# A Hierarchical Bayes Ensemble Filter

Michael Tsyrulnikov and Alexander Rakitko

*HydroMetCenter of Russia*

Email: michael.tsyrulnikov@gmail.com

## Introduction

In the EnKF, the prior covariances (the **B** matrix) cannot be accurately estimated from a small affordable background ensemble. Errors in the ensemble sample covariances are normally mitigated by a number of ad-hoc devices like covariance localization, variance inflation, and mixing with climatological covariances. In a new Hierarchical Bayes Ensemble (Kalman) Filter, termed HBEF, we propose an objective Bayesian estimation technique for **B**. The HBEF advances the EnKF methodology by introducing a full-fledged secondary filter, which treats the prior covariances like the traditional EnKF treats the state vector.

## Methodology

The HBEF introduces a forecast-analysis update cycle for the prior covariances. At the analysis step, the **B** matrix is estimated using a background for **B** ($\mathbf{B}^f$, provided by the forecast step) and the ensemble. To combine the information from $\mathbf{B}^f$ and from the ensemble members, a hyperprior probability distribution for **B** is introduced. The Inverse Wishart matrix variate probability distribution is used as the hyperprior distribution. The hyperprior distribution is updated in the analysis using the Bayes theorem, with the ensemble members treated as generalized observations on **B**. This leads to an EnVar like analysis algorithm. The posterior mean **B** is computed and propagated at the forecast step to the next analysis time using persistence or regression to climatology.

## Results

The below Figures present results for the one-variable doubly stochastic model of truth. The model is a first-order auto-regression forced by the white noise; the coefficients of the auto-regression are random processes by themselves governed by their own first-order auto-regressions but with constant coefficients. The model is designed to exhibit intermittent instability. The solution of the model equation is a non-stationary random process with the tunable degree of non-stationarity.

Figure 1 shows the analysis RMSEs (w.r.t. the known "truth", the averaging was over $2 \cdot 10^5$ assimilation cycles) as functions of the ensemble size $N$ for several filters (with the analysis RMSE of the unbeatable benchmark Kalman Filter subtracted). One can see that the HBEF was by far better than the (stochastic) EnKF and the variational (Var) filter. For small $N < 5$, the Var became more competitive than the EnKF, but still substantially worse than the HBEF.

The HBEF has also been tested with a doubly stochastic advection-diffusion-decay model on the circle, with similar results.

An important advantage of doubly stochastic models is that they allow the estimation of the *true* signal, forecast-error, and analysis-error probability distribution—in particular, signal and error variances—for each time instant separately (signal and error "statistics of the day"). The estimated true error variances can be compared with the respective variances produced by the filter. The estimated systematic errors (biases) in the forecast-ensemble sample variances are presented in Fig.2. One can see that, first, the biases in the ensemble variances were mostly

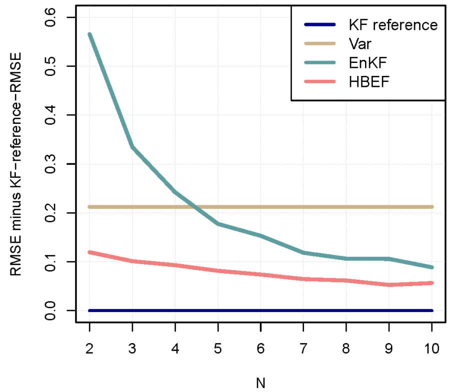


Figure 1: Analysis RMSEs for several filters as functions of the ensemble size $N$.

negative for both filters. Second, they were larger when the true variances were larger. Third, the biases for the HBEF were significantly less than those for the EnKF.

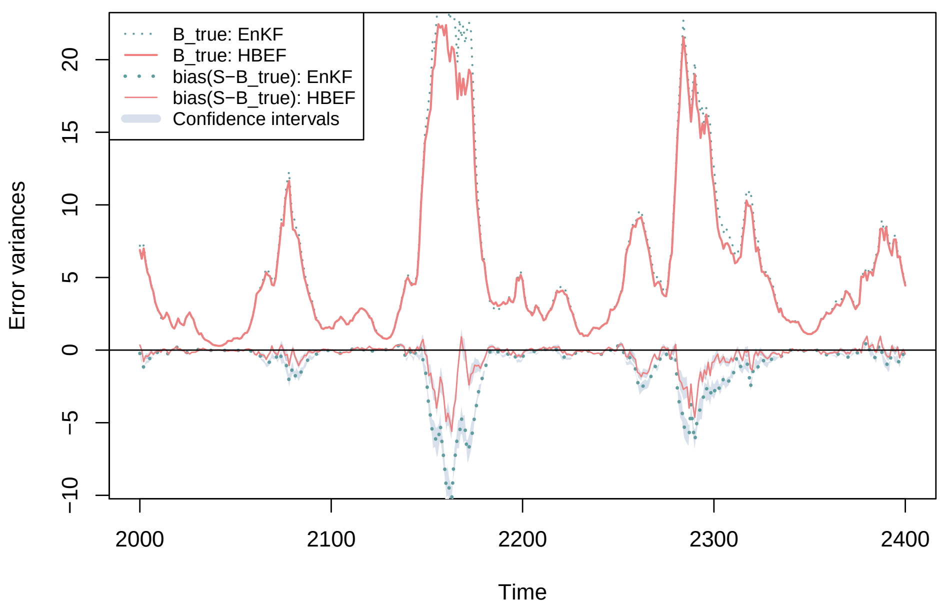


Figure 2: Segments of the time series: the true error variances for EnKF and HBEF (two upper curves) and the *biases* (with the 95% confidence intervals) in the forecast ensemble variances (two lower curves).

For a full exposition, see Tsyrulnikov and Rakitko (2017).

## Bibliography

M. Tsyrulnikov and A. Rakitko. A hierarchical Bayes ensemble Kalman filter. *Physica D (Nonlinear Phenomena)*, 338:1–16, 2017.