EAGE

The Ensemble Kalman Filter adapted for Time-Lapse seismic data

Inge Bjørn Myrseth

Summary

Time-Lapse Seismic data are frequently used in History Matching of reservoir production data. The ensemble Kalman Filter (EnKF) is a method that conditions on observations as the appear. It uses an n-member ensemble of representations of reservoir characteristics and conditions them on production data. However, if the dimension of the production data is greater than the number of ensemble members, rank problems may arise and the EnKF solution might break down. The dimension of seismic data is vast and exceeds the number of ensemble members in realistic models. We propose a hierarchical model, phrased in a Bayesian setting, to account for these problems. The model incorporates prior information about the expected level and heterogeneity of the solution. The hierarchical model is implemented on simple synthetic examples and they show promising results.

Introduction

History Matching (HM) is done by conditioning a representation of the reservoir characteristics on historical production data to forecast future reservoir behavior. The forecast should include uncertainty, and due to the nonlinear nature of the reservoir fluid flow models, this is done by repeated reservoir simulations. Traditionally HM is done by tuning individual variables in the representation to minimize the difference between observed production history and forecasted production. This approach does not account for or quantify uncertainty and the conditioning is made on all available observations simultaneously. Hence an entirely new representation has to be matched when new observations are available.

The Ensemble Kalman Filter (EnKF) was introduced in Evensen (1994) as a method to condition sequentially on observations as they appear. The method uses an initial ensemble of reservoir representations which are simulated forward in time. Each representation is sequentially conditioned on observations. The covariance matrix between representations and observations is estimated based on the ensemble at any time. However, if the number of representations in the ensemble is smaller than the number of observations this matrix will be a low rank estimate of the true covariance matrix, and there might be loss of information. Presently, high-dimensional time-lapse seismic data must also be conditioned on, and hence the rank problem in EnKF appears crucial. We propose a Hierarchical Ensemble Kalman Filter that ensures a full rank estimate of the covariance matrix and hence avoids the information loss. Prior information on the covariance matrix is used to enforce realistic solutions.

Model formulation

Consider the data assimilation state space equations

$$x_i = \Phi(x_{i-1}) \qquad i = 1, ..., m+1 \tag{1}$$

$$d_i = H_i x_i + \varepsilon_i \qquad i = 1, ..., m \tag{2}$$

where $x_i \in \mathbb{R}^p$ are the state variables following a Markov process, $\Phi(.)$ is a known, possibly non-linear, transfer function, $d_i \in \mathbb{R}^k$ are the observations, $H_i \in \mathbb{R}^{k \times p}$ is a matrix relating the observations to the state variables and $\varepsilon_i \in \mathbb{R}^k$ is the observation error following some arbitrary pdf. In the HM setting x_i are the reservoir characteristics with very large p and $\Phi(x_i)$ is the reservoir fluid flow simulator. The observations, d_i are usually production history, but it may also be time-lapse seismic data, hence very large k. The model is represented in the graph in Figure 1.

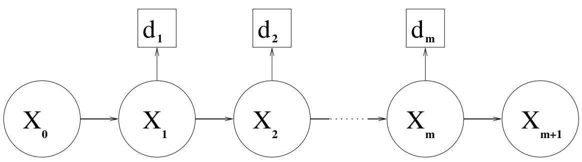

Figure 1: A graph depicting the dependency structure in a state space model.

Assign a prior model to the initial state $f(x_0)$ and let the deterministic transfer function be represented by the Dirac pdf $f(x_{i+1}|x_i)$. The prior model for all times considered up to time $t = m+1$ is then

$$f(x_0, ..., x_{m+1}) = f(x_0) \prod_{i=1}^{m+1} f(x_i|x_{i-1}). \tag{3}$$

Assume conditionally independent Gaussian likelihood models for the observations

$$[D_i|x_i] \sim f(d_i|x_i) = \mathcal{N}_k(H_i x_i, \Sigma_i^o) \text{ for } i = 1, ..., m, \tag{4}$$

where $\mathcal{N}_k(\mu, \Sigma)$ denotes a k-variate normal distribution with expectation μ and covariance matrix Σ, and Σ_i^o is the observation error covariance matrix. The prior and likelihood models define the posterior model

$$f(x_0, ..., x_{m+1}|d_1, ..., d_m) = const \times f(x_0) f(x_{m+1}|x_m) \prod_{i=1}^{m} f(d_i|x_i) f(x_i|x_{i-1}). \tag{5}$$

The objective is to obtain the marginal posterior pdf for x_{m+1} conditioned on $d_1, ..., d_m$ as it provides a forecast. It can be defined from the above expressions:

$$f(x_{m+1}|d_1, ..., d_m) = \int ... \int f(x_0, ..., x_{m+1}|d_1, ..., d_m) \mathrm{d}x_0 ... \mathrm{d}x_m \tag{6}$$

The forecast can be computed sequentially via the relations:

$$f(x_i|d_1, ..., d_{i-1}) = \int f(x_i|x_{i-1}) f(x_{i-1}|d_1, ..., d_{i-1}) \mathrm{d}x_{i-1} \tag{7}$$

$$f(x_i|d_1, ..., d_i) = const \times f(d_i|x_i) f(x_i|d_1, ..., d_{i-1}). \tag{8}$$

In general the steps in expressions (7)-(8) are hard to determine, except for some special cases discussed below.

Kalman Filter (KF) Consider the following constraints on the general data assimilation problem. Let $\Phi(x)$ be a linear function in x and let $f(x_0)$ be Gaussian. In this case the marginal posterior pdf in expression (6) will be Gaussian and analytically tractable. The solutions are given by the KF approach, which determines the posterior mean and covariance in a sequential manner and thereby fully specifies the distribution.

Ensemble Kalman Filter (EnKF) Consider the general case where $\Phi(x)$ is nonlinear. The EnKF uses an ensemble, $\mathbf{e}_0 : \{x_{0,1}, ..., x_{0,n}\}$, of n realizations sampled from the prior distribution at time $t = 0$. The ensemble members are forecasted via expression (1) and thus define an ensemble at time $t = 1$. The ensemble mean $\bar{x}_1$ and covariance S_1 are defined and the ensemble members are conditioned on the observations at time $t = 1$ via the expression:

$$x_{1,j}^c = x_{1,j} + S_1 H_1' (H_1 S_1 H_1' + \Sigma_1^o)^{-1} (d_1 + \varepsilon_{1,j}^o - H_1 x_{1,j}) \qquad j = 1, ..., n \tag{9}$$

where $\varepsilon_{1,i}^o$ is a random sample from $\mathcal{N}(0, \Sigma_1^o)$ added to ensure correct error structure. This procedure is repeated sequentially in time t.

The EnKF solution can be expressed as a linear combination of the initial ensemble. As a result, the EnKF solution is dependent on the rank of S_t, $t = 1, ..., m$. If the number of ensemble members n is smaller than the number of observations k, the $k \times k$ matrix $H_i S_i H_i'$ in expression (9) will not have full rank, and numerical problems arise. Recent results show that the EnKF may provide reliable results if the inversion in expression (9) is replaced by a pseudo-inversion, see Evensen (2004). Note however that estimation uncertainty of the covariance matrix and numerical approximations are left unspecified. If the number of observation k is far greater than the number of ensemble members n the traditional EnKF might break down due to lack of information in the ensemble. This is typically the situation when time-lapse seismic data with one observation in each grid node of the reservoir representation is available. We propose a hierarchical model that solves this problem.

Hierarchical Ensemble Kalman Filter (HEnKF)

Consider the EnKF setting as defined above. Assume the forecast at time $t = i$ follows a Gaussian distribution, $f(x_i|d_1, ..., d_{i-1}) \sim \mathcal{N}_p(\mu_i, \Sigma_i)$. Assign the following prior distributions for the parameters μ_i and Σ_i:

$$[\mu_i|\Sigma_i] \sim \mathcal{N}_p(\Phi^i(\eta), \xi \times \Sigma_i), \tag{10}$$
$$\Sigma_i \sim \mathrm{W}_p^{-1}(\Psi, \nu), \tag{11}$$

where $\Phi^i(x)$ is the transfer function applied i times to x and $\mathrm{W}^{-1}(\Psi, \nu)$ is the inverse Wishart distribution with $p \times p$ scale matrix Ψ and $\nu > p + 1$ degrees of freedom. The expectation of an inverse Wishart distributed variable is $\mathrm{E}\{\Sigma\} = \frac{\Psi}{\nu-p-1}$. The model parameters η, ξ, Ψ and ν can be specified such that they add information about the expected level and heterogeneity of the solution. Moreover, uncertainty associated with this information can be assigned. The model is represented in the graph in Figure 2.

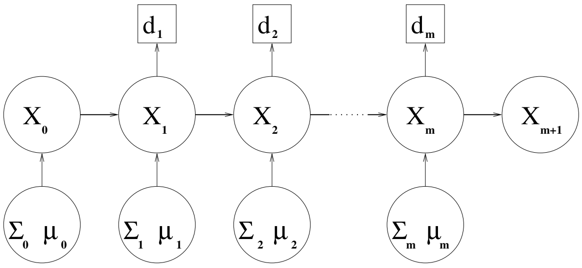

Figure 2: A graph depicting the dependency structure in a state space model.

At time $t = i$ suppose we have an n-member ensemble $\mathbf{e}_i : \{x_{i,1}, ..., x_{i,n}\}$ representing a forecast from $f(x_i|d_1, ..., d_{i-1})$ with ensemble mean $\bar{x}_i$ and covariance S_i. Conditioning on the ensemble defines the following posterior model:

$$[\mu_i|\mathbf{e}_i, \Sigma_i] \sim \mathcal{N}_p((1-\alpha)\Phi^i(\eta) + \alpha\bar{x}_i, \xi(1-\alpha)\Sigma_i) \tag{12}$$
$$[\Sigma_i|\mathbf{e}_i] \sim \mathrm{W}_p^{-1}(\Psi_i, \nu + n), \tag{13}$$

where $\alpha = \frac{n\xi}{1-n\xi}$ and $\Psi_i = \Psi + (n-1)S_i + (\frac{1}{n} + \xi)(\bar{x}_i - \Phi^i(\eta))'(\bar{x}_i - \Phi^i(\eta))$. A hierarchical version of the conditioning in expression (9) uses a covariance estimate from the posterior model:

Generate	$\Sigma_j \sim f(\Sigma_i|\mathbf{e}_i)$	$j = 1, ..., n$
Update	$x^c_{i,j} = x_{i,j} + \Sigma_j H_i'[H_i \Sigma_j H_i' + \Sigma_i^o]^{-1}(d_t + \varepsilon_{i,j} - H_i x_{i,j})$	$j = 1, ..., n$

All realizations from $\mathrm{W}_p^{-1}(\Psi_i, \nu + n)$ are full rank covariance matrices and as a result both Σ_j and $H_i \Sigma_j H_i$ will be real and of full rank. The uncertainty in the matrix estimate is accounted for, and it decreases as the number of ensemble members n increases and tends to the true covariance as $n \to \infty$. The HEnKF is an extension of the EnKF in a true Bayesian spirit.

Examples Consider two simple synthetic transfer functions, one linear and the other non linear, where x is a vector of dimension $p = 100$. The linear function is $\Phi_1(x_i) = Ax_i$ and the non-linear function is $\Phi_2(x_i) = Ax_i + \arctan(x_i)$, where $A \in \mathbb{R}^{100\times100}$ is a banded matrix where elements $A_{ij} = \frac{1}{3}$ if $i \in [j-1, j+1]$ and zeros elsewhere. Synthetic data at time $t = 0$ are generated from $\mathcal{N}_{100}(0, \sigma^2 I_{100\times100})$ and subjected to the transfer function, where $\sigma^2 = 10$ for the linear model and $\sigma^2 = 1$ for the non linear model. The left most displays of Figure 3 show the synthetic data at time t=40. Observations, are generated with $H_i = I_{100\times100}$ and error model $\mathcal{N}_{100}(0, 30I_{100\times100})$ at each time step. Hence $k = p = 100$. Prior parameters are $\eta = 0$, $\xi = 0.1$ and $\nu = 102$, and Ψ is an exponential covariance matrix with correlation length 20 scaled such that the expected marginal variance of the inverse Wishart distribution is 30. The middle displays show the ensemble mean from EnKF and HEnKF with associated 95% marginal confidence bounds based on an ensemble of size $n = 8 << p = k = 100$. The right most displays show ensemble means from 10 reruns of the EnKF and HEnKF with different initial ensembles. The middle displays show that HEnKF provides more stable solutions than EnKF both when $\Phi(.)$ is linear and non-linear, and that forecast uncertainty is more realistically assessed. The right most displays confirm that HEnKF provides more stable results than EnKF.

Summary In situations when the number of observations is far greater than the number of ensemble members the EnKF might not give reliable results due to rank issues. We have proposed the HEnKF model phrased in a consistent Bayesian setting, to account for these problems. HEnKF provides encourageing result on two simple examples.

References

Evensen, G. [1994] Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Geophys. Res. 99, 10143-10162.

Evensen, G. [2004] Sampling strategies and square root analysis schemes for the EnKF. Ocean Dynamics 54, 539-560.

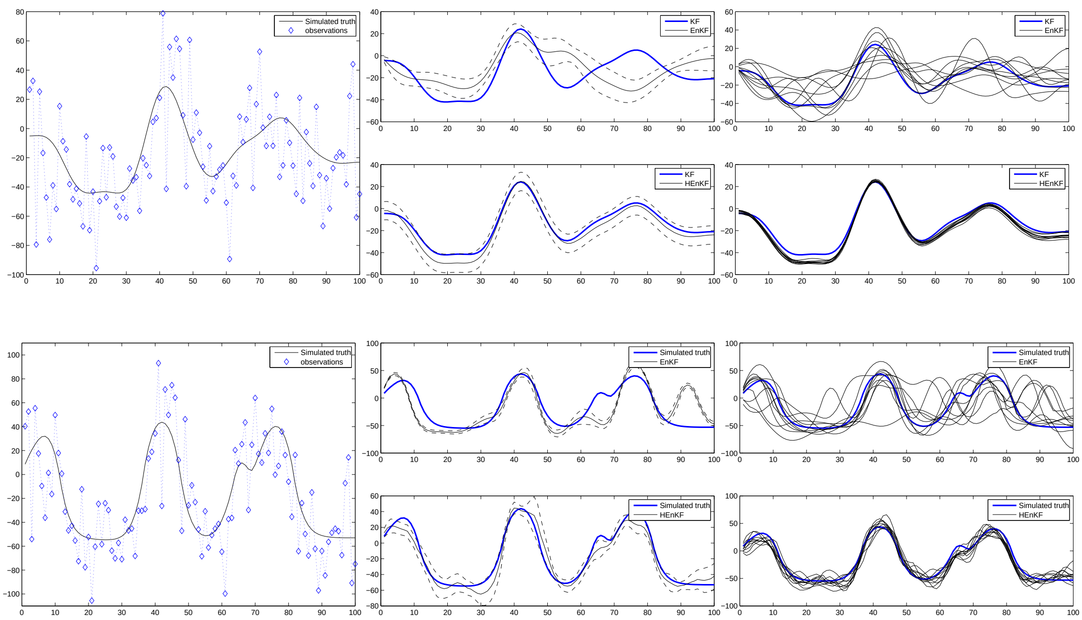

Figure 3: Linear model (top row) and non-linear model (bottom row). Simulated truth and observations (left), 10 realizations of EnKF and HEnKF (middle) and one realization of EnKF and HEnKF within 95% marginal confidence bounds (right). KF solution for the linear model and simulated truth for the non-linear model plotted for reference (blue).