Introduction

Seismic reflection data is still the primary source of data for subsurface imaging purposes, reservoir characterisation and during reservoir development. Here the objective is seismic inversion.

Modern fast computing power has made the methodology known as Full Waveform Inversion (FWI) feasible for seismic inversion. In a decision making situation it is natural to question the reliability of a (optimized) solution which is a matter of assessing the uncertainty of the estimate, and this is not straight forward to answer. Here the inversion problem is ported into a probabilistic setting where the concept of uncertainty quantification is conceptually related to the estimation itself.
In this setting the solution is a probability distribution with the estimated state being the most probable state and the distribution supplies information on its uncertainty. The approach taken here is an ensemble-based method where the probability distribution is represented by the samples within the ensemble. The basis is the Ensemble Kalman Filter (EnKF) which has been successfully applied in several areas with PDE based models. Its implementation is simple and can handle nonlinearity to some degree, see Evensen (2009) for derivation and examples of applications. Here it is demonstrated how the EnKF framework can be used for seismic data inversion.

Model

The parameter estimation is approached via Bayesian inversion with posterior distribution $p(\boldsymbol{m}|\boldsymbol{d}) \propto p(\boldsymbol{d}|\boldsymbol{m})\,p(\boldsymbol{m})$, where $p(\boldsymbol{m})$ is the prior distribution for the subsurface parameters and $p(\boldsymbol{d}|\boldsymbol{m})$ the likelihood function of the seismic reflection data.

The likelihood function is defined through the forward model $\boldsymbol{d}|\boldsymbol{m} = g(\boldsymbol{m}) + \boldsymbol{e}$ where $\boldsymbol{e}$ represents noise, primarily viewed as measurement noise. The data vector $\boldsymbol{d}$ is the vectorized shot gather, as it collects recorded wave amplitude at discrete time points in the measurement period, over several recording locations, as well as possibly several shot configurations (source positions). The observational or forward model operator $g(\boldsymbol{m})$ is the output from a wave equation solver with prescribed boundary and initial conditions, and source term function. The forward, black box used here is the spectral element solver SPECFEM2D (see e.g. Komatitsch and Tromp (1999)).
The prior distribution is a subjective choice and a difficult one, but in principle no constraints applies to it. Here we assume a multivariate Gaussian prior model $\boldsymbol{m} \sim \mathrm{N}(\boldsymbol{\mu_m}, \boldsymbol{\Sigma_m})$ whereas the error in the measurements is supposed to be zero-mean Gaussian noise so that $\boldsymbol{e} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma_e})$.

Data Assimilation

Inspired by the work that has been done in reservoir characterisation and history matching (see e.g. Oliver and Chen (2010)), an Ensemble Kalman Smoother (EnKS) is used for sequential data assimilation of seismic waveform data. The EnKS partitions all data available into N_c blocks and conditioning on these takes place sequentially. The application of Bayes' rule is rearranged according to the data blocking scheme, using the product form of the likelihood, such that the posterior from each conditioning cycle is the prior for the next cycle. The block of data used in each conditioning step can contain data from a single or multiple shots as this is a question of how one partitions the entire data set. After each conditioning the forward model is rerun with updated parameters.

The process is initiated with N_e samples from the prior distribution $\boldsymbol{m}_i^{\mathrm{f}} \sim p(\boldsymbol{m})$ and the forward model is run for each of these samples. In each cycle the linear update scheme which underlies all Ensemble Kalman Filter methodology is applied

$$\boldsymbol{m}_i^{\mathrm{a}} = \boldsymbol{m}_i^{\mathrm{f}} + \boldsymbol{K}\left(\boldsymbol{d}^{\mathrm{o}} - \boldsymbol{d}_i^{\mathrm{f}}\right) \text{ with } \boldsymbol{d}_i^{\mathrm{f}} = g(\boldsymbol{m}_i^{\mathrm{f}}) + \boldsymbol{e}_i,\, \boldsymbol{e}_i \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma_e}),\, i = 1, \ldots, N_e\,. \quad (1)$$

where $g(\cdot)$ here denotes the observation model applied so that it matches the data block of the current assimilation cycle. Then the updated ensemble samples $\boldsymbol{m}_i^{\mathrm{a}}$ are used as prior for the next cycle. When all data has been conditioned on, the resulting spread of the parameter ensemble constitutes an estimate of the uncertainty.

Estimation of the Kalman gain matrix $\boldsymbol{K}$ in Equation 1 by using the ensemble of forecast state- $\{\boldsymbol{m}_i^{\mathrm{f}}\}_{i=1}^{N_e}$ and observation realisations $\{\boldsymbol{d}_i^{\mathrm{f}}\}_{i=1}^{N_e}$ is the essence of EnKF. Using these ensembles for centered ensemble matrices $\boldsymbol{M}$ and $\boldsymbol{D}$ and finding the empirical covariance matrices, the Kalman gain estimation is $\widehat{\boldsymbol{K}} = \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{m},\boldsymbol{d}} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{d}}^{-1}$.

The EnKF methodology is known to have problems with ensemble collapse for high dimensional data vectors and spurious updates and some of the techniques to alleviate these issues are dimension reduction and localisation.

Dimension Reduction

The Kalman gain applied in the conditioning step can be equally regarded as the least squares estimation of multivariate regression coefficients $\widehat{\boldsymbol{K}} = \arg\min_{\boldsymbol{K}} (\boldsymbol{M} - \boldsymbol{K}\boldsymbol{D})^T (\boldsymbol{M} - \boldsymbol{K}\boldsymbol{D})$ which means other regression methods can be used to form the Kalman gain. Here Partial Least Squares (PLS) regression is used, which forms the regression coefficients $\boldsymbol{K}$ based on (projected) latent variables, derived from $\boldsymbol{M}$ and $\boldsymbol{D}$. The construction of regression coefficients based on lower-dimensional spaces makes the conditioning less exposed to severe over fitting which can cause the ensemble to collapse. So the conditioning still gets regressed based on principal features of the data but not all the high-frequency details of the noise. The PLS method generates projections that maximize covariance between $\boldsymbol{M}$ and $\boldsymbol{D}$ in contrast to Principal Components regression that only uses the information in $\boldsymbol{D}$. Also the PLS has shown very useful in situations with collinear data, which is the case with seismic data.

Selection of suitable PLS order i.e. the number of latent variables, is based on cross-validation techniques, see e.g. Sætrom and Omre (2010). The statistic upon which is selected the optimal order can be the minimum Predicted Residual Error Sum-of-Squares (PRESS), but it can also be other meaningful statistic. The noise samples present in $\boldsymbol{D}$ influences the statistic and hence the optimal choice, so a Monte Carlo approach to select the optimal order is taken by resampling the noise and repeating the cross-validation procedure. Among the Monte Carlo simulations the most frequently found optimal PLS order is chosen for the gain matrix.

Localisation

Localisation refers to limiting the region of influence between state variables, or between state variables and data, often justified by a spatio-temporal relationship. When at a given time instance a state variables describes a quantity distributed spatially at some mesh points and this quantity depends on other quantities discretised at same mesh, one can derive this region of influence based on understanding of dynamics.

In the considered problem context, the localisation should prevent a particular parameter being regressed on data points (time sample at specific recording location given a source position) that the parameter could not have influenced. I.e. an early time point at a recording location far from the source should not be influenced by a deeply located velocity parameter, as the source pulse simply would not had time to reach the particular parameter location, being reflected and propagated back to a distant recording location.
By this logic, a simple procedure to produce a mask matrix of the same dimensions as the Kalman gain matrix is created and applied though Schur multiplication with the gain matrix.

Example

An example using a simple layered model is presented. The domain is 2.5×2.5 km and only acoustic velocity v_p is considered. In the top of the domain is a water layer with depth 117 m and properties $(v_p, \rho) = (1500\mathrm{m/s}, 1\mathrm{g/cm^3})$. The domain is discretised into a uniform mesh of quadrilateral elements each of size ≈11 m×11 m, and the layered model assumption is reflected in the wave velocity being constant across rows of elements. Density is constant throughout the subsurface. Absorbing edges are

used along sides and bottom and perfect reflection boundary condition at top.
The water column has fixed parameters and so has the first layer just below the sea floor. The remaining 195 layers are variable and considered parameters to be inverted. At the bottom some rows are excluded as variable as these are used for the absorbing edge method.

The receiver array located in the water layer spans most of the domain width, with in-between spacing 25 m and depth 10 m. This gives a total of 93 receivers. A series of shots are positioned in 9 of these receiver locations, see Figure 1. Source type is a Ricker wavelet with dominant frequency 20 Hz and recording period is 3 sec.

The prior distribution is Gaussian with a smooth mean and long correlation range. The marginal variance increases linearly with depth and the correlation function is taken as a squared exponential with a range parameter of 500 m, such that parameters 500 m apart have 5% correlation.

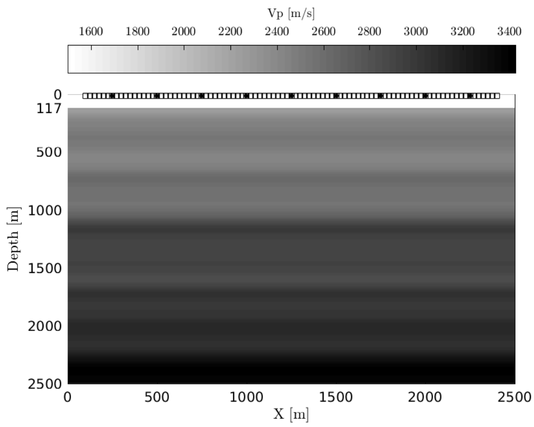


Figure 1: Domain setup with true velocity model. Empty squares at top are receiver, filled are source positions.

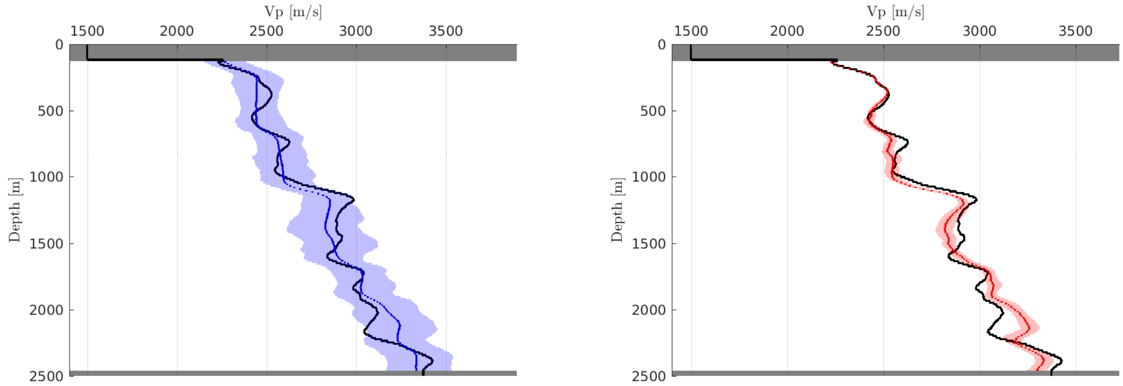
The true model is a single sample from the prior distribution. This is of course a benign true model but this practice is used in this simulation study, as the goal is to study whether the EnKS methodology is able to estimate this true model reasonably well from the seismic data.

The observations are the collection of the 9 shots, resampled at 150 Hz, so that the entire data set consists of $9 \times 93 \times 450$ data points. This set is partitioned into 3 groups of shots and 3 groups of time intervals (1 sec each) resulting in 9 assimilation cycles. The time groups form the outer loop within which the 3 groups of shots gather data is looped over. With this layered model and grouping some distributed shots, there will be a lot of redundant data. But the approach is used here as it is the considered strategy for horizontally varying subsurface models.

The measurement noise is here considered independent and constant across time samples, receivers and shots so that $\boldsymbol{\Sigma_e} = \sigma_e^2 \boldsymbol{I}$. The variance σ_e^2 is set rather low to ensure a reasonably high signal to noise ratio across the time trace(s). As no processing of recordings is performed the signal power will decay with time due to the spreading of energy. So the noise variance is kept low in order for the signal not to be masked completely by noise in the later part of the period.

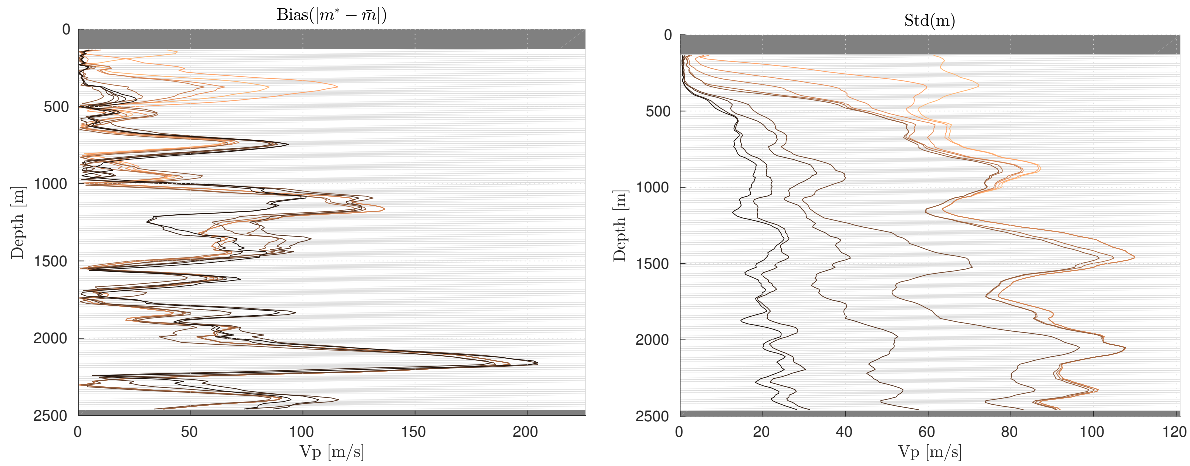
The ensemble size is set to $N_e = 50$ and sampled from the prior distribution. The empirical mean and empirical percentiles are plotted in Figure 2a. The effect of a small ensemble is clearly visible with the limits not being symmetric.
The final ensemble is depicted in Figure 2b where it is seen that the true model is recovered well down to ≈700 m. Further down the ensemble has generally moved in the right direction but has not closed sufficiently in on the true solution.

In Figure 3 is plotted the estimation bias and the standard deviation of the ensemble during the assimilation cycles. The first 3 cycles are by localisation only updating parameters down to ≈1 km. The first second of the recordings give sufficient information for the inference to be successful and the estimation is already close to the true model in this region after these 3 cycles. The PLS order selection finds a relatively low order in the first 5 cycles and then in the 6th a much high order. When reviewed it was seen that it could just as well had been a lower order, the PRESS statistic curve was quite flat and a higher order came out in a marginally higher number of MC samples.
The effect of this higher order PLS regression is seen as a sudden reduction in the ensemble spread, see Figure 3b. Once the ensemble collapses the parameter space stops being explored and no significant update takes place in following cycles.



(a) Prior (b) Posterior

Figure 2: Ensemble distributions: black is true model, coloured line is ensemble mean and area is (2.5, 97.5) percentiles. Gray areas are fixed.



(a) Bias of ensemble mean (b) Ensemble standard deviation

Figure 3: Ensemble statistics over cycles: colour goes progressively from light (prior) to dark (posterior).

Conclusion

With the example presented it is shown that seismic inversion using full waveform data is fundamentally possible using the Ensemble Kalman framework. There are still issues to address and limitations to identify, nevertheless it is expected that the described approach of sequentially processing within a probabilistic setting can lead to a useful technique for full waveform inversion.

References

Evensen, G. [2009] *Data Assimilation*. Springer.

Komatitsch, D. and Tromp, J. [1999] Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical Journal International*, **139**(3), 806–822.

Oliver, D.S. and Chen, Y. [2010] Recent progress on reservoir history matching: a review. *Computational Geosciences*, **15**(1), 185–221.

Sætrom, J. and Omre, H. [2010] Ensemble Kalman filtering with shrinkage regression techniques. *Computational Geosciences*, **15**(2), 271–292.