Introduction

Velocity model building is an essential yet challenging task in seismic imaging. Inversion techniques such as MVA and FWI have made significant progress recently, but they are deterministic and model uncertainties are not quantified. Here we present a probabilistic approach for estimating velocity and its uncertainty. In a Bayesian framework, our approach aims at assessing the posterior probability distribution of P-wave velocity conditional on seismic reflection data.

A method for sequential data assimilation is presented, where ensembles of velocity models are updated from subsets of seismic data grouped according to offsets and traveltimes. Ensemble-based methods are very useful in various applications, but they require careful tuning. Several practical techniques are outlined in this paper to optimize the inversion results. We demonstrate this approach on a layered earth model with elastic parameters adapted from a set of well logs. For computation efficiency, we use the reflectivity method to generate elastic seismic data. With adequate spatial constraints and different numerical solvers, this inversion method can be used for more complex models.

Inversion method

The subsurface parameters are denoted $\boldsymbol{m}$, while the seismic data are denoted $\boldsymbol{d}$. A Bayesian formulation is used here, with a prior model for parameters and a likelihood model for the seismic data, given the parameters. The goal of seismic inversion is to obtain the posterior model for subsurface model parameters, given the seismic data.

Prior model The prior model is described by a probability density function $p(\boldsymbol{m})$, which incorporates a priori knowledge about depth trends and uncertainties in the seismic velocity parameters. Furthermore, it is important to impose a realistic spatial correlation structure within the parameters. In this work a prior model based on Gaussian processes with mixed correlation ranges is used.

Forward modelling The seismic data are produced by forward modelling using the reflectivity method in the manner of Kennett (2011). This method allows fast computation of seismograms for a 1D medium, which in turn allows our algorithm to use a large ensemble size. We assume an isotropic and elastic medium that is parameterized in terms of layer thickness, velocities v_p and v_s, density ρ, and attenuation coefficients Q_p and Q_s. Gaussian noise $\boldsymbol{e} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma_e})$ is added to synthetic data for a stochastic formulation so that $\boldsymbol{d}|\boldsymbol{m} = g(\boldsymbol{m}) + \boldsymbol{e}$, thus resulting in the likelihood model $p(\boldsymbol{d}|\boldsymbol{m}) = \mathrm{N}(g(\boldsymbol{m}), \boldsymbol{\Sigma_e})$.

Sequential inversion Instead of using an entire seismic record at once, a record is divided into smaller data blocks and inversion is performed over these subsets of data sequentially. Such a way of partitioning and assimilation of data allows one to express Bayes' rule as

$$p(\boldsymbol{m}|\boldsymbol{d}) \propto p(\boldsymbol{d}|\boldsymbol{m})p(\boldsymbol{m}) = p(\boldsymbol{m})\prod_{i=1}^{N_c} p(\boldsymbol{d}_i|\boldsymbol{d}_{1:i-1}, \boldsymbol{m}) \propto p(\boldsymbol{m}|\boldsymbol{d}_1)\prod_{i=2}^{N_c} p(\boldsymbol{d}_i|\boldsymbol{d}_{1:i-1}, \boldsymbol{m}) \tag{1}$$

in which a data record $\boldsymbol{d}$ is partitioned into N_c blocks, and $\boldsymbol{d}_{1:i-1}$ represents $\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_{i-1}$. This approach aims to allow inversion to progress from the top of the subsurface region and downwards. In doing so, the solution is stabilized, and the uncertainty is gradually reduced from top to bottom.

Ensemble Smoother In the ensemble-based inversion scheme suggested, all ensemble members are sequentially updated using subsets of seismic data at every assimilation step. Data are integrated using a linear update at every step in the sequence. Without going into mathematical details, the basic form of the ensemble Kalman filter (EnKF) update is

$$\begin{aligned} &\boldsymbol{m}_i^{\mathrm{a}} = \boldsymbol{m}_i^{\mathrm{f}} + \boldsymbol{K}\left(\boldsymbol{d}_i^{\mathrm{o}} - \boldsymbol{d}_i^{\mathrm{f}}\right) \text{ with} \\ &\boldsymbol{d}_i^{\mathrm{f}} = g(\boldsymbol{m}_i^{\mathrm{f}}), \quad \boldsymbol{d}_i^{\mathrm{o}} \sim \mathrm{N}(\boldsymbol{d}^{\mathrm{o}}, \boldsymbol{\Sigma_e}), i = 1, \ldots, N_e, \end{aligned} \tag{2}$$

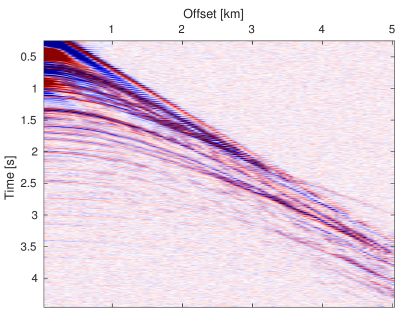
where $\boldsymbol{m}_i^{\mathrm{f}}$ represents the forecast (prior) model and $\boldsymbol{m}_i^{\mathrm{a}}$ is the updated/analysed (posterior) ensemble. Moreover, $\boldsymbol{d}_i^{\mathrm{f}}$ is the forecast data obtained by running the forward model and extracting a data subset. The Kalman gain $\boldsymbol{K}$ is based on correlations in the model and data ensembles.

In this context the data forecast is recomputed at every step, so the update resembles an Ensemble Kalman Smoother/asynchroneous EnKF (Sakov et al., 2010), with a forward model restart at each update cycle. As the state vector $\boldsymbol{m}$ holds no dynamic variables, the concept of time indexing has no meaning, instead it represents the assimilation cycles for the blocks of seismic data in this case. In each cycle, the posterior from previous cycle is used as prior for the current step when assimilating the current data block. In doing so, the end result is the posterior model conditioned to all the seismic data.

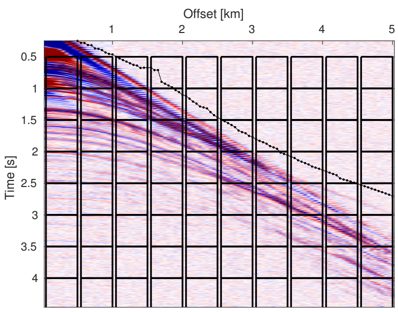
Adaptive processing The uncertainty in the predicted model parameter tends to shrink with every ensemble update in Equation 2. When all the data are used at once, the result of ensemble-based methods tends to underestimate the uncertainty, and the predicted distributions will only rarely cover the true variable or observations. Here, we discuss several adaptive approaches for addressing issues associated with the ensemble update.

A choice must be made regarding the structuring of subsets of data, and the order in which they are assimilated. As shown in Figure 1b data partitioning is done in time intervals and offset groups for a CMP gather. The size and order of these blocks have an effect of regularization in the sequential inversion scheme. The figure also shows a line which represents an adaptive mute region. When an assimilation cycle is about to process a particular data block, this line excludes data samples within the block that are above the line. The ensemble data forecast is also used to determine T-X location at which there are significant variations to be exploited during an update step. For an example, when the model ensemble for layers in the top region is estimated and the spread is very small, the resulting data forecast will also have a lower variation at earlier time samples and as a result the mute line is pushed forward in time. In doing so, only data points that have a certain spread in the data forecast are included in the conditioning.

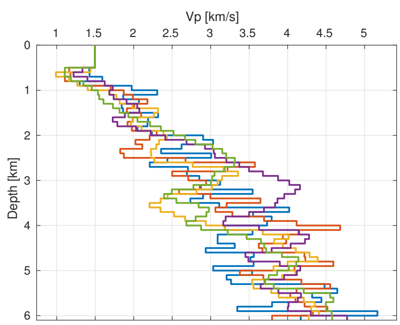
Given the conditioning on a particular data partition, the resulting analysed state can be categorised as: i) ineffective when the update does not bring much update to the ensemble, ii) divergent when the parameter ensemble gets updated in such a wildly manner that the posterior is useless as prior for the next iteration, or iii) satisfactory. The main reason for diverging updates is low predictive quality in the predicted data $\boldsymbol{d}^{\mathrm{f}}$. The residuals $\left(\boldsymbol{d}_i^{\mathrm{o}} - \boldsymbol{d}_i^{\mathrm{f}}\right)$ are large and the ensemble members are overly corrected. This issue is due to two reasons: data forecasts do not "cover" observations; or they do not "represent" the observations. When observations are mostly outside of the span of the forecast, it can be attributed to the inadequate coverage of the true parameter space by the ensemble, leading to a poor prediction. In this case, ensemble inflation is used to expand the sampling of the parameter space. In the case of non-representativeness, Multiple Data Assimilation (MDA) can be used (Emerick and Reynolds (2013)). This technique iterates over the same data block and performs relatively small updates with a larger



(a) Synthetic gather plot



(b) Data partition and mute line



(c) Initial prior model members

Figure 1: Configuration for sequential inversion

measurement variance. Thus a series of more restrained updated can be obtained at the cost of additional ensemble evaluations.

A set of scores are used to indicate potential issues of ineffective and divergent ensemble updates. Together with some decision rules based on these indicators, an adaptive response is enabled to alleviate unexpected and unwanted behaviour of the update. These scoring rules reflect the predictive capability of the forecast observations. In enabling an action, none of these score rules stand alone. The energy score (Gneiting and Raftery, 2007) is employed as a measure of the overall quality of the forecast. This score rule is a multivariate generalization of the Continuously Ranked Probability Score and has a form that permits using ensemble forecasts. In addition, the variogram score is used to quantify the dissimilarity in the correlation structure between the observation and the forecast. Lastly, the concept of coverage probability is used to evaluate whether the observations are within the span of the forecast ensemble.

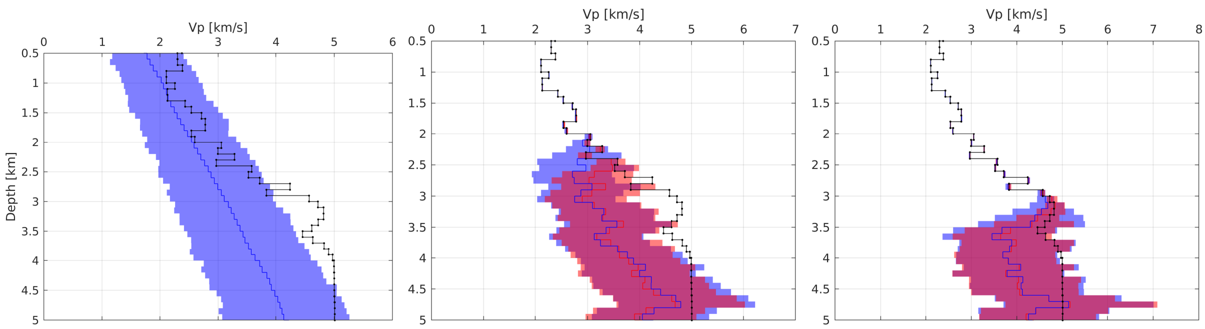
Numerical example

A synthetic example is presented here to demonstrate the validity of this ensembled-based inversion concept. A 1D earth model is produced from field borehole measurements. Although only P-wave velocity is inverted, elastic forward modelling is performed.

The P-wave velocity profile consists of a top water layer of depth 500 m with a free surface boundary condition. The subsurface region targeted for inversion is between 0.5–5.0 km. This region is discretised into 51 layers, each 100 m thick. The shear velocity, density and attenuation factors are fixed. Synthetic traces are recorded at 100 receivers located from 50 m to 5 km with a uniform spacing of $\Delta x = 50$ m. All sources and receivers are placed at a depth of 5 m. Time sampling is 2 ms and record length is 4.5 s. The source wavelet is a minimum phase Butterworth wavelet. Frequency content in the output from forward modelling is set to 3–30 Hz. The noise variance is set as a constant so that $\boldsymbol{\Sigma_e} = \sigma_{\boldsymbol{e}}^2 \boldsymbol{I}$ in Equation 2. This variance is $\sigma_{\boldsymbol{e}}^2 = 1.8 \times 10^{-4}\,\mathrm{m}^2$, which corresponds to a noise level that has a power of -10 dB below the average signal power over a window between 1–3 s in time and 0–3 km in offsets. The resulting seismic gather used as true observations is shown in Figure 1a.

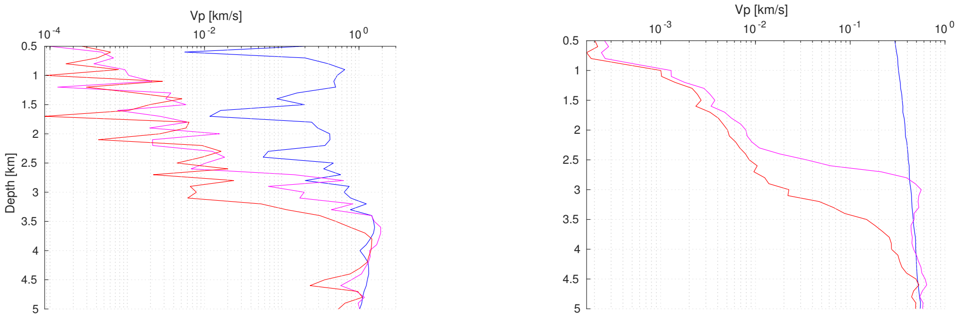
The seismic gather is partitioned into 80 blocks in the T-X domain, with 10 offset groups and 8 time intervals of 500 ms duration. The total analysis window is from 0.5 to 4.5 s. This data partitioning is illustrated in Figure 1b as the square boxes. Inversion progresses from early to late arrivals and far to near offsets.

To obtain an initial ensemble of $\boldsymbol{m} = \ln \boldsymbol{v}_p$, a multivariate Gaussian distribution $\mathrm{N}\left(\boldsymbol{\mu}_{v_p}, \boldsymbol{\Sigma}_{v_p}(\phi)\right)$ in the v_p-domain is sampled followed by taking the logarithm. This allows physical units to be used when setting the mean and bounds (expressed through marginal variances for each depth location) while ensuring positive velocities during inversion. The initial prior ensemble is thus skewed due to the logarithmic



(a) Initial prior (b) Intermediate result (c) Final posterior

Figure 2: Ensemble coverage; black line is true model, blue is prior and red is posterior in particular cycle. Colored lines are ensemble mean, shaded area is empirical 95% band.



(a) Error $|\boldsymbol{m}_{\text{true}} - \widehat{\boldsymbol{m}}|$ (b) Standard deviation of $\boldsymbol{m}$ ensemble

Figure 3: Ensemble statistics from results in Figure 2. Blue is initial prior in 2a, magenta is intermediate (posterior) solution in 2b and red is final posterior in 2c.

function. The mean $\boldsymbol{\mu}_{v_p}$ is chosen as a linear function of depth, and so is the marginal variance such that the spread of the ensemble increases with depth. The spatial correlation structure is chosen to be of a Matern type $\frac{5}{2}$ with range parameter ϕ. The correlation structure and marginal variances form the covariance matrix $\boldsymbol{\Sigma}_{v_p}(\phi)$. A strategy of using a mixed ensemble is applied, so the initial ensemble holds samples based on different range parameters. The ensemble of size 500 is split in 5 groups of different range parameter, from 200 to 1000 m in steps of 200, so that the ensemble holds 100 samples for each range. A collection of velocity profiles from different range parameters is displayed in Figure 1c, illustrating various levels of smoothness in the initial ensemble.

Among the 80 data blocks outlined, only 30 are used for the ensemble updates. MDA is applied 7 times, and the total number of ensemble forward model evaluations is 37. Skipping a data block happened 19 times because data points was found to be the mute region, and 31 times the update was evaluated to be ineffective. The ensemble distribution of the initial prior, an intermediate assimilation cycle and the final posterior update are shown in Figure 2. The intended behaviour of inverting from top down is evident. Even if the true model at depth 3–3.5 km does not lie within the span of the initial prior, the sequential inversion coupled with adaptive steps of inflation/MDA, is still able to get a reasonable prediction at those depths. A more detailed view of these results are shown in Figure 3, where the bias (3a) of the velocity profile and the standard deviation (3b) of the ensemble are shown. At shallow depths the ensemble has a very low uncertainty and a small bias, indicating that the solution is reliable. At greater depths, the inversion result shows rather large uncertainties. This is due to the relatively low signal to noise in the late arrivals as data are no longer supporting the estimation of velocities and ensemble range is not reduced. We aim to extend this method to higher spatial dimensions using real field data.

Conclusions

An outline for probabilistic seismic inversion has been presented for 1D vertical profile modelling. The approach is based on the ensemble-based framework and sequential assimilation of time-offset blocks of seismic data. Various adaptive measures are used to make the sequential processing robust. The synthetic example shows promising results, where the uncertainty in predictions are gradually reduced in the depth domain, while keeping small prediction bias.

References

Emerick, A.A. and Reynolds, A.C. [2013] Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, **55**, 3–15.

Gneiting, T. and Raftery, A.E. [2007] Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, **102**(477), 359–378.

Kennett, B. [2011] Seismic Wave Propagation in Stratified Media. *ANU Press*.

Sakov, P., Evensen, G. and Bertino, L. [2010] Asynchronous data assimilation with the EnKF. *Tellus A*, **62**(1), 24–29.