Toward background error covariance hybridization for climate prediction

Sébastien Barthélémy[1,3], François Counillon[1,2,3] & Noel Keenlyside[1,3]

[1]Universitetet i Bergen, [2]Nansen Center,
[3]Bjerknes Center for Climate Research

Contact Information:
Geophysical Institute
Universitetet i Bergen
5008, Bergen, Norway

Phone: +47 955 51 294
Email: sebastien.barthelemy@uib.no

Introduction

The Norwegian Climate Prediction Model (NorCPM) combines the Norwegian Earth System Model (NorESM) with the Ensemble Kalman Filter (EnKF) and aims at providing seasonal to decadal climate predictions. On nowadays supercomputer, it is not computationally tractable to run more than 30 members (and 5 members with the high resolution version of NorCPM), which results in sampling issues when estimating the background error covariance matrix.

To overcome these issues, an hybridization method derived from previous work from (Hamill and Snyder, 2000) has been used and led to the implemeantion of 2 methods: climatological hybridization and dual resolution. These 2 methods allow for a reduction of sampling error when compared to standard EnKF.

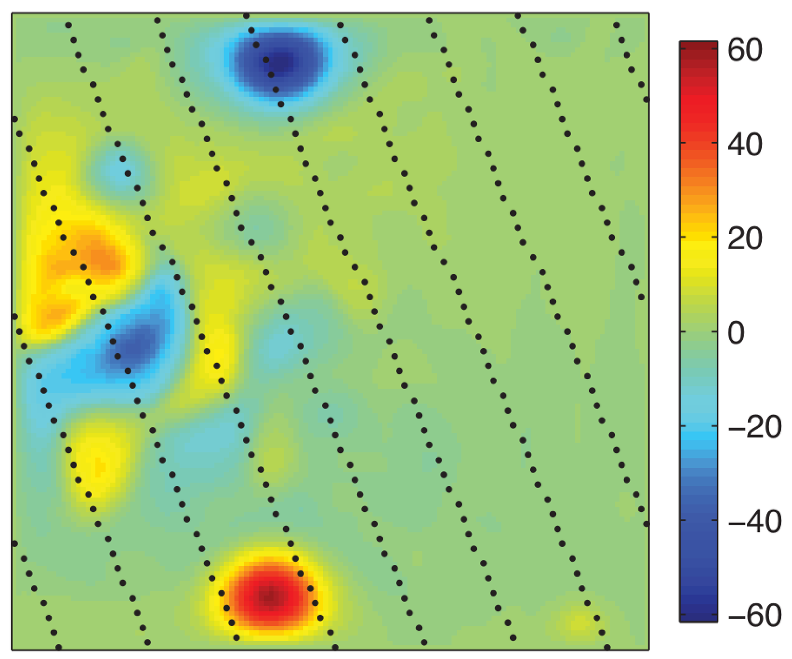
The hybrid covariance method are tested with the quasi-geostrophic model within the DAPPER package. It is shown that the method outperforms the standard implementation of the EnKF in particular for small ensemble size.

Further work will assesses the performance of the two methods with NorCPM in the context of twin experiments.

The quasi-geostrophic (QG) model

"The QG model is a derivative of the 1.5-layer reduced-gravity quasi-geostrophic model with a double gyre wind forcing" (Sakov & Oke, 2008) with following parameters for data assimilation:

- 300 observation points with $\sigma_o = 2$;
- Observation frequency : every 10 time step.



Surface elevation in QG model with obs. points (black dots)

Hybrid DEnKF

Notations

Let $\mathbf{E}$ be an ensemble of N model states, $\mathbf{x}$ the ensemble average and $\mathbf{A}$ the ensemble anomalies:

$$\mathbf{x} = \frac{1}{m}\mathbf{E}\mathbf{1}, \qquad \mathbf{A} = \mathbf{E}\left(\mathbf{I} - \frac{1}{m}\mathbf{1}\mathbf{1}^T\right)$$

where m is the number of observation point, $\mathbf{1}$ is a vector with all elements equal to 1, $\mathbf{I}$ is the identity matrix.

DEnKF Algorithm

Mean update: $\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(\mathbf{d} - \mathbf{H}\mathbf{x}^f\right)$

Anomaly update: $\mathbf{A}^a = \mathbf{A}^f - \frac{1}{2}\mathbf{K}\mathbf{H}\mathbf{A}^f$

where:

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \qquad (1)$$

$$\mathbf{P}^f = \frac{1}{N-1}\mathbf{A}^f\left(\mathbf{A}^f\right)^T \qquad (2)$$

and $\mathbf{d}$ is the observation vector.

Hybridization

Let us consider 2 ensembles $\mathbf{E_1}$ and $\mathbf{E_2}$ of N_1 and N_2 model states and their respective background error covariance matrix $\mathbf{P}_1^f$ and $\mathbf{P}_2^f$. Hybridization aims at overcoming under-sampling issues in the estimation of $\mathbf{P}^f$ (eq. (2)) by replacing $\mathbf{P}^f$ in eq. (1) by an hybrid matrix $\mathbf{P}_h^f$ that is the linear combination of matrices $\mathbf{P}_1^f$ and $\mathbf{P}_2^f$:

$$\mathbf{P}_h^f = (1-\alpha)\mathbf{P}_1^f + \alpha\mathbf{P}_2^f, \qquad \alpha \in [0;1]$$

Rewriting DEnKF in scaled ensemble observations hybrid anomalies, Sakov *et al.* 2009

Mean update:

$$\mathbf{x}_1^a - \mathbf{x}_1^f = \mathbf{A}_h^f\mathbf{G}_h\mathbf{s}_1 \qquad (3)$$

Anomaly update:

$$\mathbf{A}_1^a - \mathbf{A}_1^f = \mathbf{A}_h^f\mathbf{T} \qquad (4)$$

where:

$$\mathbf{s}_1 = \mathbf{R}^{-1/2}\left(\mathbf{d} - \mathbf{H}\mathbf{x}_1^f\right), \qquad \mathbf{S}_h = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{A}_h^f$$

$$\mathbf{A}_h^f = \sqrt{N_1+N_2-1}\left[\sqrt{\tfrac{1-\alpha}{N_1-1}}\mathbf{A}_1^f;\sqrt{\tfrac{\alpha}{N_2-1}}\mathbf{A}_2^f\right]$$

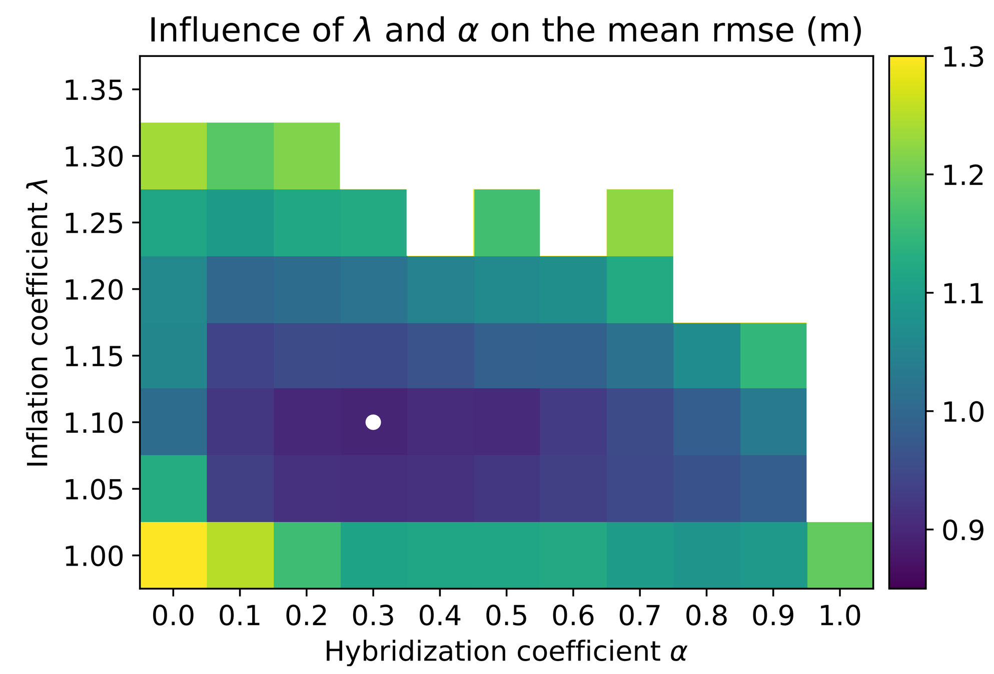
$$\mathbf{G}_h = \mathbf{S}_h^T\left(\mathbf{I} + \mathbf{S}_h\mathbf{S}_h^T\right)^{-1}, \qquad \mathbf{T} = -\tfrac{1}{2}\mathbf{G}_h\mathbf{S}_1$$

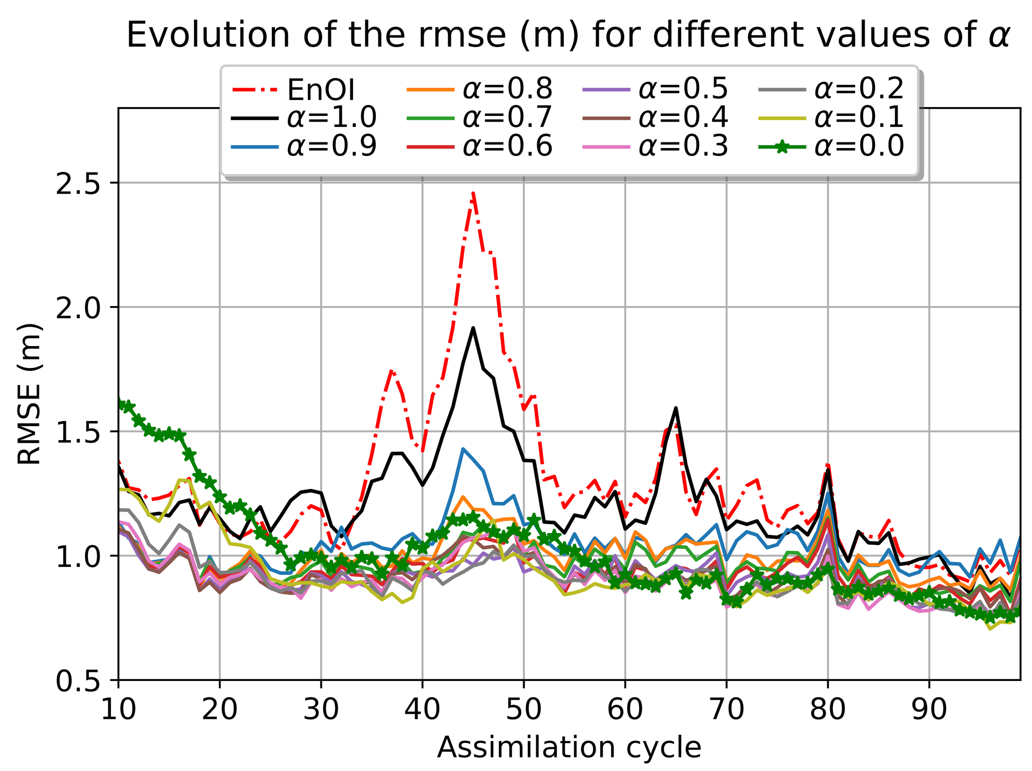
$$\mathbf{S}_1 = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{A}_1^f$$

Derived algorithms

Climatological hybridization, DEnKF-OI - Counillon *et al.*, 2009

- $\mathbf{E}_1 = \mathbf{E}_d$: small set of dynamic members $\approx$ 20 members ;
- $\mathbf{E}_2 = \mathbf{E}_s$: large set of static members $\approx$ 200 members not updated at the analysis step nor propagated forward by the model ;
- $\alpha = 0$: full dynamic $\sim$ DEnKF ;
- $\alpha = 1$: full static $\sim$ set of EnOI.





Dual resolution - Rainwater & Hunt, 2013

Use of 2 models with different resolutions and 2 different ensembles:

- High resolution model (HR): 16641 grid points ;
- Low resolution model (LR): 4225 grid points ;
- $\mathbf{E}_H$: small set of high resolution members $\approx$ 5 members ;
- $\mathbf{E}_L$: large set of low resolution members $\approx$ 120 members ;
- which is equivalent, from the point of view of computational resources, to 20 HR members or 160 LR members.

Resolution of eq. (3) and (4) on both model HR and LR with interpolated anomalies from HR (LR) model to LR (HR) model respectively:

- **HR model:** $\mathbf{E}_1 = \mathbf{E}_H$, $\mathbf{E}_2 = \mathbf{E}_L$ and

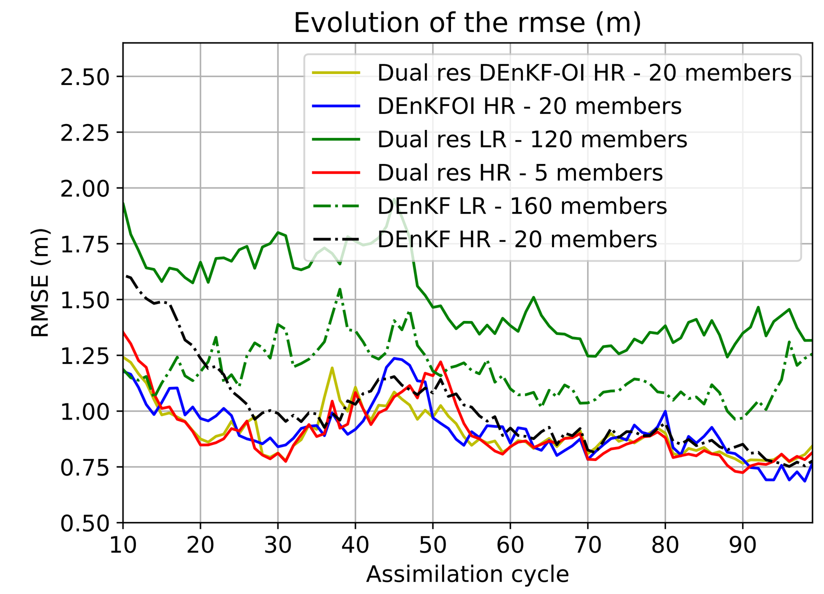
$$\mathbf{A}_h^f = \sqrt{N_H+N_L-1}\left[\sqrt{\frac{1-\alpha}{N_H-1}}\mathbf{A}_H^f;\sqrt{\frac{\alpha}{N_L-1}}\pi_{LH}\left(\mathbf{A}_L^f\right)\right]$$

- **LR model:** $\mathbf{E}_1 = \mathbf{E}_L$, $\mathbf{E}_2 = \mathbf{E}_R$ and

$$\mathbf{A}_h^f = \sqrt{N_H+N_L-1}\left[\sqrt{\frac{1-\alpha}{N_H-1}}\pi_{HL}\left(\mathbf{A}_H^f\right);\sqrt{\frac{\alpha}{N_L-1}}\mathbf{A}_L^f\right]$$

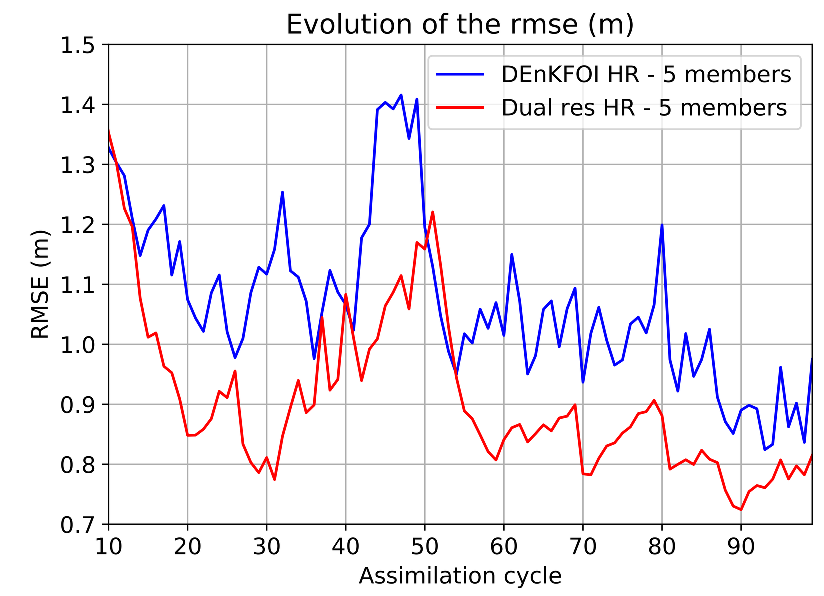
where π_{LH} (π_{HL}) is the interpolation operator from the LR (HR) model to the HR (LR) model.

Here we compare all the methods at equivalent computational cost in a configuration such that standalone HR EnKF converges.



- HR dual resolution (red line) has lower error than standard HR EnKF (black dashed line);
- Hybrid EnKF (blue line) has lower error than standard HR EnKF (black dashed line) ;
- With large dynamical ensemble size climatological hybrid (blue line) and dual resolution (red line) perform similarly.
- Combining the 2 methods (yellow line) does not yield improvement.

Here we compare the hybridization of a HR ensemble with a HR climatological ensemble (blue line) and a LR dynamic ensemble (red line).



- With low dynamical ensemble size the dual resolution (red line) outperforms the climatological hybrid (blue line) ;

Conclusions and forthcoming research

- Climatological hybridization improves the results with no additional computational cost, compared to a standalone EnKF ;
- Dual resolution method improves the results of the HR model with no additional computational cost ;
- Assessment of the climatological hybridization in NorCPM in the context of twin experiments (ongoing) ;
- Initialization of the HR version of NorCPM with the dual resolution method ;
- Determine the optimal localization and hybridization coefficients automatically, Ménétrier and Auligné, 2015.