# Tu P1 07

# Ensemble-Based Uncertainty Estimation in Full Waveform Inversion

J. Thurin* (Univ. Grenoble Alpes), R. Brossier (Univ. Grenoble Alpes), L. Métivier (Univ. Grenoble Alpes - CNRS)

## Summary

Uncertainty Quantification is a major issue for most tomography problems. In this work, We propose an original application of ensemble-based methods in Full Waveform Inversion (FWI). The method relies on classical FWI-schemes coupled with a deterministic version of an Ensembl Kalman Filter (EnKF). This approach allows accessing a low-rank estimation of the posterior covariance matrix, giving us access to quantitative uncertainty measurements as long as we are close enough to the global minimum. Applications to 2D FWI in the frequency domain have proven to be successful, and the variance maps obtained from joint EnKF-FWI are encouraging premise for this method

## Introduction

Full Waveform Inversion (FWI) is a powerful technique aiming at retrieving high-resolution subsurface physical properties. The FWI literature shows that the primary research topic for decades has been to make the concept working. Combining the understanding of the FWI concept, the developments of computational capacities and the design of acquisitions, FWI has become a routine for many academic and industrial targets. However, the quality control and error estimations are still major issues, and the literature shows that only a few studies have tackled the problem of quantitative uncertainty estimation in FWI (Fichtner and Trampert, 2011; Fichtner and van Leeuwen, 2015), mainly because of the large scale of the problem and its computational cost.

In parallel of the seismic/seismology community, the Data Assimilation (DA) community has designed a set of methods to solve inverse problems related to dynamic problems in time, such as weather forecasting or climatology. From the very beginning of these developments, DA community has focused on methods allowing to estimate both the solutions of their inverse problems and their uncertainties. One standard approach is the Kalman Filter (KF) algorithm (Kalman, 1960), aiming to combine a set of recorded data and forecasted models, taking into account for the bias introduced by modeling and measurement errors. However, the KF and its extension to non-linear problems can not handle large scale problems and tackling high-dimension problems has become possible with the Ensemble Kalman Filter (EnKF) (Evensen, 1994). The main idea behind EnKF is to avoid any explicit storage or manipulation of covariance matrices, by using an ensemble which contains this information (through a low-rank representation) implicitly. It is noteworthy that EnKF is an operational tool commonly employed by the DA community in weather forecasting which can cope with problem sizes of the same order than those associated with realistic FWI applications.

Some ensemble-based approaches have been proposed by Du et al. (2012); Jordan (2015) for some tomography problems for instance, but without using the least-square analysis step behind EnKF. Jin et al. (2008) also proposed an EnKF scheme for 1D prestack FWI based on a convolutional model. In this work, we propose and apply a concept of combining the EnKF frame and the FWI problem, for accessing a low-rank approximation of the covariance matrix, allowing to obtain valuable information on the uncertainty.

## Theory

The general purpose of Kalman filtering is, as any alpha-beta filter, to determine what is the best tradeoff between information brought by the modeling and the data that we measure. In EnKF, one must define an ensemble $\mathbf{m}$, which is a matrix composed of $N_e$ state vectors of size $N$ (number of state parameters). In FWI applications, $N_e$ will be in the range of $10^1$ to $10^2$ while $N$ will typically be $10^6$ to $10^9$. From this ensemble, the mean

$$\bar{m} = \frac{1}{N_e}\sum_{i=1}^{N_e} m_i \tag{1}$$

and the perturbation matrix

$$\mathbf{M} = [m_1 - \bar{m}, m_2 - \bar{m}, \ldots, m_{N_e} - \bar{m}], \tag{2}$$

can be deduced from the ensemble repartition to compute its covariance matrix

$$\mathbf{P}_e = \frac{1}{N_e - 1}\sum_{i=1}^{N_e}(m_i - \bar{m})(m_i - \bar{m})^T = \frac{1}{N_e - 1}\mathbf{M}\mathbf{M}^T, \tag{3}$$

where the superscript $^T$ denotes the transpose operation.

EnKF is classically divided in two main steps : (1) During the forecast part (superscript $f$), each ensemble member is forecasted independently from time/state $k$ to $k+1$ by applying a modeling operator $\mathscr{M}$, such as $m^f_{k+1} = \mathscr{M}(m_k)$. There is no interaction between members during this step. (2) During the analysis part (superscript $a$), the objective is to combine forecasts and observations, in the least-square sense, to obtain the best estimation of the system state at $k+1$. Among the various implementations of EnKF, we choose here to follow the formalism of the deterministic Ensemble Transform KF (ETKF, Bishop et al., 2001). During the analysis step, we compute the analyzed perturbation matrix

$$\mathbf{M}^a = \mathbf{M}^f\mathbf{T}, \tag{4}$$

where $\mathbf{T}$ is a symetric transformation matrix of size the number of elements in the ensemble, which can be deduced from a square root of the operator

$$\mathbf{T}\mathbf{T}^T = \left(\mathbf{I}_{N_e} + \frac{1}{N_e - 1}\mathbf{Y}^{f^T}\mathbf{R}^{-1}\mathbf{Y}^f\right)^{-1}, \tag{5}$$

where $\mathbf{R}$ is the measurement noise matrix, $\mathbf{I}_{N_e}$ an identity matrix of size the number of elements in the ensemble and $\mathbf{Y}$ is the perturbation observation matrix defined as

$$\mathbf{Y} = \left[d_1 - \bar{d}, d_2 - \bar{d}, ..., d_{N_e} - \bar{d}\right], \tag{6}$$

with the observation mean

$$\bar{d} = \frac{1}{N_e}\sum_{i=1}^{N_e} d_i. \tag{7}$$

For a linear observation operator $\mathbf{H}$, we have $d_i = \mathbf{H}m_i$, while we have $d_i = \mathscr{H}(m_i)$ for a non-linear version. Wang et al. (2004); Ott et al. (2004) have shown that using a truncated singular value decomposition (SVD) of equation (5), we can have

$$\mathbf{T} = \mathbf{C}\Gamma^{-1/2}\mathbf{C}^T, \tag{8}$$

where $\mathbf{C}$ is the singular vectors matrix and $\Gamma$ the diagonal matrix containing the truncated singular values of $\mathbf{T}\mathbf{T}^T$. If the ensemble members ar 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. It has to be highlighted that the linear algebra operations are quite negligible here, as the SVD involves a matrix the size of the ensemble and not linked to the observation or state sizes. From the definition of $\mathbf{T}$, we can compute the updated $\mathbf{M}^a$ and $\bar{m}^a$ with

$$\mathbf{M}^a = \sqrt{N_e - 1}\mathbf{M}^f\mathbf{C}\Gamma^{-1/2}\mathbf{C}^T, \tag{9}$$

$$\bar{m}^a = \bar{m}^f + \mathbf{M}^f\mathbf{C}\Gamma^{-1}\mathbf{C}^T\mathbf{Y}^f\mathbf{R}^{-1}(d_{obs} - d), \tag{10}$$

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

**ETKF-FWI scheme**

FWI, as other tomographic problem does not rely on a dynamic time-evolution problem as classically found in DA. Therefore, we have to formulate how to combine ETKF and FWI. Classical FWI approaches rely on a hierarchical scheme with frequency-continuation for limiting cycle-skipping (Bunks et al., 1995; Sirgue and Pratt, 2004). As a first proposition, we choose to replace the usual time-evolution of DA by the frequency-band selection. We also have to choose what to include in the state vector $m$, in the observation vector $d$, and in the observation $\mathscr{H}$ and forecasting $\mathscr{F}$ operators. In this work, we have considered to put only the physical properties of the Earth in $m$, and to consider the seismic observation in $d$. That leads us to consider a non-linear observation $\mathscr{H}$

$$d = \mathscr{H}(m), \tag{11}$$

which embeds the wave equation modeling and the extraction of the wavefield at the receiver positions.

Finaly, in our formulation, we associate the forecasting operator $\mathscr{F}$ to the non-linear FWI process for a a given initial model, and a given frequency-band, giving

$$m_{k+1} = \mathscr{F}(m_k). \tag{12}$$

This whole ETKF-FWI scheme can be seen in Figure 1.

The usual way of generating an ensemble from a given covariance matrix is to factorize the covariance matrix with a Cholesky decomposition as $\mathbf{P} = \mathbf{L}\mathbf{L}^T$, such that a vector $v$ satisfying the covariance can be built as $v = \mathbf{L}u$, from a random vector $u$. Considering that we want our ETKF-FWI scheme to be applied to large-scale problems, a Cholesky decomposition is out of reach. Therefore, we use the following straightforward and pragmatical way to generate the initial ensemble. We built a population from our desired initial model $m_0$ (which would be used as starting model for regular FWI). We consider $N_e$ zero mean random vectors $u_i$ (white noise), smoothed out by a Gaussian filter with realistic correlation length, with respect to our starting frequency for inversion. Each ensemble member can, therefore, be considered as $m_{0,i} = m_0 + \mathscr{G}u_i$, with $\mathscr{G}$ the operator associated to the convolution with the Gaussian filter. Finally, the prior covariance associated with this ensemble generation is a Gaussian squared as $\mathbf{P} = \mathscr{G}\mathscr{G}^T$.

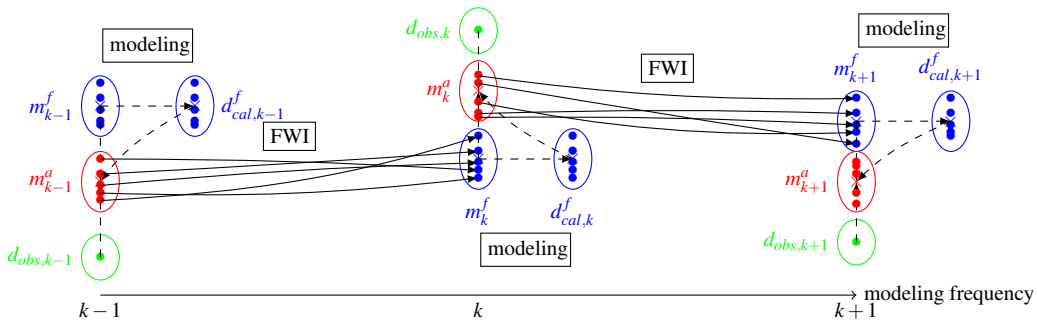


***Figure 1*** *Joint FWI-EnKF algorithm schematics. The assimilation is done over modeling frequencies. The forecasted ensemble is presented in blue, the analysed ensemble is in red, the observed data are in green and the modeled data from the forecast are depicted in gray. Modeling and FWI are the observation $\mathscr{H}$ and forecast $\mathscr{M}$ operators.*

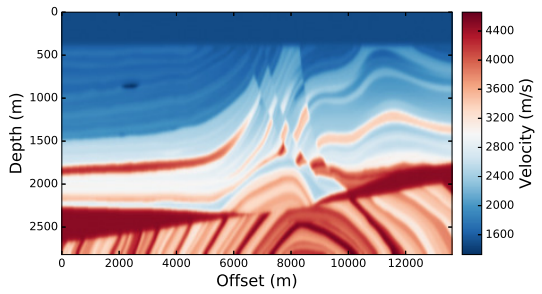


***Figure 2*** *Marmousi reference velocity model.*

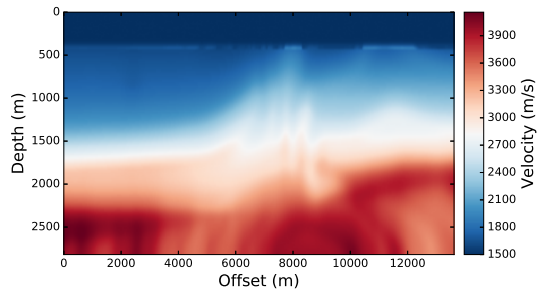


***Figure 3*** *Marmousi initial mean velocity model.*

## Application

Our example is an application to the Marmousi 2D model (Fig. 2). The goal is to evaluate the performances expected from a realistic experiment, with noisy data (25db of noise). A surface acquisition is used with 144 sources and 660 receivers at 25-meter depth in the water layer. The ETKF-FWI scheme has been applied from 3 to 15 Hz, using the smooth initial model of Figure 3. An ensemble of 200 members has been used, using a Gaussian filter of 500 m correlation length and ensuring perturbations from -100 m.s$^{-1}$ to 100 m.s$^{-1}$ around the initial mean model. Figure 6 is showing examples of starting perturbation models for 4 members of the ensemble. The measurement noise matrix $R$ has been set to maximize the effect of the analysis step and thus improve the convergence rate. The optimal value is still an open question, as the values in **R** have a strong impact on the assimilation result (low measurement noise).

Figure 4 shows the mean result after the ETKF-FWI workflow. This result is very similar to an FWI results that could have been obtained without any ETKF scheme using the mean initial model and frequencies from 3 to 15Hz. This result shows that the choice of the initial ensemble has allowed converging in the same minimum of the misfit function. The extraction of the diagonal of the posterior covariance matrix of the ETKF-FWI displayed as 2D "variance map", is presented in Figure 5. As predicted by the physical interpretation of the inverse Hessian, this variance is minimum close to sources and receivers, and progressively increase with distance (depth and bottom corners) due to the geometrical spreading and the decrease of wavefield coverage. In addition to this smoothly increasing component, a higher wavenumber component appears at each interfaces positions, highlighting again the region less constrained by the data. These features are quite consistent with expectations, showing again the useful information embedded in the ensemble covariance.

## Discussion and conclusions

Joint FWI-EnKF seems to provide an appealing method to quantify uncertainties in FWI. Variance maps are indeed a very convenient way to evaluate inversion results, while extraction of a line of the covariance matrix should allow quantifying the resolution and trade-offs. Still, this work set-up many questions related to the method itself that will require further and deeper investigations, the most obvious one being

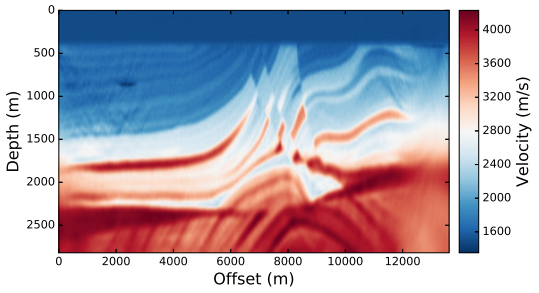


***Figure 4*** *Marmousi final mean velocity model after 30 assimilation step from* $3Hz$ *to* $15Hz$.

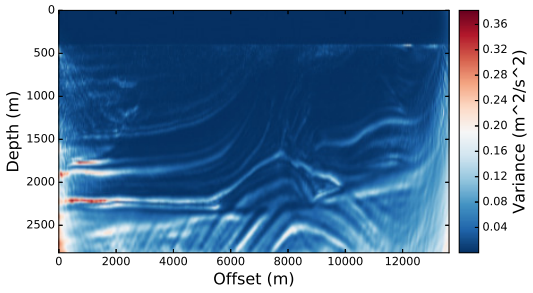


***Figure 5*** *Marmousi final variance (in* $m^2.s^{-2}$*) after 30 assimilation step from* $3Hz$ *to* $15Hz$.

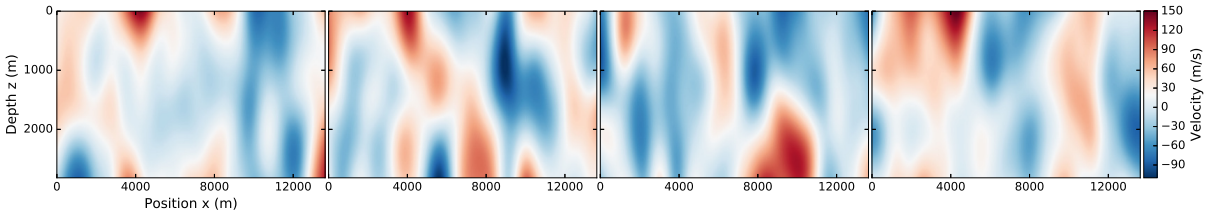


***Figure 6*** *Example of random gaussian perturbation applied to the starting model during the Ensemble generation.*

related to the design of the filter parameters: What would be the ideal number of ensemble members? What is the effect of the analysis step? How to set-up the measurement noise matrix? How to introduce a modeling noise matrix? How to prevent ensemble collapse? How to generate the initial ensemble?

Of course, ensemble-based methods would require one to two orders of magnitude more computations than the original FWI problem, but it is worth mentioning that the process is highly parallel and the development of hardware capacities toward the exascale should make this kind of approach valuable even for large scale problems in the future.

**Acknowledgements**

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 WOODSIDE. This study was granted access to the HPC resources of CIMENT infrastructure (*https://ciment.ujf-grenoble.fr*) and CINES/IDRIS under the allocation 046091 made by GENCI

**References**

Bishop, C.H., Etherton, B.J. and Majumdar, S.J. [2001] Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly weather review*, **129**(3), 420–436.

Bunks, C., Salek, F.M., Zaleski, S. and Chavent, G. [1995] Multiscale seismic waveform inversion. *Geophysics*, **60**(5), 1457–1473.

Du, Z., Querendez, E. and Jordan, M. [2012] Resolution and Uncertainty in 3D Stereotomographic Inversion. In: *Expanded Abstracts, 74th Annual EAGE Meeting (Copenhagen)*.

Evensen, G. [1994] Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **C5**(99), 143–162.

Fichtner, A. and van Leeuwen, T. [2015] Resolution analysis by random probing. *Journal of Geophysical Research: Solid Earth*, n/a–n/a. 2015JB012106.

Fichtner, A. and Trampert, J. [2011] Resolution analysis in full waveform inversion. *Geophysical Journal International*, **187**, 1604–1624.

Jin, L., Sen, M.K. and Stoffa, P.L. [2008] One-dimensional prestack seismic waveform inversion Using Ensemble Kalman Filter. In: *SEG Technical Program Expanded Abstracts 2008*. SEG, 1920–1924.

Jordan, M. [2015] Estimation of Spatial Uncertainties in Tomographic Images. In: *Expanded Abstracts, 77th Annual EAGE Meeting (Madrid)*.

Kalman, R. [1960] A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, **82**(1), 35–45.

Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D. and Yorke, J.A. [2004] A local Ensemble Kalman filter for atmospheric data assimilation. *Tellus A*, **56**, 415–428.

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

Wang, X., Bishop, C.H. and Julier, S.J. [2004] Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Monthly Weather Review*, **132**(7), 1590–1605.