An Ensemble-Transform Kalman Filter - Full Waveform Inversion scheme for Uncertainty estimation

J. Thurin,[1] *R. Brossier*[1] *and L. Métivier*[2,1]
[1] *Univ. Grenoble Alpes, ISTerre,* [2] *Univ. Grenoble Alpes, CNRS, LJK*

SUMMARY

Uncertainty Quantification is a major topic for most geophysical tomography techniques, in particular for large-scale problems. In this work, we present an original application of ensemble-based methods to Full Waveform Inversion. This approach relies on a deterministic Ensemble-Transform Kalman Filter borrowed from the Data Assimilation community, and a frequency-domain Full Waveform Inversion. This methodology gives access to a low-rank version of the posterior covariance matrix of our inverse problem, thanks to the ensemble repartition. We can thus extract information from this covariance matrix to assess uncertainty in the Bayesian sense. This proof-of-concept is applied to a 2D Marmousi case, before discussing many questions associated with the design of the scheme.

INTRODUCTION

Geophysical tomography allows retrieving subsurface physical properties from discrete surface measurements. While many different techniques exist, Full Waveform Inversion (FWI) has become, for a decade, a widely used method thanks to its high-resolution and quantitative outputs, both in the academia and in the exploration industry (Plessix, 2009; Sirgue et al., 2010; Plessix et al., 2012; Warner et al., 2013; Fichtner et al., 2013; Zhu et al., 2015; Operto et al., 2015). These detailed outputs come, however, at the expense of a more complex inverse problem to be solved, compared to travel time tomography, as the entire recorded wavefield is used as data to be fit. Since the advent of the technique (Lailly, 1983; Tarantola, 1984), most of the research has been dedicated to make the approach working, focusing on theoretical and methodological developments, computational hardware capacities and seismic acquisition design. The method is becoming now more mature, but uncertainty estimation has largely been left aside and very few studies in the literature have tried to tackle this problem. The quantification of uncertainties remains a significant challenge for most targets. Up to now, most quality control relies on data fit, *in-situ* comparisons with well-log data, common-image gathers flatness or by comparing with other geophysical methods.

In order to assess uncertainties, the Bayesian Inference Framework for general inverse problems, as developed by Tarantola (2005), may provide a good solution. Tarantola (2005) states that close to the global minimum of an optimization problem, the inverse Hessian operator is also the posterior covariance matrix. In a least-squares sense and under the assumption that the global minimum is reached, these operators would, therefore, be crucial for uncertainty estimation. For realistic large-scale FWI problems, however, computing those operators is out of reach, as they grow larger with the problem size. While the matrix-free methods proposed by Fichtner and Trampert (2011a); Métivier et al. (2013, 2014) allow assessing the effect of the Hessian matrix on a vector, their use for estimating the whole inverse Hessian remains challenging. Promising methods have been proposed to decrease the computational cost of Hessian (or inverse Hessian) estimation by selecting prior parameterization of the Hessian or employing random probing strategies. This enables some uncertainty estimation (Fichtner and Trampert, 2011b; Fang et al., 2014; Zhu et al., 2016; Fichtner and van Leeuwen, 2015).

Aside those works, the Data Assimilation (DA) community has developed tools to solve complex non-linear inverse problems, with similar difficulties as encountered in FWI. Typical applications are found in weather forecasting, climatology, and oceanography. As opposed to FWI, where the goal is parameter estimation in a static case, DA often aims at recovering the state of a dynamic system given a set of sparse measurements trough time. Despite this difference, we believe some of the DA methods could be useful for uncertainty estimation in FWI. Indeed, due to the nature of their applications, DA community focused early on uncertainty quantification and quality control, with schemes such as the Kalman Filter (KF) algorithm (Kalman, 1960), which combines cleverly modeling and measurements while accounting for their errors. However, KF is only suited for small linear problems due to its extensive manipulation of covariance and noise matrices. While Extended KF proposed to cope with non-linearity, it is still not relevant for high-dimensional cases. Thanks to Evensen (1994, 2009) and his formulation of Ensemble Kalman Filter (EnKF), explicit manipulation of such large matrices can be avoided, relying on low-rank approximations contained through the repartition of an ensemble of states. Thus, EnKF appears to be well suited for large scale applications and is nowadays an operational tool in the DA community for solving problems with similar scales that found in FWI.

Du et al. (2012); Jordan (2015) have suggested ensemble-based approaches for tomographic problems. Their approaches do not include, however, least-squares analysis, characteristic of KF algorithms. Jin et al. (2008) also proposed an EnKF scheme for 1D prestack FWI based on a convolutional model. In this abstract, we intend to propose a combination of EnKF with the FWI problem, in order to access a low-rank approximation of the posterior covariance matrix. This allows reaching quantitative uncertainty information for large scale FWI problems. The difficulty to adapt this peculiar method used in dynamic applications, to the static FWI case, is discussed. In a first part, we propose a short review of the EnKF and Ensemble Transform Kalman Filter (ETKF) formalism. The details of our ETKF-FWI method are then exposed before showing some preliminary results of this proof of concept. Finally, we address the challenges and open questions associated to this original scheme.

BRIEF ENKF REVIEW

Kalman Filter allows retrieving the state of a linear dynamic system, using the modeling and measurement properties, and their respective errors. Schemes of the KF family are used to define the best tradeoff between modeling and measurements. KF-schemes are classically formalized as alternated two steps algorithms (figure 1) :

Prediction step: During this step, the state m of a dynamic system is forecast from time/state k to $k+1$ by applying a modeling operator $\mathscr{F}$,

$$m^f_{k+1} = \mathscr{F}(m_k), \tag{1}$$

where the superscript f denotes the forecast system.

Analysis step: Using measured data and forecast state at $k+1$, the analyzed system state at $k+1$ is computed in the least-squares sense, weighted over modeling and measurement errors ratio. The analyzed system state is generally defined by the superscript a.

The EnKF application requires the definition of an ensemble $\mathbf{m} = \left[m_{1,k}, ..., m_{N_e,k}\right]$ Matrices will be denoted by bold letters and transpose operator by the superscript ^{T} in the following. This matrix contains N_e state vectors, each of them the size N, the number of state parameters. From the ensemble, the mean $\bar{m}$ and the perturbation matrix $\mathbf{M}$ are given by

$$\bar{m}_k = \frac{1}{N_e}\sum_{i=1}^{N_e} m_{i,k}, \tag{2}$$

$$\mathbf{M}_k = \left[m_{1,k} - \bar{m}_k, ..., m_{N_e,k} - \bar{m}_k\right], \tag{3}$$

From equations (2) and (3) the ensemble covariance matrix $\mathbf{P}_{e,k}$ is computed as

$$\mathbf{P}_{e,k} = \frac{1}{N_e - 1}\sum_{i=1}^{N_e}(m_{i,k} - \bar{m}_k)(m_{i,k} - \bar{m}_k)^T = \frac{1}{N_e - 1}\mathbf{M}_k\mathbf{M}_k^T. \tag{4}$$

This covariance can be extracted from the ensemble repartition during both the forecast and analysis steps.

Each ensemble member is forecast independently by applying the operator $\mathscr{F}$. It is worth mentioning that it is an embarrassingly parallel process as members are independents. During the analysis, the forecasts and measured observations are combined, in the least-squares sense. We choose here to follow the formalism of the deterministic Ensemble Transform KF (ETKF, Bishop et al., 2001), among the variety of existing EnKF formalisms. Deterministic EnKF is chosen, as they converge toward the solution faster than their stochastic counterparts and ETKF analysis phase also has a lower algorithmic cost (Tippett et al., 2003). EnKF relies on the definition of $\mathbf{P}_e$ as a product of square root matrices, based on equation (4) (subscript k is removed for clarity in the following equations). The update during the analysis requires to first compute $\mathbf{M}^a$ according to definition

$$\mathbf{M}^a = \mathbf{M}^f\mathbf{T}, \tag{5}$$

where $\mathbf{T}$ is a transformation matrix of size (N_e, N_e). The least-squares formalism tells us that

$$\mathbf{T}\mathbf{T}^T = \left(\mathbf{I}_{N_e} + \frac{1}{N_e - 1}\mathbf{Y}^{f^T}\mathbf{R}^{-1}\mathbf{Y}^f\right)^{-1}, \tag{6}$$

$\mathbf{R}$ is the measurement noise matrix, $\mathbf{I}_{N_e}$ the identity matrix of size (N_e, N_e) and $\mathbf{Y}$ is the perturbation observation matrix defined as

$$\mathbf{Y} = \left[d_1 - \bar{d}, d_2 - \bar{d}, ..., d_{N_e} - \bar{d}\right], \tag{7}$$

with the observations mean

$$\bar{d} = \frac{1}{N_e}\sum_{i=1}^{N_e} d_i. \tag{8}$$

For a non-linear observation operator $\mathscr{H}$, we have $d_i = \mathscr{H}(m_i)$.

However, determining the square root of the operator $\mathbf{T}\mathbf{T}^T$ is a nonunique problem. Wang et al. (2004); Ott et al. (2004) propose to use the truncated singular value decomposition (SVD) of $\mathbf{T}\mathbf{T}^T$, giving

$$\mathbf{T}\mathbf{T}^T = \mathbf{C}\Gamma\mathbf{C}^T \rightarrow \mathbf{T} = \mathbf{C}\Gamma^{-1/2}\mathbf{C}^T. \tag{9}$$

Here, $\mathbf{C}$ is the singular vectors matrix and Γ the diagonal matrix containing the truncated singular values of $\mathbf{T}\mathbf{T}^T$. If the ensemble members are uncorrelated, the rank of $\mathbf{T}\mathbf{T}^T$ can be shown to be $\min(N_e - 1, N_{obs})$ with N_{obs} the number of observations. Recall that these steps have a low computational cost, since the operator in (6) is only of size N_e.

From the definition of $\mathbf{T}$, the updated $\mathbf{M}^a$ and $\bar{m}^a$ are given by

$$\mathbf{M}^a = \sqrt{N_e - 1}\mathbf{M}^f\mathbf{C}\Gamma^{-1/2}\mathbf{C}^T, \tag{10}$$

$$\bar{m}^a = \bar{m}^f + \mathbf{M}^f\mathbf{C}\Gamma^{-1}\mathbf{C}^T\mathbf{Y}^{f^T}\mathbf{R}^{-1}(d_{obs} - \bar{d}), \tag{11}$$

giving the new analysed ensemble $\mathbf{m}^a = \bar{m}^a + \mathbf{M}^a$.

The whole analysis phase is equivalent to the following variational minimization problem (Hunt et al., 2007)

$$\mathscr{C}(m) = \frac{1}{2}(m - \bar{m}^f)^T\mathbf{P}^{-1}(m - \bar{m}^f) + \frac{1}{2}(d_{obs} - \bar{d})^T\mathbf{R}^{-1}(d_{obs} - \bar{d}), \tag{12}$$

which combines, in the least-squares sense, the forcast state and the data, considering their respective uncertainties.

ETKF-FWI SCHEME

FWI is a static inverse problem, which does not relate directly to dynamic/evolutive problems specific to DA. Our idea is to use hierarchical strategies commonly used in FWI could conveniently replace temporal evolution in ETKF-FWI. The most common strategy in FWI is the frequency-continuation, originally employed to limit cycle skipping (Bunks et al., 1995; Sirgue and Pratt, 2004). As a first proposition, we chose to replace the time-evolution of DA by increasing frequency band selection. Then, the state vector m must be defined, as well as the content of the data vector d, and the observation $\mathscr{H}$ and forecasting $\mathscr{F}$ operators. In our application, m encapsulate the physical properties of the subsurface. From the non-linear operator $\mathscr{H}$, d is defined as the seismic wavefield recorded at the receivers. Finally, our chosen forecasting operator $\mathscr{F}$ corresponds to the non-linear FWI process for a given initial model $m_{i,k}$ at one specified frequency band. Thus we have,

$$m^f_{i,k+1} = \mathscr{F}(m_{i,k}) = \underset{m_{i,k}}{\operatorname{argmin}}\frac{1}{2}\|\mathscr{H}(m_{i,k}) - d_{obs,k}\|^2 \tag{13}$$

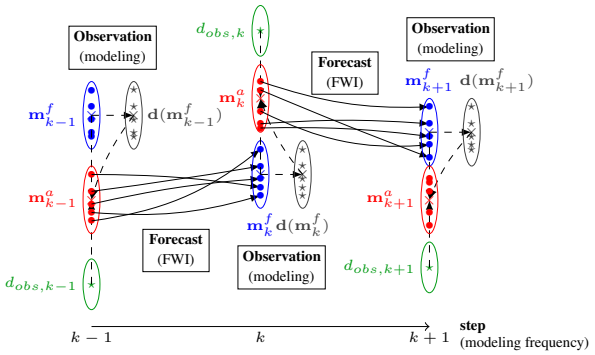

Figure 1: EnKF algorithm schematics. Ensemble's members are represented by dots, data by stars and ellipses represent uncertainty. The forecast ensemble is denoted in blue, the analyzed ensemble in red, the observed data in green, and the modeled data from the forecast are depicted in gray. In bold we have the general EnKF operations while in parenthesis we have the associated case for our ETKF-FWI application. The dashed lines denotes the Analysis steps.

which provides the model minimizing the *l2 norm* of the misfit between modeled data $\mathscr{H}(m_{i,k}) = d_{i,k}$ and measured data $d_{obs,k}$. The ETKF-FWI scheme is represented in figure 1.

The common way to generate an ensemble with a given statistics would be to first factorize the desired covariance matrix with a Cholesky decomposition as $\mathbf{P} = \mathbf{L}\mathbf{L}^T$, and then building a vector v satisfying this covariance by

$$v = \mathbf{L}u, \tag{14}$$

from a random vector u. However, targeting large-scale applications, a Cholesky decomposition is not achievable. A straightforward and pragmatical way to generate the initial ensemble is considered instead. The population is built from a consistent starting model m_0, by considering N_e zero mean random vectors u_i (white noise), convolved with a 2D Gaussian filter with realistic correlation lengths. Each ensemble member can be considered as $m_{0,i} = m_0 + \mathscr{G}u_i$, with $\mathscr{G}$ the convolution operator with the 2D Gaussian filter. Thus, the prior covariance obtained with this ensemble generation strategy is a Gaussian squared as $\mathbf{P} = \mathscr{G}\mathscr{G}^T$.

APPLICATION - MARMOUSI EXAMPLE

In this part, a synthetic experiment is conducted on the 2D Marmousi model (fig.2-A) with our ETKF-FWI strategy. 2D visco-acoustic frequency-domain is chosen for this application. The formulation and operators are set as described in the previous section and frequency evolution replace the dynamic evolution. 25 dB of white noise have been applied to the data, preventing noise-free inverse-crime and reviewing the sensibility of the technique to noise. The acquisition is a fixed spread surface geometry with 144 sources and 660 receivers at a 25m depth position under water surface. ETKF-FWI has been applied from 3 to 15Hz, each 0.4Hz, from m_0 initial model (fig.2-B). The 200 ensemble members are generated by using a 500m correlation length Gaussian filter and ensuring perturbations around the mean with amplitudes ranging from $-100m.s^{-1}$ to $100m.s^{-1}$. Measurement noise matrix $\mathbf{R}$ is considered as diagonal with small values compatibles with the set level of noise.

Fig. 2-C shows the result of the ETKF-FWI workflow. This result is similar to the model that could be obtained from FWI alone with the same settings, starting from m_0. This also implies that our initial model was sufficiently good to ensure convergence.

The approximated covariance matrix is extracted as a low-rank version from the ensemble repartition. The covariance for the velocity is given in $m^2.s^{-2}$ and represents the local diversity through the ensemble members. The variance (diagonal of the covariance matrix) can be displayed as a 2D map in fig. 2-D. This result corresponds to expectations from the theoretical understanding of the FWI problem with a surface acquisition setup, in term of uncertainty quantification. The variance map can be interpreted as the superimposition of a low-wavenumber background and a high-wavenumber perturbation. The low-wavenumber background has low variance values near the acquisition and progressively increases with distance and depth due to the decrease of wavefield coverage and wave amplitude with the geometrical spreading. The high-wavenumber component of the variance map highlights the interfaces. This can be attributed to band-limited data, which does not constrain the solution enough.

DISCUSSION AND CONCLUSION

ETKF-FWI seems to be a powerful and straightforward method allowing uncertainty quantification in FWI. Variance maps are easily readable to evaluate inversion results, and resolution could be studied from lines of $\mathbf{P}$. Still, this original application set-up many questions that will require extensive work.

First, the actual meaning of "uncertainty" as extracted from the ensemble must be understood. Working with finite-frequency waves propagation and limited coverage, cause a filter-like effect. Thus can the quantitative uncertainties be associated with direct uncertainty on real physical parameters? Alternatively, will it only be able to account for the apparent macro-model seen by the waves, as questioned by Capdeville et al. (2010) for homogenization and down-scaling problems?

Considering the current state of development of our method, many points need to be explored: How to design the measurement noise matrix $\mathbf{R}$? This parameter should be simple to consider and be related with the recording noise level and sensor design. Up to now, the process noise matrix has been left aside but should be associated with the noise and error of the forecasting operator. Classically, it is a troublesome parameter to estimate even for linear operators in DA. In the frame of ETKF-FWI, for which the FWI process is considered as forecasting, estimating this matrix could be a challenge but will ultimately be needed in practical applications. The initial ensemble repartition, directly linked to the prior covariance $\mathbf{P}$, is also an open question. A pragmatical approach is to use a Gaussian filter, leading to a Gaussian squared covariance. However,

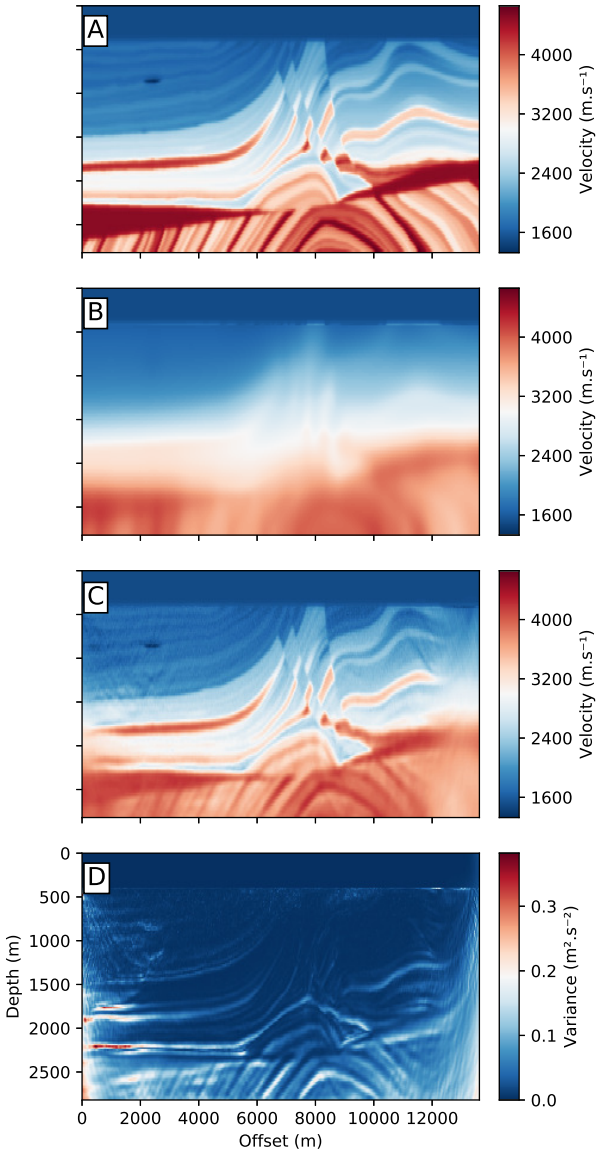

Figure 2: Velocity models and variance map associated with the experiment depicted in the Application section. A) 2D Marmousi true velocity model. B) ETKF-FWI initial model m_0. C) ETKF-FWI final mean model after 30 assimilation steps from $3Hz$ to $15Hz$ each 0.4Hz. D) ETKF-FWI final variance map after 30 assimilation step from $3Hz$ to $15Hz$ each 0.4Hz.

other filters as Laplace or Bessel filter (Trinh et al., 2017) may be relevant if used with the same strategy. These filters directly affect the spatial shape of the covariance between parameters, but also its amplitude. This amplitude should be cautiously set, as too large values could lead to significant kinematics differences in the data resulting in cycle-skipping. However, the values should be sufficiently large to ensure satisfying exploration of the model space and provide meaningful information about the misfit function's local curvature. Dramatically low values should also be avoided to ensure enough diversity in the ensemble repartition and avoid ensemble collapse. The number of ensemble members is also an important point to be assessed. EnKF usually involves few tenths to hundreds of members for large scale problems. The optimal number of members for ETKF-FWI scheme is still to be determined and may vary according to model complexity/size and acquisition design.

The "dynamic" strategy of the proposed method may also be re-designed. For frequency-domain FWI, frequency selection seems a natural hierarchy. This proxy for dynamic evolution may be extended or replaced with frequency bands and/or time-windows for time-domain FWI. The combination with data subsampling, shot decimation or source assembling (Krebs et al., 2009; Warner et al., 2013) may prove to be pertinent. This would also result in a reduction of the technique's cost.

A more global view of the approach also leads to questioning the variables and observations of the filter itself. Only the velocity properties have been accounted for up to now, but multi-parameters unknowns are inherently easy to consider. Well-log data, for instance, may be used as a direct observations or constraint in addition to the seismic wavefield. The entire wavefield may also be considered an as unknown variable in the EnKF, making some link with Wavefield Reconstruction Inversion proposed by van Leeuwen and Herrmann (2013), as both the physical parameter and the wavefield would be considered as unknowns.

Finally, the differences and added value of the proposed approach must be explicitly evaluated on other methods from the literature. The interest of using the analysis step of EnKF, compared to a purely independent ensemble approach as used by Du et al. (2012); Jordan (2015) needs to be determined. How does the method compare to a Markov chain Monte Carlo sampling of the misfit function at the convergence point as proposed by Fang et al. (2014), or the random sampling of the Hessian at the convergence point as suggested by Fichtner and van Leeuwen (2015)?

Of course, one point to carefully consider is the computational cost, as each ensemble members requires to solve the FWI problem on its own, increasing the cost of one or two order of magnitude compared to classical FWI. Nonetheless, reminding that this scheme is embarrassingly parallel and thanks to the development of hardware capacities towards the exascale and the current trends in grid computing, ETKF-FWI applications may be promptly achievable even for large scale FWI problems, as it is the case for actual DA problems.

ACKNOWLEDGMENTS

This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by CGG, CHEVRON, EXXON-MOBIL, JGI, SHELL, SINOPEC, STATOIL, TOTAL and WOOD-SIDE. This study was granted access to the HPC resources of the Froggy platform of the CIMENT infrastructure (https://ciment.ujf-grenoble.fr), which is supported by the Rhône-Alpes region (GRANT CPER07_13 CIRA), the OSUG@2020 labex (reference ANR10 LABX56) and the Equip@Meso project (reference ANR-10-EQPX-29-01).

EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2017 SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001, Adaptive sampling with the ensemble transform kalman filter. Part I: Theoretical aspects: Monthly Weather Review, **129**, 420–436, http://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Bunks, C., F. M. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473, http://doi.org/10.1190/1.1443880.

Capdeville, Y., L. Guillot, and J.-J. Marigo, 2010, 2-D non-periodic homogenization to upscale elastic media for P-SV waves: Geophysical Journal International, 182, 903–922, http://doi.org/10.1111/j.1365-246X.2010.04636.x.

Du, Z., E. Querendez, and M. Jordan, 2012, Resolution and uncertainty in 3D stereotomographic inversion: 74th Annual International Conference and Exhibition, EAGE, Extended Abstracts, http://doi.org/10.3997/2214-4609.20148576.

Evensen, G., 1994, Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics: Journal of Geophysical Research, **99**, 143–162, http://doi.org/10.1029/94JC00572.

Evensen, G., 2009, Data assimilation: The ensemble Kalman filter: Springer.

Fang, Z., F. J. Herrmann, and C. D. Silva, 2014, Fast uncertainty quantification of 2D full-waveform inversion with randomized source subsampling: 76th Annual International Conference and Exhibition, EAGE, Extended Abstract, http://doi.org/10.3997/2214-4609.20140715.

Fichtner, A., and J. Trampert, 2011a, Hessian kernels of seismic data functionals based upon adjoint techniques: Geophysical Journal International, **185**, 775–798, http://doi.org/10.1111/j.1365-246X.2011.04966.x.

Fichtner, A., and J. Trampert, 2011b, Resolution analysis in full waveform inversion: Geophysical Journal International, **187**, 1604–1624, http://doi.org/10.1111/j.1365-246X.2011.05218.x.

Fichtner, A., J. Trampert, P. Cupillard, E. Saygin, T. Taymaz, Y. Capdeville, and A. V. nor, 2013, Multiscale full waveform inversion: Geophysical Journal International, 194, 534–556, http://doi.org/10.1093/gji/ggt118.

Fichtner, A., and T. van Leeuwen, 2015, Resolution analysis by random probing: Journal of Geophysical Research: Solid Earth, **120**, 5549–5573, http://doi.org/10.1002/2015JB012106.

Hunt, B., E. Kostelich, and I. Szunyogh, 2007, Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter: Physica D: Nonlinear Phenomena, 230, 112–126, http://doi.org/10.1016/j.physd.2006.11.008.

Jin, L., M. K. Sen, and P. L. Stoffa, 2008, One-dimensional prestack seismic waveform inversion using ensemble kalman filter: 78th Annual International Meeting, SEG, Expanded Abstracts, 1920–1924, http://doi.org/10.1190/1.3063815.

Jordan, M., 2015, Estimation of spatial uncertainties in tomographic images: 77th Annual International Conference and Exhibition, EAGE, Extended Abstracts, http://doi.org/10.3997/2214-4609.201413555.

Kalman, R., 1960, A new approach to linear filtering and prediction problems: Journal of basic Engineering, **82**, 35–45, https://doi.org/10.1115/1.3662552.

Krebs, J., J. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M. D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, **74**, no. 6, WCC105–WCC116, http://doi.org/10.1190/1.3230502.

Lailly, P., 1983, The seismic problem as a sequence of before-stack migrations: Presented at the Conference on Inverse Scattering: Theory and Applications, SIAM, Philadelphia.
Métivier, L., F. Bretaudeau, R. Brossier, S. Operto, and J. Virieux, 2014, Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures: Geophysical Prospecting, **62**, no. 6, 1353–1375, http://doi.org/10.1111/1365-2478.12136.
Métivier, L., R. Brossier, J. Virieux, and S. Operto, 2013, Full waveform inversion and the truncated Newton method: SIAM Journal On Scientific Computing, **35**, B401–B437, http://doi.org/10.1137/120877854.
Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti, and J. Virieux, 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: Geophysical Journal International, **202**, 1362–1391, http://doi.org/10.1093/gji/ggv226.
Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. Patil, and J. A. Yorke, 2004, A local Ensemble Kalman filter for atmospheric data assimilation: Tellus A, **56**, 415–428, http://doi.org/10.1111/j.1600-0870.2004.00076.x.
Plessix, R. E., 2009, Three-dimensional frequency-domain full-waveform inversion with an iterative solver: Geophysics, **74**, no. 6, WCC53–WCC61, http://doi.org/10.1190/1.3211198.
Plessix, R.-E., G. Baeten, J. W. de Maag, and F. ten Kroode, 2012, Full waveform inversion and distance separated simultaneous sweeping: a study with a land seismic data set: Geophysical Prospecting, **60**, 733–747, http://doi.org/10.1111/j.1365-2478.2011.01036.x.
Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010, Full waveform inversion: The next leap forward in imaging at Valhall: First Break, **28**, 65–70, http://doi.org/10.3997/1365-2397.2010012.
Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies: Geophysics, **69**, 231–248, http://doi.org/10.1190/1.1649391.
Tarantola, A., 1984, Linearized inversion of seismic reflection data: Geophysical Prospecting, **32**, 998–1015, http://doi.org/10.1111/j.1365-2478.1984.tb00751.x.
Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: Society for Industrial and Applied Mathematics.
Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003, Ensemble square root filters: Monthly Weather Review, **131**, 1485–1490.
Trinh, P.-T., R. Brossier, L. Métivier, J. Virieux, and P. Wellington, 2017, Bessel smoothing filter for spectral element mesh: Geophysical Journal International, **209**, 1489-1512, https://doi.org/10.1093/gji/ggx103.
van Leeuwen, T., and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, http://doi.org/10.1093/gji/ggt258.
Wang, X., C. H. Bishop, and S. J. Julier, 2004, Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble?: Monthly Weather Review, **132**, 1590–1605, https://doi.org/10.1175/1520-0493(2004)132%3C1590:wibaeo%3E2.0.co;2.
Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3D full-waveform inversion: Geophysics, **78**, R59–R80, http://doi.org/10.1190/geo2012-0338.1.
Zhu, H., E. Bozdag, and J. Tromp, 2015, Seismic structure of the European upper mantle based on adjoint tomography: Geophysical Journal International, **201**, 18–52, https://doi.org/10.1093/gji/ggu492.

Zhu, H., S. Li, S. Fomel, G. Stadler, and O. Ghattas, 2016, A bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration: Geophysics, **81**, R307–R323, https://doi.org/10.1190/geo2015-0641.1.