Local uncertainty quantification for 3-D time-domain full-waveform inversion with ensemble Kalman filters: application to a North Sea OBC data set

Alexandre Hoffmann, Romain Brossier, Ludovic Métivier, Alizia Tarayoun

Geophysical Journal International

Geophys. J. Int. (2024) **237,** 1353–1383
Advance Access publication 2024 March 21
GJI Seismology

https://doi.org/10.1093/gji/ggae114

Local uncertainty quantification for 3-D time-domain full-waveform inversion with ensemble Kalman filters: application to a North Sea OBC data set

Alexandre Hoffmann [1,2] Romain Brossier,[1] Ludovic Métivier[1,3] and Alizia Tarayoun[1]

[1] *Univ. Grenoble Alpes, ISTerre, F*-38000 *Grenoble, France. E-mail: alexandre.hoffmann.etu@gmail.com*
[2] *Univ. Grenoble Alpes, CEA, List F*-38000 *Grenoble, France*
[3] *Univ. Grenoble Alpes, CNRS, LJK, F*-38000 *Grenoble, France*

Accepted 2024 March 15. Received 2024 March 14; in original form 2023 June 19

SUMMARY

Full-waveform inversion (FWI) has emerged as the state-of-the art high resolution seismic imaging technique, both in seismology for global and regional scale imaging and in the industry for exploration purposes. While gaining in popularity, FWI, at an operational level, remains a heavy computational process involving the repeated solution of large-scale 3-D wave propagation problems. For this reason it is a common practice to focus the interpretation of the results on the final estimated model. This is forgetting FWI is an ill-posed inverse problem in a high dimensional space for which the solution is intrinsically non-unique. This is the reason why being able to qualify and quantify the uncertainty attached to a model estimated by FWI is key. To this end, we propose to extend at an operational level the concepts introduced in a previous study related to the coupling between ensemble Kalman filters (EnKFs) and FWI. These concepts had been developed for 2-D frequency-domain FWI. We extend it here to the case of 3-D time-domain FWI, relying on a source subsampling strategy to assimilate progressively the data within the Kalman filter. We apply our strategy to an ocean bottom cable field data set from the North Sea to illustrate its feasibility. We explore the convergence of the filter in terms of number of elements, and extract variance and covariance information showing which part of the model are well constrained and which are not. Analysing the variance helps to gain insight on how well the final estimated model is constrained by the whole FWI workflow. The variance maps appears as the superposition of a smooth trend related to the geometrical spreading and a high resolution trend related to reflectors. Mapping lines of the covariance (or correlation matrix) to the model space helps to gain insight on the local resolution. Through a wave propagation analysis, we are also able to relate variance peaks in the model space to variance peaks in the data space. Compared to other posterior-covariance approximation scheme, our combination between EnKF and FWI is intrinsically scalable, making it a good candidate for exploiting the recent exascale high performance computing machines.

Key words: Tomography; Computational seismology; Waveform inversion.

1 INTRODUCTION

Full-waveform inversion (FWI), introduced by Lailly (1983); Tarantola (1984), is a high resolution tomographic technique which aims at reconstructing the 3-D distribution of geophysical subsurface parameters such as the wave velocities, density, attenuation or anisotropy parameters (Virieux & Operto 2009; Virieux *et al.* 2017). Its resolution power compared to other seismic tomographic techniques, such as seismic ray tomography, makes it popular both in academic seismology (Fichtner *et al.* 2009, 2013; Tape *et al.* 2010; Yuan *et al.* 2014; Fichtner & Villaseñor 2015; Bozdağ *et al.* 2016; Górszczyk *et al.* 2017; Karaoğlu & Romanowicz 2018; Lei *et al.* 2020; Lu *et al.* 2020; Thrastarson *et al.* 2022) for a better understanding of complex mechanisms and structures at depth, and in the seismic imaging industry, mainly for crustal-scale exploration (Plessix 2009; Sirgue *et al.* 2010; Plessix & Perkins 2010; Warner *et al.* 2013; Stopin *et al.* 2014; Vigh *et al.* 2014; Operto *et al.* 2015; Raknes *et al.* 2015; Solano & Plessix 2019).

FWI typically involves solving a non-linear optimization problem where one minimizes a given distance between *observed* seismic data and *simulated* seismic data computed through the solution of a wave propagation equation. The minimization is performed over a set of parameters controlling this wave propagation equation, which depends on the subsurface mechanical properties. The non-linearity of FWI arises from the non-linear relationship between these parameters and the wavefield. Such problem is generally solved using quasi-Newton techniques (Pratt *et al.* 1998; Shin *et al.* 2001; Operto *et al.* 2006; Brossier *et al.* 2009b; Métivier & Brossier 2016b, for instance). FWI is a computationally intensive technique as for typical applications it requires to repeatedly solve 3-D wave equations in domains containing several tens to several hundred of wavelengths in each spatial direction (leading to hundreds of discretization of points in each spatial direction), for a high number of different source terms.

Historically, this cost has been somehow controlled by formulating FWI in the frequency-domain. In this context the solution of the wave equation gets back to the solution of a linear system where the left-hand side involves a complex-valued impedance matrix whose values depends on both the medium parameters and the frequency (Pratt *et al.* 1998; Operto *et al.* 2006) and the right-hand side represents the source term at a given frequency. As soon as the impedance matrix can be factorized, using for instance an LU factorization (Virieux & Operto 2009), the wave equation can be solved for each right hand side in linear complexity which reduces largely the overall computational cost of FWI. While this approach can be used for 2-D FWI, the memory requirement and associated lack of scalability of direct solvers make them difficult to use for 3-D FWI (Li *et al.* 2020). For such problems, the FWI community has now switched to a time-domain formulation of the wave equation (Sirgue *et al.* 2010; Hu *et al.* 2012; Warner *et al.* 2013; Kamath *et al.* 2021), which has a limited memory requirement but is more computationally demanding, especially when a large number of sources is involved (Brossier *et al.* 2013).

The non-linearity of FWI coupled with its high computational cost makes it a challenging problem to solve for practical applications. For this reason, uncertainty estimation of FWI output remains a very challenging task. While simple quality controls, such as direct comparison between synthetic observed data via direct seismograms comparison and zero-lag cross-correlation between the observed and synthetic data, are possible (e.g. Pladys *et al.* 2022), they might not be sufficient as different velocity models can lead to similar data-fit (Prieux *et al.* 2011). In practice, most of the quality control assessment is either conducted by cross-validation with other geophysical techniques or well-log data, which are, at best, costly, at worst, impossible to realize past the shallow crustal scale. Uncertainty quantification is however all the more crucial that FWI is a highly non-linear problem with a significant 'numerical' null-space (ensemble of model parameters having a weak influence on the data then poorly constrained). In addition FWI models serves as a basis for geological interpretation in seismology, and more and more in the exploration industry as the frequency content of the data used to derive these models reach the frequency range of migration (Shen *et al.* 2018).

The uncertainty quantification of the FWI output is still weakly tackled by the literature as most of the research focuses on diminishing the method cost and on improving its robustness and resolution. While recent work propose a Bayesian approach for uncertainty quantification (Gebraad *et al.* 2020; Zhang & Curtis 2020; Zhang *et al.* 2023) the whole topic of uncertainty quantification is thus still challenging (Rawlinson *et al.* 2014). Nonetheless this topic is hardly new for the inverse problem community. Work on uncertainty quantification for seismic inverse problems was already considered more than decades ago by (Tarantola 2005). Using a Bayesian inference framework, Tarantola states that, for weakly non-linear problems, when the solution is close to the global minimum, the inverse of the Hessian operator is equivalent to the posterior covariance operator (Tarantola 2005, chapter 3). We should mention that, when dealing with non-linear inverse problems, Tarantola recommends to explore the model space and variance space with global methods such as Monte Carlo.

Alternatively the data assimilation (DA) community has been relying on Kalman filter (KF) to estimate the uncertainty of weather forecast. KF are used in dynamical process where the initial state is not exactly known. The state variable is *forecasted* by a forward modelling equation and then corrected or *filtered* to bring it closer to observations of the state variable. An interesting feature of KF is that they can be used on Gaussian processes and provide a tool to estimate the random variable mean and covariance after the filtering step (also called the *analysis* step; Kalman 1960). Recent attempt to estimate the uncertainty of the FWI took inspiration from theses two communities and can be divided into three categories.

A first category of methods rely on global or semi-global optimization methods (instead of local optimization) such as simulated annealing (Sen & Stoffa 1991; Tran & Hiltunen 2011), and genetic algorithms (Sen & Stoffa 1992), or methods aiming at sampling the posterior probability density function (PDF), that is Markov Chain Monte Carlo (Sen & Stoffa 1996; Martin *et al.* 2012; Bardsley *et al.* 2014; Biswas & Sen 2017) or more recently Hamiltonian Monte Carlo (Sen & Biswas 2017; Fichtner *et al.* 2018; Gebraad *et al.* 2020). This allows for the sampling of potentially multimodal PDF. These studies show promising results, and while the number of unknowns in the underlying FWI problem seems to hamper their applicability for realistic scale 3-D FWI problems, some recent works start to tackle 3-D synthetic problems (Zhang *et al.* 2023). Let us also mention stochastic optimization which seems also promising but has, so far, only been applied to small 2-D synthetic examples (Zhang & Curtis 2020).

Methods from the second category use information from the Hessian operator in the final FWI model. This approach is especially interesting in the case of multiparameter FWI where inter-parameters cross-talk are involved. In such a case, the Hessian gives an understanding of the contribution of each parameters and of the inter-parameter interactions (Operto *et al.* 2013; Métivier *et al.* 2014b). Due to the large scale aspect of FWI the Hessian operator is not directly accessible. It is however possible to evaluate its effect on a given vector at the price of the solution of four wave propagation equation (Fichtner & Trampert 2011a; Métivier *et al.* 2013, 2014a; Métivier *et al.* 2017; Yang *et al.* 2018; Matharu & Sacchi 2019). This yields the possibility to build a low-rank approximation of the inverse Hessian via a matrix-free Lanczos method (Bui-Thanh *et al.* 2013) or a randomized singular value decomposition (SVD; Zhu *et al.* 2016; Eliasson & Romdhane 2017). This low-rank approximation is motivated by the rapid decay of the eigenvalues of the Hessian operator observed by Bui-Thanh *et al.* (2013) and by Zhu *et al.* (2016) indicating that the Hessian operator has an *effective null-space*, that is there is a set of parameter perturbations with negligible influence on the FWI cost function. To further reduce the computation cost, it is possible to project the model parameters onto a low-dimensional subspace thus drastically reducing the dimension of the Hessian (Du *et al.* 2012; Jordan 2015).

A general limitation of these methods is that it is assumed that FWI has reached, or at least is very close from the global minimum

of the cost function. This makes it difficult, as noted by Keating & Innanen (2021), to quantify how close the inverse Hessian operator is to the actual posterior covariance operator. This is especially true, when an approximate Hessian is used to reduce the computational burden of the SVD/Lanczos iterations (e.g. Liu & Peter 2019). Furthermore, while constructing a low-rank approximation of the Hessian operator by omitting the *effective null-space* makes sense from a physical point of view, it also means that the model perturbations corresponding to substantial parameter uncertainty have been removed from the resulting truncated inverse Hessian operator as the neglected small eigenvalues of the Hessian correspond to large eigenvalues in the inverse Hessian. It is thus difficult, in practice, to select a threshold to truncate the SVD that preserves the accuracy of the truncated correlation matrix. Some authors thus try to quantify the uncertainty of FWI results without relying on approximating the inverse of the Hessian matrix. Hessian probing (Fichtner & Trampert 2011b; Fichtner & van Leeuwen 2015), as an example, has been used to assess the quality of FWI results obtained from field data (Tao *et al.* 2018; Lei *et al*. 2020). Alternatively, *null-space shuttles* can be viewed as the dual to SVD/Lanczos methods as one tries to find a subset of model perturbations that preserves the data misfit, thus exploring the effective null space neglected by the SVD/Lanczos based methods. In other words, null-space shuttles explores the set of model features that are not properly constrained by the data. The concept was first introduced in the case of a linear overdetermined data fitting problem (Deal & Nolet 1996) and later adapted to FWI. A direct approach is to perform a quadratic approximation of the FWI cost function and to explore the set of vectors that preserves the value of the approximated cost function, either using SVD and random probing (Liu & Peter 2020) or by investigating only specific feature of the model making the uncertainty quantification more tractable (Keating & Innanen 2021). Another elegant approach to explore the null-space is to design an artificial Hamiltonian system where the model is treated as a high-dimensional particle and where the potential energy of the system is the FWI cost function, which forces the system to evolves along trajectories that preserves the value of the cost function (Fichtner & Zunino 2019).

The third general category of strategies relies on methods developed by the DA community to estimate the uncertainty of the FWI. DA has a wide range of applications including land surface (Rodell *et al.* 2004), weather prediction (Navon 2009), oceanography (Cosme *et al.* 2010) and, more recently, reservoir characterization (Lee *et al.* 2016). The DA community has developed a wide set of tools to tackle such large scale inverse problems with scarce and noisy data. Some of their tools directly integrate the uncertainty quantification within their problem-solving schemes. The first DA modern tool is the above mentioned KF (Kalman 1960). It is a predictor-corrector method which computes a random variable mean and covariance.

However, KF only work with *linear* forecast and observation operators. Besides it is restricted to low dimensional problems as it involves manipulations of dense matrices preventing to go beyond a few hundred to thousands of parameters. Theses limitations have been the motivation for the introduction of the ensemble KF (EnKF; Evensen 1994). EnKF rely on a low-rank approximation of the system state based on an ensemble of unknowns, the benefit of which is twofold. First, it avoids the inversion of large covariance matrices. Secondly, the forecast covariance can be approximated from the forecasted ensemble, removing the need for a linear forecast operator (Evensen 2003; Hunt *et al.* 2007).

Today, EnKF are used in weather forecasting codes that involves up to 10^9 degree of freedom (DoF) such as the MOGREPS (Bowler *et al.* 2008) or the ICON (Wang *et al.* 2018) codes, which is a testimony to their capacity to tackle large-scale problems. There has been quite a few application of DA techniques to seismic tomography and seismic imaging, including 1-D pre-stack waveform inversion (Jin *et al.* 2008), 1-D velocity profile inversion (Gineste & Eidsvik 2017; Gineste *et al.* 2019, 2020), elastic and petro-physical rock properties estimation by maping petrophysical properties onto classical elastic FWI parameters (Liu & Grana 2018), 2-D frequency domain FWI (Thurin *et al.* 2019) and 3-D time-lapse FWI (TLFWI; Huang & Zhu 2020). The methodology proposed by Huang & Zhu (2020) is very similar to a standard DA method where the data is assimilated progressively using a FWI scheme.

In this study, we are interested in the ensemble transform KF-FWI (ETKF-FWI) scheme proposed by Thurin *et al.* (2019). Assimilating progressively the data in time is adapted to 4-D/time-lapse imaging but not for general inversion applications. In Thurin *et al.* (2019), an alternative is proposed, where the forecast operator becomes the whole FWI operator. The conventional multi-scale approach, inverting for low-frequency to high frequency, is used to define a dynamic evolution over the frequency axis. This method has yielded satisfactory results on 2-D synthetic and field data. The ETKF-FWI scheme makes it possible to estimate reliable and interpretable mean model, variance model, and covariance matrix, using ensembles containing few tens to few hundred of models, for problems going up to hundred of thousands of unknowns. This implies being able to compute the solution of as many FWI problems as the number of ensemble members. However, the solution of these FWI problems can be computed in parallel, each of them being run independently from the other during the forecast step. The analysis step, which couples the information from the different runs, requires only basic linear algebra operations on squares matrices whose size is the number of elements in the set. This is to us the main interest of the ETKF-FWI strategy: it is by essence scalable, making it suitable for running on the new generation of exascale machines. Note that posterior covariance estimation through inverse Hessian estimation does not enjoy such properties as Lanczos/randomized SVD algorithms are by essence sequential. This limitation adds to the above mentioned difficulty to set an appropriate threshold to truncate the Hessian estimation.

The main limitation in the work of Thurin *et al.* (2019) is the 2-D frequency-domain framework within which it is performed. While this framework is adapted to develop a proof of concept, we want in this study to extend it to a 3-D time-domain framework adapted to the application to 3-D field data, and provide the first 3-D field data application of the ETKF-FWI technique. After discussing two possible adaptations of the ETKF-FWI technique to time-domain FWI, namely a time-offset based scheme and a source subsampling scheme, we apply the latter one to the 3-D North Sea Ocean-Bottom-Cable (OBC) data set in a 3–5 Hz frequency band. We show how variance and covariance estimated from this scheme can contribute to interpret the final estimation of the model. We link the variance of the model to the variance of the data, showing that seismic arrivals with high variance travel through model structures with high variance. We also provide quantitative variance estimates, making it possible to derive error bars from the final estimation (corresponding to the ensemble mean). We show that the variance follows the illumination pattern, with higher values in less illuminated zones at depth and on the side of the acquisition. Local peaks of variance are also located on interfaces. Correlation maps derived from the

covariance estimation can give an information about the local resolution of the model. We show how this resolution is high in the shallow part of the model and decreases with depth.

We shall highlight here that the uncertainty estimation workflow we design remains local: we explore locally the minimum found by FWI. Our method can thus be understood as a sensitivity analysis scheme for FWI based on EnKF, yet quantitative (by how much do we deviate from the mean model?). It means that the specific goal of designing a global uncertainty quantification method in the Bayesian sense, with error bars with respect to the ground truth, is still not tackled by our strategy. Nonetheless, our local uncertainty estimation scheme yields a significant added value to the common practice of FWI where single best fitting models are used. Using our strategy, confidence indices on specific zones of the FWI estimated model can be derived, which shall greatly help the interpretation of these models.

The study is structured as follows. In Section 2, we provide a short recall of EnKF and describe the specific EnKF scheme we are using in this work. In Section 3, we describe how two standard time-domain FWI workflow can be adapted to the ETKF-FWI formalism. We illustrate and apply theses workflows to a synthetic 2-D example. In Section 4, we illustrate how the method works on a 3-D field data set from the North Sea. We conclude by a discussion about our results and future work.

2 ETKFs

2.1 General framework

The EnFK is well known in the DA community (e.g. Bowler *et al.* 2008; Miyoshi 2011; Chen *et al.* 2020) and has been introduced to the seismic imaging community notably by Thurin *et al.* (2019). Therefore we propose here only a quick summary, and refer interested readers to Thurin *et al.* (2019) and Evensen (2009) for more details. The main drawback of *deterministic* EnKF is that the covariance analysis does not satisfy the best linear unbiased estimator (BLUE) equation. ETKFs (Bishop *et al.* 2001; Ott *et al.* 2004) where introduced to overcome this limitation. While EnKF directly update the covariance matrix, ETKK update the *square root* factorization of the covariance matrix to compute the appropriate transformation and to perform the analysis with small size reduced covariance matrices (Tippett *et al.* 2003; Nerger *et al.* 2012). ETKF, very much like the standard KF, works in two steps, forecast and analysis, which we recall below.

The *forecast* step consists in evolving the state variable from time t_k to time t_{k+1} by a forecast operator $\mathcal{F}_k$. Let us assume that the system under study has N_{dof} DoFs and we have N models in our ensemble. The essence of ensemble methods is to have $N \ll N_{\text{dof}}$. Typically, N reaches few hundreds, even sometimes few tens, while N_{dof} can be as large as 10^{10}. Mathematically, the forecast operation is described by

$$m^i_{f,k} = \mathcal{F}_k(m^i_{k-1}) + \eta^i_k \quad i = 1 \dots N, \tag{1}$$

where $m_{f,k} = [m^1_{f,k} \dots m^N_{f,k}] \in \mathbb{R}^{N_{\text{dof}} \times N}$ is our forecasted state variable (ensemble of models) and where η_k is the process-noise at step k, typically $\eta_k \sim \mathcal{N}(0,\ Q_k)$ with Q_k being the forecast error covariance matrix. Note that in practice, very few information about Q_k are available and in such a case, one generally neglects the process noise η_k as it cannot be properly simulated. This is equivalent to consider that $Q_k = 0$. This is what is considered in this study.

The first and second order Gaussian moments of $m_{f,k}$, namely mean $\bar{m}_{f,k}$ and covariance $P_{f,k}$, can be computed as

$$\begin{aligned} \bar{m}_{f,k} &= \frac{1}{N}\sum_{i=1}^{N} m^i_{f,k}, \\ P_{f,k} &= \frac{1}{N-1} M_{f,k} M^T_{f,k}, \end{aligned} \tag{2}$$

where $M_{f,k} = [m^1_{f,k} - \bar{m}_{f,k} \dots m^N_{f,k} - \bar{m}_{f,k}]$ is the *square root* of the covariance matrix $P_{f,k} \in \mathbb{R}^{N_{\text{dof}} \times N_{\text{dof}}}$. Such representation makes it possible to store and manipulate only $M_{f,k} \in \mathbb{R}^{N_{\text{dof}} \times N}$, which has N_{dof} rows and N columns, rather than storing $P_{f,k}$, a square matrix of size N_{dof}. This is much more efficient both computationally and memory wise. On the other hand, with such representation, the rank of $M_{f,k}$ is, at best, $N-1$. $P_{f,k}$ is thus a low-rank approximation, of rank, at best, $N-1$ of the covariance behind ensemble schemes.

The second step of an ETKF is the so-called *analysis* step. The core idea of the analysis is to compute a new variable $m_{a,k}$ which is a balance between the model state $m_{f,k}$ predicted by the forecast operator and the synthetic data state y^{obs}_{k+1} that is available at time t_{k+1}. First, one computes the synthetic data $y_{f,k}$ from the state variable:

$$y^i_{f,k} = \mathcal{H}_{k+1}\left(m^i_{f,k}\right) \quad i = 1 \dots N, \tag{3}$$

where $\mathcal{H}_{k+1}$ is the non-linear observation operator at step $k+1$, which will be further detailed in the next section. We then compute the average synthetic data $\bar{y}_{f,k} \in \mathbb{R}^{N_{\text{obs}}}$ and the zero-mean observation matrix $Y_{f,k} \in \mathbb{R}^{N_{\text{obs}} \times N}$ (N_{obs} rows and N columns) as follows:

$$\begin{aligned} \bar{y}_{f,k} &= \frac{1}{N}\sum_{i=1}^{N} y^i_{f,k}, \\ Y_{f,k} &= \left[y^1_{f,k} - \bar{y}_{f,k} \ \dots \ y^N_{f,k} - \bar{y}_{f,k}\right]. \end{aligned} \tag{4}$$

Using these two sources of information, the first and second order Gaussian moments of our analysed state variable, $m_{a,k}$, can be computed as:

$$\begin{aligned} \bar{m}_{a,k} &= \bar{m}_{f,k} + M_{f,k} A_k Y^T_{f,k} R^{-1}_{k+1}\left(y^{\text{obs}}_{k+1} - y_{f,k}\right), \\ P_{a,k} &= \frac{1}{N-1} M_{f,k} A_k M^T_{f,k}, \end{aligned} \tag{5}$$

where $R_{k+1} \in \mathbb{R}^{N_{\text{obs}} \times N_{\text{obs}}}$, a square matrix of size N_{obs}, is the measurement noise matrix expressing the data uncertainty, and where $A_k \in \mathbb{R}^{N \times N}$, a square matrix of size N, is our reduced covariance matrix, which can be computed as

$$A_k = \left((N-1)\, I_N + Y^T_{f,k} R^{-1}_{k+1} Y_{f,k}\right)^{-1}, \tag{6}$$

where I_N is the identity matrix of size N. The state variable after analysis can then be computed as

$$m_{a,k} = m_{f,k} + \sqrt{N-1} M_{f,k} T_k, \tag{7}$$

where $T_k \in \mathbb{R}^{N \times N}$ is the so-called *transformation matrix*, which is computed from the square root of A_k to preserve the ensemble mean (Wang *et al.* 2004; Sakov & Oke 2008)

$$T_k = A_k^{\frac{1}{2}} \Lambda, \tag{8}$$

with Λ an arbitrary orthogonal mean-preserving matrix (Sakov & Oke 2008; Livings *et al.* 2008).

In the current work, we chose $\Lambda = I_N$. It is worth mentioning that, instead of directly evaluating A_k and then $A_k^{\frac{1}{2}}$, it is easier to compute the eigendecomposition of A_k^{-1} from which we can easily compute

both A_k and $A_k^{\frac{1}{2}}$. For the measurement noise matrix R_{k+1}, we assume in this work it has a diagonal structure, which implies uncorrelated noise. We estimate for each shot-gather a level of noise, and use a shot-dependent weighting strategy for build the diagonal entries of the measurement noise matrix R_{k+1}.

Because all underlying Gaussian assumptions hold for the ETKF, the analysis step as described above is equivalent to solve the following least-squares minimization problem (Hunt *et al.* 2007):

$$\underset{\bar{m}_{a,k} \in \mathbb{R}^{N_{\text{dof}}}}{\text{minimize}} \; \mathcal{C}(\bar{m}_{a,k}) := \frac{1}{2} \left\| y_{k+1}^{\text{obs}} - \mathcal{H}_{k+1}(\bar{m}_{a,k}) \right\|_{R_{k+1}^{-1}}^2 + \frac{1}{2} \left\| \bar{m}_{f,k} - \bar{m}_{a,k} \right\|_{P_{f,k}^{-1}}^2, \tag{9}$$

where the cost function $\mathcal{C}$ realizes a balance between the forecast model and the available observations.

If we define, for a given *weight vector*, $\omega \in \mathbb{R}^N$

$$\begin{aligned} \bar{m}_{a,k}(w) &= \bar{m}_{f,k} + M_{f,k} w, \\ \mathcal{H}_{k+1}(\bar{m}_{a,k}(w)) &\approx \bar{y}_{f,k} + Y_{f,k} w. \end{aligned} \tag{10}$$

$\bar{m}_{a,k}$ as defined in eq. (5) is a solution of eq. (9) (Hunt *et al.* 2007; Harlim & Hunt 2007).

In terms of computational cost, the ETKF involves N application of the forecast and observation operators as well as the manipulation of $N_{\text{dof}} \times N$, $N_{\text{obs}} \times N$ and $N \times N$ matrices and the diagonalization of a $N \times N$ matrix to compute A_k and $A_k^{\frac{1}{2}}$.

2.2 Balancing the analysis

We should mention that, in the particular case of time-domain ETKF-FWI, by defining our data space over the (physical) time-domain, we end up with a data space that is significantly larger than our model space. As an example, when working with the North Sea OBC data set in Section 4, our data space is approximately 3000 times larger than our model space.

This dimensionality difference may be problematic because, as mentioned in the previous section and as illustrated by eq. (9), the analysis step is supposed to find a model that minimizes the data misfit while remaining relatively close to the forecasted model. However, a huge dimensionality difference between the data space and the model space can induce a bias in the analysis step where the minimization of the data misfit will be prioritized over the minimization of the background misfit.

In order to illustrate this, let us assume there exists a $\bar{m}_{a,k}$ such that both background and observation residuals are constant vectors with all entries equal to ϵ_{model} and ϵ_{obs}, respectively, such that

$$\begin{aligned} m_{f,k} - \bar{m}_{a,k} &= \epsilon_{\text{model}} \mathbf{1}, \\ y_{k+1}^{\text{obs}} - \mathcal{H}_{k+1}(\bar{m}_{a,k}) &= \epsilon_{\text{obs}} \mathbf{1}, \end{aligned} \tag{11}$$

with $\mathbf{1}$ denoting a vector with constant entries equal to 1. In this case the values taken by the cost functions in eq. (9) would be

$$\mathcal{C}(\bar{m}_{a,k}) = N_{\text{obs}} \frac{\epsilon_{\text{obs}}}{2} \sum_{i=1}^{N_{\text{obs}}} \sum_{j=1}^{N_{\text{obs}}} (R_{k+1})_{i,j}^{-1} + N_{\text{dof}} \frac{\epsilon_{\text{model}}}{2} \sum_{i=1}^{N_{\text{dof}}} \sum_{j=1}^{N_{\text{dof}}} (P_{f,k})_{i,j}^{-1}. \tag{12}$$

Thus if $N_{\text{dof}} \gg N_{\text{obs}}$, the minimization of $\mathcal{C}$, will tend to favour the diminution of the background error. On the other hand, if $N_{\text{obs}} \gg N_{\text{dof}}$, the diminution of the observation error will be favoured.

In our first trials, this dimensional difference resulted in an analysis that collapsed the ensemble into a single model that minimized the data misfit, even when the ensemble after the first forecast had a very high variance.

We solved this problem by replacing the L^2 norms in eq. (9) by mean squared deviation (MSD), which leads to the following optimization problem:

$$\underset{\bar{m}_{a,k} \in \mathbb{R}^{N_{\text{dof}}}}{\text{minimize}} \; \tilde{\mathcal{C}}(\bar{m}_{a,k}) := \frac{1}{2N_{\text{obs}}} \left\| y_{k+1}^{\text{obs}} - \mathcal{H}_{k+1}(\bar{m}_{a,k}) \right\|_{R_{k+1}^{-1}}^2 + \frac{1}{2N_{\text{dof}}} \left\| \bar{m}_{f} - \bar{m}_{a,k} \right\|_{P_{f,k}^{-1}}^2, \tag{13}$$

which is equivalent to scaling the observation error matrix by $\frac{N_{\text{dof}}}{N_{\text{obs}}}$. Our analysis step is thus performed by simply modifying eq. (6) as follows:

$$A_k = \left((N-1) I_N + \frac{N_{\text{dof}}}{N_{\text{obs}}} Y_{f,k}^T R_{k+1}^{-1} Y_{f,k} \right)^{-1}. \tag{14}$$

Note that, with our modified analysis, if we go back to the previous schematic example, we get the following equality

$$\tilde{\mathcal{C}}(\bar{m}_a) = \frac{\epsilon_{\text{obs}}}{2} \sum_{i=1}^{N_{\text{obs}}} \sum_{j=1}^{N_{\text{obs}}} (R_{k+1})_{i,j}^{-1} + \frac{\epsilon_{\text{model}}}{2} \sum_{i=1}^{N_{\text{dof}}} \sum_{j=1}^{N_{\text{dof}}} (P_{f,k})_{i,j}^{-1}. \tag{15}$$

Thus, the minimization of $\tilde{\mathcal{C}}$ should not favour the reduction of one term against another. Note that, instead of using the MSD, we could simply divide the data misfit by the number of observations, which would result in a similar scaling and would make sense from a physical point of view. It is worth mentioning that the DA literature proposes tests to check the consistency of the analysis and proposes method to balance it (e.g. Desroziers *et al.* 2005).

3 ETKFs FOR FWI

3.1 General ETKF-FWI scheme

FWI can be formulated as

$$\min_{m} \quad J(m) := \mathcal{D}(d_{\text{cal}}(m), d_{\text{obs}}) \tag{16a}$$

$$\text{s.t.} \quad d_{\text{cal}}(m) = \mathcal{R} u(m) \tag{16b}$$

$$\mathcal{A}(m) u = s, \tag{16c}$$

where $\mathcal{D}$ is a function measuring the misfit between the observed data d_{obs} and the calculated data $d_{\text{cal}}(m)$ (typically the least-squares distance). The latter is computed through the solution of the wave propagation eq. (16c), where $\mathcal{A}(m)$ stands for a general wave propagation operator (from acoustics to viscoelastic), and the application of a restriction operator $\mathcal{R}$, which extracts the values of the wavefield $u(m)$ solution of eq. (16c) at the receiver positions.

As mentioned in the introduction, FWI problems are solved through local optimization methods due to the large number of degrees of freedom involved in the discretization of the wave equation. This implies the definition of an initial guess m_0 from which the solution of eq. (16) is computed. For this reason, we introduce the operator $\mathcal{I}(m_0)$, which provides the solution of the FWI problem starting from an initial guess m_0.

A ETKF-FWI scheme is constructed by defining the following forecast operator

$$m_{k+1} = \mathcal{F}_k(m_k) := \mathcal{I}(m_k) \tag{17}$$

and observation vector $\mathcal{H}_k$

$$\mathcal{H}_k(m_k) := \mathcal{R} u(m_k). \tag{18}$$

As mentioned in the introduction, the evaluation of both the forecast and observation operator can be performed in parallel. The implementation of a parallel ETKF-FWI is quite challenging as a single FWI is already computationally expensive. The Appendix presents the framework we developed to run our ETKF-FWI in parallel.

What remains to be defined is a form of dynamic along which the models m_k evolve, that is a specific splitting of the data so that each piece of data is interpreted step-by-step in the ETKF-FWI scheme.

3.2 Frequency-domain ETKF-FWI

The key idea from Thurin *et al.* (2019) is to define this dynamic from the conventional multiscale approach used in frequency-domain FWI. More precisely, the ETKF steps correspond to the inversion of the data in different frequency bands, starting from the lowest frequencies and progressively interpreting higher and higher frequencies.

In this configuration, the forecast operator $\mathcal{F}_k$ represents a FWI over the kth frequency band and the observation operator $\mathcal{H}_k$ returns the synthetic observations for the kth frequency band. The method is sketched in Fig. 1 that we reproduce from Thurin *et al.* (2019).

The method has proved successful to produce uncertainty quantification for frequency-domain FWI. However, memory requirement of frequency domain FWI is prohibitive for tackling 3-D applications. While it is possible to work in the frequency-domain using a time-domain modelling engine (Sirgue *et al.* 2010), it is computationally expensive, especially when a large number of sources is involved (Brossier *et al.* 2013). The industry standard is now to work in the time-domain both for modelling and inversion. This is our motivation to extend the ETKF-FWI formalism to time-domain FWI.

3.3 Time-domain ETKF-FWI

Following Thurin *et al.* (2019), we must restore a dynamic with a hierarchical aspect to the time-domain FWI. A straightforward way of doing that would be to perform a time-domain FWI with data filtered within a certain list of frequency bands and then re-use the formalism proposed by Thurin *et al.* (2019). However this is not optimal as this would be computationally demanding. For ETKF-FWI to produce reliable uncertainty quantification, it is important to ensure a sufficient number of assimilation steps k. Running an entire FWI (meaning with all the sources) on a large number of filtered data sets would be prohibitive. Moreover, while time-domain FWI typically loop over a certain frequency range, an increase in the frequency content of the data is typically accompanied by a finer discretization of the model parameters. Such a scheme would thus require the dimensionality of our ETKF's state variable to change with each iteration, which is not possible within the ETKF framework.

3.3.1 Time-offset windowing scheme

Frequency windowing is not the only hierarchy which can be considered in a FWI scheme. As an example, it is possible to use a hierarchy of time-offset windowing to avoid cycle skipping (Shipp & Singh 2002; Wang 2009; Brossier *et al.* 2009a). Such strategy has recently been applied to a North Sea OBC data set to avoid cycle skipping when starting from a 1-D initial velocity model (Pladys *et al.* 2022). These strategies directly echo the DA framework where data are added iteration after iteration. A noticeable difference between this strategy and the DA framework is that data corresponding to low-offset and early arrival time will be used several times in the inversion process, which can introduce a bias in the reconstruction, where the upper part of the model are better reconstructed than the lower parts.

3.3.2 Source subsampling scheme

When dealing with a huge number of sources, time-domain FWI also typically uses techniques to limit the number of sources simultaneously inverted so as to reduce the computational cost of the inversion. Such techniques includes source encoding (Capdeville *et al.* 2005; Vigh & Starr 2008; Krebs *et al.* 2009; Baumstein *et al.* 2011; Ben Hadj Ali *et al.* 2011; Schiemenz & Igel 2013; Castellanos *et al.* 2015) and source subsampling (Warner *et al.* 2013; Kamath *et al.* 2018, 2019; Pladys *et al.* 2022). In the latter case, the original pool of sources is divided in non-overlapping subensembles randomly constituted. If Ω is the ensemble of all the data set, we build a partition of N_s subensembles Ω_k, $k = 1, \ldots, N_s$, such that

$$\bigcup_{k=1}^{N_s} \Omega_k = \Omega, \quad \text{and} \quad \forall k, l, \quad k \neq l, \quad \Omega_k \cap \Omega_l = \emptyset. \tag{19}$$

Such subsampling strategy can also be used in the frame of ETKF-FWI: each step k of the assimilation consists in inverting for the data in the subensemble Ω_k.

3.4 Application to a 2-D synthetic model: the Marmousi II model

We perform a first experiment under the 2-D constant-density acoustic approximation. We use the SEISCOPE code TOYxDAC_TIME to model and invert the data, which is based on a 2nd-order in time and a 4th-order in space finite-difference discretization of the wave equation. More details on its implementation can be found in Yang *et al.* (2018).

We illustrate how the two time-domain ETKF-FWI strategies we have described above apply on the 2-D acoustic Marmousi-II synthetic model (Martin *et al.* 2006). The model we take as exact is presented in Fig. 2(a). Using this model, we generate a data set using a fixed spread surface acquisition, with 128 sources and 169 receivers evenly spaced each 110 and 100 m, respectively. We use a spatial discretization step of 25 m yielding $141 \times 681 = 96\,021$ DoFs. We use a Ricker wavelet centred on 5 Hz to generate the data, yielding signal up to 12.5 Hz approximately. We add a Gaussian random noise to the data. The noise is filtered in the frequency band of the data (0–12.5 Hz) before it is added to it. The signal-to-noise ratio (SNR) is set to 10.

The TOYxDAC_TIME is coupled with the SEISCOPE optimization toolbox (Métivier & Brossier 2016b), which enables the use of a conventional limited-memory Broyden–Fletcher–Goldfarb–Shanno (l-BFGS) algorithm to solve the minimization problem (Nocedal 1980). A depth pre-conditioner and a non-stationary Gaussian smoothing strategy are used to precondition the gradient. The gradient is filtered using a non-stationary Gaussian filter which correlation lengths are equal to a fraction of the dominant wavelenght λ_{dom}, defined as

$$\lambda_{\text{dom}}(\mathbf{x}) = \frac{V_P(\mathbf{x})}{f_0}, \tag{20}$$

where $V_P(\mathbf{x})$ is the pressure wave velocity model and f_0 is a reference frequency we select equal to 5 Hz here. The correlation lengths are set equal to $0.2 \times \lambda_{\text{dom}}$ in both x and z directions. We consider an ensemble of $N = 50$ velocity models. According to

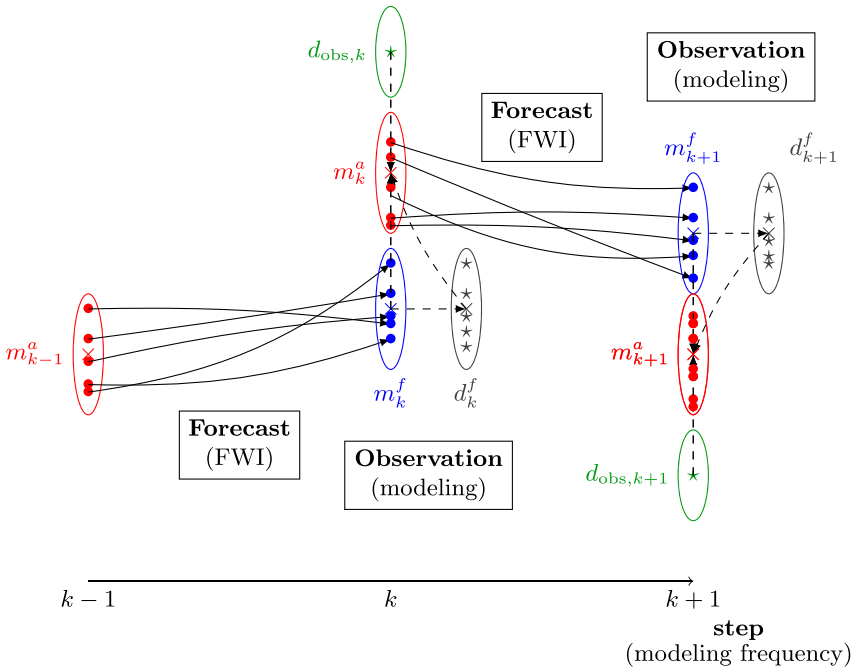

Figure 1. Representation of a ETKF-FWI scheme. Ellipses represent the Gaussian distributed ensembles. The ensembles are discretized by a set of vectors denoted by dots for the state ensemble and by stars for observational ensemble. The forecasted (resp. analysed) state ensemble are denoted in blue (resp. red) and the forecasted (resp. analysed) data ensemble are denoted in grey (resp. green). Figure from Thurin *et al.* (2019).

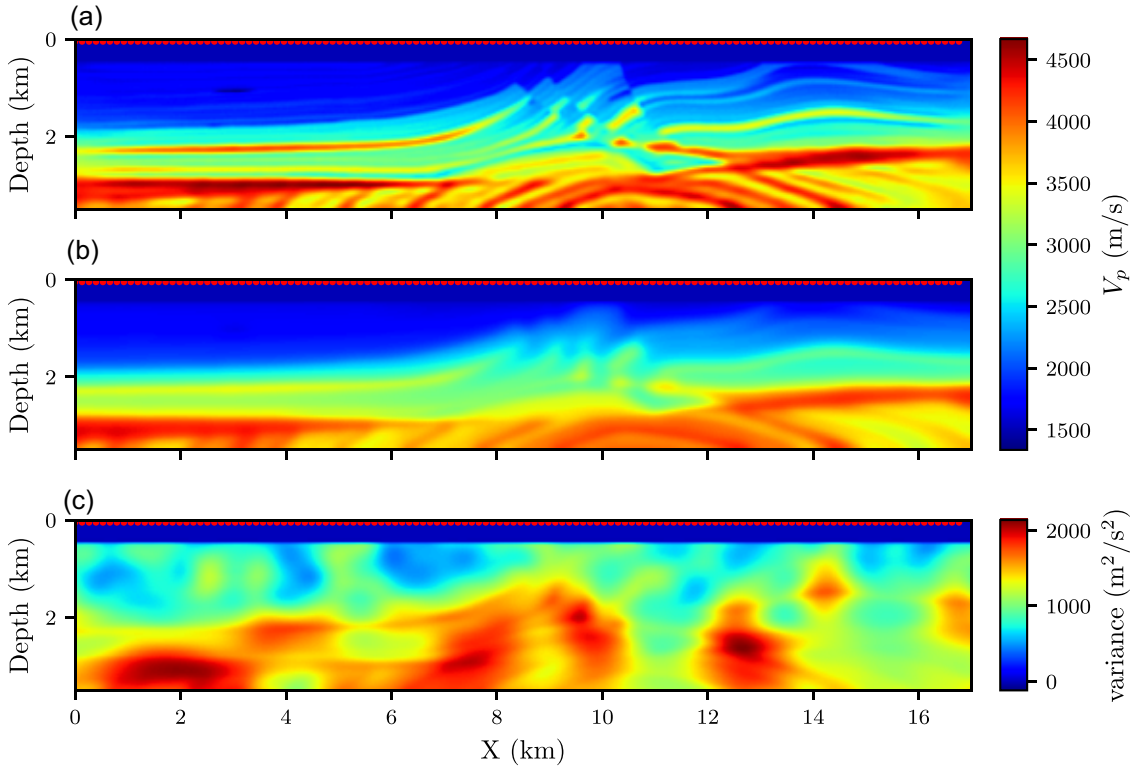

Figure 2. Marmousi II experimental settings. (a) The exact Marmousi II velocity model considered here. The shot positions are displayed as red dots. (b) Initial mean model obtained by smoothing the true velocity model with a Gaussian filter. (c) Initial variance of the model ensemble.

Thurin *et al.* (2019) this number should be enough to reach a correct estimation of the variance and covariance attached to the inversion of this data. We design our initial ensemble of velocity models $m_{f,0}$ in two steps. We first set its mean by applying a Gaussian smoothing filter to the exact Marmousi-II model as is usually done in conventional synthetic FWI experiments. The correlation lengths of the smoothing filter are set to few hundreds of meters in x and z directions. We then add random perturbations following the

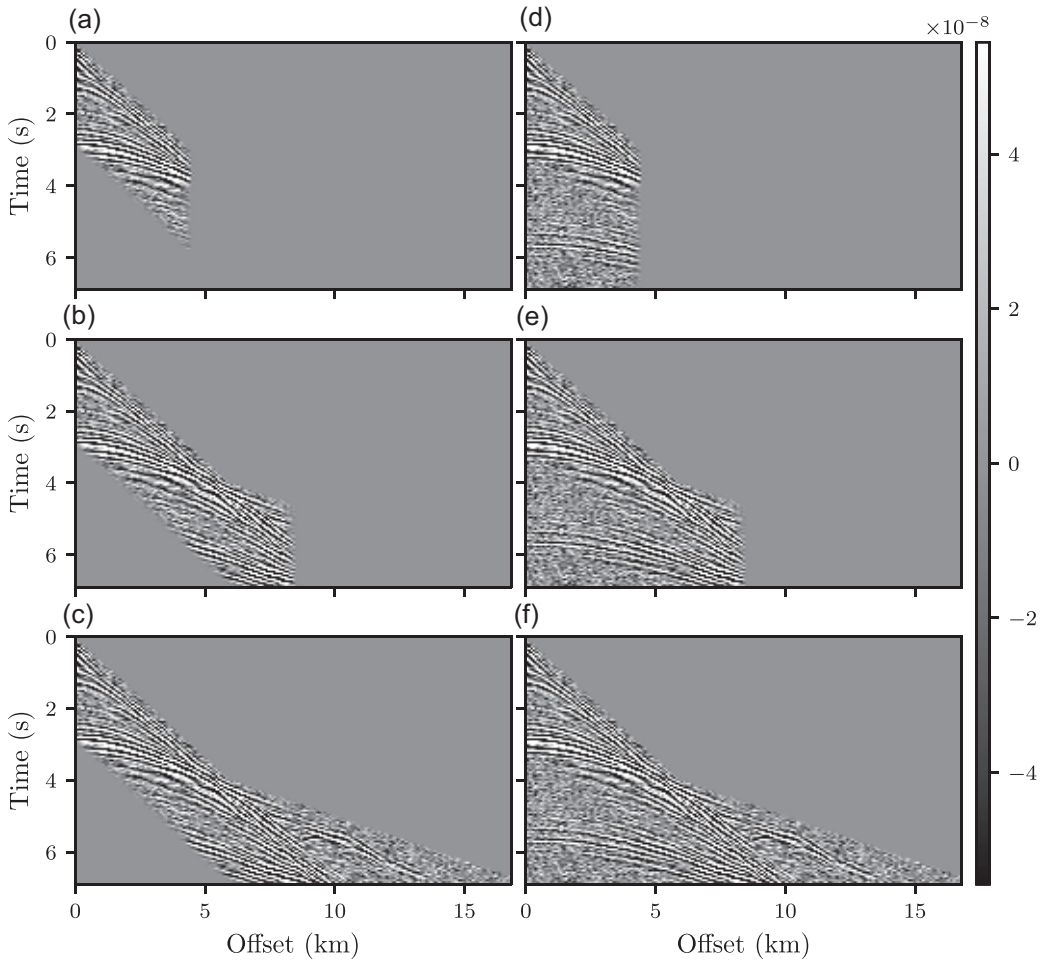

Figure 3. The data set used. We first use a (a–c) 3 s time-window from FAT and then use a (d–f) 7 s time-window. Additionally, for each of our two time-windows, we use an offset windowing of 4 km (a and d), 8 km (b and e) and 17 km (c and f). The time-offset windows are designed such that the last window (f) includes all of our data.

methodology proposed by Thurin *et al.* (2019). More precisely, we add zero-mean normally distributed random perturbations to V_p with a covariance designed such that the spatial extent of the perturbations matches the expected resolution power of FWI in the considered frequency band (Devaney 1984; Wu & Toksöz 1987; Sirgue & Pratt 2004).

The amplitude of the random perturbations must be carefully chosen. If the amplitudes of the perturbations are too small, all our models may converge towards a single model after a few FWI, thus collapsing the ensemble. On the other hand, if the amplitudes are too strong, cycle-skipping may occur in some of our models, making it impossible to satisfy the Gaussian model distribution hypothesis required by the ETKF. This is a major concern since our successive FWI work on a smaller data set, increasing the ill-posedness of the inversion. While we rely here on a visual inspection of specifically chosen shot-gather to control the size and amplitude of the velocity perturbation to avoid cycle-skipping, a more systematic approach would consist in computing first arrival time (FAT) for each shot and make sure that the differences between computed FAT and picked FAT on observed data remains in the limit of cycle skipping (half a period of at the dominant frequency).

We show, in Fig. 2(a) the exact model with the shot position in red dots. The mean model and the variance of the initial ensemble used for our two ETKF-FWI schemes are presented in Figs 2(b) and (c), respectively. Note that the variance is not completely uniform spatially, this is a side effect of using a small ensemble size compared to the 96 021 DoFs in our model space. Note that although the model has 96 021 DoFs, it has a comparatively low number of geological features, such as the high velocity layer at ∼2 km depth. Thus we can anticipate that the number of effective DoFs required to describe the model could be actually much smaller than 96 021.

We first implement a time-offset windowing ETKF-FWI scheme. More precisely we iteratively solve the FWI problem by first considering a narrow window in the time-offset domain and then progressively increasing it to finally consider the entire seismogram. In this context, all the seismograms are considered in the inversion (all the shots).

This category of scheme is linked to what is called ‘layer stripping approach’ (see Shipp & Singh 2002; Wang & Rao 2009; Brossier *et al.* 2009a, for instance). Data contained in shorter offsets and early arrival times constrain the shallow part of the domain while longer offsets and later arrival times constrain deeper parts of the domain. This type of iterative scheme comes handy when the problem is ill-posed, the layered reconstruction acting as a regularization.

We show, in Fig. 3 the time-offset windows used by our ETKF-FWI scheme. We use a sequence of increasingly wide time-offset windows. More precisely, we start our scheme by limiting our data

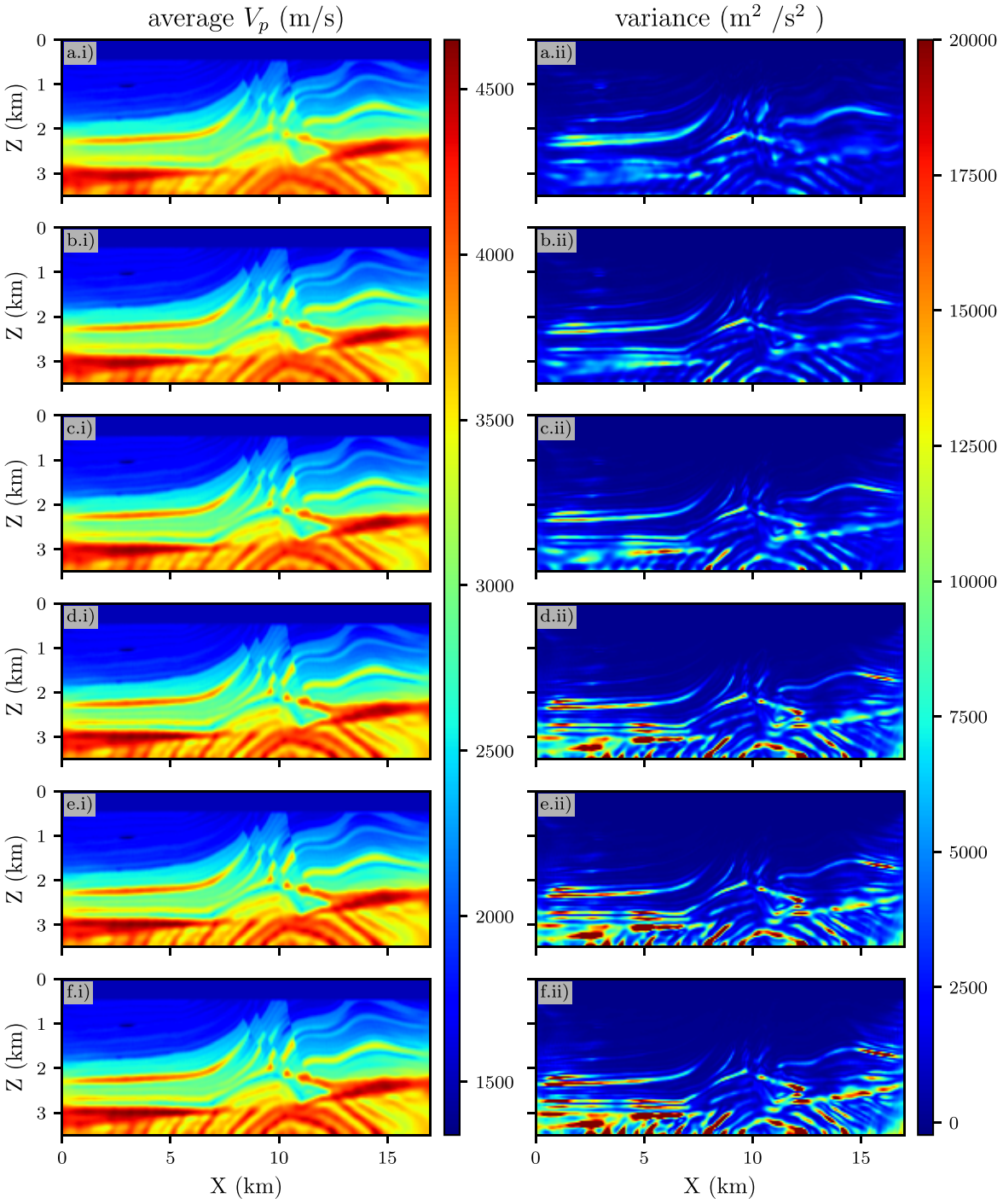

Figure 4. Mean model (left-hand panel) and variance (right-hand panel) obtained after each step of the ETKF-FWI scheme with the time-offset windowing strategy.

set to short arrival time window and short offset range (4 km), which allows us to recover the upper layer of our model. We then gradually increases the offset to 8 km and then to 17 km, which is the maximum offset for our data in this experiment. Increasing the offset range allows us to recover deeper layers of our model. We then include longer time windows but restrict our data set to short offsets range (4 km) and gradually increase again the offset to 8 km and then to 17 km. The final time offset-window includes the whole data set.

We show, in Fig. 4, the mean and variance of the velocity models obtained by our ETKF-FWI scheme. We see that, as expected, the inversion of the first time-offset window makes it possible to reconstruct the shallower part of the model. Increasing the offset and time ranges enables the reconstruction of the deeper parts of the model. The estimation of the variance follows the same path. It tends to concentrate on the bottom part of the model as we progress in our time-offset windowing scheme. We see that our final variance is similar to the one recovered by Thurin *et al.* (2019), with two

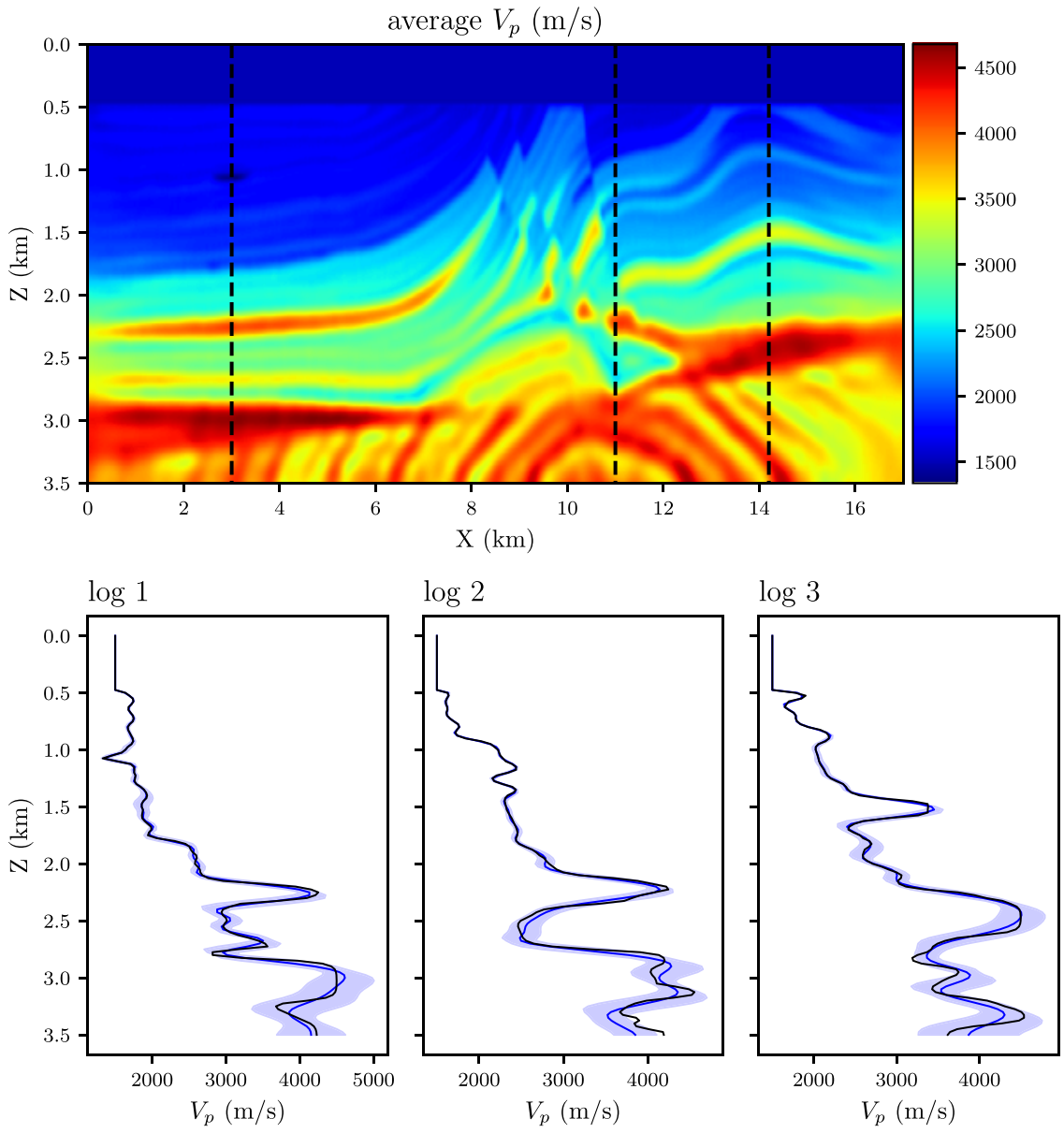

Figure 5. Top panel: mean V_p model obtained with the time-offset windowing ETKF-FWI scheme and position of the three logs displayed as dashed vertical lines. Bottom panels: comparison of the V_p profiles extracted from the mean model (in blue) and the profiles extracted from the true model (in black). Additionally, we show the confidence intervals $\pm 3.5\sigma$, with σ the square root of the computed variance.

features. The first is associated with the illumination of the model and energy spreading of the wave propagation, with higher variance values close to the lateral and bottom boundaries. The second is associated with higher variances values near the high velocity layers where rapid changes of the velocity occur. It was confirmed in Thurin *et al.* (2019) that local peaks of variance are located precisely at the interface and not within the layers. We observe the same behaviour here. Please note that variance in the bottom corners might appear misleadingly low. This is due to the colour scale choice, adapted to the peak of variance occurring at deep interfaces around 3 km depth. The variance in the bottom corner remains significantly larger than the variance in the upper parts of the model which are better sampled by the wavefield and are thus expected to be better constrained.

We then compare, in Fig. 5 three profiles extracted from average model to the corresponding profiles extracted from the true the Marmousi-II model. Additionally, we show the confidence interval $[\bar{V}_p - 3.5\sigma, \bar{V}_p + 3.5\sigma]$, where σ is the square root of the variance computed with our time-windowing ETKF-FWI scheme. This interval contains 99.95 per cent of the models following the PDF computed with our scheme. We see that, while the average inverted model with our time-offset windowing ETKF-FWI scheme and the true model may differ, the true model is approximately in the 3.5σ confidence interval. This means that, when the illumination is sufficient, we can hope that indeed the FWI model is close from the true model. In this case the local uncertainty estimation we perform gives insight also on the deviation from the ground truth. Of course, this well controlled 2-D synthetic case where the ground truth is known, is a particularly favourable case. We will see in the next section that with bandlimited field data, the ground truth is known only from localized well logs and that the FWI model fits only marginally these well logs. The information we get from our

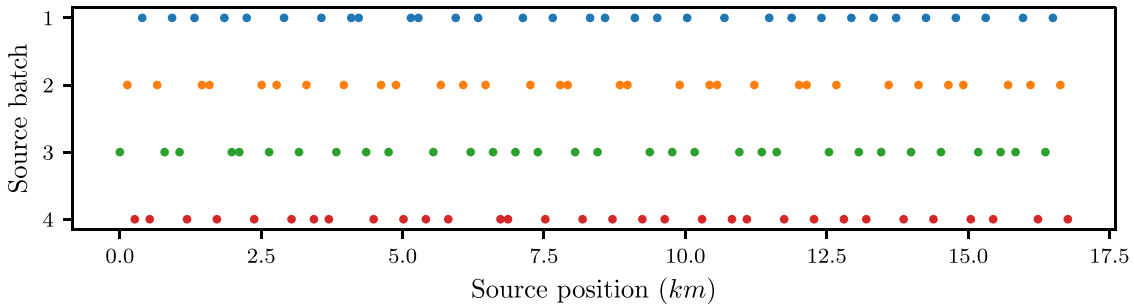

Figure 6. Partition of the sources in four batches for the source-subsampling ETKF-FWI scheme. In blue, the first batch. In orange, the second batch. In green, the third batch. In red the fourth batch. Note that each source belongs to a single batch.

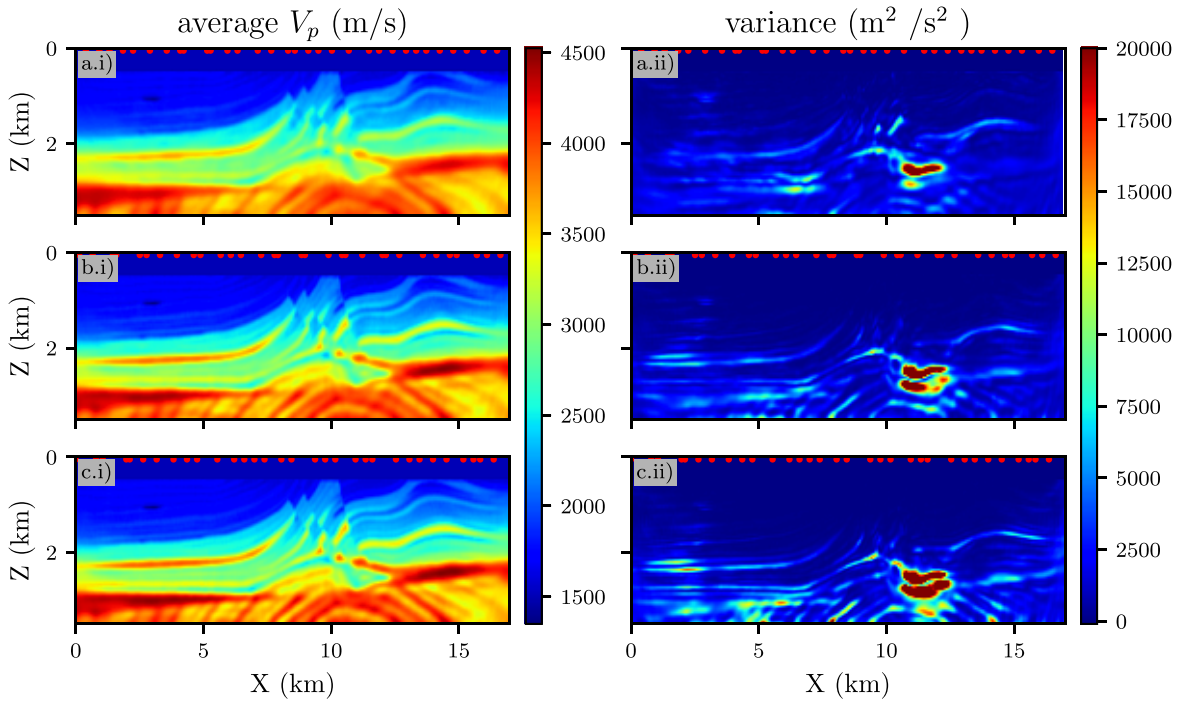

Figure 7. Mean (left-hand panel) and variance (right-hand panel) models obtained after each step of the source-subsampling ETKF-FWI scheme. The red dots indicate the position of the sources at each step of the process.

uncertainty quantification scheme thus will remain in this case only local.

Such scheme requires the same computational cost for each individual FWI, making it especially challenging for large scale problems. We are thus interested to implement another scheme based on source decimation in order to propose a cheaper ETKF-FWI scheme able to consider larger 3-D problems.

We implement a source-subsampling ETKF-FWI scheme, which, instead of using increasingly wide time-offset windows, utilizes, for each iteration a distinct set of shot gathers, thus reducing the computational cost of each individual FWI. Such source subsampling strategy is standard in FWI to decrease the computational cost of large 3-D applications (Warner *et al.* 2013). We present, in Fig. 6 the source batches used in our ETKF-FWI scheme. The sources are randomly selected. The resulting uneven illumination is compensated along the iterations of FWI by going through all the batches to finally account for all the sources. We present, in Fig. 7 the mean and variance of the velocity models obtained by our source subsampling ETKF-FWI scheme. We see between 10 and 12 km offset and around 2.5 km depth, an area with a noticeable variance, which corresponds to a part of the model with stark velocity change in both depth and offset, making this area difficult to properly recover.

Again, further insight can be gained by comparing profiles extracted from average model to the corresponding profiles extracted from the true the Marmousi-II model. We see in Fig. 8, the confidence intervals $[\bar{V}_p - 3.5\sigma, \bar{V}_p + 3.5\sigma]$ computed with our source subsampling ETKF-FWI scheme. We see, in the shallower parts of the model, similar intervals to the one obtained with the time-windowing scheme. However, below 2 km depth we start seeing differences. First, the confidence intervals do not always include the true model. This is especially the case when the true model exhibit strong variation. Moreover, we see, on the second log, around 2.5 km depth an outstandingly large confidence interval. Looking at the model distribution here, we see a bi-modal distribution, which means that the convergence towards two different local minima for this particular voxel has occurred and the confidence intervals and variance estimated here loses their meaning.

It is worth mentioning that the variance map obtained with the time-window based ETKF-FWI scheme do not exhibit such a high variance in this area, even-though both scheme started from the

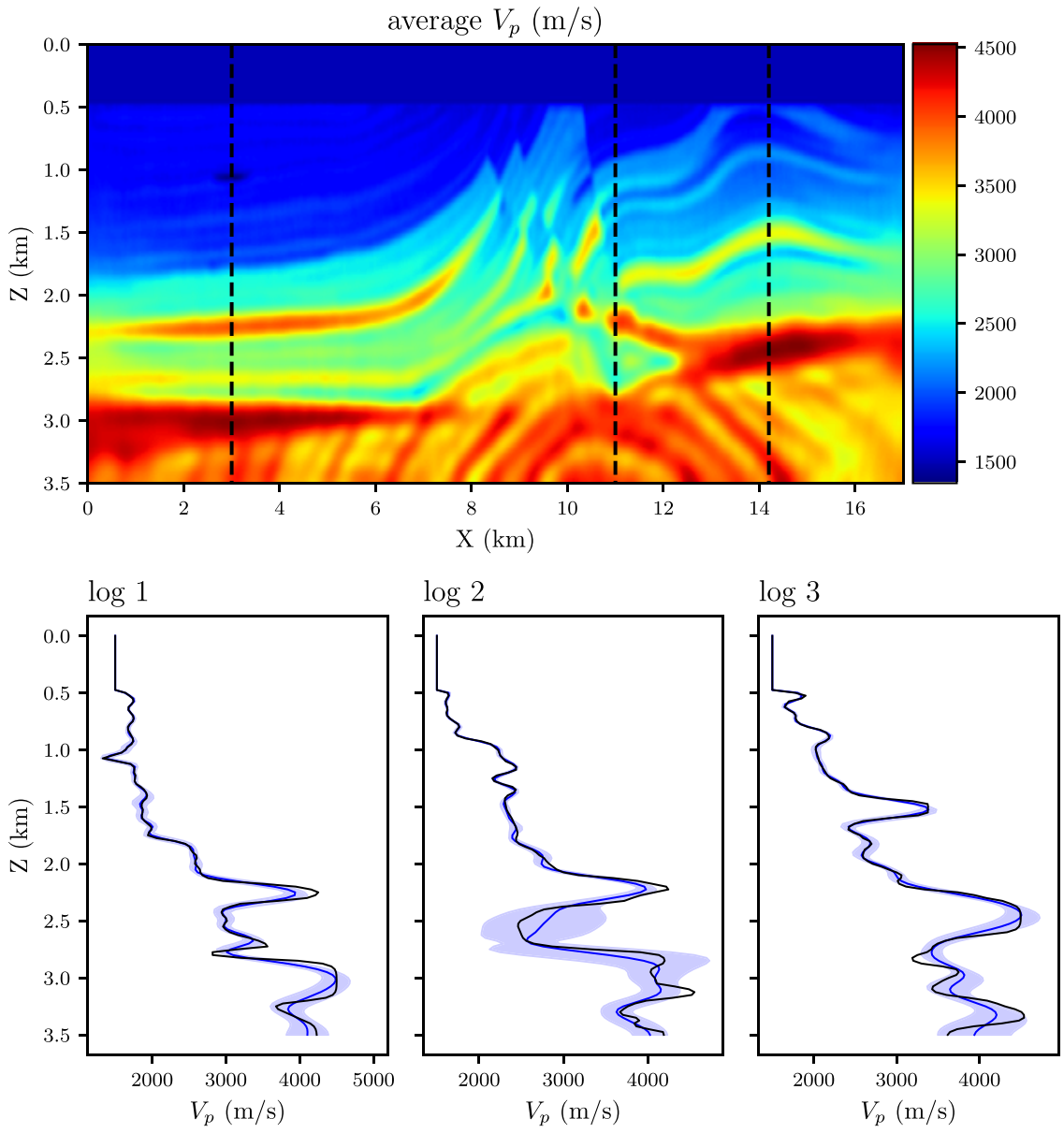

Figure 8. Top panel: mean V_p model obtained by the source subsampling ETKF-FWI scheme and position of the three logs displayed as dashed vertical lines. Bottom panels: comparison of the V_p profiles extracted from the mean model (in blue) and the profiles extracted from the true model (in black). Additionally, we show the confidence intervals $\pm 3.5\sigma$, with σ the square root of the computed variance.

same initial ensemble. This is due to the regularization property of time-offset windowing based FWI schemes. More precisely, by first recovering the shallower parts of the model, the time-windowing scheme is less sensitive to the random perturbations located in the deeper parts of our initial model.

We conclude from this experiment that the variance maps obtained by a specific ETKF-FWI scheme is dependent on said scheme. As an example, here, the variance estimated with a ETKF-FWI scheme based on time-offset windowing appears smaller than when it is based on a source-subsampling approach. This reflects the fact that time-offset windowing strategies makes use of more redundant information, yielding a "stabilizing" effect on the inverse problem, thus reducing the variance of the FWI workflow based on such time-offset windowing. On the contrary, a FWI scheme based on source subsampling show less data redundancy as the source group design is mutually exclusive. We could hope that by taking more groups, with potential overlaps in the source selection, could help reduce the variance of the process.

Moreover, the source subsampling scheme drastically reduces the computational cost of the forecast operator by diminishing the number of simultaneously inverted sources, allowing us to run more FWI simultaneously on the high performance computing (HPC) facilities we have access to. This is especially important if we want to run our scheme with larger ensembles or if we want to tackle standard industrial 3-D problems. Until now, we only considered a small 2-D problem with 96 021 DoFs and 128 sources, allowing us to perform a FWI in a few minutes on a single node of our local HPC facility. However, the field data FWI we consider in the next section involves more than 20 times more DoFs and more than 15 times more sources, making it difficult to fit a single FWI on

our local HPC facility. A source subsampling strategy is thus very appealing to fit our ensemble-based FWI on our local HPC facility and this is the one we select.

4 APPLICATION TO A NORTH SEA OBC DATA

4.1 Environment data and pre-processing

The data we consider in this study has been acquired in the North Sea, offshore Norway. It is a shallow water oil and gas field, with nearly constant water depth of 70 m. The target is composed of an anticlinal in chalk in the upper Cretaceous Hod and Tor formations, at a depth of approximately 2400 m, and of tertiary sediments above it (Sirgue *et al.* 2009, 2010; Prieux *et al.* 2011, 2013; Operto *et al.* 2015; Pladys *et al.* 2019). The data set we consider has been recorded by a 4-components OBC device. A total of 2048 receivers are deployed on the seabed along twelve cables which cover a surface of 145 km^2 with an inline spacing of 50 m and a cable spacing of 300 m. 50 824 shots located 5 m below the sea surface have been recorded. In this work, we only use the pressure component. This is the same data set as the one used in previous studies from Kamath *et al.* (2021) and Pladys *et al.* (2022). We benefit from these studies regarding both the preprocessing of the data and the setting of the FWI workflow regarding parametrization, mesh design, regularization and pre-conditioning.

The pre-processing applied to the data is minimal. First, source–receiver reciprocity is applied to reduce the global cost of FWI (the number of considered sources for modelling and inversion goes from 50 824 to 2048). After despiking the data, we compute the root mean square (RMS) amplitude of each gather. Gathers with large values of RMS amplitude are considered faulty and thus removed: four of them are excluded from the 2048 available gathers. We apply a minimum-phase band-pass filter to restrict our data to the frequency band 2.5–5 Hz. A time decimation from Δt =4 ms to Δt =8 ms is also applied. A muting window is applied to remove Schölte waves identifiable with a linear velocity of approximately 400 m s^{-1}: as we are going to invert the data in the acoustic approximation, these interface waves propagating along the seabed cannot be predicted, and are therefore considered as noise in the inversion.

4.2 FWI settings

The first application of FWI on this data set used an isotropic acoustic modelling engine (Sirgue *et al.* 2010). However, an acoustic vertical transverse isotropic (VTI) modelling has been rapidly proposed to avoid bias in the velocity inversion (Prieux *et al.* 2011). Accounting for attenuation has also been shown crucial to correctly model the data in this environment containing both unconsolidated sediment and low velocity layers (Operto *et al.* 2015; Operto & Miniussi 2018; Kamath *et al.* 2021). We therefore perform the inversion of this data set in the 3-D visco-acoustic VTI approximation, relying as previously on the TOYxDAC_TIME full waveform modelling and inversion code (Yang *et al.* 2018). We use as modelling parameters the vertical *P*-wave velocity V_p, the density ρ, the quality factor Q_p and the two Thomsen's parameters ϵ and δ. In this study, we only invert for V_p and keep the other parameters constant (passive parameters).

The initial V_p model, as well as the ϵ and δ models we use are provided courtesy of AkerBP. The V_p model has been obtained by reflection traveltime tomography and has been proven sufficiently accurate to allow for convergence starting from a 2.5–5 Hz frequency band (Operto *et al.* 2015; Operto & Miniussi 2018; Kamath *et al.* 2021; Pladys *et al.* 2022). The density model ρ is computed from this initial V_p model through the Gardner's law

$$\rho = 309.6 V_p^{0.25}. \tag{21}$$

For the attenuation, it was first shown by Operto *et al.* (2015) and further studied by Kamath *et al.* (2021) that a homogeneous quality factor Q_p set to 200 in the whole domain (except in the water layer where it is set to 1000) is sufficient to capture most of the attenuation and dispersion effects induced by the viscosity of the medium in the 2.5–5 Hz frequency band. Finally, we estimate a source wavelet common to all the gathers, using the frequency domain approach from Pratt (1999), from 128 randomly selected gathers.

The inversion scheme is based on a subsampling strategy where we split the initial ensemble of shots in 15 batches of 128 sources + 1 batch of 124 sources, following the strategy depicted in eq. (19). For each batch of sources 3 iterations of *l*-BFGS are performed, therefore in total a whole inversion represents 48 *l*-BFGS iterations with 128 sources (124 sources for the final 3 iterations). It is shown in Kamath *et al.* (2021) that this strategy guarantees the convergence towards a model similar to what would be obtained by inverting simultaneously for the 2044 sources for the same number of iterations, therefore reducing the computational cost by approximately a factor 16.

At each iteration the gradient is smoothed using the same non-stationary Gaussian filter we use for the Marmousi II test case. Here the reference frequency f_0 is chosen as f_0=3.75 Hz. The correlation lengths are chosen equal to $0.4 \times \lambda_{\text{dom}}$, $0.4 \times \lambda_{\text{dom}}$ and $0.3 \times \lambda_{\text{dom}}$, respectively in the x, y and z directions, where λ_{dom} is defined in eq. (20).

We also use a preconditioning strategy. The gradient is multiplied by an approximation of the inverse Hessian to speed-up the convergence of the minimization algorithm. In practice, this preconditioning helps to remove the strong imprint of the acquisition by rebalancing the energy in the gradient so as to amplify perturbation at depth and decrease the gradient values close from the acquisition, in the shallow part. This is achieved based on wavefield information: the strategy is fully described in Kamath *et al.* (2021) and introduced as wavefield pre-conditioner.

Additionally, we incorporate the post-processing proposed by Pladys *et al.* (2022) into our FWI scheme. More precisely, we precompute a stencil based on the acquisition shape. The part of the gradient outside of the stencil is extrapolated radially from the parts of the gradients located within the stencil using a nearest-neighbour strategy. This extrapolation aims at removing the areas on the edge of the model that, because of the lack of illumination, are weakly updated by FWI, and end up in creating artificial reflectors surrounding the illuminated zone. This kind of *ad hoc* strategy is complemented with a test to ensure that the smoothed, pre-conditioned and extrapolated gradient remains a descent direction (we test if the scalar product between the gradient and the post-processed gradient is still positive).

Now that the subsampling FWI scheme is presented, we move to uncertainty quantification using our ETKF-FWI strategy.

4.3 Initial ensemble design

We generate an initial ensemble m_0 following the methodology proposed by Thurin *et al.* (2019), similarly to what we have performed for the 2-D synthetic Marmousi II case study. Both the initial model

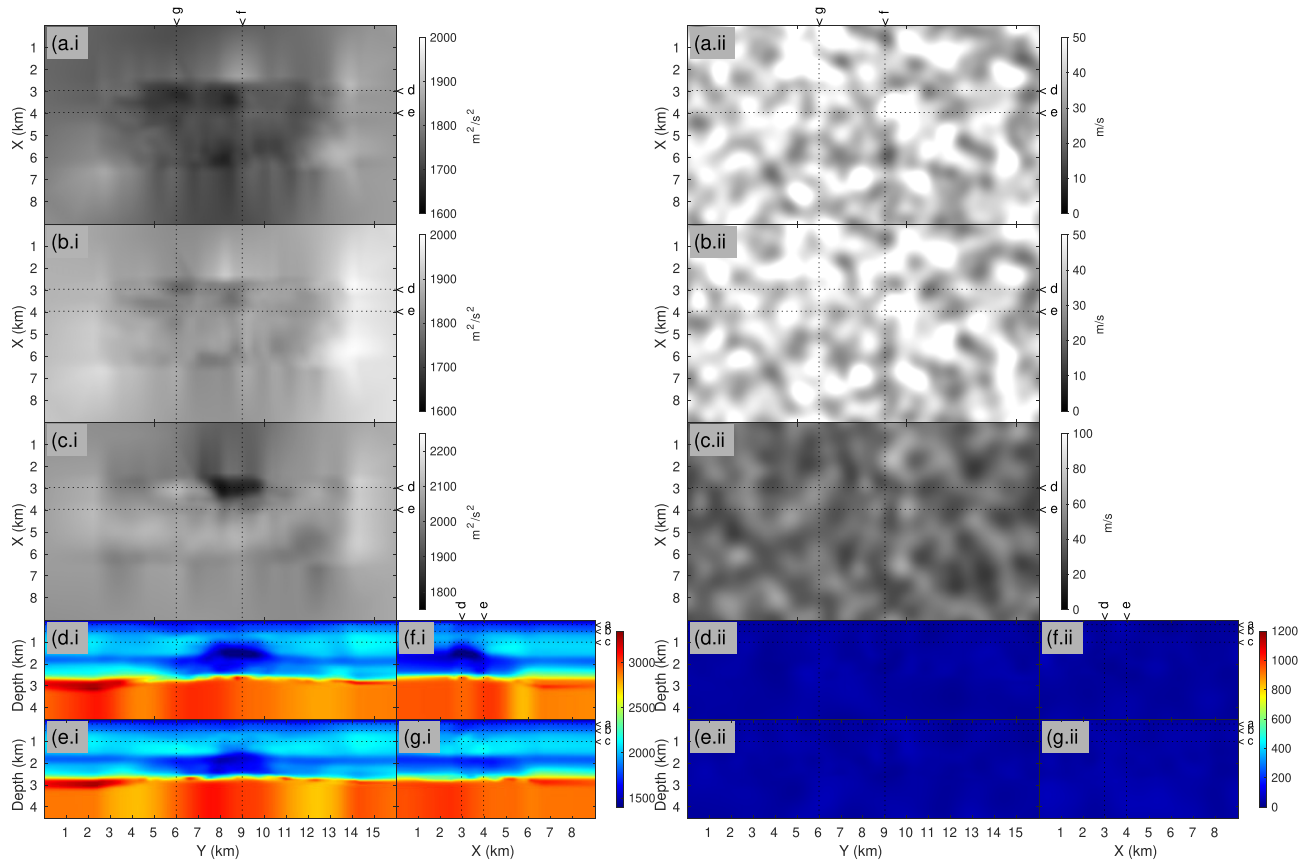

Figure 9. Left-hand panel: sections of the average initial model obtained by reflection tomography. Right-hand panel: sections of the variance of the initial normally distributed random models. (a–c) Horizontal sections at (a) 200 m depth, (b) 500 m depth and (c) 1 km depth. (d and e) Inline vertical sections for (d) $x = 2.95$ km and (e) $x = 3.95$ km. (f and g) Crossline vertical sections at (f) $y = 9$ km and (g) $y = 6$ km. Note that the depth sections use a greyscale colour-map with two different velocity and variance ranges, whereas the vertical sections use a 'jet' type colour-map with a fixed velocity and variance range for all sections.

mean and the diagonal of the initial covariance matrix (initial variance) are displayed in Fig. 9. As discussed previously, the initial ensemble design is a crucial step for ETKF-FWI. The diversity of the ensemble needs to be sufficient to avoid an ensemble collapse, which occur when all the models converge to the same point in the model space. However, the perturbations added to the mean model should be sufficiently small so that all the models sample the same global minimum valley.

To generate this initial ensemble, we therefore rely on trial and error, and use as quality check the computation of the first arrival travel time within several of the ensemble members. We ensure that the time shifts induced on the first-arrival travel time by the velocity perturbation added to the initial model are smaller than half a reference period. This reference period is computed by considering the higher end of the frequency-band in which the inverted data are filtered. This makes sure cycle skipping will not occur during FWI starting from any of the members of the ensemble.

Another *a posteriori* check consists in looking at the distribution of the ensemble for randomly selected points in the model space. The distribution should resemble a Gaussian and in any case be unimodal. A multimodal distribution is indicative of cycle skipping and/or violation of the Gaussian assumption behind the ETKF formalism.

Until now we have not discussed the critical question of the size of the ensemble and of the quality of the estimated variance and covariance resulting from our ETKF-FWI. If our ensemble is too small we are at risk of both underestimating the variance of our models and to find spurious correlation within our models. In the next section, we are going to asses the quality of the result obtained by studying how certain feature, such as variance estimation, model distribution and correlation map in our models evolves with the size of the ensemble N. Observing some sort of 'convergence' of theses features for $N \geq 50$ indicates that it should be possible to interpret the results obtained by the ETKF-FWI scheme for this size of ensemble.

4.4 Convergence of the ETKF scheme

We asses the quality of a reference ensemble with $N = 50$ models by comparing it to a 'coarse' ensemble with only $N = 10$ models and a 'fine' ensemble with $N = 200$ models. We want to see how certain features, such as average, variance and model distribution vary with the size of the ensemble.

We first show, in Fig. 10 both average model and model variances computed as well as the model distribution at a certain voxel within the low-velocity anomaly for $N = 10$, 50 and 200. We do not see any difference in the model average. However, we see that for

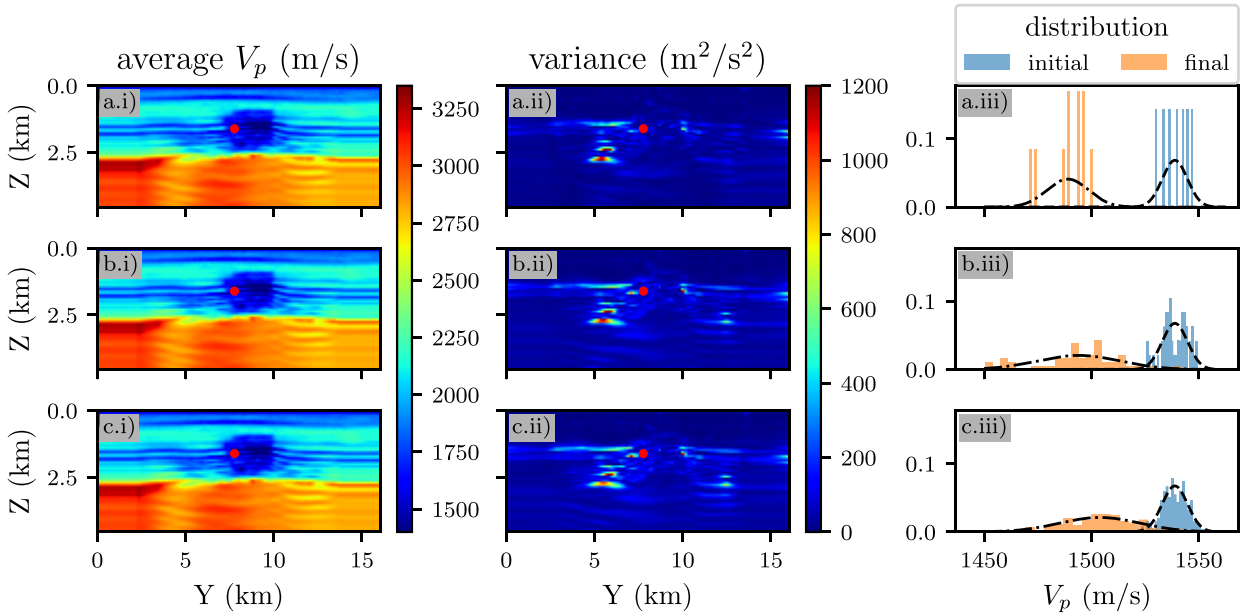

Figure 10. (a.i–c.i) Final mean V_p model obtained by ETKF-FWI, (a.ii–c.ii) final variance, (a.iii–c.iii) distribution of the models at a given point marked in red. The results are obtained with three ensemble sizes N: (a) $N = 10$, (b) $N = 50$ and (c) $N = 200$.

$N = 10$, the variance is underestimated, especially at depth below 2.5 km and within the low-velocity anomaly. The two ensembles with $N = 50$ and $N = 200$ yield similar variances. More interestingly, looking at the final model distribution within the low-velocity anomaly, we see a stark contrast between the ensembles with $N = 50$ and $N = 200$ models. We see, for the ensemble with 50 models, what could be interpreted, either as a bimodal distribution with a first peak around 1460 m s^{-1} and a second peak around 1500 m s^{-1}, or as a poor sampling of the Gaussian distribution with an average of ≈1500 m s^{-1}. The former hypothesis would make it impossible to properly analyse our results as it would violate a fundamental hypothesis of the ETKF, and potentially indicate some cycle skipping occurred during the successive FWI. However, the distribution obtained with the ensemble with 200 models appears to be Gaussian, indicating the latter hypothesis to be valid.

Following up theses observations, we then show, in Fig. 11, the model distribution of our three ensembles along a vertical profile for which a sonic log is available. Once again, we see that the velocity models from our coarse ensemble are not as spread as the velocity models from our reference and fine ensembles. This is especially true when the V_p models exhibit sudden changes. We come back later on the interpretation of this spread and the estimated variance.

Using a higher number of models always allows for a better approximation of the covariance matrices. However, like for other low-rank approximation techniques such as SVD, a certain 'cut-off' has to be selected in order to reduce the computational burden of our method. The results we show indicate the variance do not drastically change when quadrupling the number of models from $N = 50$ to $N = 200$. We can thus assume our reference model ($N = 50$) provides already a meaningful estimate of both mean model and variance.

We now want to study how N influences the off-diagonal terms of the *correlation matrix*. This is of particular interest as the correlation matrix can be used to gain understanding of the interdependence between a given DoF and the other DoFs of the velocity model, thus allowing one to characterize the local resolution of the FWI (Thurin *et al.* 2019).

The correlation matrix is obtained by scaling the covariance matrix, making it dimensionless ranging from −1 to 1 (Feller 2008, chapter 3):

$$C = \mathrm{diag}\,(P)^{-\frac{1}{2}}\, P\, \mathrm{diag}\,(P)^{-\frac{1}{2}}\,. \tag{22}$$

A correlation coefficient close to 1 indicates a strong positive link between two parameters, implying they evolve in a similar manner. On the other hand, a coefficient close to −1 indicates a strong negative link between the two parameters, implying they evolves in an opposite manner. A coefficient close to 0 indicates the absence of link between the two parameters. We show, in Fig. 12, correlation maps between a certain voxel of interest near the low-velocity anomaly and the rest of the velocity model. We observe that our coarse ensemble exhibits long range correlations that are not present in our reference and fine model. This leads us to believe that theses are spurious correlations, caused by the low rank approximation of the correlation matrix. We can also see that the correlation maps obtained both from the reference ensemble and from the fine ensemble are much better localized. One can still observe some spurious correlations in our reference ensemble with $N = 50$ members, but these are localized and far from the voxel of interest. The resulting correlation maps are still interpretable as long as we restrict our analysis to neighbouring voxels.

We conclude that our reference ensemble ($N = 50$) provides good enough results to be interpreted and is thus a good trade-off between computational cost and quality of the final ensemble. In the remainder, we interpret the results obtained with $N = 50$.

4.5 Mean model obtained by ETKF-FWI

The mean model obtained with our ETKF-FWI scheme is displayed in Fig. 13. We can see that the mean velocity model resembles the

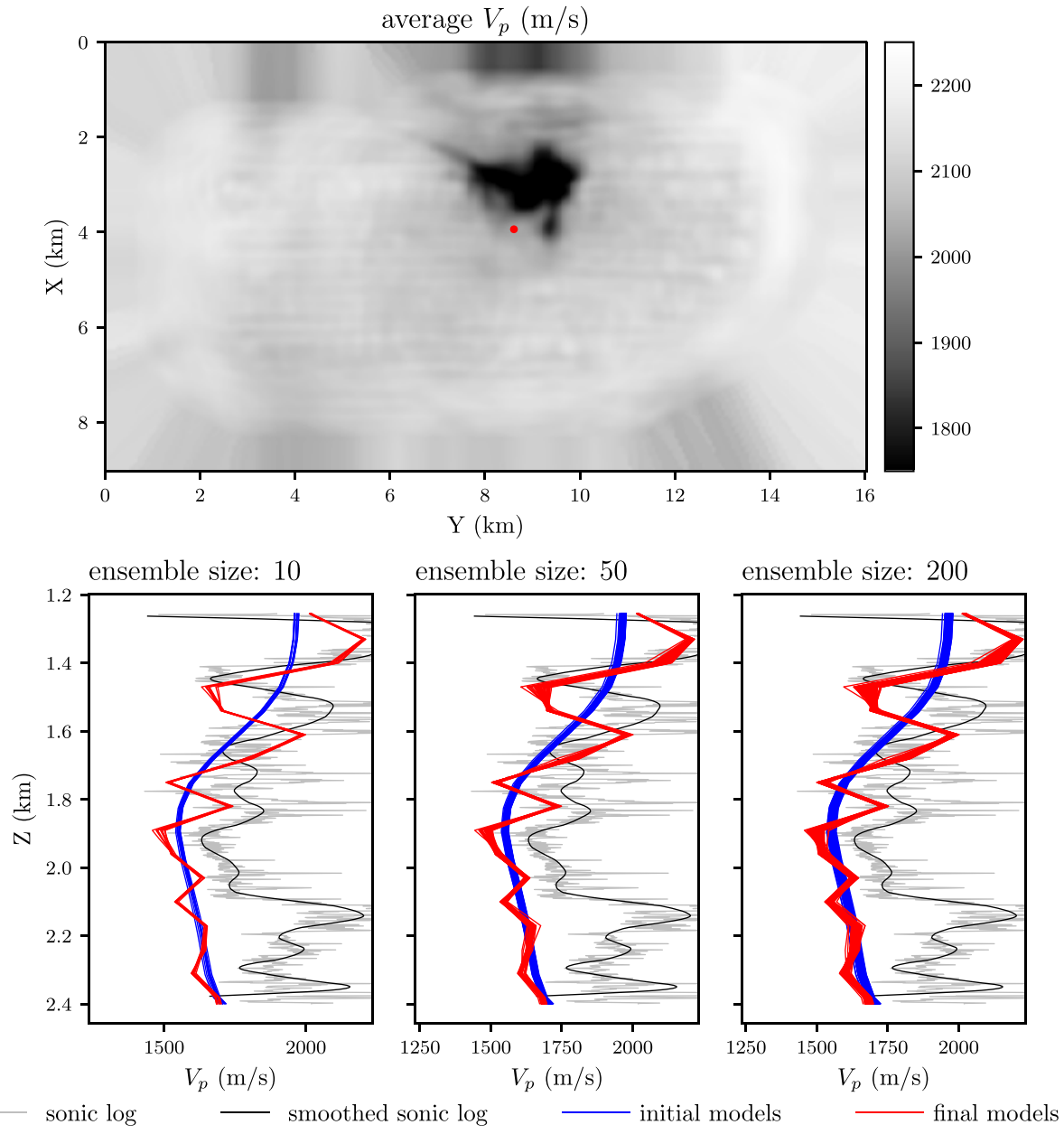

Figure 11. Top panel: the positions of the V_p sonic log of interest overlapped with an horizontal section at 1 km depth of the mean V_p model obtained by ETKF-FWI for $N = 50$. Bottom panels: comparison of V_p profiles extracted from the mean model (in red), the sonic log (in grey) and the smoothed sonic log filtered in the 0–12 Hz frequency band (in black)n for three ensemble sizes, $N = 10$, 50 and 200.

inverted velocity model that can be found in the literature (Sirgue *et al.* 2010; Operto *et al.* 2015; Kamath *et al.* 2021; Pladys *et al.* 2022). More precisely, we can see features such as a shallow channel at 200 m depth, scrapes presumably left by drifting icebergs at 500 m (Sirgue *et al.* 2010) and at 1 km depth, a low-velocity anomaly. This low-velocity anomaly can be visualized more in details in sections (d–g).

In order to further confirm the quality of our mean model, we compare the synthetic data recorded along two cables to the field data. We show, in Fig. 14 the data-fit for two common-receiver gathers in the initial tomography model and in the mean model. Similarly as what is observed in Kamath *et al.* (2021) and Pladys *et al.* (2022), we can see that the data fit from the initial to the final mean model is greatly improved. The transmitted events are fit even at large offsets and for arrival times until 1.5 s from the first arrival time. The reflected energy is also mostly correctly fit, apart from late reflections at short offsets. The visco-acoustic approximation with a fixed smooth density might explain why these events remain not predicted. Further assessment of our mean model can be done by comparison with sonic logs. We show in Fig. 15 both the positions of the wells in which the logs where drilled and a comparison between our model and the logs. Theses results are consistent with results found in the literature which can confirm that ETKF-FWI can yield a quality mean model, as long as the initial ensemble has a good

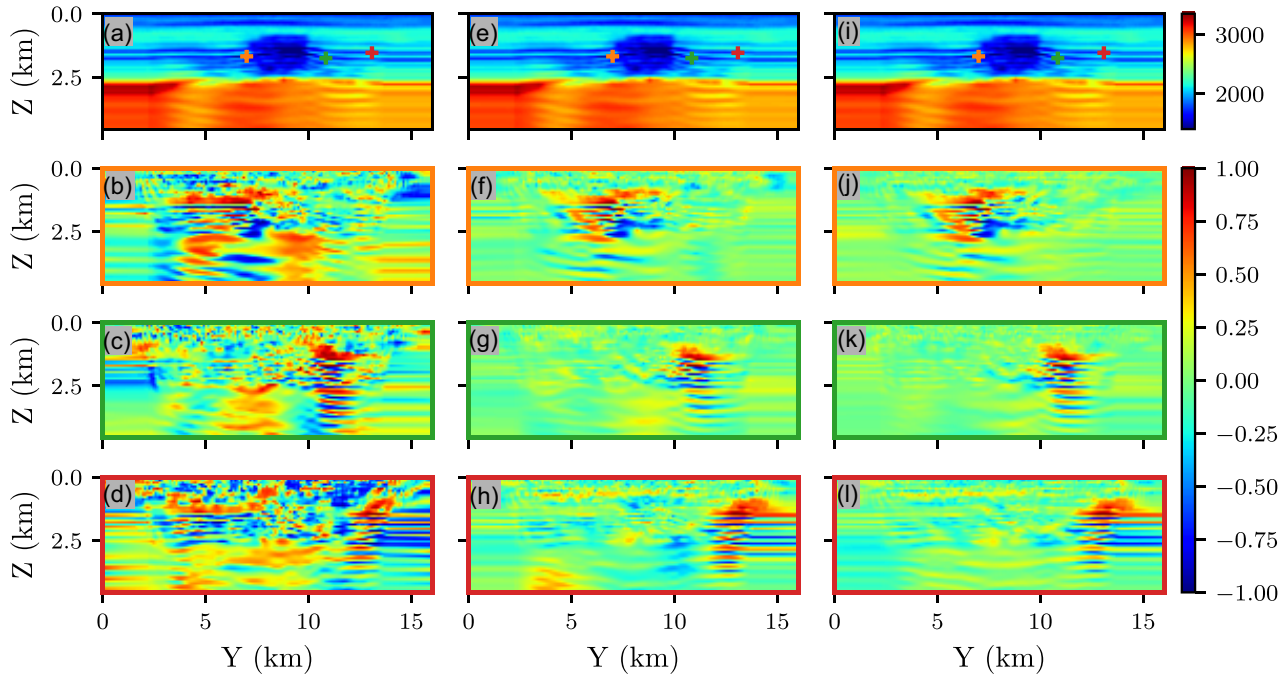

Figure 12. For different ensemble sizes, we show: (a, e and i) inline vertical sections of the final mean model obtained with ETKF-FWI for $x = 2.95$ km, lines of the correlation matrix corresponding to the points in orange, green and red boxes for (b–d) $N = 10$, (f–h) $N = 50$, (i-l) $N = 200$.

enough initial mean and the initial models of the ensemble are not too far from the initial mean to avoid cycle-skipping. But the mean model is only one output from ETKF-FWI. Now we analyse the uncertainty attached to the reconstruction of this model thanks to the ensemble statistics.

4.6 Variance estimation by ETKF-FWI

4.6.1 Variance maps

We first compute the variance of our ensemble, which can be mapped onto the physical space in order to give us an idea of how much our V_p models differs from each other. The variance is measured in m^2s^{-2} and gives a measure of how a randomly sampled model will differ from the mean model.

We display, in Fig. 16, the horizontal sections of the variance. We can see, as expected, the acquisition footprint at 200 m depth in the form of highly localized variance peaks. Moreover, we see at 500 m depth, two zones with higher variance, the first one is between $x = 3$ km to $x = 4$ km and $y = 6$ km to $y = 9$ km and the second one just below the largest scrape seen in Fig. 13(a) at $x \sim 7$ km and $y \sim 4$ km.

Additionally, at 1 km depth, we can see two very high variance zones on top of the section. The first one is located at $x < 2$ km and y between 7 and 11 km and the second at $x \leq 1.5$ km and y between 3 and 5 km. These high variance zones highlight a strong variability of the reconstruction of the velocity in these zones. This might be induced by the fact that these two areas are in the limit of the illumination zone. Also, the initial tomography model is extrapolated in these zones (as can be seen in Fig. 9), which indicates that it is probably far from correct values there. These two factors explain the larger diversity of the model values in these two zones, yielding a stronger variance. This is an indication that the low velocity value appearing on the mean velocity model on top of the depth section at 1 km (*cf.* Fig. 13c) should be interpreted with care.

We show, in Fig. 17, both inline and crossline vertical sections of the variance. We see, for the inline vertical sections (a–b), that high variances are mostly distributed along vertical velocity contrasts, especially on top of the two low velocity layers near the low-velocity anomaly. This third low velocity layer at ∼1.25 km depth seems to be difficult to properly characterize, at least with the resolution attached to the inversion of 5 Hz data. Moreover, we see that the interface below punctually exhibits high variances, notably at depth ∼2.75 km for $x = 2.95$ km (section a) and y between 4 and 7 km. Theses local high variances indicate that it is difficult to properly characterize the interface between the low-velocity anomaly and below. Nevertheless, the low-velocity anomaly exhibits a fairly low variance, aside from some very localized peaks at the interface, indicating it is well characterized by FWI.

Similar observations can be made on the crossline vertical sections (c–d) within the illuminated area. However, the boundary of the illuminated area is more visible, especially for $y = 6$ km (d) and $x \leq 1.5$ km. This is in part due to the lack of illumination, which leads to an unconstrained model, in theses areas. However, this alone does not explain the stark difference between the $x \leq 1.5$ km and $x \geq 8$ km for $y = 6$ km (d).

This confirms what we observe in the horizontal sections (*cf.* Fig. 16c) and indicates that some parts where the initial model was extrapolated might be too far from the ground truth, leading to cycle skipping occurring, at least for some models.

We can get a better idea of how our variances changes according to geological features by looking at the locations of local variance maximums. We thus also show, in Fig. 17, the maximum variance peaks extracted with a maximum filter of radius 500 m. First, we can

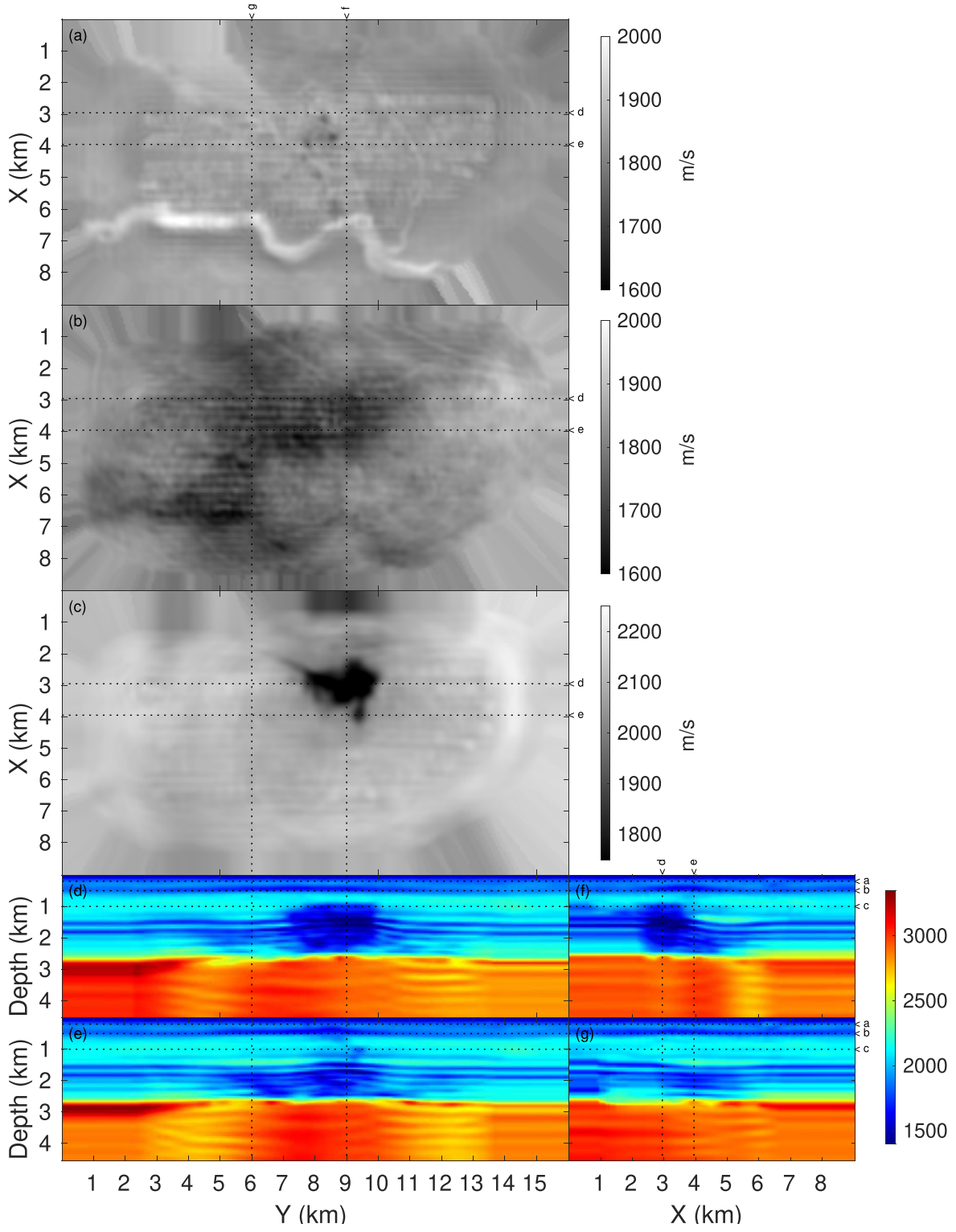

Figure 13. Slices of the final mean model obtained by ETKF-FWI. (a–c) Horizontal sections at (a) 200 m depth, (b) 500 m depth and (c) 1 km depth. (d and e) Inline vertical sections for (d) $x = 2.95$ km and (e) $x = 3.95$ km. (f and g) Crossline vertical sections at (f) $y = 9$ km and (g) $y = 6$ km.

see the variance peaks are focused on the two low velocity layers located in the shallower parts of the model. We also see peaks being distributed along the low velocity layers at ∼2 km depth. We can also see several of variance peaks within the low-velocity anomaly. More precisely, the peaks are mostly distributed along the features present within the low-velocity anomaly, indicating said features are difficult to properly recover. Finally, we see a lot of variance peaks along the layers within the reservoir, which is expected due to both the depth and the lack of illumination of this part of the model.

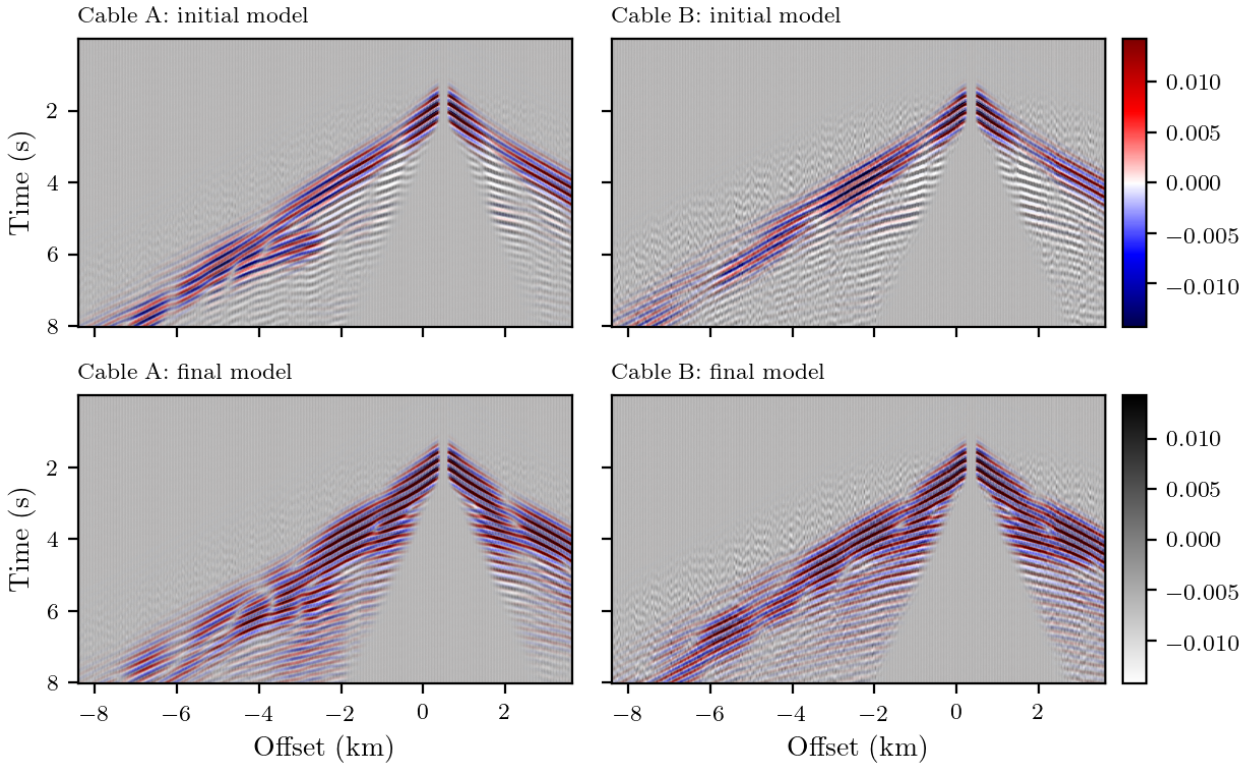

Figure 14. Common-receiver gathers along two cables aligned with the y-axis at (left-hand panel) $x = 2.95$ km and (right-hand panel) $x = 5.5$ km. The first cable crosses the low-velocity anomaly while the second is adjacent to it. The synthetic data, appearing in blue/white/red colour scale, is overlapped with the field data, in greyscale with transparency. On the top row, the synthetic data is computed in the initial tomography model. On the bottom row, the synthetic data is computed in the final mean model.

4.6.2 Model distribution

A variance map gives us an idea of where and how much each model differ from each other. However, it does not gives us an idea of how the models are distributed and what is their actual range in $\mathrm{m\,s^{-1}}$. We first compare, in Fig. 18, our various models to the sonic logs. On the first log, which is far from the low-velocity anomaly, we see that our model distribution tends to spread below 1.5 km depth as the V_p models exhibit sudden changes. However, our models remain relatively close to the mean model, even when the fit is relatively bad. On the other hand, we see on the second log, which is located near the low-velocity anomaly, that the models tend to disagree more, especially around 1.5 km depth, where we see a difference between the largest and smallest V_p values, of $\sim$100 $\mathrm{ms^{-1}}$. By looking at the section (c) of the average model and of the model variance, we see that this part of the model correspond to the boundary of the low-velocity anomaly (*cf.* Fig. 13), which exhibited a relatively high variance (*cf.* Fig. 16).

Finally, as previously mentioned, the third log is near the boundary of the acquisition, making it difficult to properly analyse. Our models remains relatively close to the average model between 0.5 and 1.5 km depth and spread below 1.5 km depth but it is unclear weather it is due to the model being extrapolated from a well illuminated area with little variance or if its because the variance is actually low in this region.

The models range and variance we show are relatively small. This may be due to our initial model perturbations, which we kept low in order to prevent cycle skipping from occurring or to an underestimation of the covariance in the data space. Moreover, we see, in all three cases, the models do not seem to spread more whenever they are far from the sonic log profiles. The question behind is related to the interpretation we shall give to the variance estimation. It could be interpreted as variance with respect to a 'ground truth' (i.e. the sonic logs). In this case, it appears the variance is underestimated, as the spread of the model do not encompass the velocity values from the log. A second interpretation, which is to us more consistent, is that the variance is estimated with respect to a 'best possible FWI model' in the considered frequency band (here 2.5–5 Hz). It is well known that such a model, due to finite-frequency effects, is only a low-pass filtered version, or more precisely an homogenized version (Capdeville & Métivier 2018) of the 'ground-truth' in the inverted frequency band. Sharp velocity jumps such as the one which can be seen on logs 1 and 3 at 2.7 km depth cannot be reconstructed from the data in this frequency band, this is why none of the models from the ensemble display such a jump. Our ETKF-FWI strategy makes it possible to sample the covariance near our local minima but not to explore models which are closer to the ground truth.

This is further illustrated in Fig. 19 where we show both initial and final model distribution for some chosen points within the low-velocity anomaly. We see that, for all of the sampled point near the low-velocity anomaly, FWI noticeably increase the model variance. This is especially true for Figs 19(b.iii) and (c.iii) where the models range approaches 150 $\mathrm{ms^{-1}}$. Additionally, we see in Fig. 19(c.iii) a sort of multimodal distribution. Since the model distribution is studied at the interface of a low velocity layer, such a distribution indicate that our FWI resolution is too low to properly characterize this layer. Fig. 19(f.iii) is very interesting as we see a clearly non

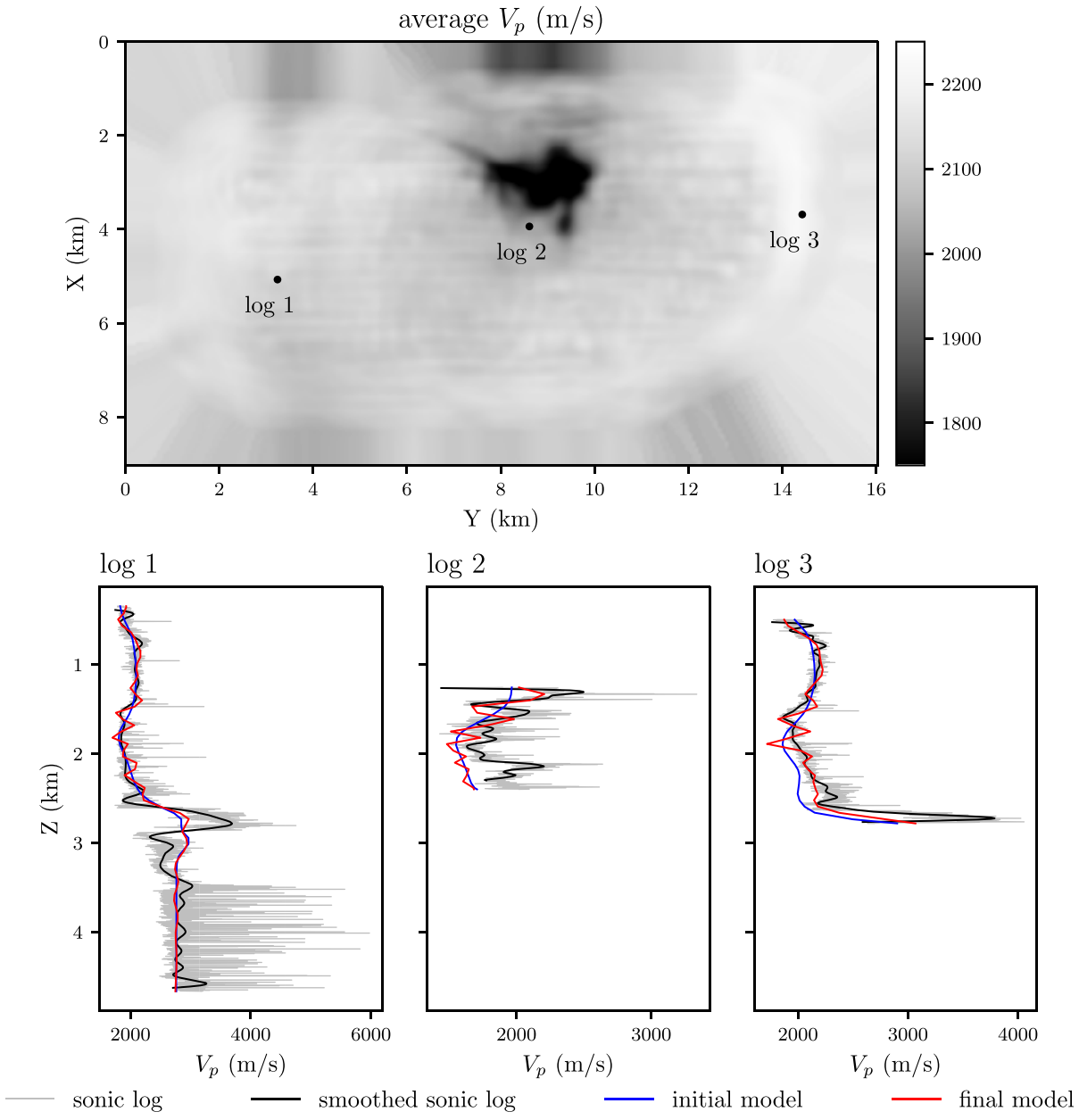

Figure 15. Top: positions of the three V_p sonic logs overlapped with an horizontal section at 1 km depth of the mean V_p model obtained by ETKF-FWI. Bottom: comparison of V_p profiles extracted from the mean model (in red), the sonic logs (in grey) and the smoothed sonic log filtered in the 0–12 Hz frequency band (in black).

Gaussian distribution as well as some outliers. The non Gaussian distribution can be explained by the fact that we bounded the values of V_p during our FWI, which resulted in a lot of models being forced to the minimal V_p value. The outliers are a more difficult to explain. On one hand they can be explained by cycle skipping occurring for some models. Alternatively, it could be caused by the analysis step of the ETKF. More precisely, ETKF first compute the ensemble's mean and variance first and then compute an ensemble that satisfy said mean and variance. This can result in the apparition of outliers that enforces the desired variance (Lawson & Hansen 2004; Leeuwenburgh *et al.* 2005). Since the initial perturbations are really small, it is unlikely that cycle skipping occurred, the most likely hypothesis is that the V_p constraints resulted in a non Gaussianly distributed forecasted ensemble, which in turn resulted into a bias in the analysis.

This goes to show that FWI tends to spread the velocity models, which can be problematic especially when the initial model is close to the extremes values of V_p. In our applications, we chose to reduce the initial model perturbation. This could also be remedied by using particle filter (PF), which do not make assumptions on the posterior distribution (Gordon *et al.* 1993), instead of KF.

4.6.3 Variance in the data-space

In this section we are going to discuss the variance in the data-space and to relate it to the variance in the model space. We show both

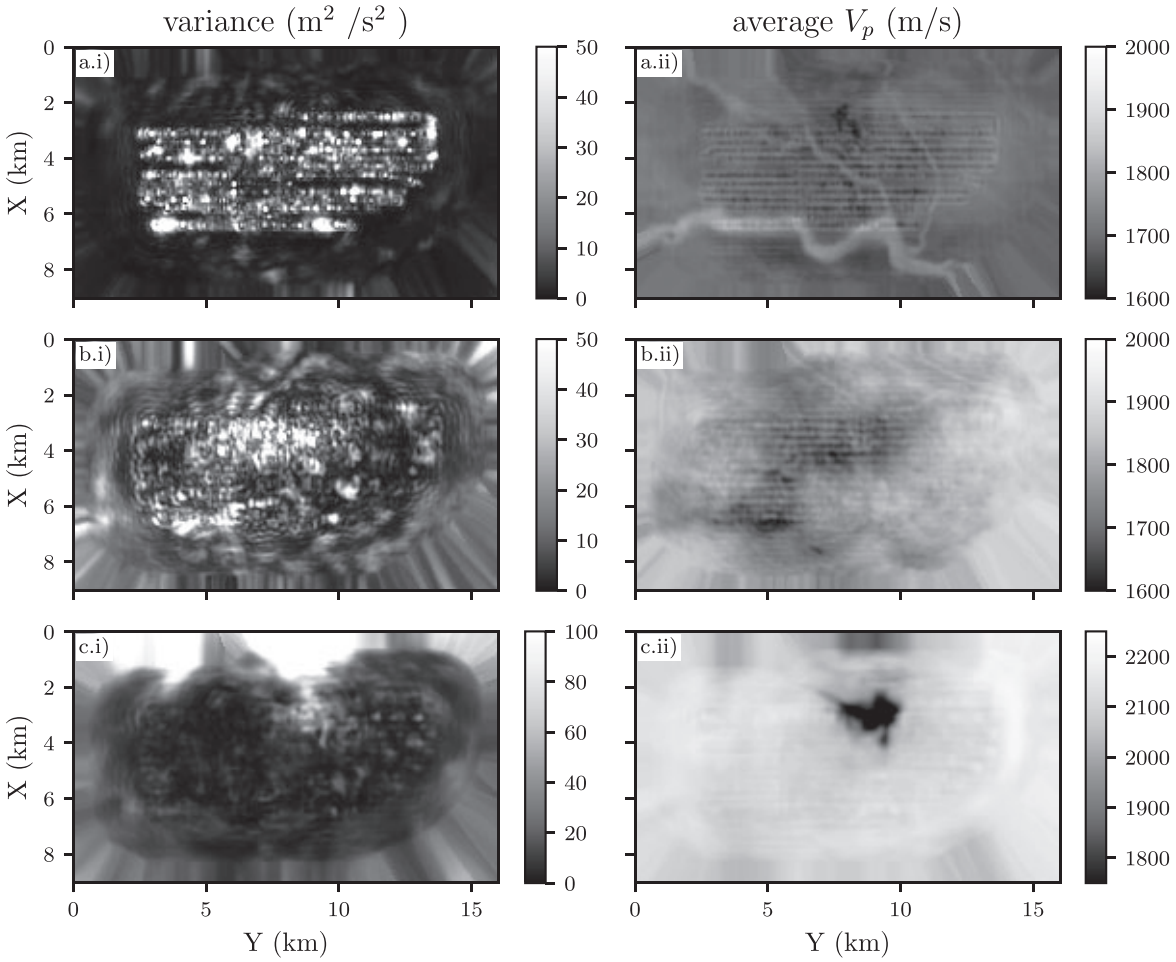

Figure 16. Horizontal sections of the variance (a.i–c.i) of the final models and the final mean model (a.ii–c.ii) obtained by ETKF-FWI at (a) 200 m depth, (b) 500 m depth and (c) 1 km depth.

the observation associated to the average model and the variance in the data space in Fig. 20. We can distinguish three batches of high variance in the data space, a first batch for short times and short offset which can be explained by artefact from the inversion in the shallow part of the model. A second batch at times between 3.5 and 4.5 s and for medium offsets, which, as we can see in Fig. 20, corresponds to the reflection of the diving wave onto the top of the low-velocity anomaly at approximately 1.5 km depth. Finally, the third batch at times between 6 and 6.5 s and for long offsets corresponds to the reflections of the diving wave onto the reservoir at approximately 3 km depth. We illustrate this by overlapping the wave propagation and the average model in Fig. 21, where we see diving waves within a low velocity layer at $t = 2$ s, the reflections arriving to the receiver line at $t = 3.3$ s, the interaction between reflected waves and diving waves at $t = 3.9$ s and finally, the diving waves reaching the reservoir interface.

The third batch of high variance in the data space seen in Fig. 20 is due to the reflection of the diving wave onto the interface of the reservoir.

4.6.4 Correlation as a measure of resolution

While studying the variance maps or, equivalently, the model spreading, can help to identify areas that are difficult to properly characterize via FWI, it only provides a point-wise metric and thus, yielding little information about the actual resolution of the FWI. On the other hand, the off-diagonal terms of the *correlation matrix* can be used to gain understanding of the interdependence between a given DoF and the other DoFs of the velocity model, thus allowing one to characterize the local resolution of the FWI (Thurin *et al.* 2019).

It is thus possible to study the local resolution of FWI by examining the rows of the correlation matrix corresponding to a given voxel of the physical space. In this work, we study the local resolution of our velocity model near the variance peaks identified in Fig. 17 thus allowing us to understand whether a variance spike is indicative of the difficulty to properly characterize a layer or to the difficulty to properly characterize the frontier between two layers. We show, in Fig. 22, the correlation map between the six variance peaks previously identified and their neighbouring voxels. We can see for the section $x = 2.95$ km (a), that low velocity layers (orange and green boxes) are well characterized, with strong positive correlation to surrounding layers and negative correlation with the rest of the model. However, the voxels above the low velocity layer in the red box have a strong positive correlation with the voxel within the low velocity layer. On the other hand, the voxels below the low velocity layer have a strong negative correlation and we even see positive correlation with the low velocity layer below

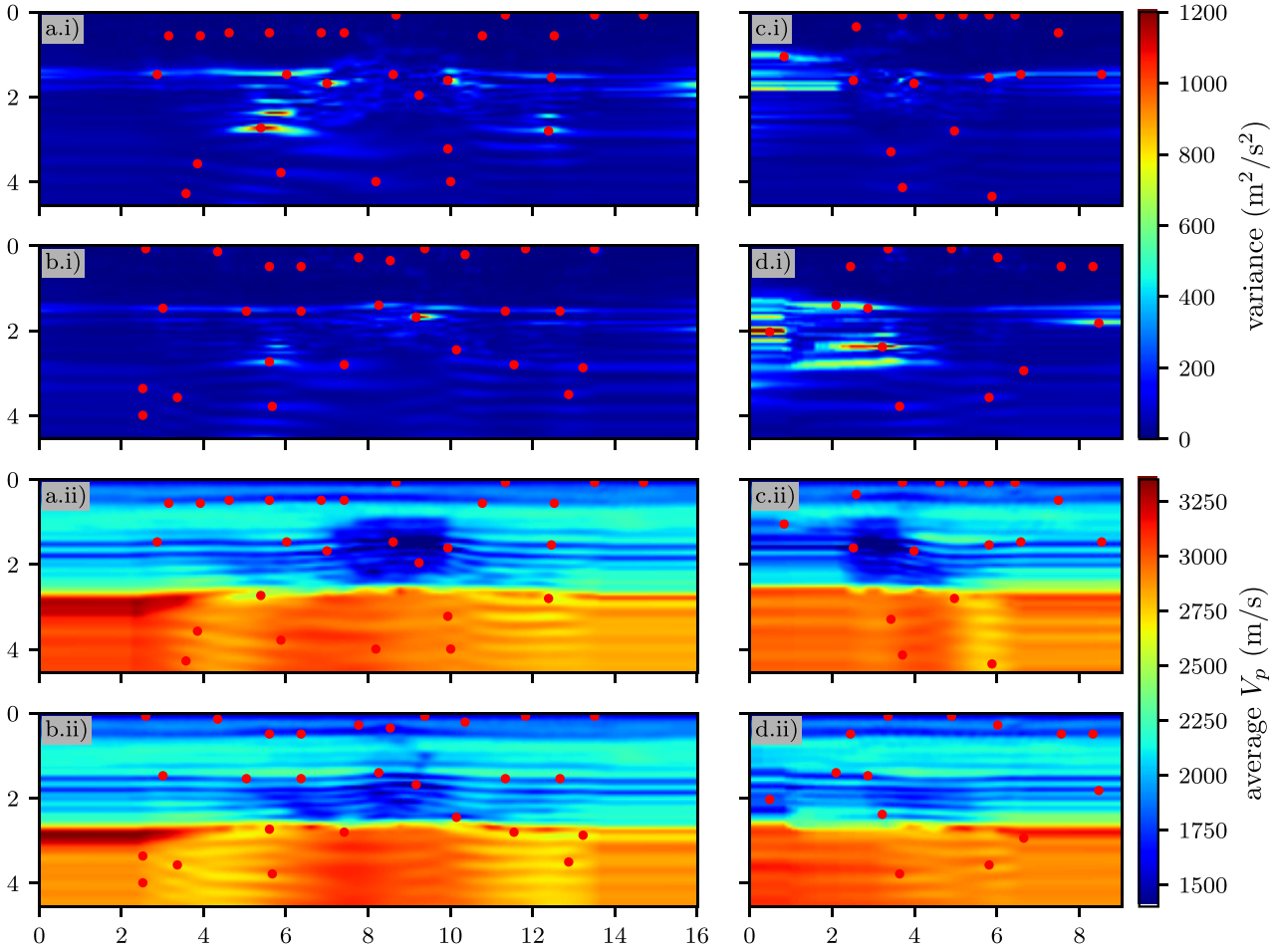

Figure 17. Slices of the variance of the models (a.i–d.i) and of the mean model (a.ii–d.ii) obtained by ETKF-FWI. Red dots denotes local maximum variance peaks, obtained with a 2-D maximum filter defined with a 1000 m radius. We show two inline vertical sections for (a) $x = 2.95$ km and (b) $x = 3.95$ km and two crossline vertical sections at (c) $y = 9$ km and (d) $y = 6$ km.

the voxel under study. This indicates that the top interface of the low velocity layer is poorly characterized by our FWI scheme. It is worth mentioning this area is located directly below the source A which might explain one of the high variance pattern observed in Figs 14 and 20. We see a similar phenomenon for the section $y = 9$ km (b) with the low velocity layer in the red box, as for the top of the low-velocity anomaly (in orange box). On the other hand, the surface of the deep reflector (green box) seems to be properly recovered. We can see that the lateral resolution of the top layer of the low-velocity anomaly is quite low, even-though we fit correctly the part of data corresponding to the reflection of the wavefield on said layer.

From theses results, we could sketch a workflow for detecting low resolution parts of the models. First we need to identify, from the variance in the data set, which components of the wavefield presents high variances. In this case, we would focus on reflections generated by the shallower part of the model. Then we can extract variance peaks and select the ones that might have the desired influence on the wavefield. In our case, we are looking at variance peaks located on the top sheet of the low-velocity anomaly and on the top of the reservoir. Finally, we can compute the correlation between the voxels corresponding to the variance and the surrounding voxels helping us to identify low resolution areas. Note that this workflow can help to find low-resolution part of the model that actually result in a good data fit, confirming our method can provide additional information to standard FWI quality control.

5 DISCUSSION

We give in this section few comments on possible extension or improvement of the scheme for a better efficiency of the ETKF-FWI scheme, and a general discussion paragraph based on the constructive exchanges with the reviewers during the revision process of this study.

5.1 Random transformation matrix

As mentioned earlier, the ETKF's transformation matrix T is not unique and can be defined as follows:

$$T = A^{\frac{1}{2}}\Lambda, \tag{23}$$

where Λ is an arbitrary orthogonal mean-preserving matrix ($\Lambda\mathbf{1} = \mathbf{1}$) (Sakov & Oke 2008; Livings *et al.* 2008). In our current work, we used $\Lambda = I$. However, a randomly generated Λ, as was proposed by Tödter & Ahrens (2015) might be a better option. Such

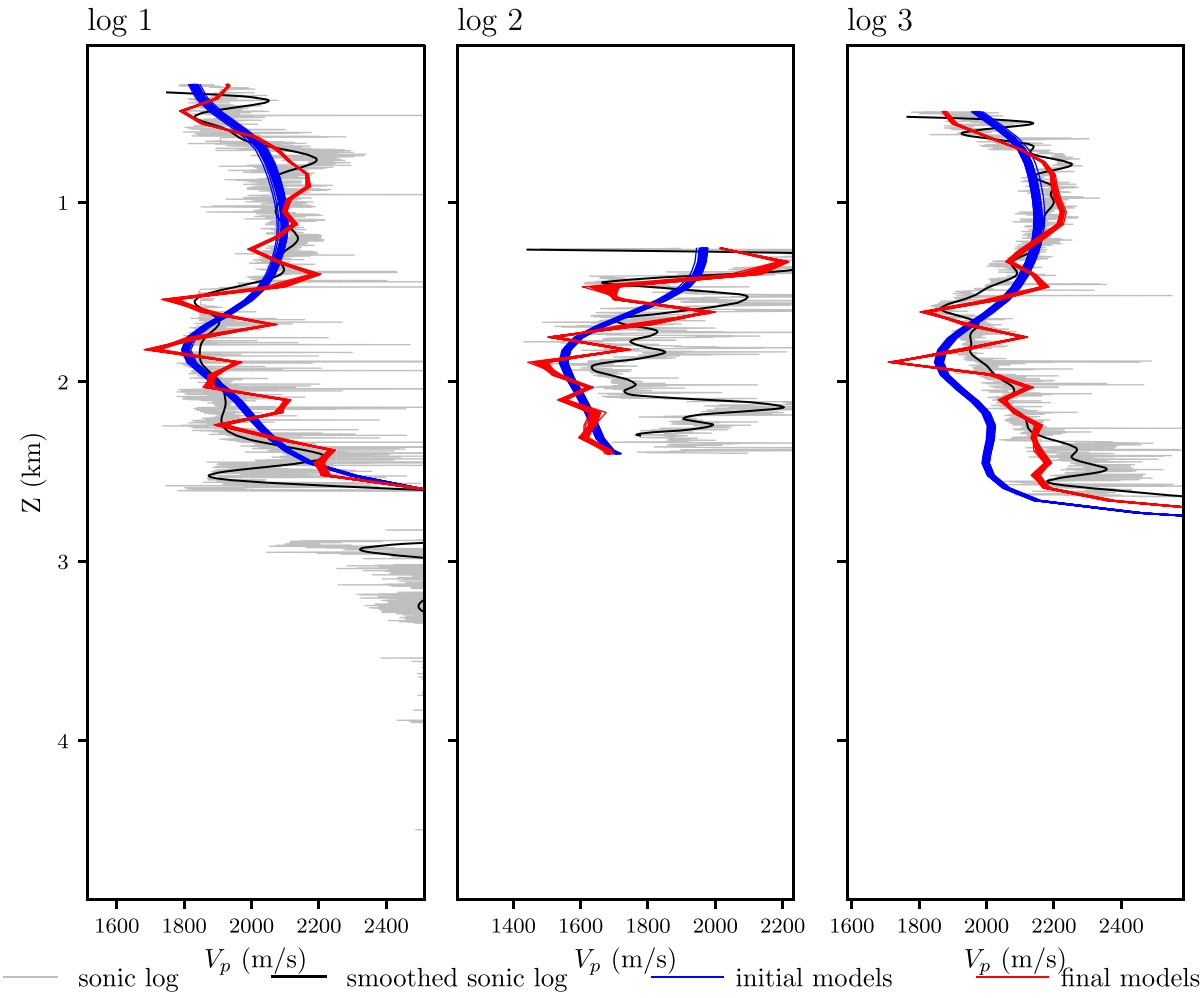

Figure 18. Comparison of V_p profiles extracted from the ETKF-FWI models (in blue), V_p profiles extracted from the mean model (in red), sonic logs (in grey) and smoothed sonic logs filtered in the 0–12 Hz frequency band (in black). The V_p range was restricted so that we can get a better understanding of the model spreading in the region located above the reservoir.

transformation matrix would allow us to effectively redraw an entirely new random ensemble with the correct mean and covariance at each ETKF iteration. Such additional transformations have been shown to potentially improve filter stability. A purely deterministic update may enhance a situation in which the ensemble is nearly collapsed and where the correct moments are enforced by a few outliers (Lawson & Hansen 2004; Leeuwenburgh *et al.* 2005).

Such a modification of the analysis step could be very beneficial to our ETKF-FWI as outlier in our velocity model may cause cycle skipping to occur during our forecast step. Conversely, the presence of outlier after the forecast step is a major indication that cycle skipping did occur. It is thus important to ensure the analysis do not create outliers in our ensemble to ensure the potential outliers are indeed caused by cycle-skipping in our forecast step.

5.2 Alternative filters

In our application, as explained in the previous section, the application of the forecast operator is by far the most expensive part of the ETKF. It is worth mentioning that certain filters such as the Hierarchical matrix ($\mathcal{H}^2$-matrices) powered KF (HiKF) (Li *et al.* 2014) may be the key to avoid performing several FWI as stated by Huang & Zhu (2020). While, HiKF are not designed for non-linear forecast and observation operators, they can be a very good tool for assimilating data in a quasi-continuous manner, as applying the forecast operator over very short period of time mitigate its non-linearities. However, when the data are scarse and when dealing with a highly non-linear operator such as FWI, EnKF are a better option as they where initially designed for non-linear forecast operator.

Another difficulty with ETKF-FWI comes from the high non-linearity of both forecast and observation operators. The non-linear nature of our forecast may result in forecasted ensemble that are not Gaussianly distributed, especially if cycle skipping occurs for some elements of our ensemble. In such a case, the forecast posterior mean and covariance becomes uninterpretable. Additionally, because the ETKF assumes the forecasted ensemble to follow a Gaussian distribution, the analysis mean and covariance can become biased (Lei & Bickel 2011).

To overcome this difficulty, Tödter & Ahrens (2015) introduced a square-root filter that uses mean and covariance computed from a PF (Gordon *et al.* 1993) for it. The resulting non-linear ensemble transform filter (NETF) only makes assumptions on the distribution

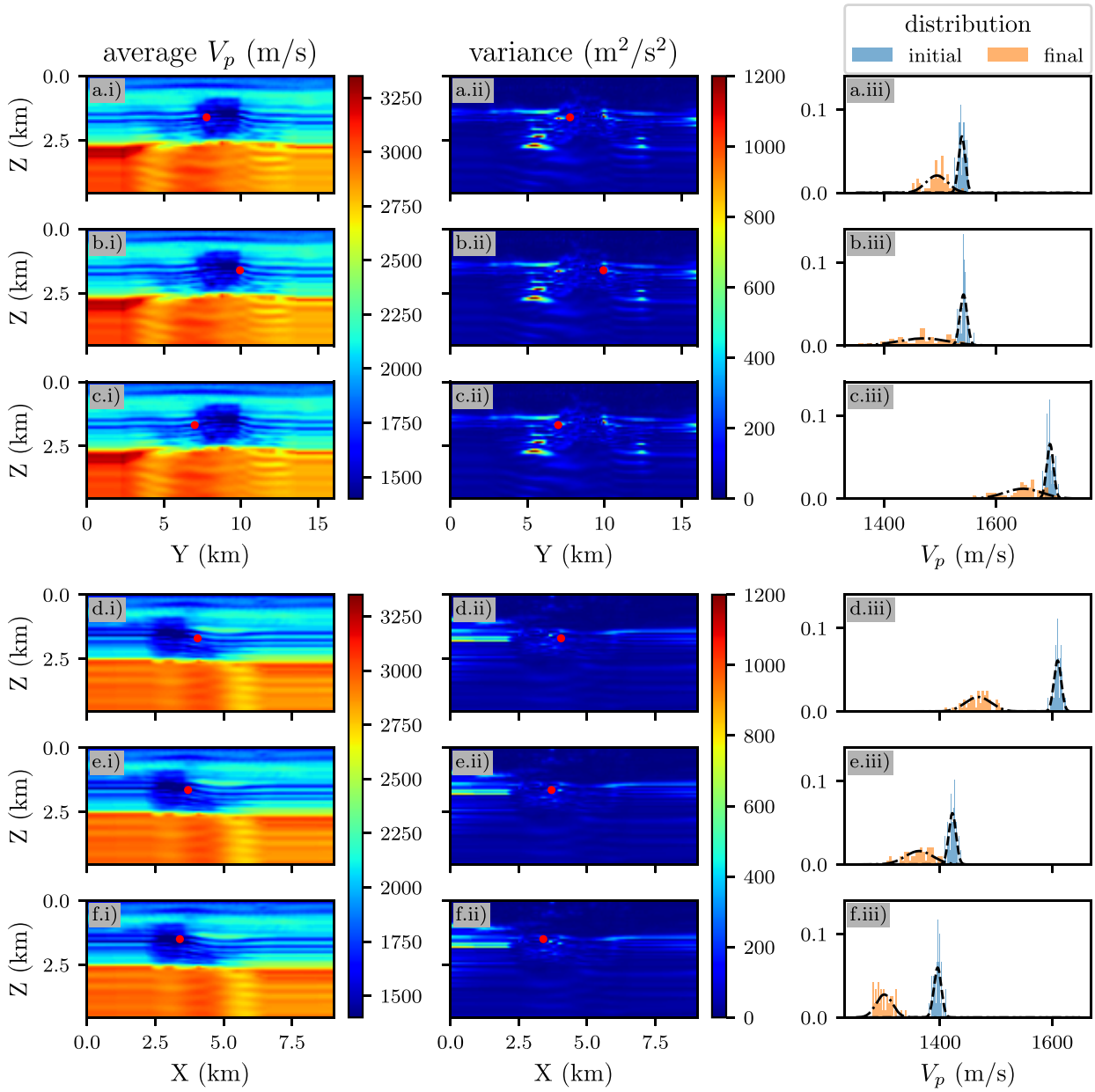

Figure 19. Different point of interest at which we display (a.i–f.i) the final mean V_p model obtained by ETKF-FWI, (a.ii–f.ii) the final variance of the models, (a.iii–f.iii) the distribution of the model at a given point marked in red. (a–c) inline vertical section at $x = 2.95$ km (d–f) crossline vertical section at $y = 9$ km.

of the ensemble in the data space, making the filter robust to outlier in the model space.

5.3 General discussion

We complete the discussion with general comments arising from the exchanges with the reviewers during the revision of this study. A first point is related to a potential underestimation of the variance, which is always a risk when using ensemble methods with relatively low number of elements. In this respect, the sonic logs comparison presented in Fig. 18 might be misleading. It appears that the spread of the models is too small to capture the values exhibited by the log, hence giving a false impression of low variability in a zone far from the 'ground truth'. This interpretation should be performed with care: indeed the 'ground truth' might simply be inaccessible from the data available in this example. We would better interpret the estimated variance with respect to a 'best achievable FWI model' which should be, in a given frequency band, a homogenized version of the subsurface mechanical parameters (Capdeville & Métivier 2018). For the 3–5 Hz data we consider here, there is no chance to get really closer to the ground truth.

This interpretation goes in hand with the fact that the ETKF-FWI scheme is dependent from the workflow which is used to invert the data. Indeed, it should be noted that the variance estimation on the Marmousi II case study, depending on the choice of the time-offset windowing workflow or the source subsampling workflow, is not strictly the same, even if it exhibits similar patterns. This is the result of the uncertainty estimation we perform being local, around a mean FWI model. The ETKF-FWI takes into account all the workflow design which is set up to invert a specific data

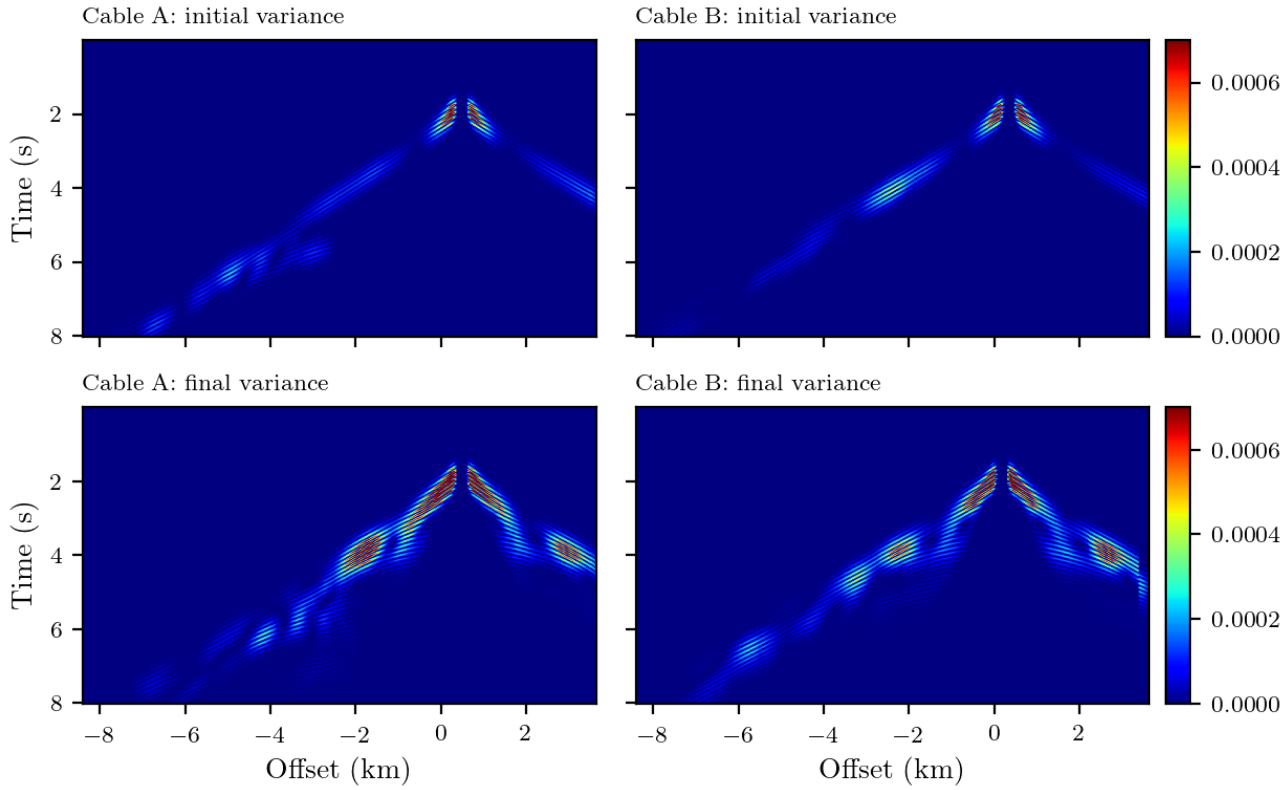

Figure 20. Variance of the common-receiver gathers, generated into the final V_P models.

set: data hierarchy, initial model design, regularization/smoothing strategy, misfit function design, not to speak about all the numerical implementation choices behind the solution of a FWI problem.

Behind this question lies a more general one, related to the interpretation we can give to the uncertainty estimation for the FWI problem. What we would qualify as an 'abstract' view is considering the FWI problem as the problem to find the model which provides the best fit to the data in an infinite-dimension space, with a non convex misfit function exhibiting numerous geologically non-informative minima. In this framework, FWI is an ill-posed inverse problem, and characterizing the uncertainty attached to this problem would require a sampling of the different minima of the misfit function. However, if we consider 'practical' applications of FWI, the situation is different. Prior to the inversion, the design of the specific workflow we just mentioned makes sure that FWI converge towards a geologically informative local minimum (potentially the global minimum). In this context, FWI, seen as a whole, becomes a relatively well-posed inverse problem. The ETKF-FWI strategy we propose makes it possible to estimate locally the uncertainty attached to FWI seen as such a process. From the more 'abstract' point of view, it amounts to explore locally the uncertainty of a given minimum of the misfit function. This local exploration is however more efficient than methods based on inverse Hessian sampling because it is inherently parallel (the FWI associated with each ensemble members can be ran in parallel), and it is based on a sound statistical theory arising from the DA community.

Of course this positioning might appear less ambitious than seeking for global uncertainty quantification. However, for us, it appears as a pragmatical choice, adapted to a large variety of problems for which such FWI workflow already exist, both in academy and industry, and applicable to 3-D FWI imaging. We should end this discussion by restating that this study presents the first 3-D field data application of a systematic uncertainty quantification scheme.

6 CONCLUSIONS AND PERSPECTIVES

We have presented two synthetic and one 3-D OBC field data application of our ETKF-FWI schemes which are based on popular iterative time-domain FWI schemes. On the synthetic example, our scheme yields similar results to the one obtained by frequency domain ETKF-FWI (Thurin *et al.* 2019). On the 3-D OBC field data data set, our ETKF-FWI produces a mean model that is similar to the one obtained by FWI (Sirgue *et al.* 2010; Operto *et al.* 2015) on the same data set and fits relatively well the available sonic logs.

Additionally, our ETKF-FWI gives us insight on which part of the model is correctly resolved and which part of the velocity model suffers from a lack of illumination and is therefore 'uncertain'. We show how we can use the estimated variance map and correlation matrix (derived from the covariance matrix) to assess the quality and reliability of the FWI solution. We can also link high variance zones in the data space to low-resolution area in the velocity model estimation. In thisse a ,relatively low frequency range which explains the low resolution of certain parts of the model. Further work thus include applying our scheme to higher frequency data with an expected challenge in terms of computational cost.

From a methodological standpoint, further work on the analysis step could be considered. As mentioned earlier, the balance between the background error and the observation error must be carefully calibrated, in order to reduce the observation error while preserving some disparity in our model ensemble. In this work, using the MSD

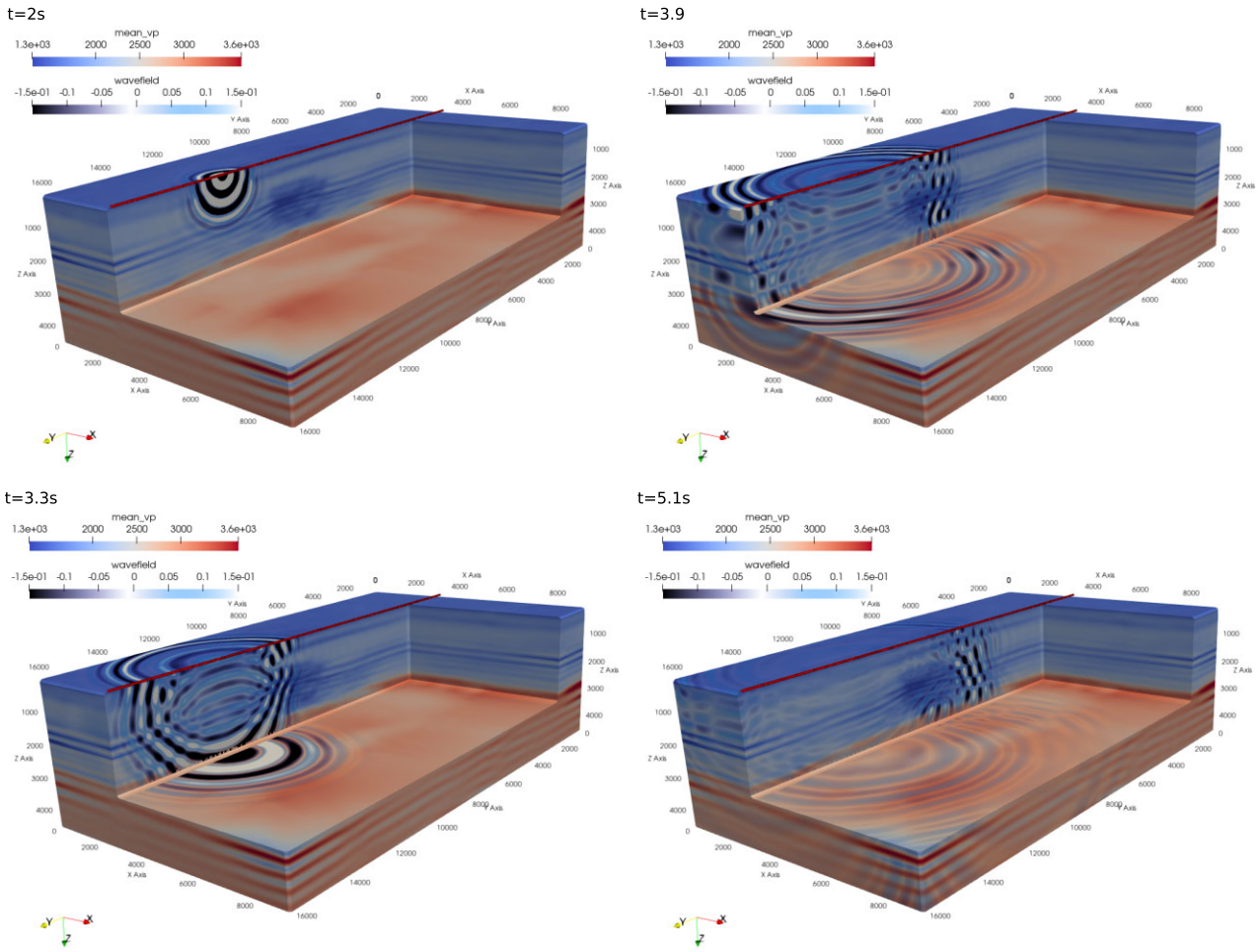

Figure 21. Wavefield super-imposed onto the average velocity model obtained by ETKF-FWI and, in red, the position of the receivers deployed along the line $x = 2.95$ km. The wavefield is displayed at times $t = 2$, 3.3, 3.9 and 5.1 s.

instead of the L^2 distance in the analysis yielded satisfactory results. It is worth mentioning that, instead of using the MSD, we could simply divide the data misfit by the number of (physical) time step, which would result in a similar scaling and would make sense from a physical point of view. While theses two options seem to make sense intuitively, the DA community has proposed several tools in order to properly calibrate the analysis (e.g. Desroziers *et al.* 2005). A more in depth sate of the art on this specific topic could bring interesting alternatives.

An open question is related to the application of the ETKF-FWI strategy in the context of multiparameter FWI. The difficult point might be the definition of the initial ensemble, to capture sufficient variety in the model space while still ensuring to converge in the same basin of attraction. This tuning is relatively easy in the context of mono-parameter velocity inversion, because natural 'physical' criterion can be devised for the velocity in terms of cycle skipping, as is always done for FWI. When secondary parameters are inverted, such as density or quality factor, the problem might become more difficult. However, there is a specific interest to go towards such multiparameter framework, because recent results show that inverting the data in a broad frequency band might yield the required constraints to stably reconstruct multiple parameters (Métivier *et al.* 2024). In this specific multiparameters context, trade-offs are expected between parameters, and the covariance estimation will provide a tool to analyse these trade-offs in a more systematic way. Accessing the correlation matrix in this case would bring additional information on these trade-offs. Preliminary results in this direction by Thurin *et al.* (2019) are encouraging.

The fact that ETKF-FWI estimates the uncertainty associated with a given 'workflow' to invert seismic data naturally enables naturally enables the uncertainty quantification of hyperparameters defining the workflow. The variance associated with regularization tuning parameters or misfit function choice for instance could bring crucial information on the robustness of a chosen workflow to invert a specific data set.

Finally, single application of 3-D FWI on field data set remains often challenging from a computational point of view, especially in the current context of broad frequency band FWI application and the beginning of 3-D elastic FWI applications. Running in parallel few tens of such FWI seems thus still computationally expensive. This calls for a rational usage of computational power, in the perspective of using upcoming exascale machines. Codes running on such architecture need to be fault-tolerant/fault-resilient. We think the ETKF-FWI algorithm could be adapted to such platforms relatively easily, by using a task managing module overarching the ETKF-FWI scheme, with a subdivision of tasks in terms of shot-by-shot gradient computation, instead of a whole FWI as it is implemented by now.

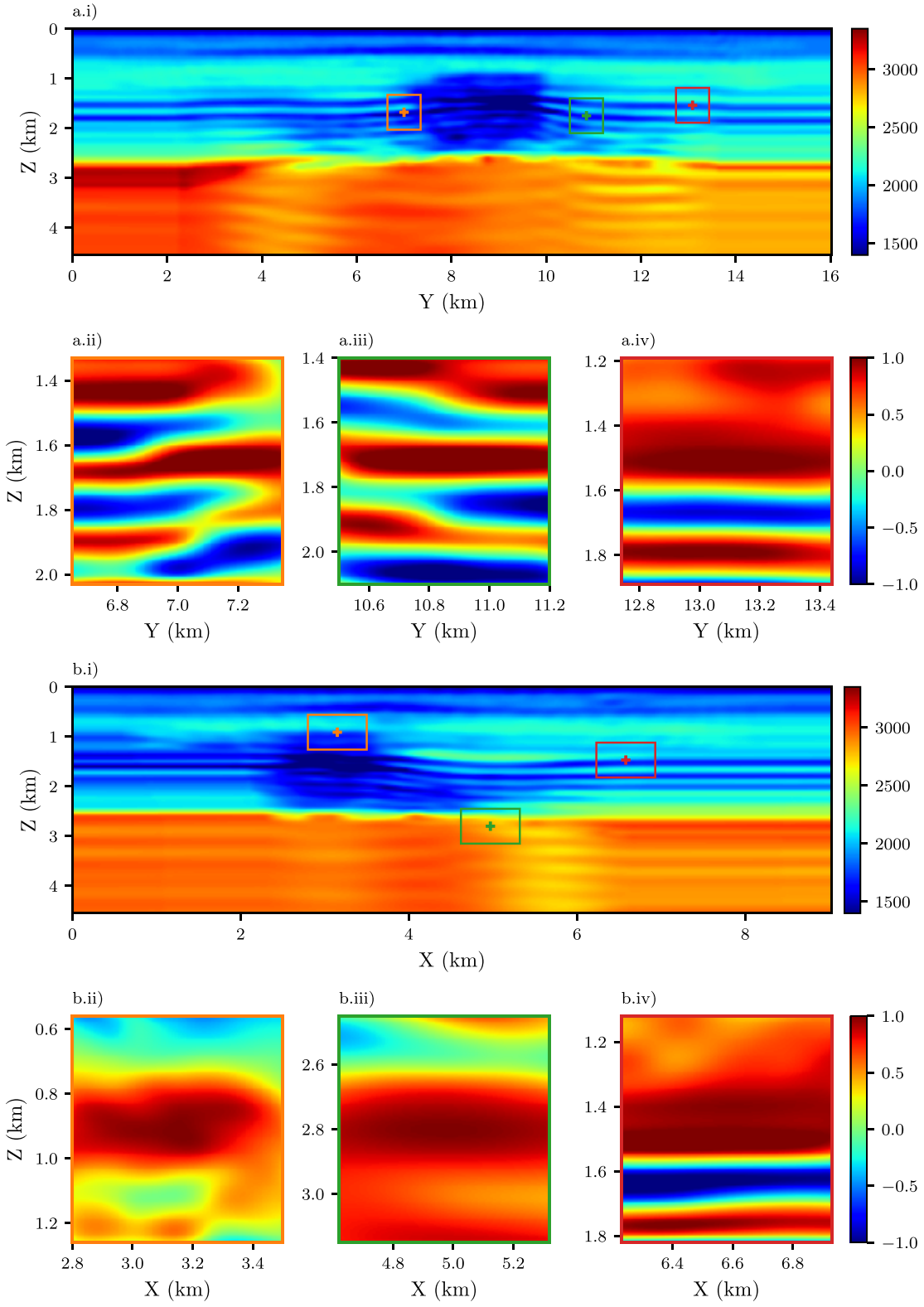

Figure 22. (a.i) Inline vertical section of the average final model obtained by ETKF-FWI for $x = 2.95$ km and (a.ii–a.iv) local correlation maps in orange, green and red boxes. (b.i) Crossline vertical section of the average final model obtained by ETKF-FWI for $y = 9$ km and (b.ii–b.iv) local correlation maps in orange, green and red boxes.

ACKNOWLEDGMENTS

This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by AKERBP, CGG, DUG, EXXONMOBIL, GEOLINKS, JGI, PETROBRAS, SHELL, SINOPEC and TOTALENERGIES. This study was granted access to the HPC resources provided by the GRICAD infrastructure (*https://gricad.univ-grenoble-alpes.fr*), which is supported by Grenoble research communities, the HPC resources of Cray Marketing Partner Network (*https://partners.cray.com*), and those of CINES/IDRIS/TGCC under the allocation 046091 made by GENCI. The authors would like to express their gratitude to the two anonymous reviewers and the associated editor Carl Tape who provided numerous valuable comments which helped us in the writing of this article.

DATA AVAILABILITY

The data that support the findings of this study are available from AKERBP. Restrictions apply to the availability of the data, which were used in this study.

REFERENCES

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.

Baumstein, A., Ross, W. & Lee, S., 2011. Simultaneous source elastic inversion of surface waves, in *Proceedings of the 73rd EAGE Conference and Exhibition incorporating SPE EUROPEC 2011,* May 2011, cp-238-00111. European Association of Geoscientists and Engineers.

Ben Hadj Ali, H., Operto, S. & Virieux, J., 2011. An efficient frequency-domain full waveform inversion method using simultaneous encoded sources, *Geophysics,* **76**(4), doi:10.1190/1.3581357.

Bishop, C.H., Etherton, B.J. & Majumdar, S.J., 2001. Adaptive sampling with the ensemble transform kalman filter. Part I: theoretical aspects, *Mon. Wea. 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.

Bowler, N.E., Arribas, A., Mylne, K.R., Robertson, K.B. & Beare, S.E., 2008. The MOGREPS short-range ensemble prediction system, *Quarterly Journal of the Royal Meteorological Society,* **134**(632), 703–722.

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.

Brossier, R., Operto, S. & Virieux, J., 2009a. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion, *Geophysics,* **74**(6), WCC105–WCC118.

Brossier, R., Operto, S. & Virieux, J., 2009b. Two-dimensional seismic imaging of the Valhall model from synthetic OBC data by frequency-domain elastic full-waveform inversion, *SEG Technical Program Expanded Abstracts,* **28**(1), 2293–2297.

Brossier, R., Etienne, V., Hu, G., Operto, S. & Virieux, J., 2013. Performances of 3D frequency-domain full-waveform inversion based on frequency-domain direct-solver and time-domain modeling: application to 3D OBC data from the Valhall field, in *Proceedings of the International Petroleum Technology Conference, IPTC 16881,* Beijing, China.

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. Sci. Comput.,* **35(6),** A2494–A2523.

Byrd, R.H., Lu, P. & Nocedal, J., 1995. A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. 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.

Capdeville, Y., Gung, Y. & Romanowicz, B., 2005. Towards global earth tomography using the spectral element method: a technique based on source stacking, *Geophys. J. Int.,* **162,** 541–554.

Castellanos, C., Métivier, L., Operto, S., Brossier, R. & Virieux, J., 2015. Fast full waveform inversion with source encoding and second-order optimization methods, *Geophys. J. Int.,* **200(2),** 720–744.

Chen, Y., Zhang, W. & Zhu, M., 2020. A localized weighted ensemble Kalman filter for high-dimensional systems, *Quart. J. R. Meteorol. Soc.,* **146**(726), 438–453.

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.

Deal, M.M. & Nolet, G., 1996. Nullspace shuttles, *Geophys. J. Int.,* **124**(2), 372–380.

Desroziers, G., Berre, L., Chapnik, B. & Poli, P., 2005. Diagnosis of observation, background and analysis-error statistics in observation space, *Quart. J. R. Meteorol. Soc.,* **131**(613), 3385–3396.

Devaney, A., 1984. Geophysical diffraction tomography, *IEEE Trans. Geosci. Remote Sens.,* **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.

Eliasson, P. & Romdhane, A., 2017. Uncertainty quantification in waveform-based imaging methods—a Sleipner CO_2 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., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation, *Ocean Dyn.,* **53,** 343–367.

Evensen, G., 2009. *Data Assimilation: The Ensemble Kalman Filter,* Springer.

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.,* **120**(8), 5549–5573.

Fichtner, A. & Villaseñor, A., 2015. Crust and upper mantle of the western mediterranean – constraints from full-waveform inversion, *Earth planet. Sci. Lett.,* **428,** 52–62.

Fichtner, A. & Zunino, A., 2019. Hamiltonian nullspace shuttles, *Geophys. Res. Lett.,* **46**(2), 644–651.

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.

Fichtner, A., Zunino, A. & Gebraad, L., 2018. Hamiltonian Monte Carlo solution of tomographic inverse problems, *Geophys. J. Int.,* **216**(2), 1344–1363.

Gebraad, L., Boehm, C. & Fichtner, A., 2020. Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo, *J. geophys. Res.,* **125**(3), e2019JB018428, doi:10.1029/2019JB018428.

Gineste, M. & Eidsvik, J., 2017. Seismic waveform inversion using the ensemble Kalman smoother, in *Proceedings of the 79th EAGE Conference and Exhibition 2017,* Vol. 2017, pp. 1–5, European Association of Geoscientists & Engineers.

Gineste, M., Eidsvik, J. & Zheng, Y., 2019. Seismic waveform inversion using an iterative ensemble Kalman smoother, in *Proceedings of the Second EAGE/PESGB Workshop on Velocities,* Vol. 2019, pp. 1–3, European Association of Geoscientists & Engineers.

Gineste, M., Eidsvik, J. & Zheng, Y., 2020. Ensemble-based seismic inversion for a stratified medium, *Geophysics,* **85**(1), R29–R39.

Gordon, N., Salmond, D. & Smith, A., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEEE Proc. F (Radar and Signal Processing),* **140,** 107–113(6).

Górszczyk, A., Operto, S. & Malinowski, M., 2017. Toward a robust workflow for deep crustal imaging by FWI of OBS data: the eastern Nankai trough revisited, *J. geophys. Res.,* **122**(6), 4601–4630.

Harlim, J. & Hunt, B.R., 2007. A non-gaussian ensemble filter for assimilating infrequent noisy observations, *Tellus A,* **59**(2), 225–237.

Hu, G., Etienne, V., Castellanos, C., Operto, S., Brossier, R. & Virieux, J., 2012. Assessment of 3D acoustic isotropic full waveform inversion of wide-azimuth OBC data from Valhall, in *Proceedings of the SEG Technical Program Expanded Abstracts 2012,* pp. 1–6.

Huang, C. & Zhu, T., 2020. Towards real-time monitoring: data assimilated time-lapse full waveform inversion for seismic velocity and uncertainty estimation, *Geophys. J. Int.,* **223**(2), 811–824.

Hunt, B., Kostelich, E. & Szunyogh, I., 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter, *Physica D,* **230**(1), 112–126.

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.

Kamath, N., Brossier, R., Métivier, L. & Yang, P., 2018. 3D acoustic/viscoacoustic time-domain FWI of OBC data from the Valhall field, in *Proceedings of the SEG Technical Program Expanded Abstracts 2018,* pp. 1093–1097.

Kamath, N., Brossier, R., Métivier, L. & Yang, P., 2019. Multiparameter full-waveform inversion of data from the Valhall field, in *Proceedings of the SEG Technical Program Expanded Abstracts 2019,* pp. 1640–1644.

Kamath, N., Brossier, R., Métivier, L., Pladys, A. & Yang, P., 2021. Multiparameter full-waveform inversion of 3D ocean-bottom cable data from the Valhall field, *Geophysics,* **86**(1), B15–B35.

Karaoğlu, H. & Romanowicz, B., 2018. Inferring global upper-mantle shear attenuation structure by waveform tomography using the spectral element method, *Geophys. J. Int.,* **213**(3), 1536–1558.

Keating, S.D. & Innanen, K.A., 2021. Null-space shuttles for targeted uncertainty analysis in full-waveform inversion, *Geophysics,* **86**(1), R63–R76.

Krebs, J., Anderson, J., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A. & Lacasse, M.D., 2009. Fast full-wavefield seismic inversion using encoded sources, *Geophysics,* **74**(6), WCC105–WCC116.

Lailly, P., 1983. The seismic inverse problem as a sequence of before stack migrations, in *Proceedings of the Conference on Inverse Scattering, Theory and Application,* pp. 206–220, Society for Industrial and Applied Mathematics, Philadelphia.

Lawson, W.G. & Hansen, J.A., 2004. Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth, *Mon. Wea. Rev.,* **132**(8), 1966–1981.

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.

Lei, J. & Bickel, P., 2011. A moment matching ensemble filter for nonlinear non-Gaussian data assimilation, *Mon. Wea. Rev.,* **139**(12), 3964–3973.

Lei, W. *et al.*, 2020. Global adjoint tomography—model glad-m25, *Geophys. J. Int.,* **223**(1), 1–21.

Li, J.Y., Ambikasaran, S., Darve, E.F. & Kitanidis, P.K., 2014. A Kalman filter powered by -matrices for quasi-continuous data assimilation problems, *Water Resources Research,* **50**(5), 3734–3749.

Li, Y., Brossier, R. & Métivier, L., 2020. 3D frequency-domain elastic wave modeling with spectral-element method using a massively parallel direct solver, *Geophysics,* **85**(2), T71–T88.

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.

Liu, Q. & Peter, D., 2020. Square-root variable metric-based nullspace shuttle: a characterization of the nonuniqueness in elastic full-waveform inversion, *J. geophys. Res.,* **125**(2), e2019JB018687, doi:10.1029/2019JB018687.

Livings, D.M., Dance, S.L. & Nichols, N.K., 2008. Unbiased ensemble square root filters, *Physica D,* **237**(8), 1021–1028.

Lu, Y., Stehly, L., Brossier, R., Paul, A. & Group, A.W., 2020. Imaging Alpine crust using ambient noise wave-equation tomography, *Geophys. J. Int.,* **222**(1), 69–85.

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. Sci. 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., 2016a. The seiscope optimization toolbox: a large-scale nonlinear optimization library based on reverse communication, *Geophysics,* **81**(2), F11–F25.

Métivier, L. & Brossier, R., 2016b. The SEISCOPE 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. Sci. Comput.,* **35**(2), B401–B437.

Métivier, L., Bretaudeau, F., Brossier, R., Operto, S. & Virieux, J., 2014a. Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures, *Geophys. Prospect.,* **62,** 1353–1375.

Métivier, L., Brossier, R., Operto, S. & Virieux, J., 2014b. Multi-parameter FWI - an illustration of the Hessian operator role for mitigating trade-offs between parameter classes, in *Proceedings of the 6th EAGE St-Petersbourg International Conference & Exhibition,* Expanded Abstracts.

Métivier, L., Brossier, R., Operto, S. & Virieux, J., 2017. Full waveform inversion and the truncated Newton method, *SIAM Rev.,* **59**(1), 153–195.

Métivier, L., Brossier, R., Hoffmann, A., Mirebeau, J.-M., Provenzano, G., Tarayoun, A. & Yong, P., 2024. Coherence-enhancing anisotropic diffusion filter for 3D high resolution reconstruction of P-wave velocity and density using full waveform inversion: application to a North Sea Ocean Bottom Cable dataset, *Geophysics,* **89**(1), R33–R58.

Miyoshi, T., 2011. The gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter, *Mon. Wea. Rev.,* **139**(9), 1519–1535.

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.

Nerger, L., Janjić, T., Schröter, J. & Hiller, W., 2012. A unification of ensemble square root Kalman filters, *Mon. Wea. Rev.,* **140**(7), 2335–2345.

Nocedal, J., 1980. Updating quasi-Newton matrices with limited storage, *Math. Comput.,* **35**(151), 773–782.

Nocedal, J. & Wright, S.J., 2006. *Numerical Optimization,* 2nd edn, Springer.

Operto, S. & Miniussi, A., 2018. On the role of density and attenuation in 3D multi-parameter visco-acoustic VTI frequency-domain FWI: an OBC case study from the North Sea, *Geophys. J. Int.,* **213,** 2037–2059.

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**(9), doi:10.1029/2005JB003835.

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), 1040–1054.

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.

Pladys, A., Brossier, R., Irnaka, M., Kamath, N. & Métivier, L., 2019. Assessment of optimal transport based FWI: 3D OBC Valhall case study, in *Proceedings of the SEG Technical Program Expanded Abstracts 2019,* pp. 1295–1299.

Pladys, A., Brossier, R., Kamath, N. & Métivier, L., 2022. Robust FWI with graph space optimal transport: application to 3D OBC Valhall data, *Geophysics,* **87**(3), 1–76.

Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.,* **167**(2), 495–503.

Plessix, R.E., 2009. Three-dimensional frequency-domain full-waveform inversion with an iterative solver, *Geophysics,* **74**(6), WCC53–WCC61.

Plessix, R.E. & Perkins, C., 2010. Full waveform inversion of a deep water ocean bottom seismometer dataset, *First Break,* **28,** 71–78.

Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain. Part I: theory and verification in a physical scale model, *Geophysics,* **64,** 888–901.

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.

Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O. & Kommedal, J., 2011. On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study, *Geophys. J. Int.,* **187,** 1495–1515.

Prieux, V., Brossier, R., Operto, S. & Virieux, J., 2013. Multiparameter full waveform inversion of multicomponent OBC data from Valhall. Part 1: imaging compressional wavespeed, density and attenuation, *Geophys. J. Int.,* **194**(3), 1640–1664.

Raknes, E.B., Arntsen, B. & Weibull, W., 2015. Three-dimensional elastic full waveform inversion using seismic data from the Sleipner area, *Geophys. J. Int.,* **202**(3), 1877–1894.

Rawlinson, N., Fichtner, A., Sambridge, M. & Young, M.K., 2014. Seismic tomography and the assessment of uncertainty, *Adv. Geophys.,* **55,** 1–76.

Rodell, M. *et al.*, 2004. The global land data assimilation system, *Bull. Am. Meteorol. Soc.,* **85**(3), 381–394.

Sakov, P. & Oke, P.R., 2008. Implications of the form of the ensemble transformation in the ensemble square root filters, *Mon. Wea. Rev.,* **136**(3), 1042–1053.

Schiemenz, A. & Igel, H., 2013. Accelerated 3-D full-waveform inversion using simultaneously encoded sources in the time domain: application to Valhall ocean-bottom cable data, *Geophys. J. Int.,* **195,** 1970–1988.

Sen, M. & Stoffa, P., 1991. Non-linear one-dimensional seismic waveform inversion using simulated annealing, *Geophysics,* **56,** 1624–1638.

Sen, M. & Stoffa, P., 1992. Rapid sampling of model space using genetic algorithms: examples from seismic waveform inversion, *Geophys. J. Int.,* **108,** 281–292.

Sen, M. & Stoffa, P.L., 1996. Bayesian inference, Gibb's sampler and uncertainty estimation in geophysical inversion, *Geophys. Prospect.,* **44,** 313–350.

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

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.

Shin, C., Jang, S. & Min, D.J., 2001. Improved amplitude preservation for prestack depth migration by inverse scattering theory, *Geophys. Prospect.,* **49,** 592–606.

Shipp, R.M. & Singh, S.C., 2002. Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data, *Geophys. J. Int.,* **151,** 325–344.

Sirgue, L. & Pratt, R.G., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies, *Geophysics,* **69**(1), 231–248.

Sirgue, L., Barkved, O.I., Van Gestel, J.P., Askim, O.J. & Kommedal, J. H.., 2009. 3D waveform inversion on Valhall wide-azimuth OBC, in *Presented at the 71th Annual International Meeting, Expanded Abstracts,* EAGE.

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.

Solano, C.P. & Plessix, R.-É., 2019. Velocity-model building with enhanced shallow resolution using elastic waveform inversion — an example from onshore Oman, *Geophysics,* **84**(6), R977–R988.

Stopin, A., Plessix, R.-E. & Al Abri, S., 2014. Multiparameter waveform inversion of a large wide-azimuth low-frequency land data set in Oman, *Geophysics,* **79**(3), WA69–WA77.

Tao, K., Grand, S.P. & Niu, F., 2018. Seismic structure of the upper mantle beneath eastern Asia from full waveform seismic tomography, *Geochem. Geophys. Geosyst.,* **19**(8), 2732–2763.

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. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49**(8), 1259–1266.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* Society for Industrial and Applied Mathematics.

Thrastarson, S., van Herwaarden, D.-P., Krischer, L., Boehm, C., van Driel, M., Afanasiev, M. & Fichtner, A., 2022. Data-adaptive global full-waveform inversion, *Geophys. J. Int.,* **230**(2), 1374–1393.

Thurin, J., Brossier, R. & Métivier, L., 2019. Ensemble-based uncertainty estimation in full waveform inversion, *Geophys. J. Int.,* **219**(3), 1613–1635.

Tippett, M.K., Anderson, J.L., Bishop, C.H., Hamill, T.M. & Whitaker, J.S., 2003. Ensemble square root filters, *Mon. Wea. Rev.,* **131**(7), 1485–1490.

Tran, K. & Hiltunen, D., 2011. Two-dimensional inversion of full waveforms using simulated annealing, *J. Geotech. Geoenviron. Eng.,* **138**(9), 1075–1090.

Tödter, J. & Ahrens, B., 2015. A second-order exact ensemble square root filter for nonlinear data assimilation, *Mon. Wea. Rev.,* **143**(4), 1347–1367.

Vigh, D. & Starr, E.W., 2008. 3D prestack plane-wave, full waveform inversion, *Geophysics,* **73,** VE135–VE144.

Vigh, D., Jiao, K., Watts, D. & Sun, D., 2014. Elastic full-waveform inversion application using multicomponent measurements of seismic data collection, *Geophysics,* **79**(2), R63–R77.

Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics,* **74**(6), WCC1–WCC26.

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, S. *et al.*, 2018. Inferring dynamic topology for decoding spatiotemporal structures in complex heterogeneous networks, *Proc. Natl. Acad. Sci.,* **115**(37), 9300–9305.

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. Wea. Rev.,* **132**(7), 1590–1605.

Wang, Y., 2009. *Seismic Inverse Q Filtering,* John Wiley & Sons.

Wang, Y. & Rao, Y., 2009. Reflection seismic waveform tomography, *J. geophys. Res.,* **114**(B3), 1978–2012.

Warner, M. *et al.*, 2013. Anisotropic 3D full-waveform inversion, *Geophysics,* **78**(2), R59–R80.

Wu, R.S. & Toksöz, M.N., 1987. Diffraction tomography and multisource holography applied to seismic imaging, *Geophysics,* **52,** 11–25.

Yang, P., Brossier, R., Métivier, L., Virieux, J. & Zhou, W., 2018. A time-domain preconditioned truncated newton approach to multiparameter visco-acoustic full waveform inversion, *SIAM J. Sci. Comput.,* **40**(4), B1101–B1130.

Yuan, H., French, S., Cupillard, P. & Romanowicz, B., 2014. Lithospheric expression of geological units in central and eastern North America from full waveform tomography, *Earth planet. Sci. Lett.,* **402,** 176–186.

Zhang, X. & Curtis, A., 2020. Variational full-waveform inversion, *Geophys. J. Int.,* **222**(1), 406–411.

Zhang, X., Lomas, A., Zhou, M., Zheng, Y. & Curtis, A., 2023. 3-D Bayesian variational full waveform inversion, *Geophys. J. Int.,* **234**(1), 546–561.

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.

APPENDIX: PARALLEL IMPLEMENTATION

ETKF-FWI's forecast and observation operator are very costly as they require to perform N FWI to forecast our model and to solve N wave equations for each sources to apply our observation operator. However both operator can be computed in a fully parallel way. This is a notable advantage of ETKF-FWI over Hessian based uncertainty quantification approaches such as the SVD and Lanczos methods whose complexity scale linearly with the rank of the covariance matrix. It should thus be possible to exploit our method on exascale machines in order to compute correlation matrices with a much higher rank, allowing for a better uncertainty quantification. We will now fully describe our HPC framework.

Both FWI and modeling where performed with the TOYxDAC_TIME code developed by the SEISCOPE project, coupled with the SEISCOPE optimization tool-box Métivier & Brossier (2016a). The code uses a fourth-order staggered grid finite difference (FD) scheme for spatial discretization and a second-order leap-frog scheme for the time discretization for the modeling. The inversion is carried out by a *l*-BFGS (Byrd *et al.* 1995; Nocedal & Wright 2006; Métivier & Brossier 2016b) and uses the adjoint state method (Plessix 2006) to compute the gradient of the cost function. This code relies on source parallelization for both modeling and FWI. More precisely, one TOYxDAC_TIME process works on a single source and then combines the results with the message passing interface (MPI) protocol.

However, implementing a code that relies on both source and model parallelization is a technical challenge in itself which can lead to require rapidly thousands of processors in a single instantiation. We circumvent this difficulty by using a *root* or *master* program that submit and monitor, on a given HPC facility, *worker* programs that runs the FWI code. The advantage of this additional level of parallelism is fourfold. First, worker codes can be run asynchronously, which drastically diminish the amount of resources required at a given time. Secondly, the resource management is delegated to the HPC facility. Thirdly, in the event of a node or cluster failure, the root code is able to track which workers finished their task and which one needs to be started again, making the code more robust to hardware failure. Fourthly, each individual worker can be launched on a given architecture, allowing us to run our ETKF-FWI on a very heterogeneous architecture.

The resulting framework is illustrated in Fig. A1.

Such methodology, while needing technical refinements, could be used to tackle exascale type problems while being resilient to potential hardware failure on the cluster.

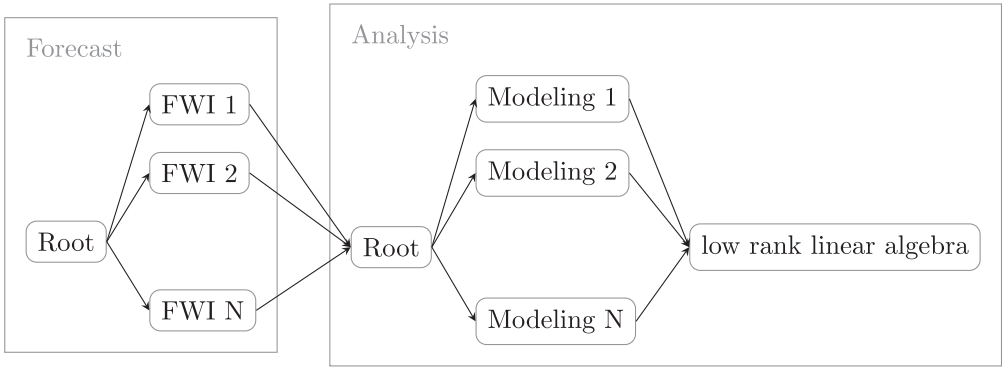

Figure A1. Illustration of our parallel implementation of the ETKF-FWI in the time domain. A root program run several FWI and modeling codes and then gather the results. An arrow from the root program symbolizes the root process running an other program. An arrow from any other program symbolizes the root process gathering the output of said program.