



Multivariate localization methods for ensemble Kalman filtering

S. Roh¹, M. Jun¹, I. Szunyogh², and M. G. Genton³

¹Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA

²Department of Atmospheric Sciences, Texas A&M University, College Station, TX 77843-3148, USA

³CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

Correspondence to: M. Jun (mjun@stat.tamu.edu)

Received: 14 April 2015 – Published in Nonlin. Processes Geophys. Discuss.: 8 May 2015

Revised: 29 October 2015 – Accepted: 18 November 2015 – Published: 3 December 2015

Abstract. In ensemble Kalman filtering (EnKF), the small number of ensemble members that is feasible to use in a practical data assimilation application leads to sampling variability of the estimates of the background error covariances. The standard approach to reducing the effects of this sampling variability, which has also been found to be highly efficient in improving the performance of EnKF, is the localization of the estimates of the covariances. One family of localization techniques is based on taking the Schur (element-wise) product of the ensemble-based sample covariance matrix and a correlation matrix whose entries are obtained by the discretization of a distance-dependent correlation function. While the proper definition of the localization function for a single state variable has been extensively investigated, a rigorous definition of the localization function for multiple state variables that exist at the same locations has been seldom considered. This paper introduces two strategies for the construction of localization functions for multiple state variables. The proposed localization functions are tested by assimilating simulated observations experiments into the bivariate Lorenz 95 model with their help.

1 Introduction

The components of the finite-dimensional state vector of a numerical model of the atmosphere are defined by the spatial discretization of the state variables considered in the model. An ensemble-based Kalman filter (EnKF) data assimilation scheme treats the finite-dimensional state vector as a multivariate random variable and estimates its probability distribution by an ensemble of samples from the distribution. To be precise, an EnKF scheme assumes that the probability

distribution of the state is described by a multivariate normal distribution, and it estimates the mean and the covariance matrix of that distribution by the ensemble (sample) mean and the ensemble (sample) covariance matrix. The estimate of the mean and the estimate of the covariance matrix of the analysis distribution are obtained by updating the mean and the covariance matrix of a background (prior) distribution based on the latest observations. The background distribution is represented by an ensemble of short-term forecasts from the previous analysis time. This ensemble is called the background ensemble.

Because the number of background ensemble members that is feasible to use in a realistic atmospheric model is small, the estimates of weak covariances (the entries with small absolute values in the background covariance matrix) tend to have large relative estimation errors. These large relative errors have a strong negative effect on the accuracy of an EnKF estimate of the analysis mean. The standard approach to alleviating this problem is to apply a physical-distance-dependent localization to the sample background covariances before their use in the state update step of the EnKF. In essence, localization is a method to introduce the empirical understanding that the true background covariances tend to rapidly decrease with distance into the state estimation process.

Data assimilation schemes treat the spatially discretized state vector, \mathbf{x} , as a multivariate random variable. We use the conventional notation \mathbf{x}^b and \mathbf{x}^a for the background and the analysis state vectors, respectively. We also use the notation \mathbf{y}^o for the vector of observations. In an EnKF scheme, the analysis mean, $\bar{\mathbf{x}}^a$, is computed from the background mean, $\bar{\mathbf{x}}^b$, by the update equation

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{K} \left(\mathbf{y}^\circ - \overline{h(\mathbf{x}^b)} \right). \quad (1)$$

The function $h(\cdot)$ is the observation function, which maps the finite-dimensional state vector into observables. Thus, $\overline{h(\mathbf{x}^b)}$ is the ensemble mean of the prediction of the observations by the background. The matrix

$$\mathbf{K} = \mathbf{P}^b \mathbf{H}^T \left(\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R} \right)^{-1} \quad (2)$$

is the Kalman gain matrix, where \mathbf{P}^b is the background covariance matrix, \mathbf{H} is the linearization of h about $\bar{\mathbf{x}}^b$, and \mathbf{R} is the observation error covariance matrix. EnKF schemes usually avoid the explicit computation of the linearized observation operator \mathbf{H} by using approximations to $\mathbf{P}^b \mathbf{H}^T$ and $\mathbf{H} \mathbf{P}^b \mathbf{H}^T$ that involve only the computation of $h(\mathbf{x}^b)$ and $\overline{h(\mathbf{x}^b)}$ (e.g., Houtekamer and Mitchell, 1998). The entry K_{ij} of \mathbf{K} determines the effect of the j th observation on the i th component of the analysis mean, $\bar{\mathbf{x}}^a$. Under the standard assumption that the observation errors are uncorrelated, the matrix, \mathbf{R} , is diagonal. Hence, the way the effect of the observations is spread from the observations to the different locations and state variables is determined by \mathbf{P}^b and \mathbf{H} . The sampling variability in the estimates of \mathbf{P}^b affects the accuracy of the information propagated in space and between the different state variables through the matrix products, $\mathbf{P}^b \mathbf{H}^T$ and $\mathbf{H} \mathbf{P}^b \mathbf{H}^T$. The goal of localization is to reduce the related effects of sampling variability on the estimates of \mathbf{K} .

Over the years, many different localization methods have been proposed.

different variables such as compactly supported correlation functions for multivariate Gaussian random fields.

The rest of the paper is organized as follows. Section 2 briefly describes EnKF and localization for the special case of two state variables. Section 3 describes the bivariate Lorenz 95 model we use to test our ideas. Section 4 summarizes the main results of the paper.

2 Methodology

2.1 Univariate localization

In principle, localization can be implemented by using filtered estimates of the background covariances rather than the raw sample covariances to define the matrix, \mathbf{P}^b , used in the computation of \mathbf{K} by Eq. (2). The filtered (localized) version of covariance matrix, $\tilde{\mathbf{P}}^b$, is obtained by computing the Schur (element-wise) product:

$$\tilde{\mathbf{P}}^b = \hat{\mathbf{P}}^b \circ \mathbf{C}, \quad (3)$$

where \mathbf{C} is a correlation matrix, which has the same dimensions as the sample covariance matrix, $\hat{\mathbf{P}}^b$. In practice, however, the localization is often done by taking advantage of the fact that localization affects the analysis through $\mathbf{P}^b \mathbf{H}^T$ and $\mathbf{H} \mathbf{P}^b \mathbf{H}^T$, or, ultimately, through \mathbf{K} . In particular – because a distance, d , can be defined for each entry, K_{ij} , of \mathbf{K} by the distance between the i th analyzed variable and the j th observation – the simplest localization strategy is to set all entries, K_{ij} , that are associated with a distance longer than a prescribed localization radius, R ($d > R$), to zero, while leaving the remaining entries unchanged (e.g., Houtekamer and Mitchell, 1998; Ott et al., 2004; Hunt et al., 2007).

Another approach is to localize $\mathbf{P}^b \mathbf{H}^T$ and $\mathbf{H} \mathbf{P}^b \mathbf{H}^T$ by a tapering function (e.g., Hamill et al., 2001; Houtekamer and Mitchell, 2001). The usual justification for this approach is that the localized matrix products provide good approximations of the products computed by using localized estimates of \mathbf{P}^b . Note that $\mathbf{P}^b \mathbf{H}^T$ is the matrix of background covariances between the state variables at the model grid points and at the observation locations, while $\mathbf{H} \mathbf{P}^b \mathbf{H}^T$ is the matrix of background covariances between the state variables at the observation locations. Thus, a distance can be associated with each entry of the two matrix products, which makes the distance-dependent localization of the two products possible. The approach becomes problematic, however, when $h(\cdot)$ is not a local function, which is the typical case for remotely sensed observations (Campbell et al., 2010).

We consider the situation where localization is applied directly to the background error covariance matrix, $\hat{\mathbf{P}}^b$. Recall that the localized covariance matrix is expressed as in Eq. (3). In particular, \mathbf{C} is a positive-definite matrix with strictly positive eigenvalues, while the sample covariance matrix, $\hat{\mathbf{P}}^b$, may have zero eigenvalues (as it is only non-negative definite). The localization in Eq. (3) helps to eliminate those zero

eigenvalues of $\hat{\mathbf{P}}^b$ and alleviates the related large relative estimation errors. The positive definiteness of \mathbf{C} ensures that localization does not introduce new zero eigenvalues in the process of eliminating the zero eigenvalues of $\hat{\mathbf{P}}^b$. The proper definition of the localization function that ensures that \mathbf{C} is positive definite has been thoroughly investigated for the univariate case ($N = 1$) in the literature (e.g. Gaspari and Cohn, 1999).

2.2 Multivariate localization

We now consider a model with multiple state variables ($N > 1$). For instance, we take a simple model based on the hydrostatic primitive equations. This model solves the equations for the two horizontal components of wind, the surface pressure, the virtual temperature and a couple of atmospheric constituents. The state of the model is represented by the state vector, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$, where $\mathbf{x}_i, i = 1, 2, \dots, N$, represents the spatially discretized state of the i th state variable in the model.

The sample background covariance matrix, $\hat{\mathbf{P}}^b$, can be partitioned as

$$\hat{\mathbf{P}}^b = \begin{pmatrix} \hat{\mathbf{P}}_{11}^b & \hat{\mathbf{P}}_{12}^b & \dots & \hat{\mathbf{P}}_{1N}^b \\ \hat{\mathbf{P}}_{21}^b & \hat{\mathbf{P}}_{22}^b & \dots & \hat{\mathbf{P}}_{2N}^b \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{P}}_{N1}^b & \hat{\mathbf{P}}_{N2}^b & \dots & \hat{\mathbf{P}}_{NN}^b \end{pmatrix}. \quad (4)$$

The entries of the submatrices, $\hat{\mathbf{P}}_{ii}^b, i = 1, \dots, N$, are called the marginal covariances for the i th state variable. In practical terms if the i th state variable is the virtual temperature, for instance, each diagonal entry of $\hat{\mathbf{P}}_{ii}^b$ represents the sample variance for the virtual temperature at a given model grid point, while each off-diagonal entry of $\hat{\mathbf{P}}_{ii}^b$ represents the sample covariances between the virtual temperatures at a pair of grid points. Likewise, the entries of $\hat{\mathbf{P}}_{ij}^b, i \neq j$, are called the sample cross-covariances between the grid point values of the i th and the j th state variables at pairs of locations, where the two locations for an entry can be the same grid point.

We thus consider matrix-valued localization functions, $\boldsymbol{\rho}(d) = \{\rho_{ij}(d)\}_{i,j=1,\dots,N}$, which are continuous functions of d . The component $\rho_{ij}(d)$ of $\boldsymbol{\rho}(d)$ is the localization function used for the calculation of the covariances included in the submatrix \mathbf{P}_{ij}^b of \mathbf{P}^b . Each entry of the localization matrix \mathbf{C} is computed by considering the value of the appropriate component of $\boldsymbol{\rho}(d)$ for a particular pair of state variables and the separation distance, d , associated with the related entry of $\hat{\mathbf{P}}^b$.

In order to get a proper matrix-valued localization function, $\boldsymbol{\rho}$, a seemingly obvious approach to extend the results of Gaspari and Cohn (1999) would be to compute the entries of \mathbf{C} based on a univariate correlation function for a multivariate variable. That is, for the pair of state variables i and j , we localize the corresponding sample background covari-

ance matrix, $\hat{\mathbf{P}}_{ij}^b$, by multiplying a localization matrix from the same correlation function for all i and j . Formally, this would be possible because the distance d is uniquely defined for each entry of $\hat{\mathbf{P}}^b$ the same way in the multivariate case as in the univariate case. This approach, however, cannot guarantee the positive definiteness of the resulting matrix, \mathbf{C} . As a simple illustrative example, consider the situation where the discretized state vector has only two components that are defined by two different scalar state variables at the same location (e.g., the temperature and the pressure). In this case, if n is the number of locations, the localization matrix for the two state variables together can be written as

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_0 & \mathbf{C}_0 \\ \mathbf{C}_0 & \mathbf{C}_0 \end{pmatrix} \quad (5)$$

independently of the particular choice of the localization function. Here \mathbf{C}_0 is an $n \times n$ localization matrix from a univariate localization function. From Eq. (5), it is clear that n eigenvalues of \mathbf{C} are zero and the rank of \mathbf{C} is n , while its dimensions are $2n \times 2n$.

As in Eq. (2), although \mathbf{C} is rank-deficient and thus so is the localized covariance matrix $\hat{\mathbf{P}}^b$ (and thus $\hat{\mathbf{P}}^b$ may be rank-deficient as well), we may still be able to calculate the inverse of $\mathbf{H}\hat{\mathbf{P}}^b\mathbf{H}^T + \mathbf{R}$, as \mathbf{R} is a diagonal matrix. The smallest eigenvalue of $\mathbf{H}\hat{\mathbf{P}}^b\mathbf{H}^T + \mathbf{R}$ is the smallest (positive) value of \mathbf{R} , and thus the matrix, $\mathbf{H}\hat{\mathbf{P}}^b\mathbf{H}^T + \mathbf{R}$, is still invertible and has positive eigenvalues. However, unless the diagonal elements of \mathbf{R} are large (which implies large observation error variance), the matrix $\mathbf{H}\hat{\mathbf{P}}^b\mathbf{H}^T + \mathbf{R}$ is seriously ill-conditioned and the computation of its inverse may be numerically unstable. Therefore, the numerical stability of the computation of the inverse of the matrix heavily relies on the observation error variance, which is an undesirable property.

We therefore propose two approaches to construct positive-definite (full rank) matrix-valued localization functions, $\rho(d)$. The first proposed method takes advantage of the knowledge of a proper univariate localization function, $\tilde{\rho}$. Instead of using the same correlation function to localize multiple state variables, for a certain distance lag, we let $\rho = \tilde{\rho} \cdot \mathbf{B}$, where \mathbf{B} is an $N \times N$ symmetric, positive-definite matrix whose diagonal entries are 1. It can be easily verified that ρ is a matrix-valued positive-definite function, which makes it a valid multivariate localization function. For instance, in the hypothetical case where the two components of the state vector are two different state variables at the same location, making the choice

$$\mathbf{B} = \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix} \quad (6)$$

for β , with $|\beta| < 1$, leads to

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_0 & \beta\mathbf{C}_0 \\ \beta\mathbf{C}_0 & \mathbf{C}_0 \end{pmatrix} \quad (7)$$

rather than what is given in Eq. (5). Since the eigenvalues of the matrix \mathbf{B} are $1 \pm \beta > 0$, it can be easily verified that

the matrix in Eq. (7) is positive definite. For the case with more than two state variables ($N \geq 3$), the matrix \mathbf{B} can be parametrized as $\mathbf{B} = \mathbf{L}\mathbf{L}^T$, where

$$\mathbf{L} = \begin{pmatrix} \ell_{1,1} & 0 & \cdots & 0 \\ \ell_{2,1} & \ell_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{N,1} & \ell_{N,2} & \cdots & \ell_{N,N} \end{pmatrix} \quad (8)$$

is a lower triangular matrix with the constraints that $\sum_{j=1}^i \ell_{i,j}^2 = 1$ and $\ell_{i,i} > 0$ for all $i = 1, \dots, N$. The constraints are used to have the diagonal entries of \mathbf{B} be 1. Other than these constraints, the elements of \mathbf{L} can vary freely in order to guarantee the positive definiteness of \mathbf{B} .

An attractive feature of this approach is that we can take advantage of any known univariate localization function to produce a multivariate localization function. However, the multivariate localization function from this approach is *separable* in the sense that the multivariate component (i.e., \mathbf{B}) and the localization function (i.e., $\tilde{\rho}$) are factored. Another limitation of the approach is that the localization radius and decay rate are the same for each pair of state variables, leaving no flexibility to account for the potential differences in the correlation lengths and decay rate for the different state vector components.

The second proposed method takes advantage of the availability of multivariate compactly supported functions from the spatial statistics literature. To the best of our knowledge, only a few papers have been published on this subject; one of them is Porcu et al. (2013). The function class they considered was essentially a multivariate extension of the Askey function (Askey, 1973), $f(d; \nu, c) = (1 - \frac{d}{c})_+^\nu$, with $c, \nu > 0$. Here, $x_+ = \max(x, 0)$ for $x \in \mathbb{R}$. For instance, a bivariate Askey function, which is a special case of the results of Porcu et al. (2013), is given by ($i, j = 1, 2$)

$$\rho_{ij}(d; \nu, c) = \beta_{ij} \left(1 - \frac{d}{c}\right)_+^{\nu + \mu_{ij}}, \quad (9)$$

where $c > 0$, $\mu_{12} = \mu_{21} \leq \frac{1}{2}(\mu_{11} + \mu_{22})$, $\nu \geq [\frac{1}{2}s] + 2$, $\beta_{ii} = 1$ ($i = 1, 2$), $\beta_{12} = \beta_{21}$, and

$$|\beta_{12}| \leq \frac{\Gamma(1 + \mu_{12})}{\Gamma(1 + \nu + \mu_{12})} \sqrt{\frac{\Gamma(1 + \nu + \mu_{11})\Gamma(1 + \nu + \mu_{22})}{\Gamma(1 + \mu_{11})\Gamma(1 + \mu_{22})}}. \quad (10)$$

Here, $\Gamma(\cdot)$ is the gamma function (e.g., Wilks, 2006) and s is the dimension of the Euclidean space where the state variable is defined. If the state is defined at a particular instant on a grid formed by latitude, longitude, and height, then $s = 3$. Here, $[x]$ is the largest integer that is equal to or smaller than x . The Askey function in Eq. (9) has the support c because it sets covariances beyond a distance c to zero. It can be seen

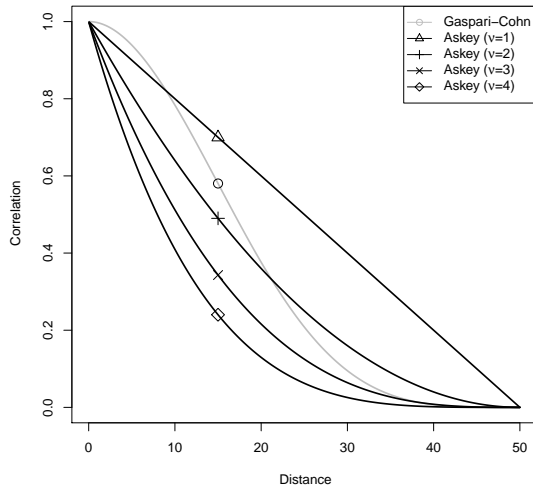


Figure 1. The Gaspari–Cohn covariance function with a localization constant $c = 25$ (support of 50) and the Askey covariance function $f(d; \nu, c) = (1 - \frac{d}{c})_+^\nu$, with a support parameter $c = 50$ and various shape parameters.

from Eq. (10) that, if the scalars, μ_{ij} , are chosen to be the same for all values of i and j , the condition on β_{12} for ρ to be valid is $|\beta_{12}| \leq 1$. (Note that the case of equality here, with the same μ_{ij} 's, reduces to the rank-deficient case where the multivariate localization matrix has zero eigenvalues, similarly to the case of $\beta = 1$ in Eq. 7.) For this choice, the second method is essentially the same as the first method with the Askey function set to $\tilde{\rho}$. The localization function given by Eq. (9) is more flexible than the functions of the first method with the Askey function set to $\tilde{\rho}$ because μ_{ij} can be chosen to be different for each pair of indexes, i and j . The localization length, however, is still the same for the different pairs of the state variables. The multivariate Askey function is formed by

$$\rho_{ij}(d; \nu, c) = c^{\nu+1} B(\mu_{ij} + 1, \nu + 1) \left(1 - \frac{|d|}{c}\right)^{\nu + \mu_{ij} + 1}, \quad |d| < c \quad (11)$$

and 0 otherwise, where $\nu \geq (s+1)/2$, $\mu_{ij} = (\mu_i + \mu_j)/2$, and $\mu_i > 0$ for all $i = 1, \dots, N$. Here, B is the beta function (Porcu et al., 2013; Genton and Kleiber, 2015).

To illustrate the differences between the shape of the Gaspari–Cohn and the Askey functions, we show the Gaspari–Cohn function for $c = 25$ and the univariate Askey function for $c = 50$ and $\nu = 1, \dots, 4$ (Fig. 1). This figure shows that, for a given support, the Askey functions are narrower.

3 Experiments

3.1 The EnKF scheme

There are many different formulations of the EnKF update equations, which produce not only an updated estimate of the mean but also the ensemble of analysis perturbations that are added to the mean to obtain an ensemble of analyses. This ensemble of analyses serves as the ensemble of initial conditions for the model integration that produce the background ensemble. In our experiments, we use the method of perturbed observations. It obtains the analysis mean and the ensemble of analysis perturbations by the equations

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^b), \quad (12)$$

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \mathbf{K}(\mathbf{y}_k^o - \mathbf{H}\mathbf{x}_k^b), \quad (13)$$

where \mathbf{x}_k^a , $k = 1, 2, \dots, M$, are the ensemble perturbations and \mathbf{y}_k^o , $k = 1, 2, \dots, M$, are random draws from the probability distribution of observation errors. As the notation suggests, we consider a linear observation function in our experiments. This choice is made for the sake of simplicity and limits the generality of our findings much less than the use of an idealized model of atmospheric dynamics.

For the case of multiple state variables, the ensemble members are considered to be in a single ensemble, that is, not being grouped into distinct subensembles.

3.2 The bivariate Lorenz model

Lorenz (1995) discussed the bivariate Lorenz 95 model, which mimics the nonlinear dynamics of two linearly coupled atmospheric state variables, X and Y , on a latitude circle. This model provides a simple and conceptually satisfying representation of basic atmospheric processes but is not suitable for some atmospheric processes. The model 3 in Lorenz (2005) made it more realistic and suitable with sacrifice of simplicity, by producing a rapidly varying small-scale activity superposed on the smooth large-scale waves. We use the Lorenz 95 model for simplicity in our following experiments.

In the bivariate Lorenz 95 model, the variable X is a “slow” variable represented by K discrete values, X_k , and Y is a “fast” variable represented by $J \times K$ discrete values. The governing equations are

$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k - (ha/b) \sum_{j=1}^J Y_{j,k} + F, \quad (14)$$

$$\frac{dY_{j,k}}{dt} = -abY_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - aY_{j,k} + (ha/b)X_k, \quad (15)$$

where $Y_{j-J,k} = Y_{j,k-1}$ and $Y_{j+J,k} = Y_{j,k+1}$ for $k = 1, \dots, K$ and $j = 1, \dots, J$. The “boundary condition” is periodic;

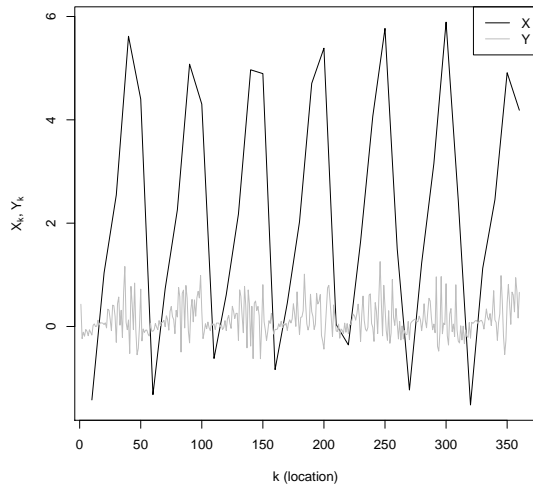


Figure 2. A snapshot of the variables X and Y from a numerical integration of the system of Eqs. (14) and (15) with $K = 36$, $J = 10$, $F = 10$, $a = 10$, $b = 10$, and $h = 2$.

that is, $X_{k-K} = X_{k+K} = X_k$, and $Y_{j,k-K} = Y_{j,k+K} = Y_{j,k}$. In our experiments, $K = 36$ and $J = 10$. The parameter h controls the strength of the coupling between X and Y , a is the ratio of the characteristic timescales of the slow motion of X to the fast motion of Y , b is the ratio of the characteristic amplitudes of X to Y , and F is a forcing term. We choose the parameters to be $a = 10$, $b = 10$, $F = 10$, and $h = 2$. These values of the model parameters are equal to those originally suggested by Lorenz (1995), except for the value of the coupling coefficient h , which is twice as large in our case. We made this change in h to increase the covariances between the errors in the estimates of X and Y , which makes the model more sensitive to the choices of the localization parameters. We use a fourth-order Runge–Kutta time integration scheme with a time step of 0.005 nondimensional units, as Lorenz (1995) did. We define the physical distances between X_{k_1} and X_{k_2} , between Y_{j_1,k_1} and Y_{j_2,k_2} , and between X_{k_1} and Y_{j_1,k_2} by $|10(k_1 - k_2)|$, $|10(k_1 - k_2) + j_1 - j_2|$, and $|10(k_1 - k_2) - j_1|$, respectively. Figure 2 shows a typical state of the model for the selected parameters. The figure shows that X tends to drive the evolution of Y : the hypothetical process represented by Y is more active (its variability is higher) with higher values of X .

3.3 Experimental design

Since the estimates of the cross-covariances play a particularly important role at locations where one of the variables is unobserved, we expect an improved treatment of the cross-covariances to lead to analysis improvements at locations where only one of the state variables is observed. This motivates us to consider an observation scenario in which X and Y are partially observed. The variable X is observed at 20 % of all locations, and Y is observed at 90 % of the locations

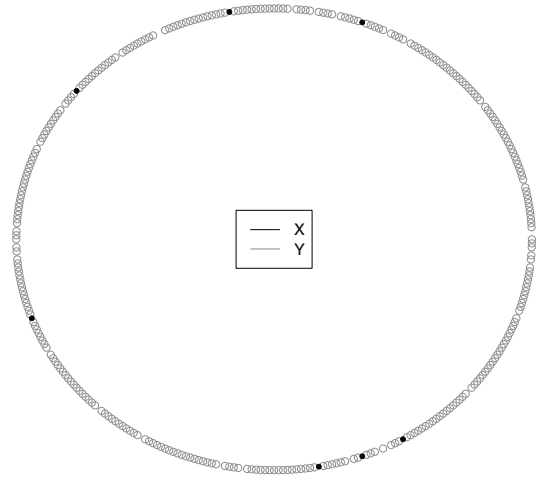


Figure 3. For the partially observed case, locations of observations of X and Y are indicated by the black dots and grey circles, respectively.

where X is not observed. These observation locations for variables X and Y are randomly chosen. Spatial locations of the partially observed X and Y are illustrated in Fig. 3. The results from this experiment are compared to those from a control experiment, in which both X and Y are fully observed.

We first generate a time series of “true” model states by a 2000-time-step integration of the model. We initialize an ensemble by adding the standard Gaussian noise to the true state; then we discard the first 3000 time steps. We then generate simulated observations by adding random observation noise of mean 0 and variance 0.02 to the appropriate components of the true state of X at each time step. We use the same procedure to generate simulated observations of Y , except that the variance of the observation noise is 0.005. Observations are assimilated at every time step by first using a 20-member ensemble with a constant covariance inflation factor of 1.015. The error in the analysis at a given verification time is measured by the root-mean-square distance between the analysis mean and the true state. We refer to the resulting measure as the root-mean-square error (RMSE). The probability distribution of the RMSE for the last 1000 time steps of 50 different realizations of each experiment is shown by a box plot. The box plot is an effective way of displaying a summary of the distribution of numbers. The lower and upper bounds of the box respectively give the 25th and 75th percentiles. The thick line going across the interior of the box gives the median. The whisker depends on the interquartile range (IQR), which is precisely equal to the vertical length of the box. The whiskers extend to the extreme values, which are no more than 1.5 IQR from the box. Any values that fall outside of the end points of whiskers are considered outliers, and they are displayed as circles.

In the box plot figures in the next section, we compare the RMSE for four different localization schemes. We use the following notation to distinguish between them in the figures:

1. S1 – the bivariate sample background covariance is used without localization;
2. S2 – same as S1 except that the cross-covariances between X and Y are replaced by zeros;
3. S3 – a univariate localization function is used to filter the marginal covariances within X and Y , respectively, while the cross-covariances between X and Y are replaced by zeros;
4. S4 – one of the bivariate localization methods described in Sect. 2.2 is used to filter both the marginal and the cross-covariances.

In the experiments identified by S4, we consider two different bivariate localization functions: the first one is $\rho^{(1)}(\cdot) = \{\beta_{ij} \rho^{(1)}(\cdot)\}_{i,j=1,2}$, with $\beta_{ii} = 1$ ($i = 1, 2$) and $\beta_{ij} = \beta$ ($i \neq j$), for some β such that $|\beta| < 1$. We use the fifth-order piecewise-rational function of Gaspari and Cohn (1999) to define the univariate correlation function, $\rho^{(1)}$, in the following form:

$$\rho^{(1)}(d; c) = \begin{cases} 1 \\ -1 \end{cases}$$

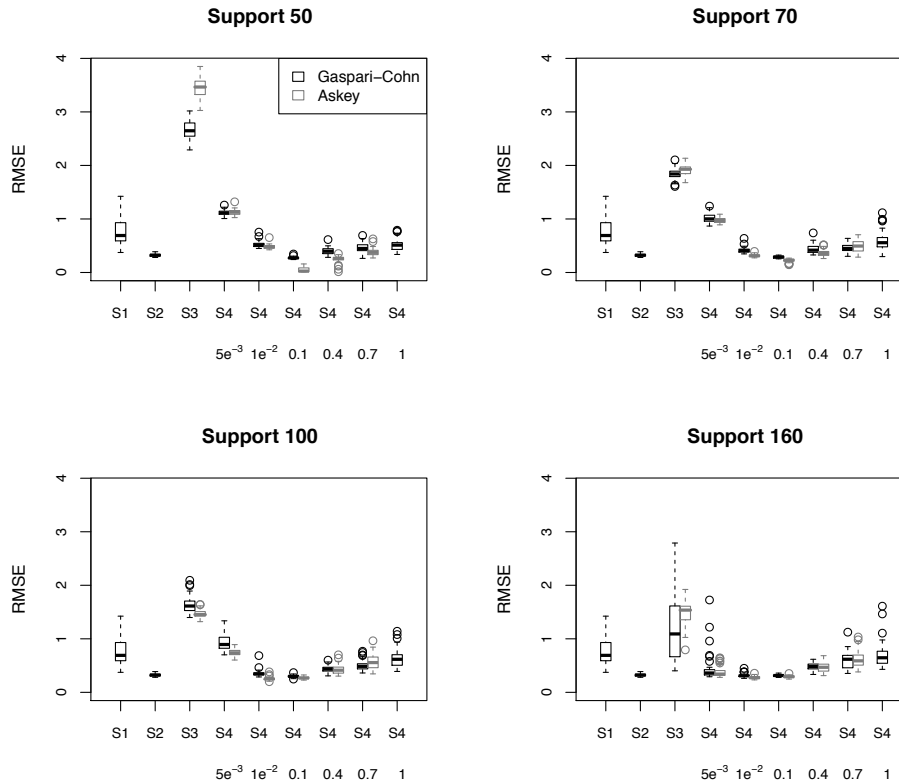


Figure 4. The box plot of RMSE for variable X in the case when the system is only partially observed. Results are shown for different localization strategies. For the definitions of localization strategies S1, S2, S3 and S4, see the text. The title of each panel indicates the localization radius (length of support). The lower and upper bounds of the box respectively give the 25th and 75th percentiles. The thick line going across the interior of the box gives the median. The whisker depends on the interquartile range (IQR), which is precisely equal to the vertical length of the box. The whiskers extend to the extreme values, which are no more than 1.5 IQR from the box. Any values that fall outside of the end points of whiskers are considered outliers, and they are displayed as circles. The numbers below S4 indicate the value of β . There is no box plot for $\beta = 1$ for the S4 with the Askey function, since the Askey function is not defined with $\beta = 1$ ($|\beta| \leq 0.79$; see Sect. 3.3).

Figure 7 shows the distribution of the errors for variable Y in the fully observed case. The best results are obtained by using a short localization radius with the Askey function, even though the variability of the error is relatively large in that case. The fact that localization strategies S3 and S4 perform similarly well shows that the estimates of the cross-covariances do not play an important role in this case; that is, X is primarily estimated based on observations of X , and Y is dominantly estimated based on observations of Y .

We also investigated the performance of EnKF with a 500-member ensemble. The results for the 500-member ensemble are shown in Figs. 8 to 11. We use an inflation factor of 1.005 for 500 ensembles, because the optimal value of the inflation factor is typically smaller for a larger ensemble. The rank of the 500-member ensemble covariance matrix is significantly larger than that of the 20-member ensemble covariance matrix, as expected.

Figures 8 to 11 show that, overall, S4 still performs better than the other localization schemes regardless of the choice of localization radius, as in the case of the 20-member en-

semble. In particular, when observations are partially observed, S4 with $\beta = 0.01$ provides the smallest RMSE. The cross-correlation between X and Y , calculated using 500-member ensembles without assimilating any observation, varies from -0.4 to 0.4 , which indicates that the cross-correlation between the two variables are not negligible. Therefore, improved treatment of cross-covariance tends to lead to an improved accuracy in the state estimation.

The results with the 500-member ensemble also show that the distribution of the state estimation error is in general less sensitive to the choice of the localization function or the localization radius, compared to the 20-member-ensemble case. Figure 8, however, shows that, for the estimation of sparsely observed X , the localization scheme S3 with smaller localization radius performs worse than that with larger localization radius. For variable Y in the partially observed case (Fig. 8) and both variables X and Y in the fully observed case (Figs. 10 and 11), the best results are obtained with S3 and S4 regardless of the localization radius. They also show

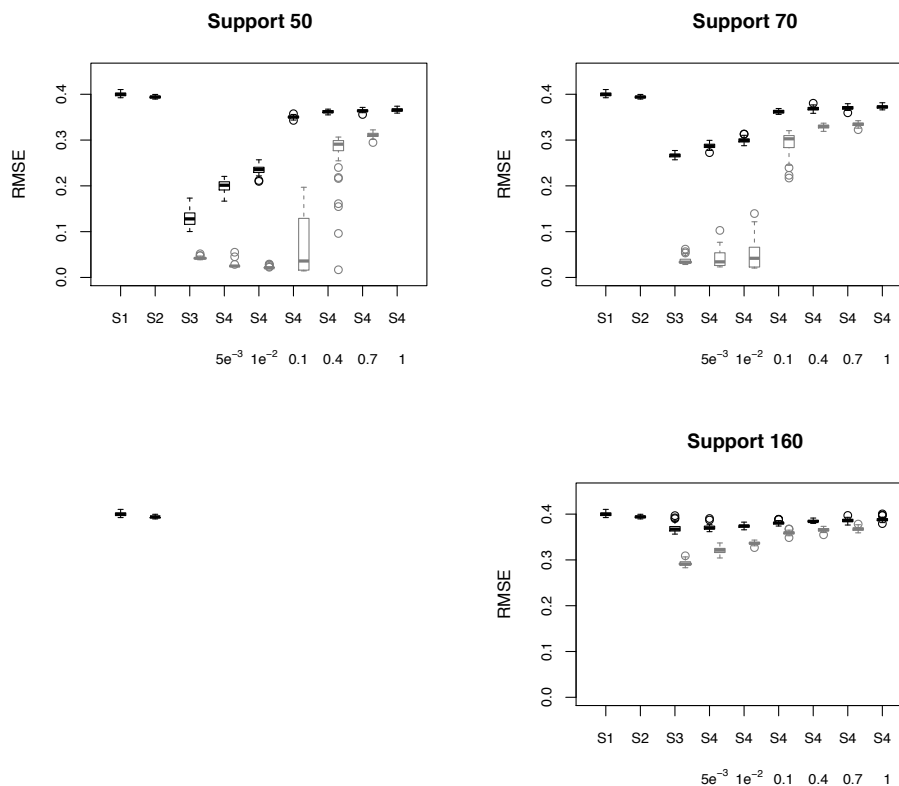


Figure 5. Same as Fig. 4 but for variable Y .

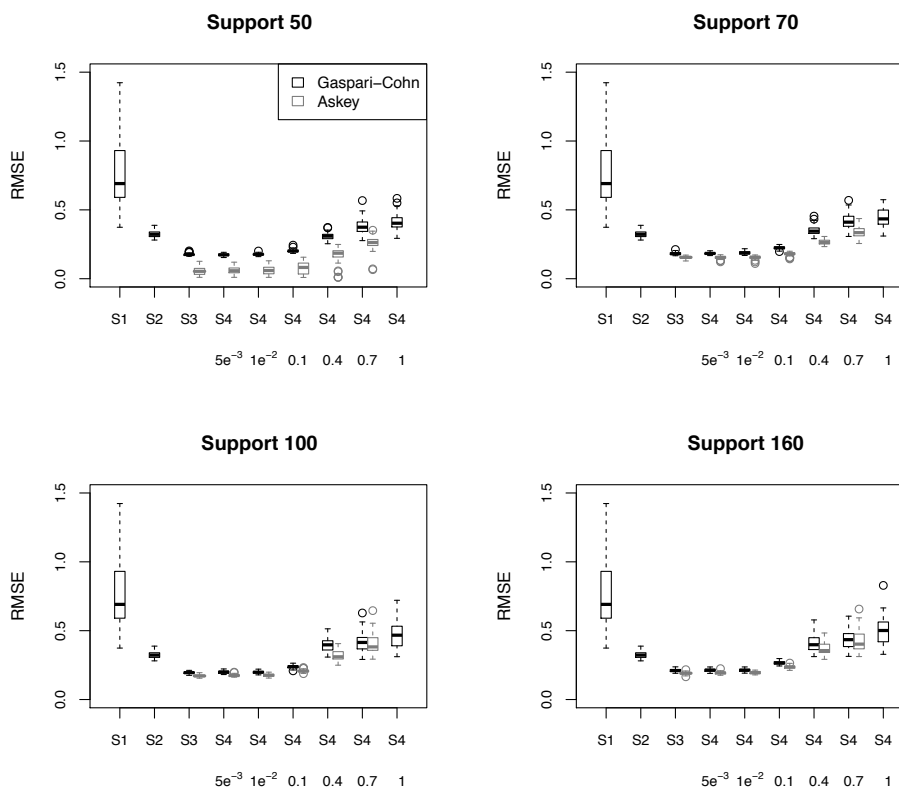


Figure 6. Same as Fig. 4 but for the case when the system is fully observed.

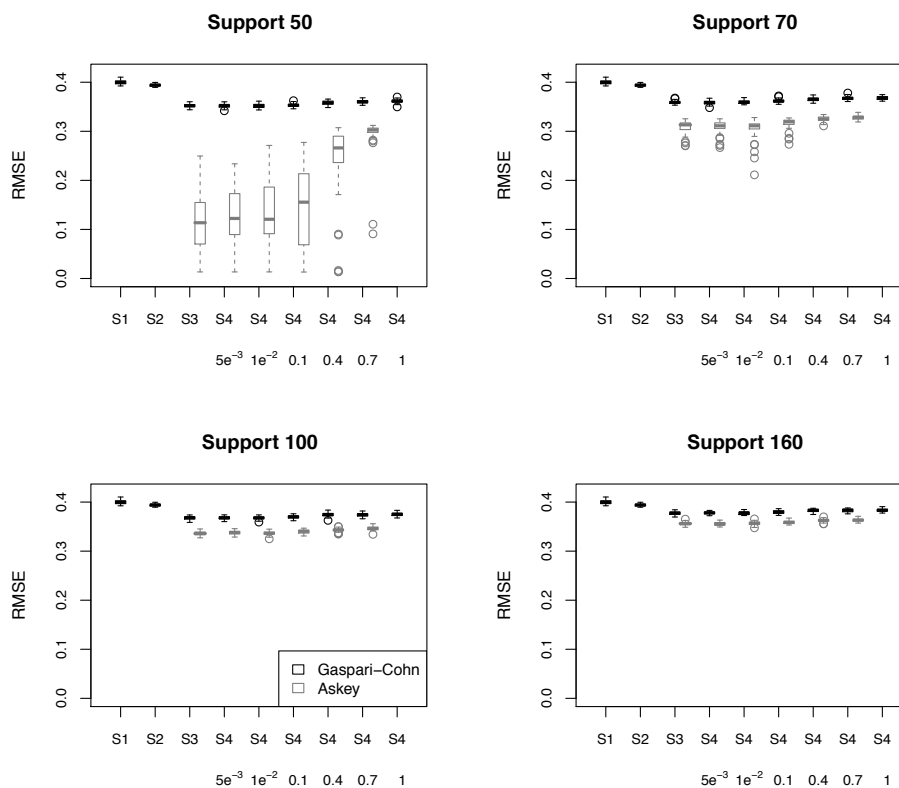


Figure 7. Same as Fig. 6 but for variable Y .

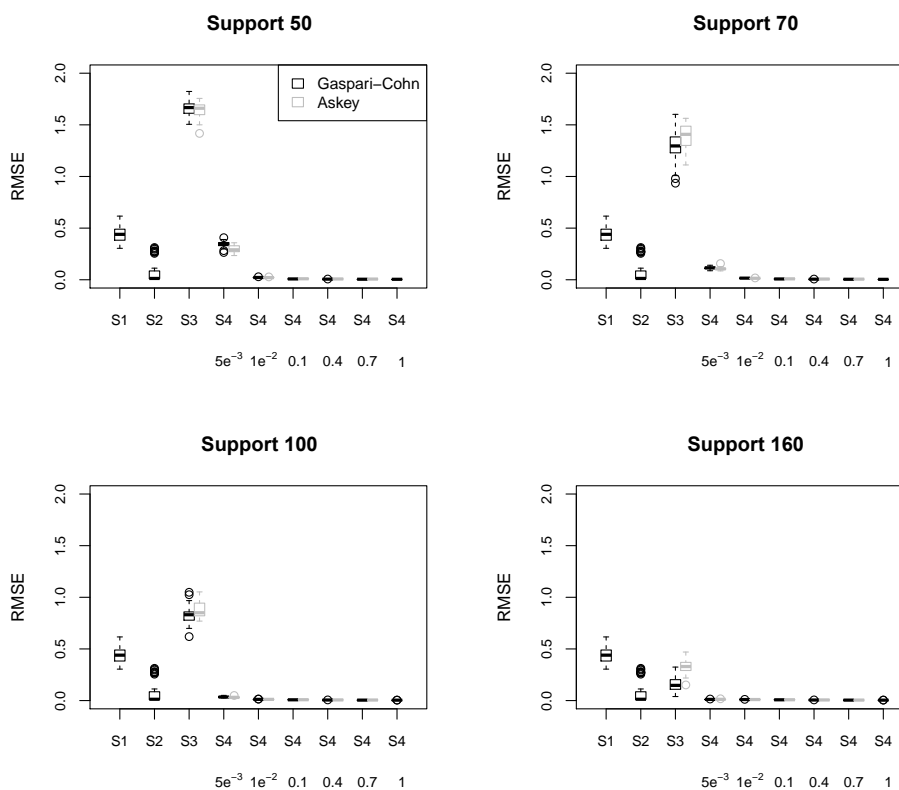
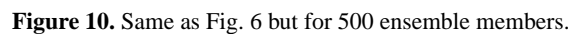
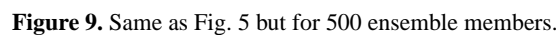


Figure 8. Same as Fig. 4 but for 500 ensemble members.



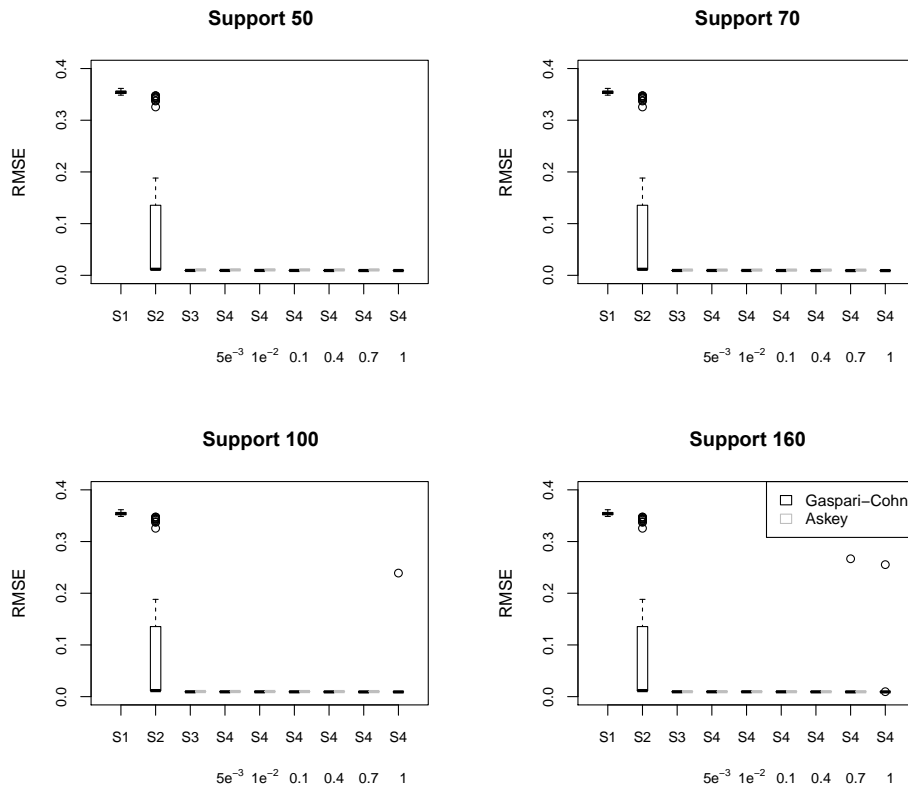


Figure 11. Same as Fig. 7 but for 500 ensemble members.

that the state estimation error is not sensitive but stable to the choice of localization radius.

Figures 10 and 11 show that the localization schemes S3 and S4 perform in a similar way, and obviously perform better than the other two localization schemes. This might imply that the cross-covariances do not have much influence on the state estimation in the fully observed case, once the covariances within each state variable are localized.

4 Discussion

The central argument of this paper is that applying a single localization function for the localization of covariances between multiple state variables in an EnKF scheme may not sufficiently increase the rank of the estimate of the background covariance matrix. In the light of this, we suggested two different approaches for the construction of positive-definite filtered estimates of the background covariance matrix. One of them takes advantage of the knowledge of a proper univariate localization function, whereas the other uses a multivariate extension of the Askey function. The results of our numerical experiments show that a mathematically proper localization function often leads to improved state estimates. The results of the numerical experiments also suggest that, of the two approaches we introduced, the one based on the Askey function produces more accurate state es-

timates than that based on the Gaspari–Cohn function. This fact, however, does not mean that the Askey function is always superior to the Gaspari–Cohn function in other chaotic models or observation networks. Which correlation function is superior depends on what the true error correlation looks like.

Acknowledgements. The authors are grateful to the reviewers for valuable comments that significantly improved presentation of the paper. M. Jun’s research was supported by NSF grant DMS-1208421, while I. Szunyogh’s research was supported by ONR Grant N000140910589. This publication is based in part on work supported by Award No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST).

Edited by: Z. Toth

Reviewed by: two anonymous referees

References

- Anderson, J. L.: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter, *Physica D*, 230, 99–111, 2007.
- Anderson, J. L. and Lei, L.: Empirical localization of observation impact in ensemble Kalman filters, *Mon. Weather Rev.*, 142, 739–754, 2013.

- Askey, R.: Radial characteristic functions, technical report no. 1262, Mathematical Research Center, University of Wisconsin-Madison, Madison, 1973.
- Bishop, C. H. and Hodyss, D.: Flow adaptive moderation of spurious ensemble correlations and its use in ensemble based data assimilation, *Q. J. Roy. Meteorol. Soc.*, 133, 2029–2044, 2007.
- Bishop, C. H. and Hodyss, D.: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models, *Tellus A*, 61, 84–96, 2009a.
- Bishop, C. H. and Hodyss, D.: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere, *Tellus A*, 61, 97–111, 2009b.
- Buehner, M. and Charron, M.: Spectral and spatial localization of background-error correlations for data assimilation, *Q. J. Roy. Meteorol. Soc.*, 133, 615–630, 2007.
- Campbell, W. F., Bishop, C. H., and Hodyss, D.: Vertical covariance localization for satellite radiances in ensemble Kalman filters, *Mon. Weather Rev.*, 138, 282–290, 2010.
- Du, J. and Ma, C.: Vector random fields with compactly supported covariance matrix functions, *J. Stat. Plan. Infer.*, 143, 457–467, 2013.
- Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, *Q. J. Roy. Meteorol. Soc.*, 125, 723–757, 1999.
- Genton, M. G. and Kleiber, W.: Cross-covariance functions for multivariate geostatistics (with discussion), *Stat. Sci.*, 30, 147–163, 2015.
- Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter, *Mon. Weather Rev.*, 129, 2776–2790, 2001.
- Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, 126, 796–811, 1998.
- Houtekamer, P. L. and Mitchell, H. L.: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation, *Mon. Weather Rev.*, 129, 123–137, 2001.
- Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, *Physica D*, 230, 112–126, 2007.
- Jun, M., Szunyogh, I., Genton, M. G., Zhang, F., and Bishop, C. H.: A Statistical Investigation of the Sensitivity of Ensemble-Based Kalman Filters to Covariance Filtering, *Mon. Weather Rev.*, 139, 3036–3051, 2011.
- Kang, J.-S., Kalnay, E., Liu, J., Fung, I., Miyoshi, T., and Ide, K.: “Variable localization” in an ensemble Kalman filter: Application to the carbon cycle data assimilation, *J. Geophys. Res.*, 116, D09110, doi:10.1029/2010JD014673, 2011.
- Kleiber, W. and Porcu, E.: Nonstationary Matrix Covariances: Compact Support, Long Range Dependence and Quasi-Arithmetic Constructions, *Stoch. Env. Res. Risk. A.*, 29, 193–204, 2015.
- Lei, L. and Anderson, J.: Comparison of empirical localization techniques for serial ensemble Kalman filters in a simple atmospheric general circulation model, *Mon. Weather Rev.*, 141, 4140–4153, 2014.
- Lorenz, E. N.: Predictability – A problem partly solved, vol. 1, ECMWF, Reading, Berkshire, UK, 1995.
- Lorenz, E. N.: Designing chaotic models, *J. Atmos. Sci.*, 62, 1574–1587, 2005.
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A.: A local ensemble Kalman filter for atmospheric data assimilation, *Tellus A*, 56, 415–428, 2004.
- Porcu, E., Daley, D. J., Buhmann, M., and Bevilacqua, M.: Radial basis functions with compact support for multivariate geostatistics, *Stoch. Env. Res. Risk. A.*, 27, 909–922, 2013.
- Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, *Mon. Weather Rev.*, 130, 1913–1924, 2002.
- Wilks, D. S.: Statistical methods in the atmospheric sciences, Academic Press, Amsterdam, 2006.
- Zhang, H. and Du, J.: Covariance Tapering in Spatial Statistics, edited by: Mateu, J. and Porcu, E., Graficas Casta \tilde{n} , s.l., Spain, 2008.