

Overview of localization techniques for ensemble based Kalman filter algorithms

T.Janjić

Alfred Wegener Institute for Polar and Marine Research

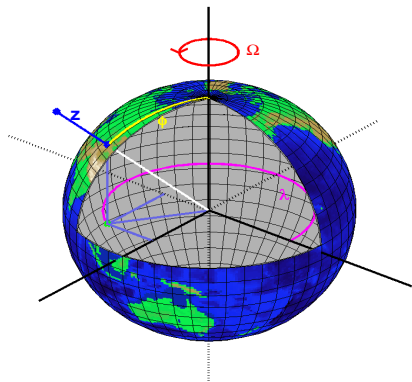
July 5th, 2011

Outline

- Large scale system
- Covariances important but not well represented through ensemble
- Localization required
- Properties of different localization schemes
- Localization and balance
- Open questions and conclusions

Dynamics of the atmosphere

- $p(\lambda, \phi, z, t) \equiv$ Pressure
- $T(\lambda, \phi, z, t) \equiv$ Temperature
- Wind field:
 - $u(\lambda, \phi, z, t) \equiv$ zonal wind
 - $v(\lambda, \phi, z, t) \equiv$ meridional wind
 - $w(\lambda, \phi, z, t) \equiv$ vertical wind
- $q(\lambda, \phi, z, t) \equiv$ specific humidity



Dynamics of the atmosphere-represented by model

$$\begin{aligned}\dot{u} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} &= -\frac{1}{\rho} \frac{\partial p}{\partial x} + 2\Omega v \sin \varphi + F_x \\ \dot{v} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} &= -\frac{1}{\rho} \frac{\partial p}{\partial y} - 2\Omega u \sin \varphi + F_y \\ \dot{w} + u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z} &= -\frac{1}{\rho} \frac{\partial p}{\partial z} + 2\Omega u \cos \varphi - g + F_z\end{aligned}$$

Thermodynamic energy equation in Cartesian coordinates

$$\frac{\partial T}{\partial t} = -u \frac{\partial T}{\partial x} - v \frac{\partial T}{\partial y} - w \frac{\partial T}{\partial z} - \frac{RT}{c_v} \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) + J,$$

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{v})$$

Dynamics of the atmosphere represented with numerical model.

Dynamics of the atmosphere

The state $\mathbf{x} \equiv \mathbf{x}(\lambda, \phi, z, t)$ of the atmosphere at time t_k :

$$\mathbf{x}(\lambda, \phi, z, t_k) \equiv \begin{bmatrix} u(\lambda, \phi, z, t_k) \\ \vdots \\ T(\lambda, \phi, z, t_k) \end{bmatrix}$$

10^9 unknowns and 10^7 observations per 24 hour period

Ocean: 10^6 unknowns and mostly surface observations per 10 days

Data assimilation algorithms

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k^o - \mathbf{H}_k\mathbf{x}_k^f),$$

\mathbf{x}_k^f is a *prior* estimate, i.e. forecast, at time t_k , of size n

\mathbf{P}_k^f is forecast error covariance matrix,

\mathbf{H}_k is the observation operator,

\mathbf{y}_k^o is the p_k -vector of observations.

\mathbf{K}_k is taken as

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1}.$$

$$\mathbf{x}_{k+1}^f = F_{k+1,k} \mathbf{x}_k^a$$

Ensemble approach

The complete error structure of a time-evolving model trajectory requires a multidimensional pdf which is impossible to know or even represent accurately

Covariances represented through

$$\mathbf{P}_k^f = \frac{1}{r} \sum_{i=1}^{r+1} [\mathbf{x}^{f,i}(t_k) - \mathbf{x}_k^f][\mathbf{x}^{f,i}(t_k) - \mathbf{x}_k^f]^T.$$

\mathbf{P}_k^f is the ensemble derived forecast error covariance;

$\mathbf{x}^{f,i}(t_k)$ are ensemble members $i = 1, \dots, r + 1$ of size n at time t_k ;

\mathbf{x}_k^f is the average over ensemble.

$$\mathbf{x}_k^f = \frac{1}{r+1} \sum_{i=1}^{r+1} \mathbf{x}^{f,i}(t_k)$$

\mathbf{P}_k^f is by definition positive semi-definite covariance with rank at most r .

Covariances are important

One observations at location j , state vector of size n

$$\mathbf{H} = [0 \dots 0 \ 1 \ 0 \dots 0]$$

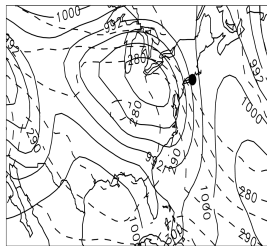
$$\begin{aligned}\mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k^o - \mathbf{H}_k \mathbf{x}_k^f), \\ \mathbf{K}_k &= \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1}\end{aligned}$$

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \frac{(y_k^o - \mathbf{H}_k \mathbf{x}_k^f)}{\sigma_{obs}} \begin{bmatrix} \mathbf{P}_{1j}^f \\ \vdots \\ \mathbf{P}_{nj}^f \end{bmatrix}$$

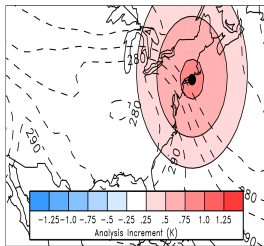
- Correction to the forecast field is proportional to a column j of matrix \mathbf{P}^f .
- This matrix determines how information from single observations is spread to neighborhood grid points
- This matrix determines how information is spread to other variables of the model

Covariances are important

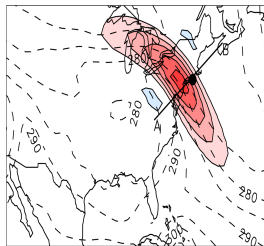
1000 hPa temperature (K) and
surface pressure (hPa)



3D-Var increment



Ensemble Filter Increment



Output from a “single-observation” experiment. The EnKF is cycled for a long time. The cycle is interrupted and a single observation 1K greater than the mean prior is assimilated. Maps of the analysis minus first guess are plotted. These “analysis increments” are proportional to the background-error covariances between every other model grid point and the background at the observation location.

Properties of ensemble derived covariances

- Cross correlations are represented naturally
- Evolutions can be calculated over long time period
- Covariances are flow dependent

However

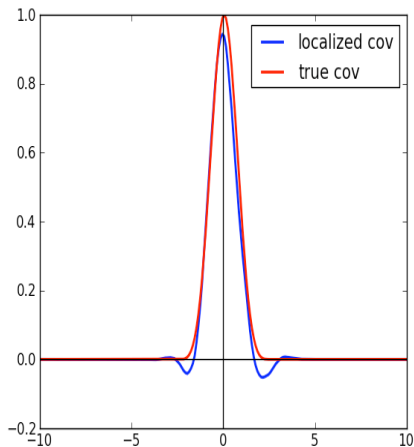
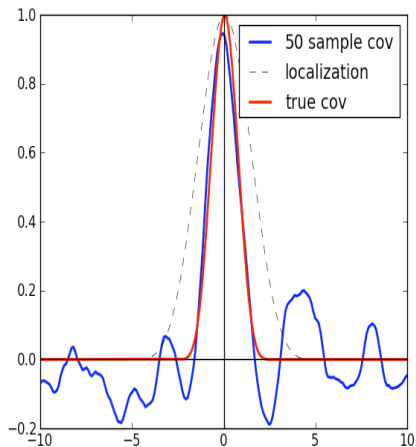
- only small number of ensembles can be evolved due to complexity of the dynamical systems
- Due to the small ensemble numbers covariances are far from representing correctly uncertainty

⇒ Localization needed

Outline localization

- What is localization?
- Two basic approaches for localization:
 - Covariance localization or direct forecast error localization (used in Houtekamer and Mitchell (1998, 2001))
 - Domain localization (used in Haugen and Evensen 2002; Brusdal et al. 2003; Evensen 2003; Brankart et al. 2003; Ott et al. 2004; Nerger et al. 2006; Hunt et al. 2007; Miyoshi and Yamane 2007)
- Simple 1D experiment
- Localization and Balance
- Conclusion

What is localization?

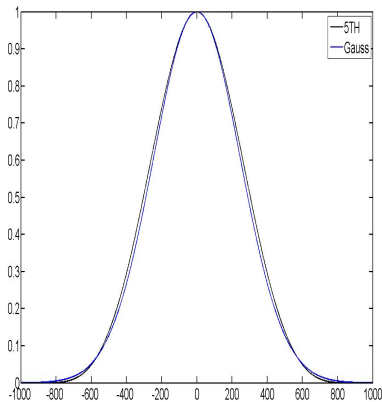


Estimates of the covariance from small ensemble size will be noisy, especially signal to noise ratio is large when covariances are small (from Hamill and Whitaker 2009).

Direct forecast error localization or covariance localization

By covariance localization we cut out distance correlation. Therefore we include two more sources of information when constructing covariance from ensemble:

- Distant correlation are not important
- Positive definite correlation matrix (Gaspari and Cohn QJ 1999)



Direct forecast error localization or covariance localization

Covariance localization: The ensemble derived forecast error covariance matrix is Schur multiplied with a stationary a priori chosen correlation matrix that is compactly supported.

Let \mathbf{C} be a matrix of rank M that is used for the Schur product. Let \mathbf{v}_j represent eigenvectors of matrix \mathbf{C} multiplied with the square root of the corresponding eigenvalue.

$$\mathbf{C} = \sum_{j=1}^M \mathbf{v}_j \mathbf{v}_j^T.$$

Let \circ denotes the element-wise product (Schur product)

Schur product theorem: If \mathbf{A} , \mathbf{B} are positive semi-definite matrices, then $\mathbf{A} \circ \mathbf{B}$ is also positive semi-definite. If \mathbf{A} , \mathbf{B} are positive definite matrices, then $\mathbf{A} \circ \mathbf{B}$ is also positive definite.

Covariance localization

For any vectors \mathbf{a} , \mathbf{b} , \mathbf{c} and \mathbf{d} :

$$(\mathbf{a} \circ \mathbf{c})(\mathbf{b} \circ \mathbf{d})^T = (\mathbf{a}\mathbf{b}^T) \circ (\mathbf{c}\mathbf{d}^T).$$

Basic properties:

- The localized error covariance $\mathbf{P}_k^f \circ \mathbf{C}$ can be represented as

$$\sum_{i,j=1}^{r+1,M} \mathbf{u}_{i,j} \mathbf{u}_{i,j}^T \quad \text{with } \mathbf{u}_{i,j} = \frac{1}{\sqrt{r}} [\mathbf{x}^{f,i}(t_k) - \mathbf{x}_k^f] \circ \mathbf{v}_j$$

This representation implies that instead of using ensemble members $\mathbf{x}^{f,i}$ for the calculation of the analysis error covariance, we can use the ensemble $\mathbf{u}_{i,j}$, and the same formulas as in original algorithms apply.

- \mathbf{C} full rank, positive definite, isotropic matrix, compactly supported. Usually 5th order polynomial correlation function (Gaspari and Cohn 1999).
- $\min(\text{diag}(\mathbf{P}_k^f)) \lambda_{\min}(\mathbf{C}) \leq \lambda_{\min}(\mathbf{P}_k^f \circ \mathbf{C}) \leq \lambda_{\max}(\mathbf{P}_k^f \circ \mathbf{C}) \leq \max(\text{diag}(\mathbf{P}_k^f)) \lambda_{\max}(\mathbf{C})$

Example

The state vector \mathbf{w} to be estimated will be taken as a realization of normally distributed random function $w(y) \sim \mathcal{N}(0, W(y_1, y_2))$ on the circle of radius $D/2\pi$, where the covariance $W(y_1, y_2)$ is either

$$W(y_1, y_2) = \left(1 + \frac{|y_1 - y_2|}{L}\right) e^{-\frac{|y_1 - y_2|}{L}}, \quad (1)$$

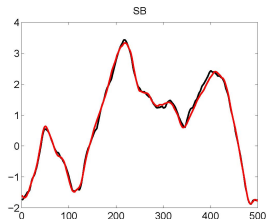
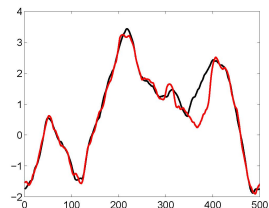
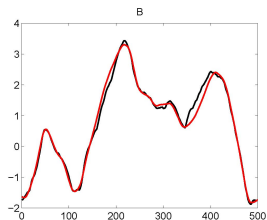
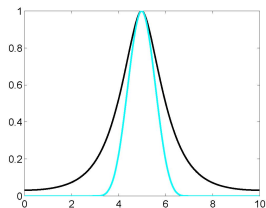
or

$$W(y_1, y_2) = e^{-\frac{|y_1 - y_2|}{L}}. \quad (2)$$

Here, $|y_1 - y_2|$ represents the chord length between the points y_1 and y_2 on the circle of radius $D/2\pi$.

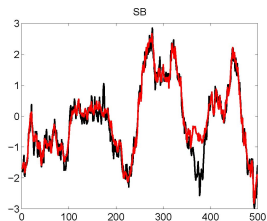
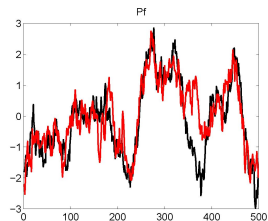
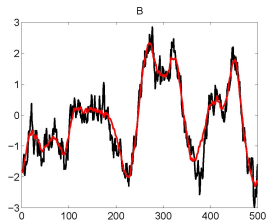
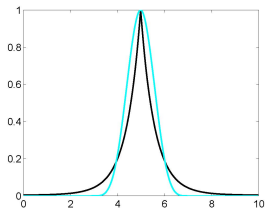
The observations are given as a vector of values of the realization at all grid points contaminated by normally distributed random noise with standard deviation of 0.05, the observations from two subdomains were removed.

Example



Upper Left: True covariance (black) and approximate B covariance (blue). **Upper Right:** True state (black) and analysis (red) after one assimilation step with approximate B covariance. **Lower Left:** True state (black) and analysis (red) after one assimilation step with ensemble covariance from 30 ensemble members. **Lower Right:** True state (black) and analysis (red) after one assimilation step with localized ensemble covariance.

Example nonsmooth field cont.



Upper Left: True covariance (black) and approximate B covariance (blue).

Upper Right: True state (black) and analysis (red) after one assimilation step with approximate B covariance.

Lower Left: True state (black) and analysis (red) after one assimilation step with ensemble covariance from 30 ensemble members.

Lower Right: True state (black) and analysis (red) after one assimilation step with localized ensemble covariance.

Covariance localization

- Distant correlation are removed
- Positive definite correlation matrix (Gaspari and Cohn QJ 1999) is introduced that increases the rank of forecast error covariance and this way
- increases space where the solution can be searched for
- usually correlation function is chosen with full rank
- More accurate solution are obtained
- Data sparse areas are estimated well

Domain localization

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

$$\mathbf{P}_k^{f,loc} = \sum_{i,j=1}^{r+1,L} \mathbf{u}_{i,j} \mathbf{u}_{i,j}^T \quad (3)$$

where $\mathbf{u}_{i,j} = \frac{1}{\sqrt{r}} [\mathbf{x}^{f,i}(t_k) - \mathbf{x}_k^f] \circ \mathbf{1}_{D_j}$ with $j = 1, \dots, L$ and L is the number of subdomains. Here $\mathbf{1}_{D_j}$ is a vector whose elements are 1 if the corresponding point belongs to the domain D_j .

Domain localization

- \mathbf{C} positive semidefinite, has block structure and is the sum of rank one matrices $\mathbf{1}_{D_j}\mathbf{1}_{D_j}^T$. The rank of matrix \mathbf{C} corresponds to the number of subdomains.
- In case that $\text{rank}(\mathbf{C})\text{rank}(\mathbf{P}_k^f) < n$, the matrix $\mathbf{C} \circ \mathbf{P}_k^f$ is singular.

Why is domain localization used?

- As for OI, one of the major advantages of using domain localization is **computational**. The updates on the smaller domains can be done independently, and therefore in parallel.
- In certain algorithms this is more natural way of localizing. Examples of such methods are the ensemble transform Kalman filter ETKF and the singular evolutive interpolated Kalman filter SEIK.

Why is domain localization used?

- In these algorithms, the forecast error covariance matrix is never explicitly calculated. Therefore, direct forecast localization as in Houtekamer and Mitchell (1998, 2001) is not immediately possible.
- In these methods an ensemble resampling in SEIK or transformation is used that ensures that the ensemble statistics represent exactly the analysis state and error covariance matrix.
- Ways of including full rank, positive definite and isotropic matrix in domain localized algorithms were developed. Two methods will be presented [Method SD+Loc](#) and [Method SD+ObsLoc](#) introduced by Hunt et al. 2007.

Method SD+Loc

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 $Dj \subseteq Dmj$.

$$\begin{aligned} \frac{1}{r} \sum_{i=1}^{r+1} \sum_{j=1}^L [\mathbf{H}_k \mathbf{x}^{f,i}(t_k) \circ \mathbf{1}_{Dmj} - \mathbf{H}_k \mathbf{x}_k^f \circ \mathbf{1}_{Dmj}] [\mathbf{x}^{f,i}(t_k) \circ \mathbf{1}_{Dj} - \mathbf{x}_k^f \circ \mathbf{1}_{Dj}]^T \\ = \sum_{j=1}^L (\mathbf{1}_{Dmj} \mathbf{1}_{Dj}^T) \circ \mathbf{H}_k \mathbf{P}_k^f \end{aligned}$$

where matrix $\sum_{j=1}^L \mathbf{1}_{Dmj} \mathbf{1}_{Dj}^T$ has entries of zeros and ones since the domains Dj are disjoint.

Method (SD+Loc): An modification to this algorithm is to 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 $\mathbf{1}_{Dmj} \mathbf{1}_{Dmj}^T \circ \mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T \circ \mathbf{H}_k \mathbf{C} \mathbf{H}_k^T$.

Observational error localization: Method (SD+ObLoc)

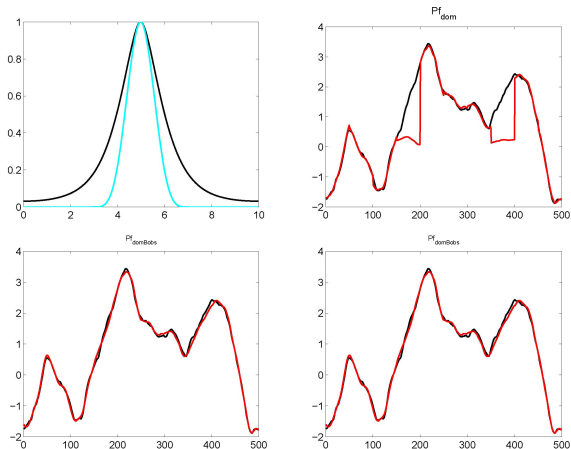
The observation localization method modifies the observational error covariance matrix \mathbf{R} .

Let us consider a single observation example, in [observation error localization method](#), the observation error σ_{obs}^2 is modified to $\sigma_{obs}^2 / weight_d$ where $weight_d$ can be calculated using any of the correlation functions.

Accordingly, the analysis increment is multiplied by $weight_d pf / (weight_d + \sigma_{obs}^2)$, where $weight_d$ depends on the distance between observation and analysis point.

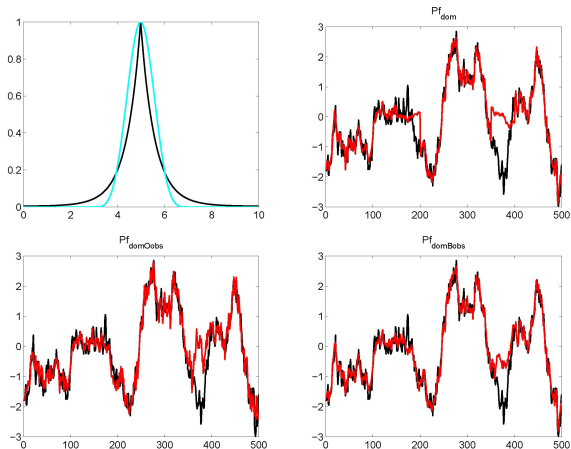
Note, for direct forecast error localization this factor is $weight_d pf / (1 + \sigma_{obs}^2)$.

Example domain localization cont.



Upper Left: True covariance (black) and approximate B covariance (blue). **Upper Right:** True state (black) and analysis (red) after one assimilation step with domain localized covariance. **Lower Left:** True state (black) and analysis (red) after one assimilation step with domain localized with overlapping observations. **Lower Right:** True state (black) and analysis (red) after one assimilation step with localized ensemble covariance with overlapping observations and B.

Example domain localization cont.

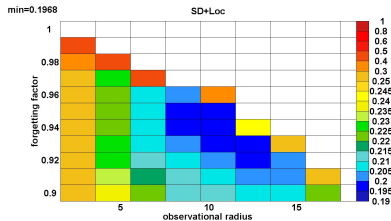
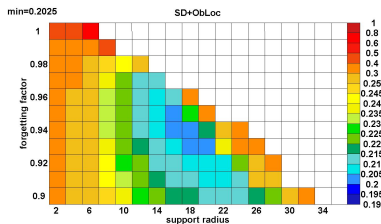
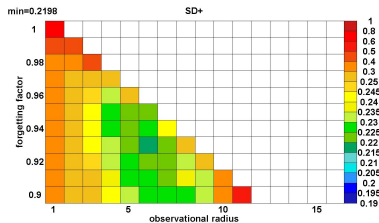
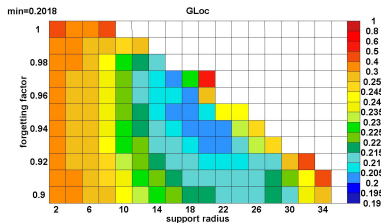


Upper Left: True covariance (black) and approximate B covariance (blue). **Upper Right:** True state (black) and analysis (red) after one assimilation step with domain localized covariance. **Lower Left:** True state (black) and analysis (red) after one assimilation step with domain localized with overlapping observations. **Lower Right:** True state (black) and analysis (red) after one assimilation step with localized ensemble covariance with overlapping observations and B.

Model Lorenz40

- $dX_i/dt = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F$
- Lorenz40 model is governed by 40 coupled ordinary differential equations in domain with cyclic boundary conditions.
- The state vector dimension is 40.
- The observations are given as a vector of values contaminated by uncorrelated normally distributed random noise with standard deviation of 1.
- The observations are assimilated at every time step.
- After a spin-up period of 1000 time steps, assimilation is performed for another 50 000 time steps.
- A 10-member ensemble is used.

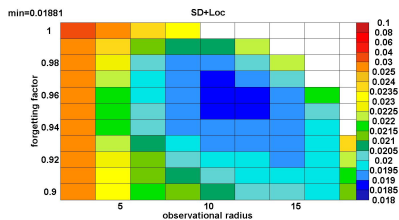
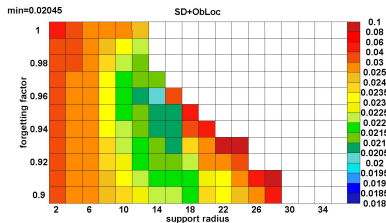
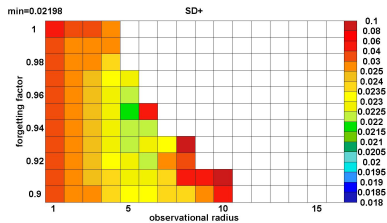
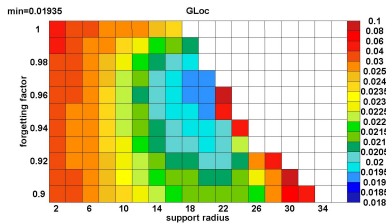
L40 results: $\sigma_{obs} = 1$



RMS error for different covariance localization techniques.

See Janjic et al. 2011 MWR for more details.

L40 results: $\sigma_{obs} = 0.1$



RMS error for different covariance localization techniques.

Domain localization conclusion

- The domain localization technique has been investigated here and compared to direct forecast error localization on simple example and 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 and is not isotropic.
- The rank of the matrix corresponding to the domain localization depends on the number of subdomains that are used in the assimilation. This matrix is positive semidefinite.
- Inclusion of positive definite matrix either through method SD+Loc or SD+ObsLoc is beneficial for domain localization methods

Localization and balance

Assume we have two variables h and v defined at the model grid points, i.e. \mathbf{h} and \mathbf{v} :

$$\mathbf{P}_k^f \equiv \begin{bmatrix} \text{cov}(\mathbf{h}, \mathbf{h}) & \text{cov}(\mathbf{h}, \mathbf{v}) \\ \text{cov}(\mathbf{v}, \mathbf{h}) & \text{cov}(\mathbf{v}, \mathbf{v}) \end{bmatrix}$$

Let us assume that we want to apply direct forecast error localization with diagonal matrix then

$$\mathbf{P}_k^f \circ \mathbf{I} \equiv \begin{bmatrix} \text{cov}(h_1, h_1) & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots \\ 0 & 0 & \text{cov}(h_n, h_n) & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 \dots 0 & \text{cov}(v_n, v_n) \end{bmatrix}$$

Localization and balance

By applying localization we destroyed correlation given by numerical model between \mathbf{h} and \mathbf{v} .

Lets look at one example from Greybush et al. MWR 2011

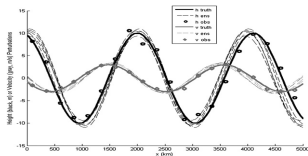
$$fv_g = g \frac{\partial h}{\partial x}$$

$$h = h_{depth} + h_{amp} \cos k(x - x_{ps})$$

$$v_g = -\frac{g}{f} h_{amp} \sin k(x - x_{ps})$$

Solve for \mathbf{h} and \mathbf{v} using observations of both \mathbf{h} and \mathbf{v} and methods (SD+ObsLoc) and direct forecast error localization with $\exp \frac{-d(i,j)^2}{2L^2}$.

Example from Greybush et al. MWR 2011



Since each ensemble is in geostrophic balance and observations are also, analysis should be too.

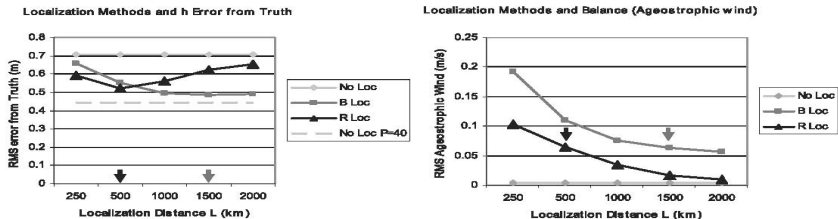
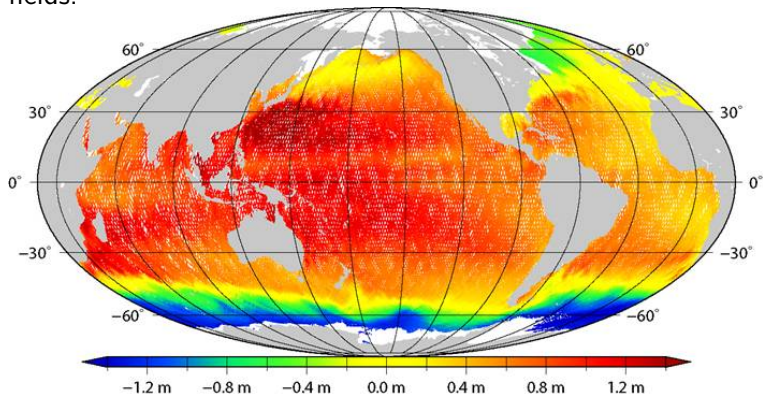


FIG. 3. (left) RMS error of the analysis from the truth for height (m) and (right) RMS ageostrophic wind (m s^{-1}) using no localization, B localization, and R localization for five ensemble members and a variety of localization distances L . For comparison, an analysis with no localization and 40 ensemble members is also plotted. Arrows depict optimum values of L .

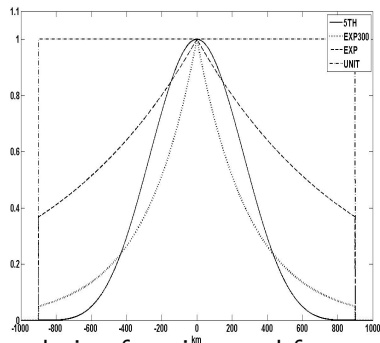
Realistic example—Assimilated Data Set

- 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)

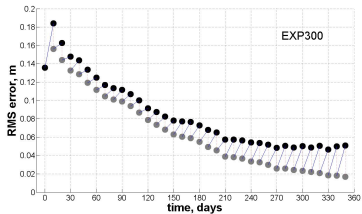
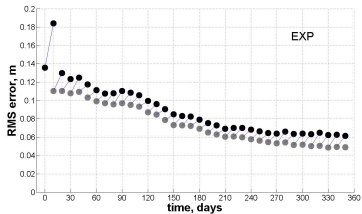
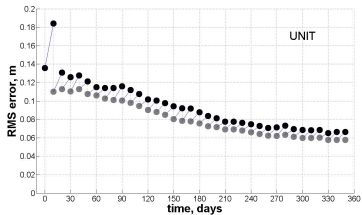
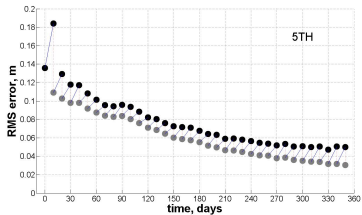
Data assimilation method



Correlation function used for localization.

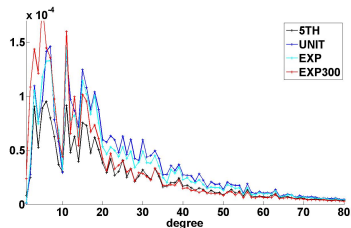
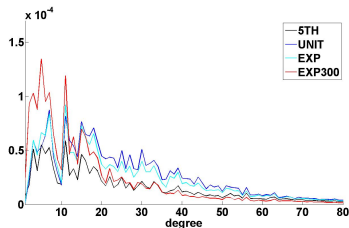
- Different correlation function are used for the method SD+ObsLoc.
- The observational error standard deviation is 5 cm.
- Observations within radius of 900 km are used.

RMS errors



RMS error for different covariance localisation 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.

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

See Janjic et al. 2011 MWR for more details

Conclusion

- Localization is necessary for application of ensemble Kalman filter algorithms for large scale problems.
- Several localization techniques are in use.
- Localization is topic of active research especially concerning the effect of localization on balance.
- Proper ways of performing multivariate localization are still not fully understood.
- Proper localization scales depend on the properties of dynamical system and observations.