Domain localization in ensemble based Kalman filter algorithms

T. Janjić[1], L.Nerger[1], A. Albertella[2], J.Schröter[1], S. Skachko[3]

(1) Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany, (2) Institute for Astronomical and physical Geodesy, TU Munich, Munich, Germany (3) University of Quebec in Montreal, Montreal, Canada.

1 Introduction

Ensemble Kalman filter methods are typically used in combination with one of two localization techniques. One technique is covariance localization, or direct forecast error localization, in which the ensemble derived forecast error covariance matrix is Schur multiplied with a chosen correlation matrix. The second way of localization is by domain decomposition. Here, the assimilation is split into local domains in which the assimilation update is performed independently. Domain localization is frequently used in combination with filter algorithms that use the analysis error covariance matrix for the calculation of the gain like the ETKF and the SEIK filter. Further, domain localization methods are used with method of weighting of the observations, or localization of the observation error covariance matrix.

In this work we focus on explaining the effects of domain localization in ensemble based Kalman filter algorithms and in particular effects of weighting of observations. We introduce a new method for the localization and compare it first to the already existing methods on Lorenz 40 system. On this simple example, the method of weighting of observations is less accurate than the new method, particularly if the observation errors are small.

We apply our finding to assimilation of geodetic dynamical ocean topography (DOT) data into the global finite element ocean model (FEOM) using the local SEIK filter. The geodetic DOT was obtained by means of the geodetic approach from carefully cross-calibrated multi-mission-altimeter data and GRACE gravity fields. We show that, depending on the correlation function used for weighting, the spectral properties of the solution can be improved.

2 Domain localization

Disjoint domains in the physical space are considered as domains on which the analysis is performed. Therefore, for each subdomain an analysis step is performed independently using observations not necessarily belonging only to that subdomain. Results of the local analysis steps are pasted together and then the global forecast step is performed.

Basic properties:

- The localized error covariance is calculated using

$$\mathbf{P}_k^{f,loc} = \sum_{i,j=1}^{r+1,L} \mathbf{u}_{i,j}\mathbf{u}_{i,j}^T$$

  $\mathbf{u}_{i,j} = \frac{1}{\sqrt{r}}[\mathbf{x}^{f,i}(t_k) - \mathbf{x}_k^f] \circ \mathbf{1}_{Dj}$ with $j = 1, \ldots, L$ and L is the number of subdomains. $\mathbf{1}_{Dj}$ is a vector whose elements are 1 if the corresponding point belongs to the domain Dj.
- Using the simple property of Schur product $(\mathbf{a} \circ \mathbf{c})(\mathbf{b} \circ \mathbf{d})^T = (\mathbf{ab}^T) \circ (\mathbf{cd}^T)$ the localized error covariance can be represented as $\mathbf{P}_k^f \circ \mathbf{C}$.
- $\mathbf{C}$ positive semidefinite, has block structure and is the sum of rank one matrices $\mathbf{1}_{Dj}\mathbf{1}_{Dj}^T$. $\mathbf{C} = \sum_{j=1}^L \mathbf{1}_{D_j}\mathbf{1}_{D_j}^T$.
- The rank of matrix $\mathbf{C}$ corresponds to the number of subdomains. In practice the subdomains may have to be made quite small, to ensure that rank($\mathbf{C}$) is large enough.
- This is in contrast to the direct localization methods where $\mathbf{C}$ is full rank, positive definite, isotropic matrix, compactly supported. Usually 5th order polynomial correlation function (Gaspari and Cohn 1999) is used.

We introduce new method (SD+Loc) and compare it to already existing methods:

Method (SD+):Let $\mathbf{1}_{Dmj}$ be a vector that has a value of 1 if the observation belongs to the domain Dm otherwise has a value of 0, and let $Dj \subseteq Dmj$. where matrix $\sum_{j=1}^L \mathbf{1}_{Dmj}\mathbf{1}_{Dj}^T$. Use for each subdomain $(\mathbf{1}_{Dmj}\mathbf{1}_{Dj}^T) \circ \mathbf{H_k}\mathbf{P}_k^f$ and $\mathbf{1}_{Dmj}\mathbf{1}_{Dmj}^T \circ \mathbf{H_k}\mathbf{P}_k^f\mathbf{H_k}^T$.

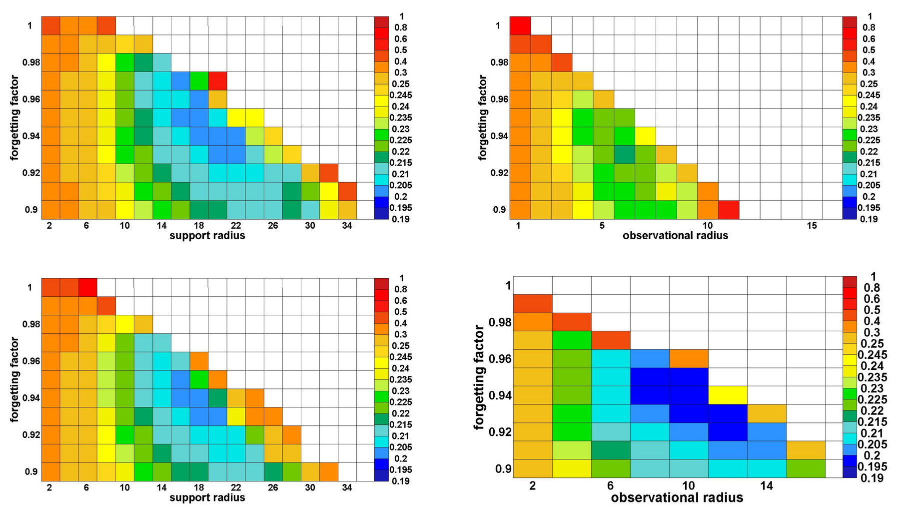
Method (SD+Loc): Use for each subdomain $(\mathbf{1}_{Dmj}\mathbf{1}_{Dj}^T) \circ \mathbf{H_k}\mathbf{P}_k^f \circ \mathbf{H_k}\mathbf{C}$ and $\mathbf{1}_{Dmj}\mathbf{1}_{Dmj}^T \circ \mathbf{H_k}\mathbf{P}_k^f\mathbf{H_k}^T \circ \mathbf{H_k}\mathbf{C}\mathbf{H_k}^T$.

Method (SD+ObsLoc): Its implementation requires for each observation a weight that depends on the distance of the observation from the analysis location (Penduff et al. 2002; Hunt et al. 2007).

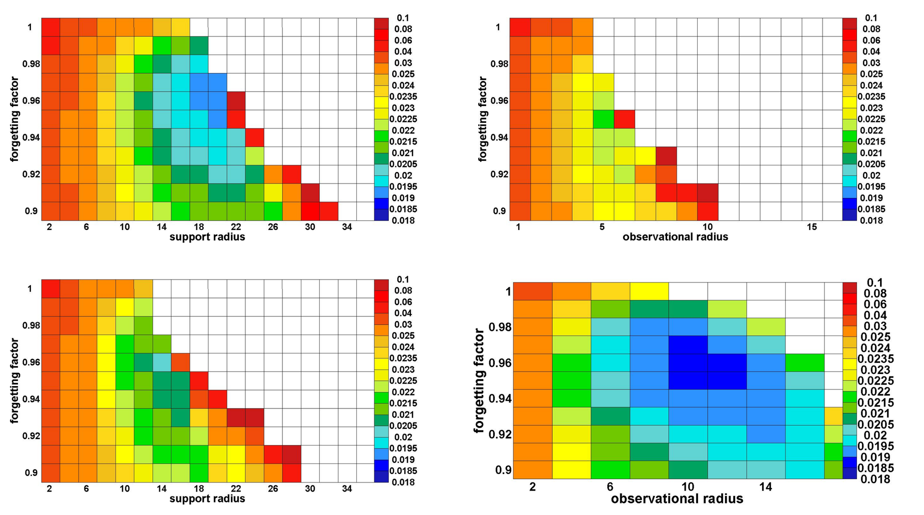
Method (GLocEn): An ensemble square root filter as in Whitaker and Hamill 2002 is applied with covariance localization is applied.

3 Lorenz40 model experiments

- Lorenz40 model is governed by 40 coupled ordinary differential equations in domain with cyclic boundary conditions.
- The experimental setup follows Whitaker and Hamill (2002).
- The observations are given as a vector of values contaminated by uncorrelated normally distributed random noise with standard deviation of 1 and 0.1.
- A 10-member ensemble is used.
- After a spin-up period of 1000 time steps, assimilation is performed for another 50 000 time steps.



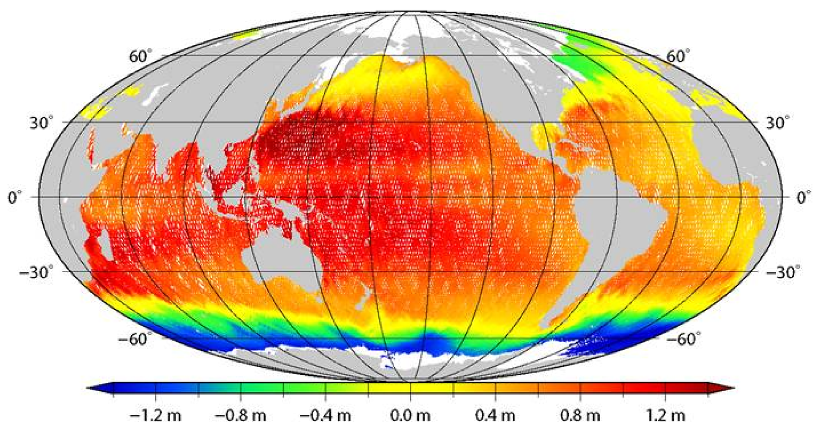
L40 experiment results for $\sigma_{obs} = 1$. RMS error for different covariance localization techniques GLocEn (upper left), SD+ (upper right), SD+ObLoc (lower left) and method SD+Loc (lower right).



L40 experiment results for $\sigma_{obs} = 0.1$. RMS error for different covariance localization techniques. GLocEn (upper left), SD+ (upper right), SD+ObLoc (lower left) and method SD+Loc (lower right).

4 DOT data assimilation experiments

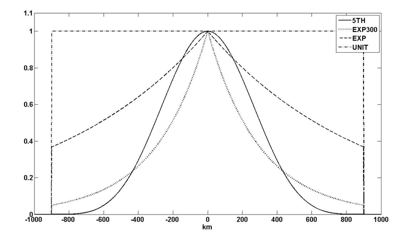
- Localization is applied to a realistic oceanographic problem: the assimilation of absolute dynamical ocean topography (DOT) into a global finite element ocean circulation model.
- Finite Element Ocean Model (FEOM) developed at the Alfred-Wegener Institut (AWI), solves the standard set of hydrostatic ocean dynamic primitive equations (Wang et al 2008).
- Prismatic mesh, 1^o resolution global ocean model, 25 levels, monthly forcing is used.
- Domain localized SEIK filter (Pham et al. 1998, Pham 2001, Nerger et al. 2006) as coded within PDAF (Nerger et al. 2005) is chosen for experiment.
- The DOT was obtained by means of geodetic approach from carefully cross-calibrated multi-mission-altimeter data and GRACE gravity fields.
- Spectral consistency is achieved by applying a Gauss-type filter (Jekeli/Wahr) on sea surface and geoid. The filter length is set to 241km. (Savcenko and Bosch 2010)



Data assimilation scheme corrects all the ocean fields, although only geodetic DOT is assimilated. Radar altimetry cannot be used for those regions where the sea-ice coverage exceeds a certain percentage during the entire year, as well as for ice shelves and near-coastal zones. Therefore it is interesting to compare the assimilation results in the area which is not well observed.

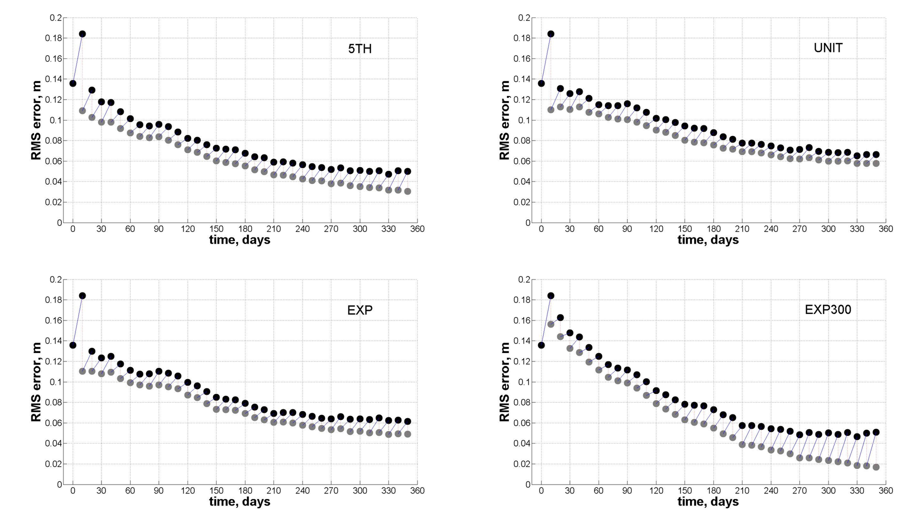
5 Choice of the correlation function

Goal: To study the filtering behavior when different correlation functions for the weighting of observations are applied using one method (SD+ObsLoc).

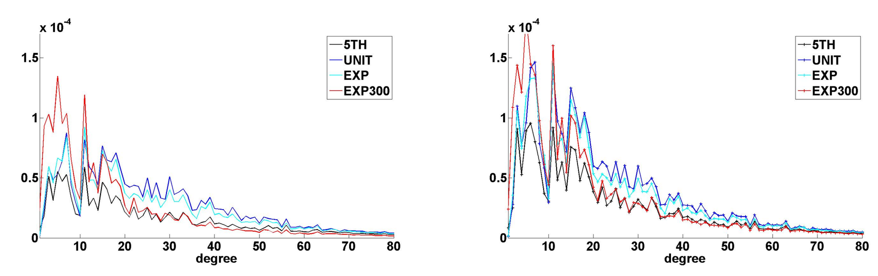


Correlation function used for localization.

- Different correlation function are used with the method SD+ObsLoc.
- The observational error standard deviation is 5 cm.
- Observations within radius of 900 km are used.

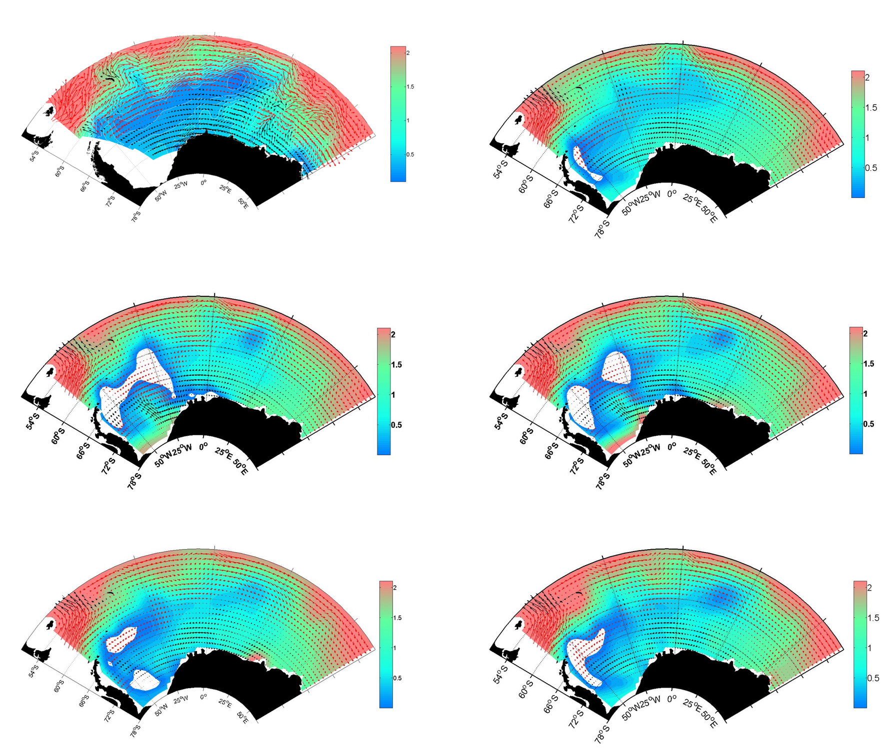


RMS error for different covariance localization techniques.



Spectral properties of the errors. Logarithm of the spectral difference between analysis and the data (left) and forecast and the data (right) depending on spherical harmonic degree.

$$\epsilon_\ell^{oi} = \sum_m \left(T_{\ell m}^o - T_{\ell m}^i\right)^2$$



In-situ temperature at 800 m depth. Composite from the ARGO data (1999 to 2010) courtesy of Dr. Klatt (upper left). Model only (upper right). As result of assimilation of geodetic DOT filtered up to 241 km and UNIT (middle left), EXP (middle right), EXP300 (lower left) and 5TH (lower right).

6 Conclusion

- The different domain localization techniques have been investigated here and compared to direct forecast error localization on L40 model.
- It was shown that domain localization is equivalent to direct forecast error localization with a Schur product matrix that has a block structure, not isotropic, and positive semidefinate.
- An algorithm is presented that for each subdomain of ensemble localization uses observations from a domain larger than the ensemble subdomain and a Schur product with an isotropic matrix on each subdomain.
- The algorithms that use the full rank covariances show superior performance for both Lorenz 40 and the example of assimilation of DOT.
- Localization function determents spectral properties and accuracy of the solution as seen for both L40 model as for the realistic global assimilation of DOT.

References

[1] Bosch, W. and R. Savcenko (2010): On estimating the dynamic ocean topography - a profile approach, *IAG Symposia,Mertikas (Ed.) Gravity, Geoid and Earth Observation*, Springer ,**135**.

[2] B. R. Hunt and E. J. Kostelich and I. Szunyogh (2007): Efficient Data Assimilation for Spatiotemporal Chaos: A local Ensemble Transform Kalman filter., *Physica D*,**230**, 112-126.

[3] T. Janjić and L. Nerger and A. Albertella and J. Schroeter and S. Skachko (2011): On domain localization in ensemble based Kalman filter algorithms, *Special collection -Intercomparisons of 4D-Variational Assimilation and the Ensemble Kalman Filter, Early online edition Mon. Wea. Rev, doi: 10.1175/2011MWR3552.1.*

[4] T. Janjić, J. Schröter, A. Albertella, W. Bosch, R. Rummel, R. Savcenko, J. Schwabe, M. Scheinert (2011): Assimilation of geodetic dynamic ocean topography using ensemble based Kalman filter, *J. of Geodynamics*,submitted.

[5] L. Nerger and W. Hiller and J. Schröter (2005): PDAF - the Parallel Data Assimilation Framework: Experiences with Kalman filtering. In W. Zwieflhofer and G. Mozdzynski, editors, Use of High Performance Computing in Meteorology - Proceedings of the 11. ECMWF Workshop, pages 63–83. World Scientific.

[6] L. Nerger and S. Danilov and W. Hiller and J. Schröter (2006): Using sea level data to constrain a finite-element primitive-equation ocean model with a local SEIK filter., *Ocean Dyn.* **56**, 634-649.

[7] T. Penduff and P. Brasseur and C.-E. Testut and B. Barnier and J. Verron (2002): A four-year eddy-permitting assimilation of sea-surface temperature and altimetric data in the South Atlantic Ocean.,*J. Mar. Res.*, **60**,805-833.

[8] D. T. Pham and J. Verron and L. Gourdeau (1998) : Singular evolutive Kalman filters for data assimilation in oceanography *C. R. Acad. Sci., Ser. II 4*, **326**, 255-260.

[9] D. T. Pham (2001): Stochastic methods for sequential data assimilation in strongly nonlinear systems, *Mon. Wea. Rev.*, **129**, 1194-1207.

[10] Q. Wang and S. Danilov and J. Schröter (2008): Ocean circulation Model based on triangular prismatic elements with application in studying the effect of topography representation, *J. Geophys. Res*, **113**, C05015.

[11] J. S. Whitaker and T. M. Hamill (2002): Ensemble Data Assimilation without Perturbed Observations. *Mon. Wea. Rev*, **130**, 1913–1924.