Hierarchical Bayes Ensemble Kalman Filtering

Michael Tsyrlunikov and Alexander Rakitko

HydroMeteorological Research Center of Russia

Melbourne, 5 December 2016
Motivation
A methodological problem in the high-dimensional EnKF

The EnKF analysis is based on the KF gain matrix
\[ K = BH^\top (HBH^\top + R)^{-1}. \]

The problem is that this \( K \) is optimal only if \( B \) is exact. But this cannot be the case in the EnKF, where \( B \) is estimated by the sample covariance matrix \( S \). The problem is especially severe in high dimensions, where the ensemble size is inevitably small:

- \( S \) is heavily contaminated by random sampling noise.
- Due to the EnKF imperfection, the forecast ensemble members shouldn’t be expected to be draws from the true background error distribution. That is, \( S \) is not only noisy, but it can (and should) be distorted (w.r.t. the true \( B \)) in a systematic way.
Remedies for the sampling noise

1. **Spurious long-distance** covariances can be reduced by a kind of **spatial localization** of $S$:
   - in model space: $L \circ S$
   - in observation space: $L \circ BH^T$, $L \circ HBH^T$
   - by inflating entries of $R$ for distant observations: $R/L$ (locally)

2. **Spatial filtering/smoothing** the sample covariances.

3. **Mixing** the sample covariances with **static** covariances (EnVar).

Side effects of the remedies

1. **Spurious short-distance** covariances remain uncorrected, non-local observations require large localization length scales, balances can be destroyed, etc.

2. The signal in the covariances is smoothed in space along with the noise.

3. The flow-dependent signal is mixed with the climatology as much as the noise.
Remedies for **systematic** errors in $S$

1. Covariance *inflation*.
2. Using sub-ensembles to avoid *inbreeding*.

**Side effects of the remedies**

1. Error **correlations** cannot be corrected using inflation. Using one scalar factor for the whole matrix seems simplistic.
2. Provides only a partial solution.
Covariance regularization techniques proposed in the statistical literature

1. Regularization by introducing a pre-specified *structure* in the covariance/precision matrix: parametric models, sparsity, factorizations, linear combinations, . . .

2. *Bayesian* estimators.
A general deficiency of the covariance regularization techniques

All the above covariance regularization techniques are not based on a theoretic optimality criterion targeting the accuracy of the resulting analysis.

The resulting common paradigm is the two-stage analysis:

1. Rectify the sample covariances using an ad-hoc device.
2. Perform the KF analysis with the rectified sample covariances assuming these are perfect.
The idea and the goal of the Hierarchical Bayes Ensemble Filter (HBEF)
Goal of the HBEF

Our basic suggestion is to explicitly admit that the true covariances are unknown, random, and subject to estimation.

We propose to switch from the two-stage analysis to a one-stage procedure in which the estimation of the sample covariances is part of the optimal analysis.

Heuristically, if we aim at the accuracy of the analysis in a procedure in which the covariances are part of the control variable, then a covariance regularization should be obtained “automatically”, because otherwise the analysis cannot be accurate.
One more methodological problem in the EnKF

In the EnKF’s analysis equations, there is no intrinsic feedback from observations to (always imperfect in high dimensions!) background-error covariances. This requires external adaptation or manual tuning.

In the HBEF, the joint estimation of the state and the covariance matrix will allow observations to influence the covariances in an optimized way.
Background: hierarchical Bayesian estimation/modeling, hierarchical filtering, and random matrices
Hierarchical Bayesian modeling

In a non-Bayesian model, the parameter of interest (in our case, the state vector $x$) is assumed to be deterministic, to be estimated using random observations. Optimal interpolation is an example.

In a non-hierarchical Bayesian model, the state $x$ is assumed to be random and having some *prior* distribution, $p(x)$.

Knowing the *likelihood function* $p(y|x)$ for the available data (observations) $y$, the *posterior* distribution $p(x|y)$ is computed using the Bayes theorem:

$$p(x|y) \propto p(x) \cdot p(y|x)$$

The KF, Var, particle filters can be viewed as examples of the non-hierarchical Bayesian models.
In a hierarchical Bayesian model, the prior is parameterized,
\[ p(x) = p(x|\gamma), \]
where the parameter \( \gamma \) is assumed to be random as well, having its own (hyper)\textit{prior} probability distribution \( p(\gamma) \).

- Switching from the simpler Bayesian model to the more complicated hierarchical Bayesian model is meaningful if \( \gamma \) is known to be highly variable.
- Online estimation of \( \gamma \) can be useful.
- Online estimation of \( \gamma \) can be successful if \( \gamma \) is reasonably observed.
In the EnKF, the prior distribution of the state is assumed to be Gaussian $\mathcal{N}(\mathbf{b}, \mathbf{B})$, where the parameters are the prior mean $\mathbf{b}$ and the prior covariance matrix $\mathbf{B}$. The prior mean is quite well approximated by the forecast or ensemble mean. As for $\mathbf{B}$,

- It is known to be highly variable. This motivates the switch from the Bayesian to the hierarchical Bayesian paradigm.
- We need to estimate $\mathbf{B}$ to perform the analysis.
- We can estimate $\mathbf{B}$ because with the advent of ensembles, the prior covariance matrix appears to be, actually, observed.

So, $\mathbf{B}$ is worth being estimated, i.e. included into the analysis control variable.
Background: Hierarchical filtering

Myrseth and Omre (SPE J. 2010) proposed to remove the assumption that the background-error covariance matrix $B$ and the background-error mean field $b$ are known deterministic quantities, replacing it in their HEnKF by the assumption that these are uncertain and random.

Bocquet (NPG 2011) assumed that $B$ is random and imposed a prior probability distribution for it—in order to change the (Gaussian) prior distribution of the state $x$ to a more realistic continuous mixture of Gaussians.
Background: random matrices and matrix variate probability distributions

The $n \times n$ matrix $\mathbf{M}$ is random if its entries $M_{ij}$ are random variables having some joint probability distribution.

**Vectorization.** A general matrix:

$$\mathbf{M} = \text{vec} \mathbf{M} := \begin{pmatrix} M_{11} \\ \vdots \\ M_{n1} \\ \vdots \\ \vdots \\ M_{n1} \\ \vdots \\ M_{nn} \end{pmatrix}$$

$p(\mathbf{M})$ is identified with $p(\text{vec} \mathbf{M})$. 
HBEF: theory, design, and properties
1. Simplified analysis of $\mathbf{B}$ using the ensemble only
Assimilation of ensemble members to update $B$

Ensemble members: $x^e(i)|B \sim \mathcal{N}(b, B)$, $i = 1, \ldots, N$, $N$ is the ensemble size.

$$p(x^e(i)|B) \propto \frac{1}{|B|^{1/2}} e^{-\frac{1}{2} (x^e(i) - b)^\top B^{-1} (x^e(i) - b)}$$

But this is the likelihood of $B$ given $x^e(i)$!

For the whole ensemble, $X^e = (x^e(1), \ldots, x^e(N))$, we have the likelihood

$$p(X^e|B) \propto \frac{1}{|B|^{N/2}} e^{-\frac{1}{2} \sum_{i=1}^N (x^e(i) - b)^\top B^{-1} (x^e(i) - b) = \frac{1}{|B|^{N/2}} e^{-\frac{N}{2} \text{tr}(B^{-1}S)}}$$

The existence of the likelihood function implies that a Bayesian update of $B$ is possible, in which the ensemble members $x^e(i)$ are used as generalized observations.
The posterior distribution of $B$ given the ensemble

1) **Prior:** Inverse Wishart with the mean $\bar{B}$:

$$ p(B) \propto \frac{1}{|B|^{\frac{\varphi}{2} + n + 1}} e^{-\frac{\varphi}{2} \text{tr}(B^{-1}\bar{B})} $$

2) **Likelihood:**

$$ p(X^e|B) \propto \frac{1}{|B|^{\frac{N}{2}}} e^{-\frac{N}{2} \text{tr}(B^{-1}S)} $$

3) **Bayes theorem $\Rightarrow$ posterior:**

$$ p(B|X^e) \propto p(B) \cdot p(X^e|B) = \frac{1}{|B|^{\frac{\varphi}{2} + n + 1}} e^{-\frac{\varphi}{2} \text{tr}(B^{-1}\tilde{B})}, $$

where

$$ \tilde{\varphi} = \varphi + N \quad \text{and} \quad \tilde{B} = \frac{\tilde{\varphi} \bar{B} + NS}{\tilde{\varphi} + N} $$
2. Full analysis
HBEF analysis

We analyze the extended control vector \((x, B)\).

1 Prior
   - \(p(B) \sim \text{IW}(\varphi, B^f)\).
   - \(p(x|B) \sim \mathcal{N}(x^f, B)\) (conditional Gaussianity).

   The background for the covariances, \(B^f\), is the persistence forecast.

   The background for the state, \(x^f\), is the traditional forecast.

2 Likelihood
   - \(p(y, X^e|x, B) = p(y|x) p(X^e|B)\)

3 Posterior = Prior \times Likelihood
HBEF analysis: the posterior distribution

\[ p^a(x, B) = p(x, B | y, X^e) \propto p(B) p(x | B) p(y | x) p(X^e | B) \]

The posterior pdf can be written hierarchically:

\[ p^a(x, B) = p^a(B) p^a(x | B) \]

Here, \( p^a(B) \) is analytically tractable but complicated (and can be approximated by an Inverse Wishart density), and, remarkably, \( p^a(x | B) \) is Gaussian that coincides with the KF’s posterior pdf.

So, the conditional Gaussianity “survives” the analysis step (and this conclusion is independent of the prior distribution of \( B \).)
Analysis algorithm

Goal:
(i) produce the deterministic analyses (point estimates) $x^a$ and $B^a$
(ii) generate the analysis ensemble.

Deterministic analyses

1. A Monte Carlo importance sampling based analysis (the Monte Carlo HBEF).

2. A two stage analysis: first, compute $B^a$ using the ensemble data only and then employ the standard EnKF technique to analyze $x$ using the resulting $B^a$ (the simplest HBEF).

Analysis ensemble

- The stochastic EnKF technique is adopted.
The Monte Carlo based HBEF analysis

The deterministic analyses of $\mathbf{B}$ (which aims to approximate $E\mathbf{B}$) is computed using importance sampling from $p^a(\mathbf{B})$ with the proposal density

$$p(\mathbf{B}|\mathbf{X}^e) \sim IW(\tilde{\phi}, \tilde{\mathbf{B}})$$

The deterministic analysis of $\mathbf{x}$ (which approximates $E\mathbf{x}$) takes advantage of the equality

$$E\mathbf{x} = E E(\mathbf{x}|\mathbf{B})$$

and the fact that in the posterior, $E(\mathbf{x}|\mathbf{B})$ is determined by the standard KF formula.
The simplest HBEF analysis

1. Analyze $B$ given the prior mean $B^f$ and the background ensemble (i.e. not using observations), getting $B^a$:

$$B^a = \frac{\varphi B^f + NS}{\varphi + N}$$

2. Analyze $x$ given $B^a$:

$$x^a = x^f + B^aH^\top(HB^aH^\top + R)^{-1}(y - Hx^f)$$

Comments:

- Localization is still needed in this version of the HBEF.
- In its simplest version, the HBEF is affordable for use in operational schemes on existing computers.
The forecast step

- \( B \): persistence: \( B^f_{k+1} = B^a_k \).

An option: regression to the climatological mean:

\[
B^f_{k+1} = B_{clim} + W \cdot (B^a_k - B_{clim}).
\]

- \( x \): the traditional ensemble forecast.

The Gaussianity of the state given the covariances is shown to be preserved at the forecast step. So, with the linear forecast model and linear observations, the conditional Gaussianity of the state is preserved in the course of filtering.
We, actually, split $B$ into the predictability error covariance matrix $P$ ("internal" error growth) and the model error covariance matrix $Q$ ("external" error growth). The forecast ensemble is split accordingly.

In the Monte Carlo based HBEF, we, actually, allow the ensemble to be imperfect. The ensemble members $x^e(i)$ are assumed to be drawn not from $\mathcal{N}(b, B)$ but from $\mathcal{N}(b, \Pi)$, where

- $\Pi$ is the (unknown) random matrix distributed around $B^f$ as $\Pi \sim IW(\varphi, B^f)$.
- The true $B$ is distributed around $\Pi$ as $B \sim IW(\theta, \Pi)$. 
Numerical experiments
(with two doubly stochastic models of truth)
Conditional vs. **unconditional** background error pdfs
Biases in the forecast ensemble variances $S$ \((\dim=1)\)
RMSEs in $B^a$ (dim=1)
Analysis RMSEs (of $x$) as functions of $N$ (dim=1)
Analysis RMSEs (of $x$) as functions of $N$ (dim=64)
Analysis RMSEs: model-error variance is distorted (dim=1)
Conclusions

Key features of the HBEF

- Covariance matrices are treated as random matrices.
- Ensemble members are assimilated as generalized observations.
- The covariance matrices are subject to sequential Bayesian update.
- The EnKF’s Gaussian assumption is replaced by the conditional Gaussianity.

Testing

- With the toy models (one-variable and 64-variable), the HBEF (even in its simplest version) significantly outperformed Var and EnKF.

Implementation

- In its simplest version, the HBEF is computationally affordable.
Conclusions -2

Future work

- A more adequate than IW prior for \( B \) (to achieve localization among other goals).
- An improved forecast for \( B \).
- A technique to better account for non-sampling errors in the covariances.

Other applications of hierarchical filtering

- Non-Gaussian parametric filters (quadratic filters, GIG).
- Non-parametric filters (particle filters, the rank histogram filter).
References


*Free access to the article until January 05, 2017:*
https://authors.elsevier.com/a/1U3FZ_3pR42554

The R code of the 1D HBEF can be found at [github.com/rakitko/hbef](https://github.com/rakitko/hbef)

Thank you!