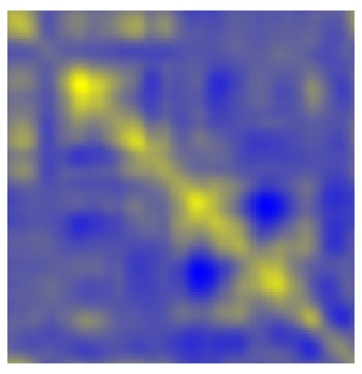


A Hierarchical Bayes based approach to ensemble-variational data assimilation

M Tsyrulnikov and A Rakitko mik.tsyrulnikov@gmail.com HydroMetCentre of Russia

Motivation

1. All existing Var, EnKF, and EnVar analysis equations assume that the effective background-error covariance matrix B is exact. But this is never the case.
2. EnVar takes a *linear combination* of static and ensemble covariances to specify B. This is ad hoc.
3. EnKF and EnVar use an ad-hoc localization. This is not theoretically optimal.
4. In the Var, EnKF, and EnVar analysis equations, there is no intrinsic feedback from observations to background-error statistics. This requires external adaptation or manual tuning.

The new technique is supposed to mitigate these problems of the existing approaches.

The proposed paradigm

In words: Acknowledge that B is uncertain and random and update it along with the state.
Observations for B: both the ensemble and the ordinary observations contain info on B.

Level	Bckg	Prior	Obs	Update
2	Static $\mathbf{B}_0$	$p(\mathbf{B}\mid\mathbf{B}_0)$	Ensm $\mathbf{X}^\mathbf{e}$	$\mathbf{B}_0 \Rightarrow \boxed{\mathbf{B}} \Leftarrow \mathbf{X}^\mathbf{e}, \mathbf{x}^\mathbf{obs}$
1	$\mathbf{x}^\mathbf{f}$ or $\overline{\mathbf{x}^\mathbf{e}}$	$p(\mathbf{x}\mid\mathbf{x}^\mathbf{b})$	$\mathbf{x}^\mathbf{obs}$	$\mathbf{x}^\mathbf{b} \Rightarrow \boxed{\mathbf{x}} \Leftarrow \mathbf{x}^\mathbf{obs}$

Level 2: extension by the new approach.
Level 1: the existing EnVar technique.

Hierarchical Bayes EnVar (HB-EnVar): principle

$$p_{post}(\mathbf{x},\mathbf{B}) \equiv p(\mathbf{x},\mathbf{B}\mid\mathbf{X}^\mathbf{e},\mathbf{x}^\mathbf{obs}) \propto p(\mathbf{B}\mid\mathbf{B}_0)\,p(\mathbf{x}\mid\mathbf{x}^\mathbf{b},\mathbf{B})\,p(\mathbf{X}^\mathbf{e}\mid\mathbf{B})\,p(\mathbf{x}^\mathbf{obs}\mid\mathbf{x})$$

where $p(\mathbf{B}\mid\mathbf{B}_0)$ is the (new) *prior* pdf for $\mathbf{B}$,
$p(\mathbf{x}\mid\mathbf{x}^\mathbf{b},\mathbf{B})$ is the traditional background-error distribution,
$p(\mathbf{X}^\mathbf{e}\mid\mathbf{B})$ is the (new) *ensemble likelihood*, and
$p(\mathbf{x}^\mathbf{obs}\mid\mathbf{x})$ is the traditional observational likelihood.

The goals are:
1) the *mode* of the *joint posterior* $p_{post}(\mathbf{x},\mathbf{B})$ (deterministic analysis).
2) the *mean* of the *marginal posterior* $p_{post}(\mathbf{x})$ (deterministic analysis).
3) a *sample* from $p_{post}(\mathbf{x})$ (ensemble analysis).

Ensemble likelihood

$$p(\mathbf{X}^\mathbf{e}\mid\mathbf{B}) \propto |\mathbf{B}|^{-N/2} e^{-\frac{1}{2}\sum_{k=1}^{N}(\mathbf{x}^\mathbf{e}_\mathbf{k}-\mathbf{x}^\mathbf{b})^\top \mathbf{B}^{-1}(\mathbf{x}^\mathbf{e}_\mathbf{k}-\mathbf{x}^\mathbf{b})},$$

– no need and no room for approximations.

So, the only term that needs explicit (and careful) specification is the prior $p(\mathbf{B}\mid\mathbf{B}_0)$.

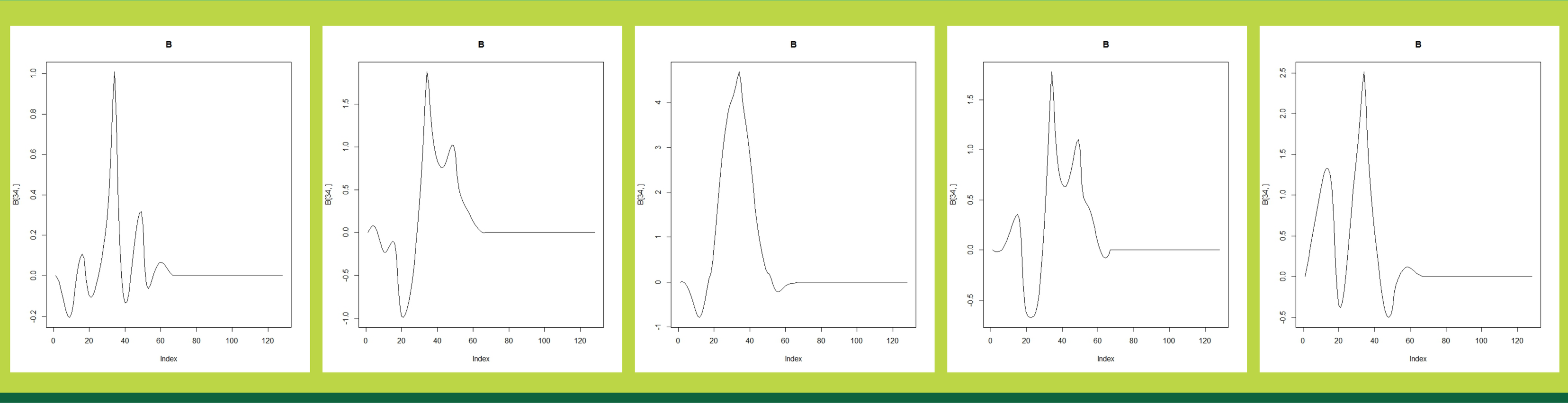
The prior pdf for B: square-root Gaussian

We decompose $\mathbf{B} = \mathbf{W}\mathbf{W}^\top$ and assume that $\mathbf{W}$ is a Gaussian random matrix with pdf

$$p(\mathbf{W}) \propto e^{-\frac{1}{2}\,\mathrm{tr}[(\mathbf{W}-\mathbf{W_0})\mathbf{U^{-1}}(\mathbf{W}-\mathbf{W_0})^\top\mathbf{U^{-1}}]}$$

Sampling: $\mathbf{W} = \mathbf{W}_0 + \mathbf{\Phi}\mathbf{Y}\mathbf{\Phi}^\top$, where $\mathbf{Y}$ is the pure-noise matrix, with $\mathcal{N}(0,1)$ independent entries.

Random samples (matrix rows) from $p(\mathrm{B})$



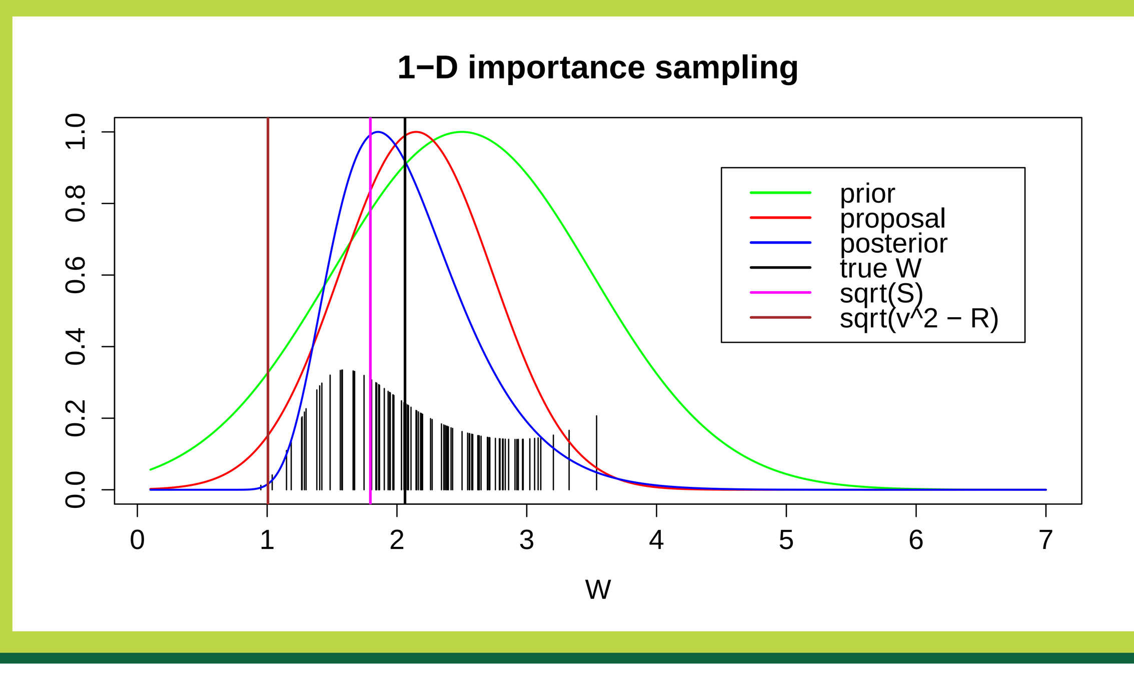
Analysis

I. **Posterior mode:** $\max p_{post}(\mathbf{x},\mathbf{B})$

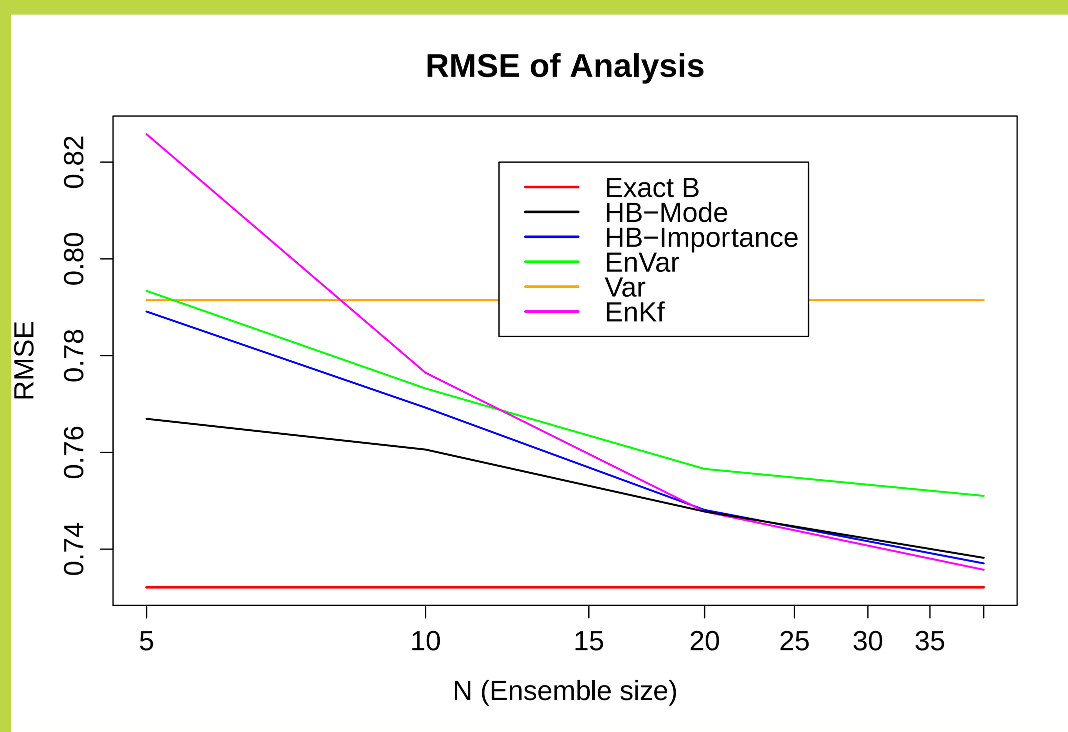
II. **Importance Sampling:** Monte-Carlo estimation of the posterior **mean**:

1. Draw M samples $\mathbf{W}^+_m$ from the proposal density $q(\mathbf{W})$.
2. Compute their non-normalized importance weights:
$$\phi'(\mathbf{W}^+_m) := \frac{p_{post}(\mathbf{W}^+_m)}{q(\mathbf{W}^+_m)}$$
3. Perform m ordinary analyses $\mathbf{x}^a_m$ with $\mathbf{B}_m = \mathbf{W}^+_m \cdot (\mathbf{W}^+_m)^\top$.
4. Average $\mathbf{x}^a_m$ with normalized importance weights w_m:
$$\overline{\mathbf{x^a}} = \sum_{m=1}^{M} w_m \cdot \mathbf{x^a}(\mathbf{B}_m)$$

Pdfs



Results



In the toy problem, the deterministic HB-EnVar analysis outperforms Var, EnKF, and EnVar.

Conclusions

Main aspects HB-EnVar

- Background-error covariance matrix $\mathbf{B}$ is treated as a sparse *random matrix* and updated in the optimal scheme along with the state.
- The key issue is the *prior distribution* of $\mathbf{B}$.
- Ensemble members are treated as *observations* on the $\mathbf{B}$ matrix and *assimilated* along with ordinary observations.
- The technique is computationally expensive.

Potential benefits of HB-EnVar

- Optimized hybridization of static and ensemble covariances.
- Optimized combination of $\mathbf{x}^f$ and $\overline{\mathbf{x}^e}$.
- Optimized covariance localization.
- Optimized feedback from $\mathbf{x}^{obs}$ to the $\mathbf{B}$ matrix.
- Uncertainty in $\mathbf{B}$ is explicitly accounted for in the generation of the analysis ensemble, resulting in *increased spread*.