media, *Geophysics*, **74** (5), T75–T95.
- Operto, S., Brossier, R., Gholami, Y., Métivier, L., Prieux, V., Ribodetti, A. & Virieux, J., 2013. A guided tour of multiparameter full waveform inversion for multicomponent data: from theory to practice, *Leading Edge*, **32**(9), 1005–1176.
- Operto, S., Miniussi, A., Brossier, R., Combe, L., Métivier, L., Monteiller, V., Ribodetti, A. & Virieux, J., 2015. Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation, *Geophys. J. Int.*, **202**(2), 1362–1391.
- Ott, E. *et al.*, 2004. A local Ensemble Kalman filter for atmospheric data assimilation, *Tellus A*, **56**, 415–428.
- Panning, M.P., Lekić, V. & Romanowicz, B.A., 2010. Importance of crustal corrections in the development of a new global model of radial anisotropy, *J. geophys. Res.: Solid Earth*, **115**(B12), doi:10.1029/2010JB007520.

- Plessix, R.E., 2009. Three-dimensional frequency-domain full-waveform inversion with an iterative solver, *Geophysics*, **74**(6), WCC53–WCC61.
- Plessix, R.-E., Baeten, G., de Maag, J.W. & ten Kroode, F., 2012. Full waveform inversion and distance separated simultaneous sweeping: a study with a land seismic data set, *Geophys. Prospect.*, **60**, 733–747.
- Pratt, R.G., Shin, C. & Hicks, G.J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.*, **133**, 341–362.
- Rawlinson, N., Fichtner, A., Sambridge, M. & Young, M.K., 2014. Seismic tomography and the assessment of uncertainty, in *Advances in Geophysics*, Vol. **55**, pp. 1–76, Elsevier.
- Rodell, M. *et al.*, 2004. The global land data assimilation system, *Bull. Am. Meteorol. Soc.*, **85**(3), 381–394.
- Sajeva, A., Aleardi, M. & Mazzotti, A., 2017. Genetic algorithm full-waveform inversion: uncertainty estimation and validation of the results, *Bollettino di Geofisica Teorica ed Applicata*, **58**(4), doi:10.4430/bgta0199.
- Schillings, C. & Stuart, A.M., 2017. Analysis of the ensemble Kalman filter for inverse problems, *SIAM J. Numer. Anal.*, **55**(3), 1264–1290.
- Shen, X. *et al.*, 2018. High-resolution full-waveform inversion for structural imaging in exploration, in *Proceedings of the SEG Technical Program Expanded Abstracts 2018*, pp. 1098–1102.
- Sirgue, L., Barkved, O.I., Dellinger, J., Etgen, J., Albertin, U. & Kommedal, J.H., 2010. Full waveform inversion: the next leap forward in imaging at Valhall, *First Break*, **28**, 65–70.
- Stewart, L.M., Dance, S.L. & LNichols, N.K., 2008. Correlated observation errors in data assimilation, *Int. J. Numer. Methods Fluids*, **56**(8), 1521–1527.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2010. Seismic tomography of the southern California crust based on spectral-element and adjoint methods, *Geophys. J. Int.*, **180**, 433–462.
- Tarantola, A., 1984. Linearized inversion of seismic reflection data, *Geophys. Prospect.*, **32**, 998–1015.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics.
- van Leeuwen, P.J., Cheng, Y. & Reich, S., 2015. *Nonlinear Data Assimilation*, Springer.
- van Leeuwen, T. & Herrmann, F.J., 2013. Mitigating local minima in full-waveform inversion by expanding the search space, *Geophys. J. Int.*, **195**(1), 661–667.
- Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.
- Virieux, J., Operto, S., Ben Hadj Ali, H., Brossier, R., Etienne, V., Sourbier, F., Giraud, L. & Haidar, A., 2009. Seismic wave modeling for seismic imaging, *Leading Edge*, **28**(5), 538–544.
- Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A. & Zhou, W., 2017. An introduction to full waveform inversion, in *Encyclopedia of Exploration Geophysics*, pp. R1–1-R1-40, eds Grechka, V., Wapenaar & K., Society of Exploration Geophysics.
- Wang, X. & Bishop, C.H., 2003. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes, *J. Atmos. Sci.*, **60**(9), 1140–1158.
- Wang, X., Bishop, C.H. & Julier, S.J., 2004. Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble?, *Mon. Weather Rev.*, **132**(7), 1590–1605.
- Warner, M. *et al.*, 2013. Anisotropic 3D full-waveform inversion, *Geophysics*, **78**(2), R59–R80.
- Weston, P.P., Bell, W. & Eyre, J.R., 2014. Accounting for correlated error in the assimilation of high-resolution sounder data, *Quart. J. R. Meteorol. Soc.*, **140**(685), 2420–2429.
- Wu, R.S. & Toksöz, M.N., 1987. Diffraction tomography and multisource holography applied to seismic imaging, *Geophysics*, **52**, 11–25.
- Yilmaz, Ö., 1993. *Seismic Data Processing*, Society of Exploration Geophysicists.
- Zhou, W., 2016. Full waveform inversion of early arrivals and reflections for velocity model building and case study with gas cloud effect, *PhD thesis*. Univ. Grenoble Alpes, France.
- Zhou, W., Brossier, R., Operto, S., Virieux, J. & Yang, P., 2018. Velocity model building by waveform inversion of early arrivals and reflections: a 2d ocean-bottom-cable study with gas cloud effects, *Geophysics*, **83**(2), R141–R157.
- Zhu, H., Bozdağ, E. & Tromp, J., 2015. Seismic structure of the European upper mantle based on adjoint tomography, *Geophys. J. Int.*, **201**(1), 18–52.
- Zhu, H., Li, S., Fomel, S., Stadler, G. & Ghattas, O., 2016. A bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration, *Geophysics*, **81**(5), R307–R323.