Response to the comments of Referee #1

We are grateful to the reviewer for his/her careful reading of the paper. The reviewer's comments helped us greatly to improve the presentation of our ideas and results. In what follows, we repeat each of the reviewer's comments in italic, and provide our response in roman.

General Comments: This manuscript deals with one of the very important topics in EnKF data assimilation. The described localization functions for multiple state variables can be used both in EnKF and Variational Ensemble data assimilation systems. The schemes are extensively described and easy to follow. In addition, OSSE results with a low-dimensional system are presented to document benefit of the new schemes. I suggest accepting this paper after addressing the following comments.

Specific comments: 1. Is the localization function defined by Eqs. 8-9 a Gaussian-like function? As shown in Fig. 1, the plotted Askey function is not a Gaussian-like function, which might make it be problematic to filter a Gaussian-like error correlation pattern.

The shape of the localization function does not affect the Gaussianity of the estimate of the distribution of the error, as it affects only second statistical moments. In other words, the estimate of the error distribution can still be Gaussian even if we use a localization function with a different shape, such as the Askey function.

2. Need more details on experiment design, especially on experiment S3 and S4 (P847-848). For experiment S3, values of localization parameters should be described. For experiment S3 and S4, please also explain the rationales used to specify localization parameters.

The values of the localization parameters μ and ν are given in the last paragraph of Section 3.c. We tried a few different values for each of these parameters and these values seem to work well. Results with various values for c and β are presented in Figs. 4-11.

3. Sensitivity to number of ensemble members can be further studied. The presented results (Page 850) showed similar results when using 20- and 40-member ensembles. Is this because the use of localization schemes? The other question is that what is the critical number of ensemble member to produce a good multi-variable background error covariance (BEC) modeling for such a low-dimensional Lorenz-95 system? The BEC with very large

number of ensemble member could be used to valid which localization scheme is a better one. I suggest showing figures of the covariance in physical space to demonstrate the sensitivity with several number of ensemble members (e.g., 100, 200, 500, ...).

We tried various ensemble sizes and we found that the state estimation error is not sensitive to the choice of ensemble size between 20 and 40. This may be due to the fact that we do localization, as the reviewer suggests.

Regarding the "critical number of ensemble members", it is not our goal in this paper to find the optimal number of ensemble members. Rather, we compare various localization schemes given an ensemble size.

We now add experiments with 500-member ensemble in Figures 8-11. Please see our texts in Section 3.d for our description. The figures show that the localization scheme S4, which localizes the variance between X and Y and within each state variable, still performs better overall than the other localization schemes.

Other minor comments: 1. Page 842, Eqs, 6-7, Please explain beta here.

β is a parameter, with magnitude less than 1, used to make the matrix $\mathbf{C}$ positive definite (c.f. $\mathbf{C}$ in (5)). We hope our explanation on that page is clear.

2. Page 843, Eqs, 8-9, for real atmospheric data assimilation, do you have any ideas on rules to specify the related parameters in Eqs. 8-9?

The parameters in Eqs. 8-9 should be determined by the characteristic of the state variables and their relationship. This obviously depends on the actual atmospheric data assimilation problem and the dynamic model.

3. Page 836, Line 13, what are values of variances for the two state variables?

We assume this comment is about page 846 (not page 836). The variances for X, variances for Y, and covariances between X and Y range from -4 to 4, -0.02 to 0.02, and -0.2 to 0.2, respectively.

4. Page 847, Line 14, beta(i,j)=beta ?, please explain beta.

β here plays the same role as in Eqs. 6-7. We modified the sentence in line 14 to make it clear.

5. Page 847, Line 21-22, Please speak more about the test results.

We presented the test results including figures in Section 3.d in the revision.

6. Page 850, Lines 14-16. Though I understand the statements are based on materials in Page 841, it seems to me the statement may confuse general readers. Actually, use of location functions is expected to alleviate the rank deficient issue in BEC modeling.

We agree the statement may be confusing, and modified it in the revision.

Response to the comments of Referee #2

We are grateful to the reviewer for his careful reading of the paper. The reviewer's comments helped us greatly to improve the presentation of our ideas and results. In what follows, we repeat each of the reviewer's comments in italic, and provide our response in roman.

In most implementations of the Ensemble Kalman Filter (EnKF), ensemble covariances between distinct variables defined at the same location in space-time are not attenuated. This paper successfully makes the point that, in some circumstances, EnKF performance would be improved by attenuating such inter-variable correlations.
The paper identifies simple ensemble covariance attenuation strategies for covariances between variables of different types such as zonal wind and temperature. The proposed strategies all guarantee the symmetry and semi-positive definiteness of the localized covariance matrix. The utility of the proposed approaches is evaluated using an idealized chaotic system with fast and slow variables. For this particular system and a fairly dense observing network, it is found that attenuating the inter-variable ensemble covariances improves EnKF performance.
The inter-variable covariance attenuation methods given by equations (6)-(9) and the associated discussion appear to pertain to the case when there are just two distinct variables at each grid-point. The paper would be of greater interest if corresponding formulae were given for the general case where there were M distinct variables at each model grid point.

We appreciate the reviewer's comments. We agree with the reviewer that extending to a more general case with M distinct variables is useful. We now give a general formula for arbitrary number of variables in Eqs. (8) and (11) of the revision.

The numerical experiments performed in this paper would be more informative if an effort had been made to establish the magnitude of the true error correlations between the slow (X) and fast (Y) variables in the system considered. It is only when that number is known that the potential gain of not zeroing out the inter-variable correlations can be quantified.

We calculated the cross-correlation between X and Y variables with 500 samples drawn from the model (without assimilation) and thus this cross-correlation can be considered as the true cross-correlation between the two variables. The magnitude varies between -0.4 and 0.4, thus the correlation between two variables seems quite strong. We added this statement in the

revision.

Another issue the authors might aim for in revising their paper is to come up with some rules of thumb to predict from the true error correlation and known random ensemble size when (a) there would be little to be gained in attenuating inter-variable ensemble correlations, or (b) there would be little to be gained in keeping any of the inter-variable ensemble correlation, or (c) partial attenuation of the inter-variable correlations would be helpful.

We appreciate the reviewer's comments. We agree that this is a great point, but given the limited time frame, we could only add an experiment with 500 ensemble members, under the same setting as the experiment with 20 ensemble members. Based on our additional experiment, we see that, in the bivariate Lorenz-95 model case (which shows non negligible cross-correlation between the two state variables), accounting for cross-correlation and at the same time partially attenuating cross-correlation give the best result. By trying various β values, we attempted to pick the best amount of partial-attenuation of cross-correlation.

3. As a check on the EnKF code and to quantify the upper limit of the usefulness of X-Y covariances, I think the paper would be stronger if you added another experiment in which the ensemble size was 16 to 32 times larger than the 40 member ensemble you considered here. With that experiment, you should be able to isolate the data assimilation value of accurately estimating X-Y covariances.

We appreciate this comment and we agree that suggested experiment can give us more insight. However, as indicated in the previous response, we could only add an experiment with 500 ensemble members.

4. The conclusions need to admit that the found superiority of the Askey localization function over the Gaspari-Cohn localization function may not extend to other chaotic models or observation networks. Askey might be better when the true correlation function looks more like the Askey function than the Gaspari-Cohn function and vice-versa?

We agree with your comment, and added in Discussion section that the Askey function may not be superior to the Gaspari-Cohn function in other models, and that which correlation function is better depends on what the true correlation function looks like.

Minor Specific Comments: 5. Abstract, line 8: Change ?entry-wise? to

?element-wise?.

Done.

6. Equation (2): Most EnKFs do not use the linearized observation operator in the definition of the gain. Thus, (2) or the discussion around it needs to be changed so that the reader is fully aware of the EnKF's ability to directly estimate covariances between forecasts of variables that are non-linear functions of the state and the state variables themselves.

We put the following sentence instead of avoiding the use of "linearization". 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).

7. P843, line 14. To more clearly define s give an example e.g ?if the state is defined at a particular instant on a latitude, longitude, height grid then s=3? ? if that?s what you mean.

We added that if the state is defined at a particular instant on a grid formed by latitude, longitude, and height, then s=3.

8. Equations (12) and (13): The relationship between this model and model 3 of Lorenz (2005, JAS) needs to be pointed out.

We compared the two models in the beginning of Section 3.b.

9. P847., lines 6-11. I think it would be clearer if you replaced "cross-covariances are set to zero" with "covariances between X and Y variables are set to zero" if if that's what you mean. Also, I'm not sure what you mean by "marginal covariances". Do you mean X-X and Y-Y covariances? Please clarify these issues in the revised text.

Cross-covariance and marginal covariance are defined in Section 2.b, but we replaced the terms as suggested by the reviewer.

10. Very little was said about Figures 8-11. Either discuss each one individually or replace some of them with a summary statement of what they indicate.

We replaced this with the results with 500-member ensemble with discussion about the results.

Multivariate Localization Methods for Ensemble Kalman Filtering

Soojin Roh

Department of Statistics, Texas A&M University, College Station, TX, USA

Mikyoung Jun*

Department of Statistics, Texas A&M University, College Station, TX, USA

Istvan Szunyogh

Department of Atmospheric Sciences, Texas A&M University, College Station, TX, USA

Marc G. Genton

CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

*Corresponding author address: 3143 TAMU, College Station, TX 77843-3143 USA.
Email:mjun@stat.tamu.edu

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 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 $\mathbf{y}^\circ$ 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

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

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

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

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 $\overline{h\left(\mathbf{x}^b\right)}$ (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. Hamill et al. (2001), Houtekamer and Mitchell (1998, 2001), Hunt et al. (2007), Ott et al. (2004), and Whitaker and

Hamill (2002) used localization functions which set the covariance to zero beyond a certain distance (localization radius). Jun et al. (2011) proposed a nonparametric statistical method to estimate the covariance. Anderson (2007) used a hierarchical ensemble filter which estimates the covariance using an ensemble of ensemble filters. Bishop and Hodyss (2007, 2009a,b) adaptively determined the width of localization by computing powers of the sample correlations. Buehner and Charron (2007) examined the spectral and spatial localization of error covariance. Anderson and Lei (2013) and Lei and Anderson (2014) proposed an empirical localization function based on the output of an observing system simulation experiment.

The focus of the present paper is on the family of schemes that localize the covariances by taking the Schur (Hadamard) product of the sample background covariance matrix and a correlation matrix of the same size, whose entries are obtained by the discretization of a distance-dependent correlation function with local (compact) support (e.g., Hamill et al. 2001; Houtekamer and Mitchell 2001; Whitaker and Hamill 2002). Such a correlation function is usually called a *localization* or *taper function*. The commonly used localization functions were introduced by Gaspari and Cohn (1999). Beyond a certain distance, all localization functions become zero, forcing the filtered estimates of the background covariance between state variables at locations that are far apart in space to zero. This property of the filtered background covariances can also be exploited to increase the computational efficiency of the EnKF schemes.

~~Eqs. (1) and (2) provide the solution of a formulation of the data assimilation problem that assumes that $\mathbf{P^b}$ is invertible (e.g., sections 4.2.1 and 4.2.3 of Szunyogh 2014).~~ A realistic atmospheric model has multiple scalar state variables (e.g., temperature, coordinates of the wind vector, surface pressure, humidity). If a univariate localization function, such as that described by Gaspari and Cohn (1999), is applied directly to a multivariate state vector, the resulting localized background covariance matrix may not be positive-definite. Because $\mathbf{P}^b$ is symmetric,

its eigenvalues are real and non-negative, which implies that it is invertible, only if it is also positive-definite. (An $n \times n$ symmetric matrix $\mathbf{A}$ is positive-definite if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all non-zero vectors $\mathbf{x} \in \mathbb{R}^n$.) Because the computation of the right-hand-side of Eq. (2) does not require the invertibility of $\mathbf{P}^b$, singularity of the localized $\mathbf{P}^b$ usually does not lead to a breakdown of the computations in practice. An ill-conditioned estimate of $\mathbf{P}^b$, however, can degrade the conditioning (increase the condition number) of $\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}$, making the numerical computation of the right-hand side of Eq. (2) less stable. This motivates us to seek rigorously-derived multivariate localization functions for ensemble Kalman filtering. As will be demonstrated, such rigorously-derived multivariate localization functions often produce more accurate analyses than those that apply the same univariate localization functions to each scalar component of the state vector. Kang et al. (2011) also introduced a multivariate localization method that zeros out covariances between physically unrelated variables. Their motivation for zeroing out such covariances, however, was to filter apparent spurious covariances rather than to preserve the positive-definiteness of the background error covariance matrix.

In our search for proper multivariate localization functions, we take advantage of recent developments in the statistics literature. In particular, we use the localization functions developed in Porcu et al. (2012), who studied the radial basis functions to construct multivariate correlation functions with compact support. Note that Section 5 in Zhang and Du (2008) described a general methodology for covariance tapering in the case of multiple state variables. Du and Ma (2013) used a convolution approach and a mixture approach to derive covariance matrix functions with compactly supported covariances. Kleiber and Porcu (2015) constructed nonstationary correlation functions with compact support for multivariate random fields. Genton and Kleiber (2015) reviewed approaches to building models for covariances between two 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

a. 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:

$$\tilde{\mathbf{P}}^b = \hat{\mathbf{P}}^b \circ \mathbf{C}, \tag{3}$$

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 ~~$\mathbf{H}$ is typically the linearization of a local interpolation function, $h(\cdot)$, for which~~ 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 (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)).

b. 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, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$, where $\mathbf{x}_i$, $i = 1, 2, \ldots, N$, 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

$$\hat{\mathbf{P}}^b = \begin{pmatrix} \hat{\mathbf{P}}^b_{11} & \hat{\mathbf{P}}^b_{12} & \cdots & \hat{\mathbf{P}}^b_{1N} \\ \hat{\mathbf{P}}^b_{21} & \hat{\mathbf{P}}^b_{22} & \cdots & \hat{\mathbf{P}}^b_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{P}}^b_{N1} & \hat{\mathbf{P}}^b_{N2} & \cdots & \hat{\mathbf{P}}^b_{NN} \end{pmatrix}. \tag{4}$$

The entries of the submatrices, $\hat{\mathbf{P}}^b_{ii}$, $i = 1, \ldots, N$, 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 $\hat{\mathbf{P}}^b_{ii}$ represents the sample variance for the virtual temperature at a given model grid point, while each off-diagonal entry of $\hat{\mathbf{P}}^b_{ii}$ represents the sample covariances between the virtual temperatures at a pair of grid points. Likewise, the entries of $\hat{\mathbf{P}}^b_{ij}$, $i \neq j$, 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, $\boldsymbol{\rho}(d) = \{\rho_{ij}(d)\}_{i,j=1,\ldots,N}$, 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 sub-matrix $\mathbf{P}^b_{ij}$ 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 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 covariance matrix, $\hat{\mathbf{P}}^b_{ij}$, 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

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_0 & \mathbf{C}_0 \\ \mathbf{C}_0 & \mathbf{C}_0 \end{pmatrix} \tag{5}$$

independently of the particular choice of the localization function. Here 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 dimension is $2n \times 2n$.

As in Eq. (2), although $\mathbf{C}$ is rank-deficient and thus so is the localized covariance matrix $\tilde{\mathbf{P}}^b$, we may still be able to calculate the inverse of $\mathbf{H}\tilde{\mathbf{P}}^b\mathbf{H}^T + \mathbf{R}$, as $\mathbf{R}$ is a diagonal matrix. The smallest eigenvalue of $\mathbf{H}\tilde{\mathbf{P}}^b\mathbf{H}^T + \mathbf{R}$ is the smallest (positive) value of $\mathbf{R}$, and thus the matrix, $\mathbf{H}\tilde{\mathbf{P}}^b\mathbf{H}^T + \mathbf{R}$, 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 $\mathbf{H}\tilde{\mathbf{P}}^b\mathbf{H}^T + \mathbf{R}$ 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, $\boldsymbol{\rho}(d)$. 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 $\boldsymbol{\rho} = \tilde{\rho} \cdot \mathbf{B}$, where $\mathbf{B}$ is an $N \times N$ symmetric, positive-definite matrix whose diagonal entries are one. It can be easily verified that $\boldsymbol{\rho}$ 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

$$\mathbf{B} = \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix}, \tag{6}$$

for β with $|\beta| < 1$, leads to

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_0 & \beta\mathbf{C}_0 \\ \beta\mathbf{C}_0 & \mathbf{C}_0 \end{pmatrix} \tag{7}$$

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 (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

$$\mathbf{L} = \begin{bmatrix} \ell_{1,1} & 0 & \cdots & 0 \\ \ell_{2,1} & \ell_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \ell_{N,1} & \ell_{N,2} & \cdots & \ell_{N,N} \end{bmatrix} \tag{8}$$

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 $i = 1, \ldots, N$. The constraints are used to have the diagonal entries of $\mathbf{B}$ to be one. ~~Furthermore, we need the constraint that for each row of $\mathbf{L}$, the sum of all of the squared elements should be 1 in order to have the diagonal entries of $\mathbf{B}$ to be one.~~ 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. (2012). The function class they considered was essentially a multivariate extension of the *Askey* function (Askey 1973), $f(d; \nu, c) = \left(1 - \frac{d}{c}\right)_+^{\nu}$, 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. (2012), is given by $(i, j = 1, 2)$

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

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

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

Here, $\Gamma(\cdot)$ is the gamma function (e.g., Wilks 2006), 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 (9) has the support c because it sets covariances beyond a distance c to zero. It can be seen from (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 $\boldsymbol{\rho}$ to be valid is $|\beta_{12}| \leq 1$. ~~Note that~~ 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 (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

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

and 0 otherwise, where $\nu \geq (s+1)/2$, $\mu_{ij} = (\mu_i + \mu_j)/2$, and $\mu_i > 0$ for all $i = 1, \ldots, N$. Here, B is the beta function (Porcu et al. 2012; 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 $\nu = 1, \ldots, 4$ (Fig. 1). This figure shows that for a given support, the Askey functions are narrower.

3. Experiments

a. 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

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

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

where $\mathbf{x}_k^{'}$, $k = 1, 2, \ldots, M$ are the ensemble perturbations and $\mathbf{y}_k^{o'}$, $k = 1, 2, \ldots, M$ 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 sub-ensembles.

b. 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.

~~The idealized model we use is the bivariate Lorenz-95 model (Lorenz 1995). The model mimics the nonlinear dynamics of two linearly coupled atmospheric state variables, X and Y, on a latitude circle.~~ 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

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

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

where $Y_{j-J,k} = Y_{j,k-1}$ and $Y_{j+J,k} = Y_{j,k+1}$ for $k = 1, \ldots, K$ and $j = 1, \ldots, J$. The "boundary condition" is periodic; 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 time scales 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 non-dimensional 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. Fig. 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.

c. 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 randomly chosen 20% of all locations and Y is observed at randomly chosen 90% of those locations where X is not observed. 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 $2,000$-time-step integration of the model. We initialize an ensemble by adding the standard Gaussian noise to the true state; then, discarding the first $3,000$ time steps. We then generate simulated observations by adding random observation noise of mean zero and variance 0.02 to the 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 $1,000$ time steps of 50 different realizations of each experiment is shown by a boxplot. The boxplot 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) that 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 boxplot 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:

i. S1–the bivariate sample background covariance is used without localization;

ii. S2–same as S1 except that the cross-covariances between X and Y are replaced by zeros;

iii. 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;

iv. S4–one of the bivariate localization methods described in Section 2.b 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 $\boldsymbol{\rho}^{(1)}(\cdot) = \{\beta_{ij}\rho^{(1)}(\cdot)\}_{i,j=1,2}$ 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,

$$\rho^{(1)}(d;c) = \begin{cases} -\frac{1}{4}(|d|/c)^5 + \frac{1}{2}(d/c)^4 + \frac{5}{8}(|d|/c)^3 - \frac{5}{3}(d/c)^2 + 1, & 0 \le |d| \le c, \\ \frac{1}{12}(|d|/c)^5 - \frac{1}{2}(d/c)^4 + \frac{5}{8}(|d|/c)^3 + \frac{5}{3}(d/c)^2 - 5(|d|/c) + 4 - \frac{2}{3}c/|d|, & c \le |d| \le 2c, \\ 0, & 2c \le |d|. \end{cases} \tag{16}$$

This correlation function attenuates the covariances with increasing distance, setting all the covariances to zero beyond distance $2c$. So this function has the support $2c$. If $|\beta| < 1$ and c is the same for both the marginal- and the cross-covariances, the matrix-valued function, $\boldsymbol{\rho}^{(1)}$, is positive-definite and of full rank. We test various values of the localization parameters c and β, and present the test results in next section.

The second multivariate correlation function we consider, $\boldsymbol{\rho}^{(2)}$, is the bivariate Askey function described in Section 2.b. In particular, we use $\mu_{11} = 0$, $\mu_{22} = 2$, $\mu_{12} = 1$, and $\nu = 3$. According to Eq. (10), for these choices of parameters, the one remaining parameter, β_{12}, must be chosen such that $|\beta_{12}| < 0.79$.

d. Results

Figure 4 shows the distribution of RMSE for variable X for different configurations of the localization scheme in the case where the state is only partially observed. This figure compares the Askey function and Gaspari-Cohn function which have the same support (localization radius), so setting all the covariances to zero beyond the same distance. We recall that because X is much more sparsely observed than Y, we expect to see some sensitivity of the analyses of X to

the treatment of the cross-covariance terms. The figure confirms this expectation. A comparison of the results for configurations S1 and S2 suggests that ignoring the cross-covariances is a better strategy than to use them without localization. This conclusion does not hold once a univariate localization is applied to the marginal covariances, as using configuration S3 produces worse results than applying no localization at all (S1).

Figure 4 also shows that the distribution of the state estimation error is less sensitive to the choice of localization strategy for the larger values of support. Of all localization schemes, S4 with $\beta = 0.1$ performs best regardless of the localization radius: the distribution of the state estimation error is narrow with a mean value that is lower than those for the other configurations of the localization scheme. For this choice of localization scheme and β, the Askey function produces smaller errors than the Gaspari-Cohn function, particularly, for smaller localization radii.

Figure 5 is the same as Fig. 4, except for variable Y rather than for variable X. A striking feature of the results shown in this figure is that the Askey function clearly performs better than the Gaspari-Cohn function. Another obvious conclusion is that using a smaller localization radius (a lower value of support) is clearly advantageous for the estimation of Y. This result is not surprising, considering that Y is densely observed and its spatial variability is much higher than that of X. In contrast to the results for variable X, configuration S3 produces much more accurate estimates of variable Y than do configurations S1 or S2. In addition, configuration S4 performs only slightly better, and only for the lowest value of support, than does configuration S3. The latter observations indicate that the marginal covariances play a more important role than do the cross-covariances in the estimation of the densely observed Y. The proper filtering of the marginal covariances can thus greatly increase the accuracy of the estimates of Y. In other words, the densely observed Y is primarily estimated based on observations of Y. Hence, the low signal-to-noise ratio for the sample estimate of the marginal covariances for Y greatly limits

the value of the observations of Y at longer distances.

Figure 6 is the same as Fig. 4, except for the case of a fully observed state. By comparing the two figures, we see that the analysis is far less sensitive to the localization radius in the fully observed case than in the partially observed case. As can be expected, the state estimates are also more accurate in the fully observed case. In the fully observed case, localization strategy S3 performs much better than do strategies S1 and S2 and similarly to S4. This result indicates that in the fully observed case, X is primarily analyzed based on observations of X, making the analysis of X more sensitive to the localization of the marginal covariances than to the localization of the cross-covariances. Similar to the partially observed case, the Askey function tends to perform better than the Gaspari-Cohn function, but the differences between the accuracy of the state estimates for the two filter functions are negligible, except for the shortest localization radius.

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 500-member ensemble. The results for the 500-member ensemble are shown in Figures 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 ensemble. 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 (Figure 8) and both variables X and Y in the fully observed case (Figures 10 and 11), the best results are obtained with S3 and S4 regardless of the localization radius. They also shows 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.

~~Figures 8 -11 show the results for the 40-member ensemble. We use an inflation factor of 1.005, because the optimal value of the inflation factor is typically smaller for a larger ensemble. Figure 8 shows that the ensemble size has little effect on the estimates of X in the case of the partially observed state. 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), however, the best results are obtained with a larger localization radius than in the case of the 20-member ensemble. This behavior is~~

~~expected, as a larger ensemble can more accurately estimate the weaker covariances associated with the longer distances. As for the 20-member ensemble, the localization schemes using the Askey function perform better than those using the Gaspari-Cohn function.~~

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 lead to a rank deficient estimate of the background covariance matrix. 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 estimates 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.

Acknowledgments.

The authors are grateful to the reviewers for valuable comments that significantly improved presentation of the paper. Mikyoung Jun's research was supported by NSF grant DMS-1208421,

while Istvan 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).

REFERENCES

Anderson, J. and L. Lei, 2013: Empirical localization of observation impact in ensemble kalman filters. *Mon. Wea. Rev.*, **142**, 739–754.

Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D*, **230**, 99–111.

Askey, R., 1973: Radial characteristic functions. technical report no. 1262, Mathematical Research Center, University of Wisconsin-Madison, Madison.

Bishop, C. H. and D. Hodyss, 2007: Flow adaptive moderation of spurious ensemble correlations and its use in ensemble based data assimilation. *Quart. J. Roy. Meteor. Soc.*, **133**, 2029–2044.

Bishop, C. H. and D. Hodyss, 2009a: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus*, **61A**, 84–96.

Bishop, C. H. and D. Hodyss, 2009b: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere. *Tellus*, **61A**, 97–111.

Buehner, M. and M. Charron, 2007: Spectral and spatial localization of background-error correlations for data assimilation. *Quart. J. Roy. Meteor. Soc.*, **133**, 615–630.

Campbell, W. F., C. H. Bishop, and D. Hodyss, 2010: Vertical covariance localization for satellite radiances in ensemble kalman filters. *Mon. Wea. Rev.*, **138**, 282–290.

Du, J. and C. Ma, 2013: Vector random fields with compactly supported covariance matrix functions. *J. Stat. Plan. Infer.*, **143**, 457–467.

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.*, **125**, 723–757.

Genton, M. G. and W. Kleiber, 2015: Cross-covariance functions for multivariate geostatistics (with discussion). *Statist. Sci.*, **30**, 147–163.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.

Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

Houtekamer, P. L. and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126.

Jun, M., I. Szunyogh, M. G. Genton, F. Zhang, and C. H. Bishop, 2011: A statistical investigation of the sensitivity of ensemble-based Kalman filters to covariance filtering. *Mon. Wea. Rev.*, **139**, 3036–3051.

Kang, J.-S., E. Kalnay, J. Liu, I. Fung, T. Miyoshi, and K. Ide, 2011: "Variable localization" in an ensemble Kalman filter: Application to the carbon cycle data assimilation. *J. Geophys. Res.*, **116**, doi:10.1029/2010JD014673.

Kleiber, W. and E. Porcu, 2015: Nonstationary matrix covariances: Compact support, long range dependence and adapted spectra. *Stoch. Environ. Res. Risk. Assess.*, **29**, 193–204.

Lei, L. and J. Anderson, 2014: Comparison of empirical localization techniques for serial ensemble kalman filters in a simple atmospheric general circulation model. *Mon. Wea. Rev.*, **141**, 4140–4153.

Lorenz, E. N., 1995: Predictability-A problem partly solved. Proceedings of the Seminar on Predictability, vol. 1. ECMWF, Reading, Berkshire, UK.

Lorenz, E. N., 2005: Designing chaotic models. *J. Atmos. Sci.*, **62**, 1574–1587.

Ott, E., et al., 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428.

Porcu, E., D. J. Daley, M. Buhmann, and M. Bevilacqua, 2012: Radial basis functions with compact support for multivariate geostatistics. *Stoch. Environ. Res. Risk. Assess.*, **27**, 909–922.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.

Wilks, D. S., 2006: *Statistical methods in the atmospheric sciences*. Academic Press, Amsterdam.

Zhang, H. and J. Du, 2008: Covariance tapering in spatial statistics. Positive Definite Functions: From Schoenberg to Space-Time Challenges, eds. J. Mateu and E. Porcu, Gráficas Castañ, Spain.

List of Figures

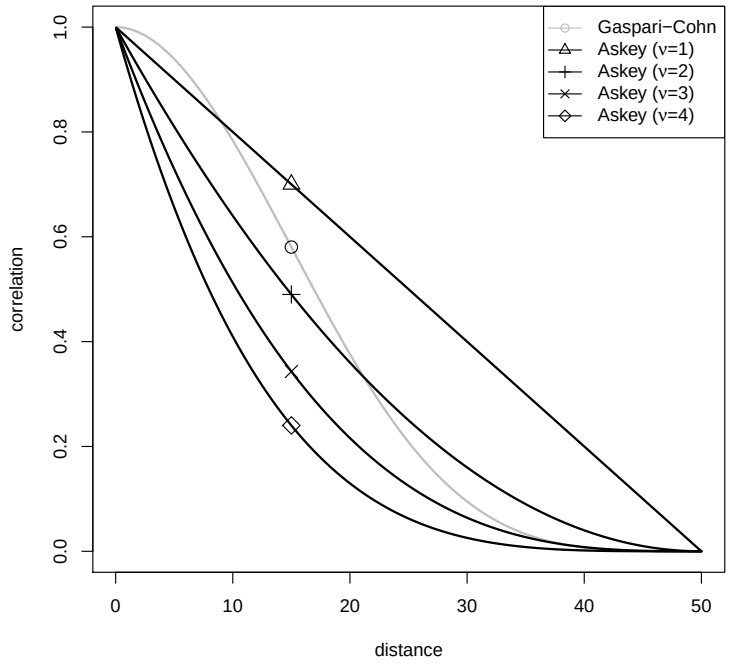

Figure 1: The Gaspari-Cohn covariance function with a localization constant $c = 25$ (support of 50) and the Askey covariance function $f(d; \nu, c) = \left(1 - \frac{d}{c}\right)_+^\nu$, with a support parameter $c = 50$ and various shape parameters.

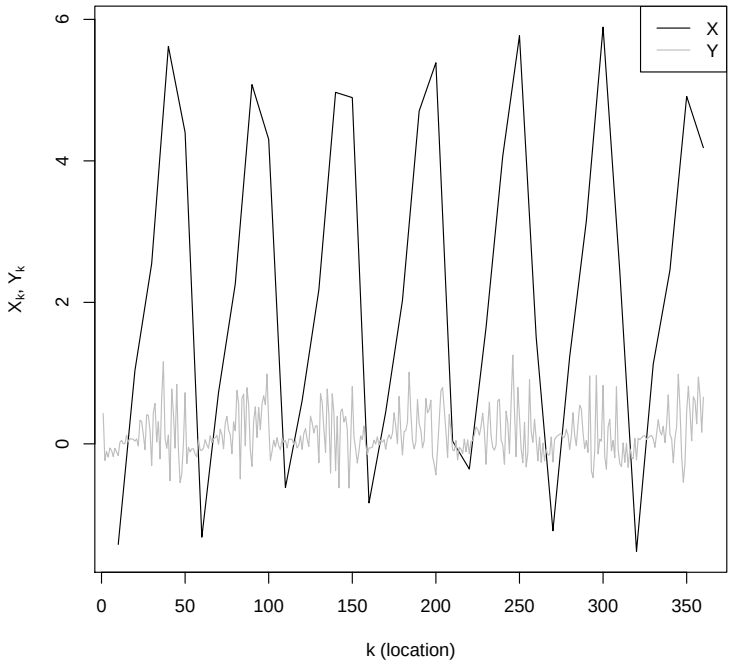

Figure 2: A snapshot of the variables X and Y from a numerical integration of the system of Eqs. (14) and (15) with $K = 36$, $J = 10$, $F = 10$, $a = 10$, $b = 10$, and $h = 2$.

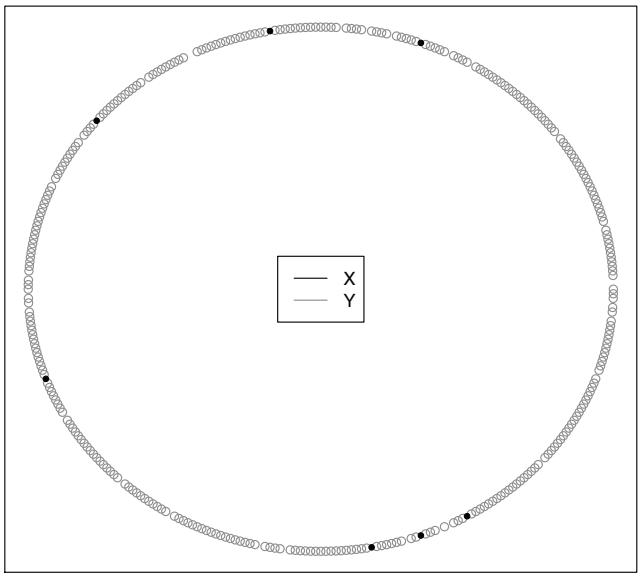

Figure 3: Spatial locations of partial observation of X and Y.

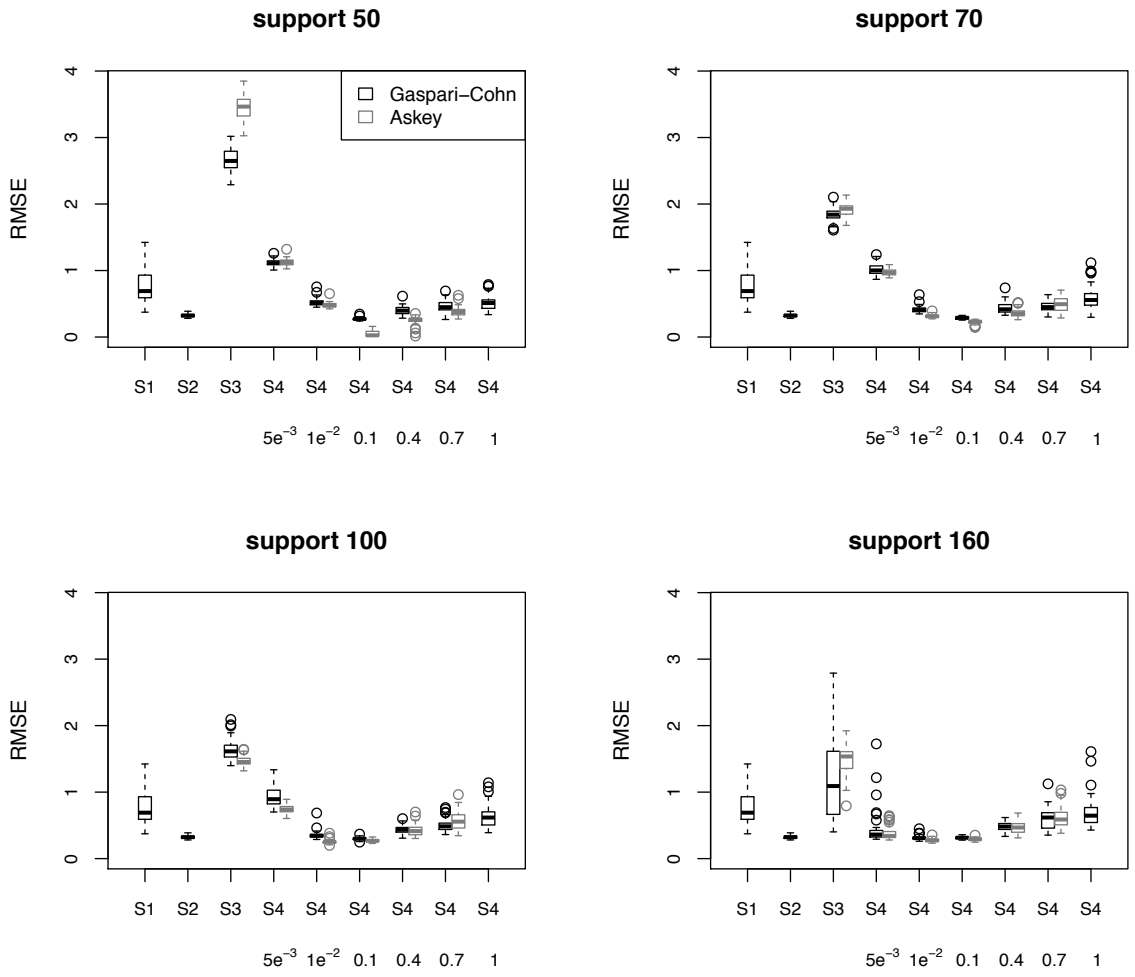

Figure 4: The probability distribution of RMSE for variable X in the case when the system is only partially observed. Results are shown for different localization strategies. For the definitions of localization strategies S1, S2, S3 and S4, see the text. The title of each panel indicates the localization radius (length of support). The numbers below S4 indicate the value of β.

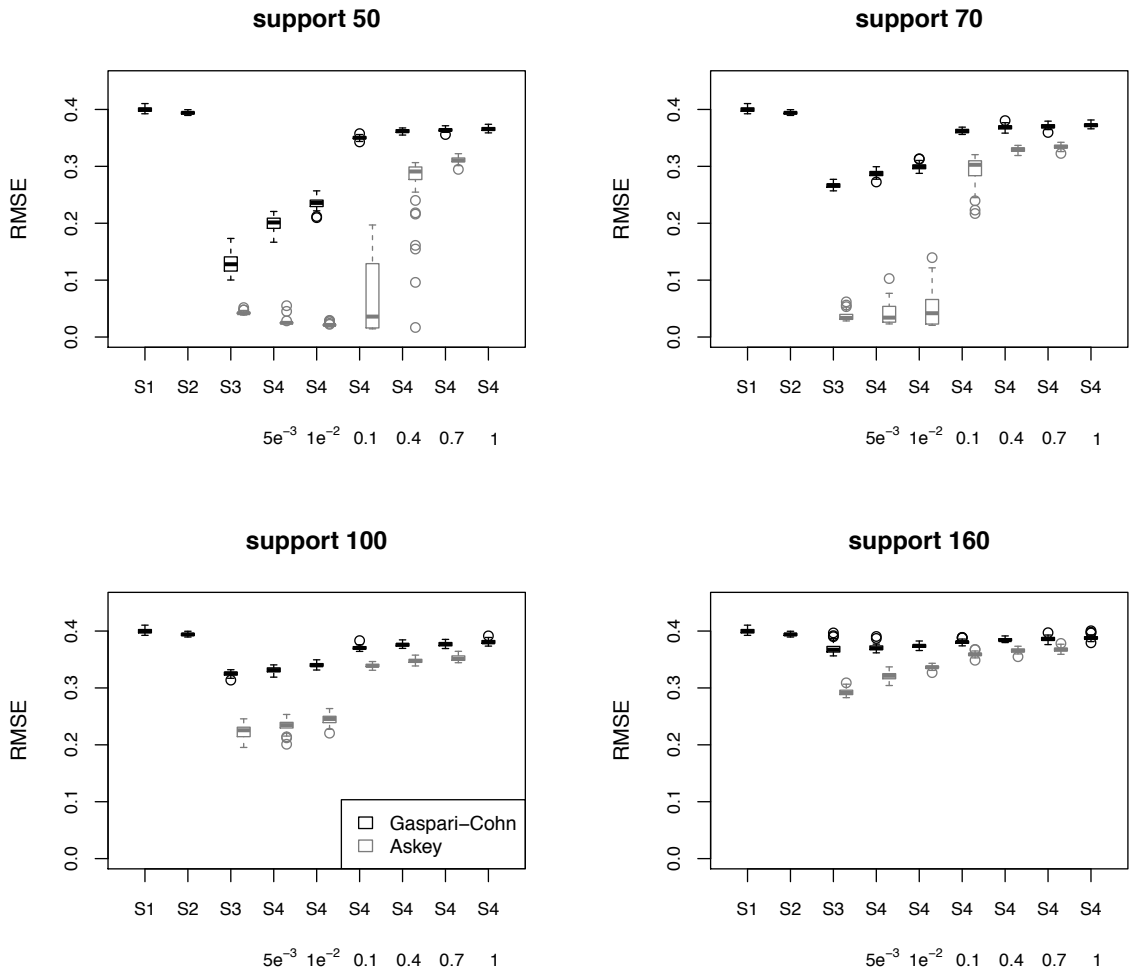

Figure 5: Same as 4, except for variable Y.

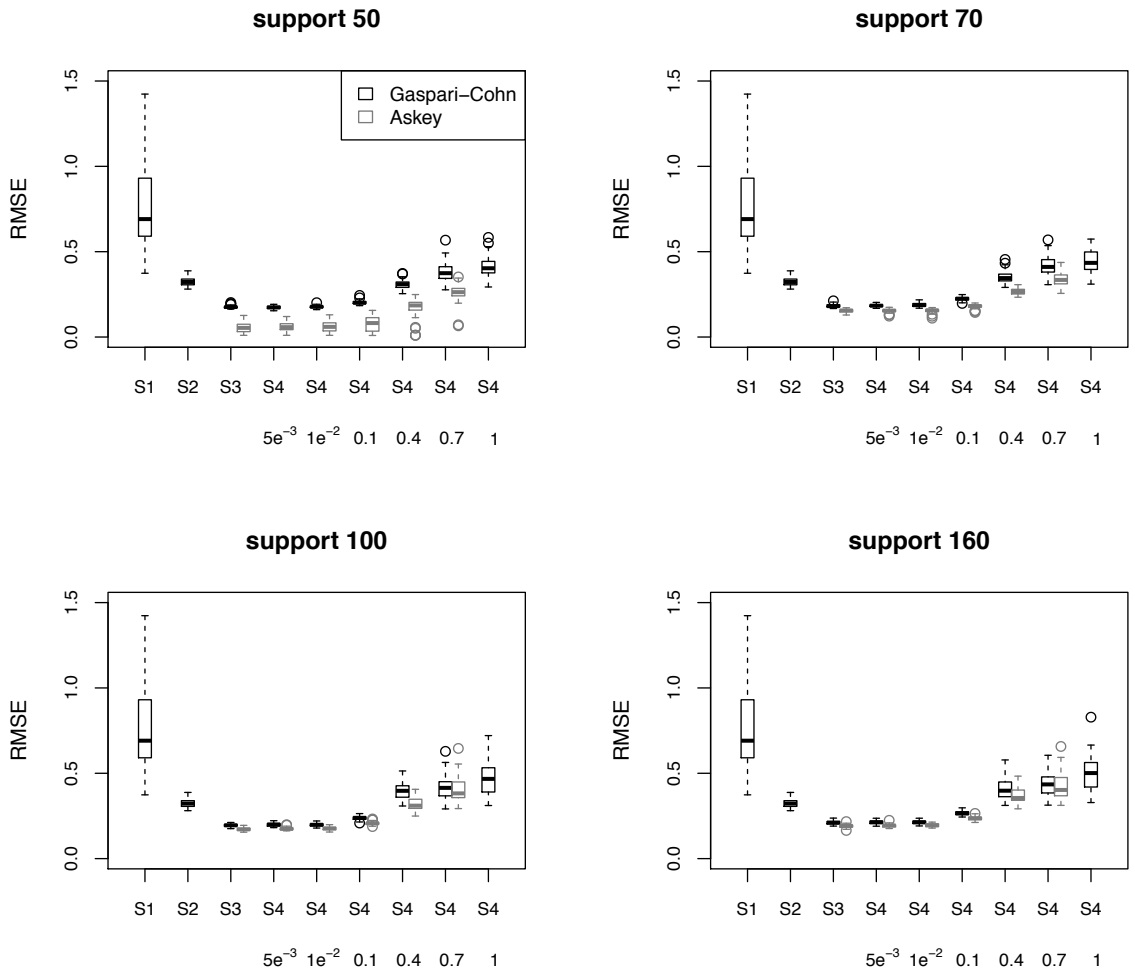

Figure 6: Same as 4, except for the case when the system is fully observed.

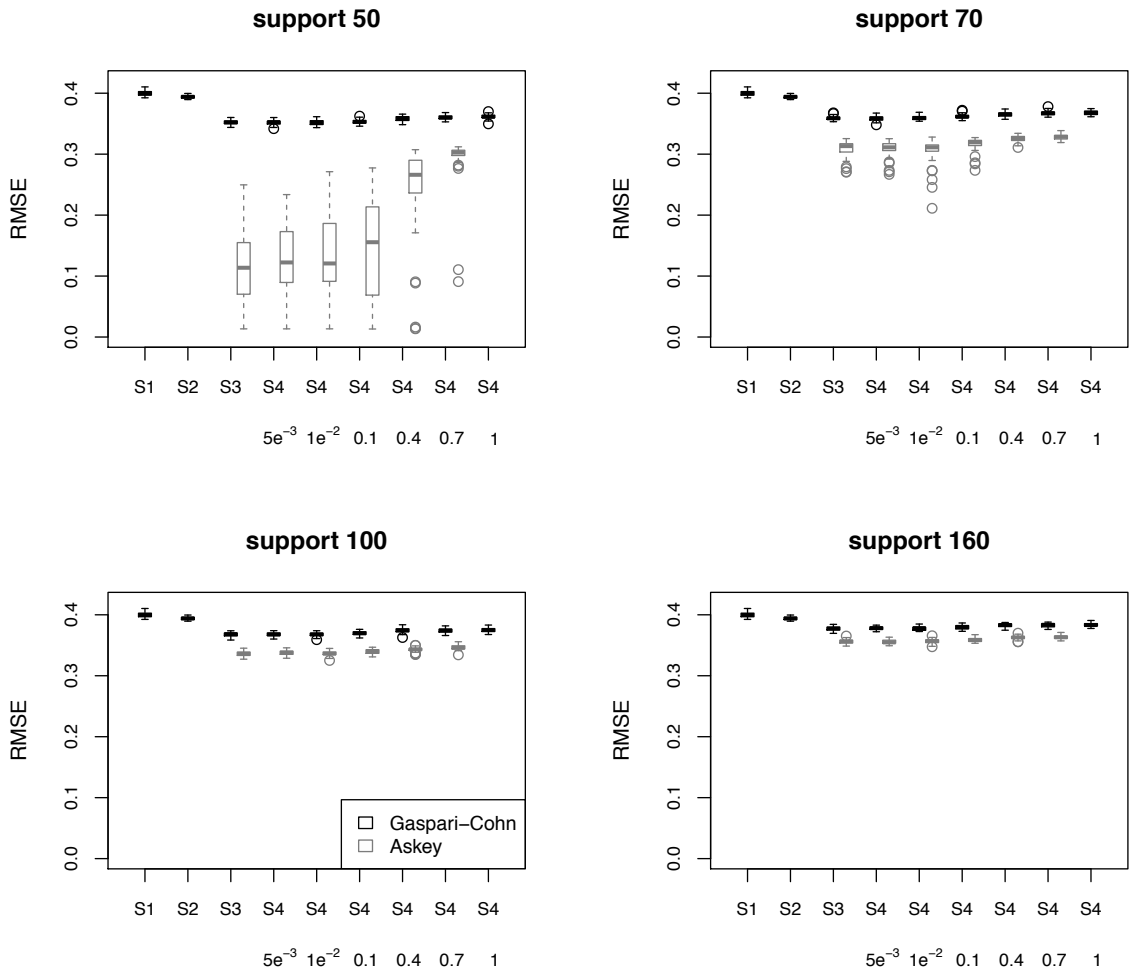

Figure 7: Same as 6, except for variable Y.

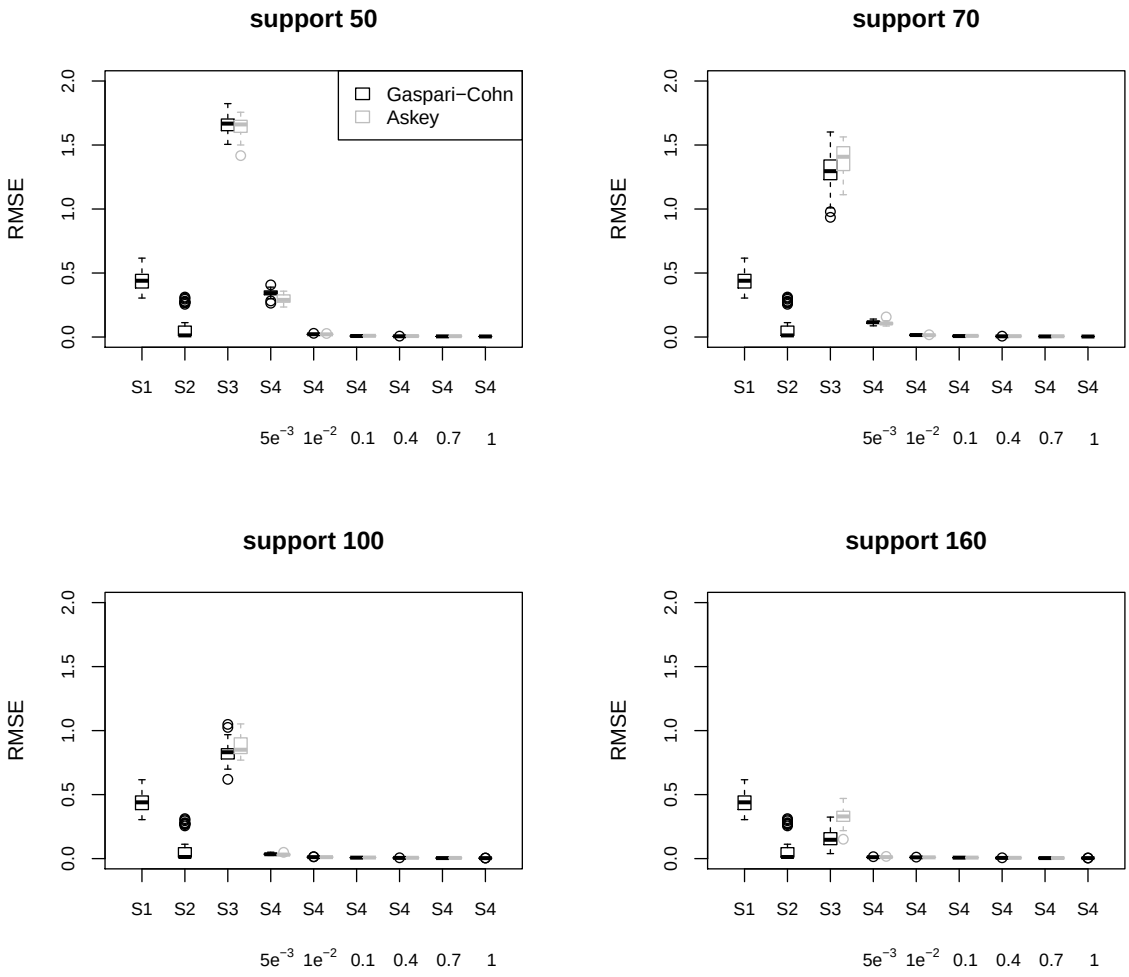

Figure 8: Same as 4, except for 500 ensemble members.

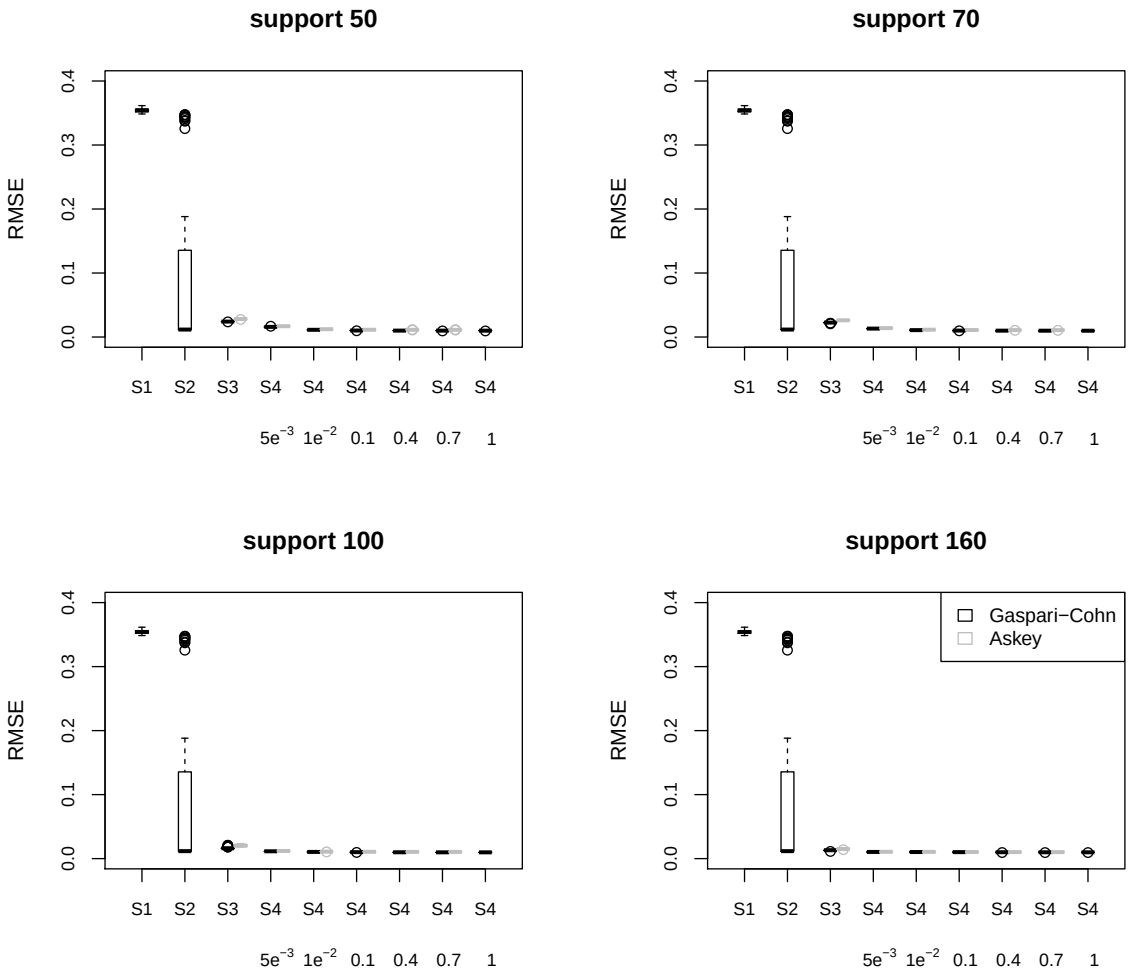

Figure 9: Same as 5, except for 500 ensemble members.

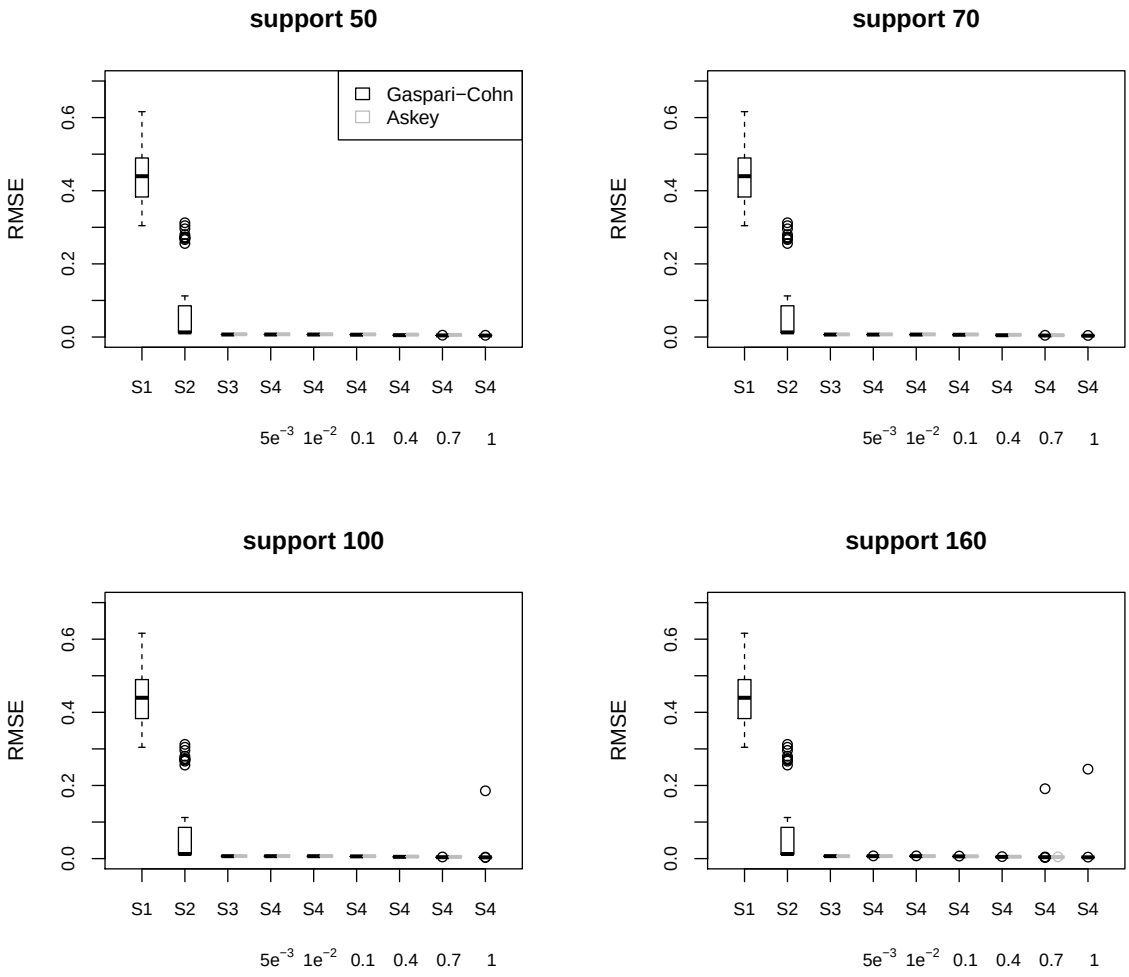

Figure 10: Same as 6, except for 500 ensemble members.

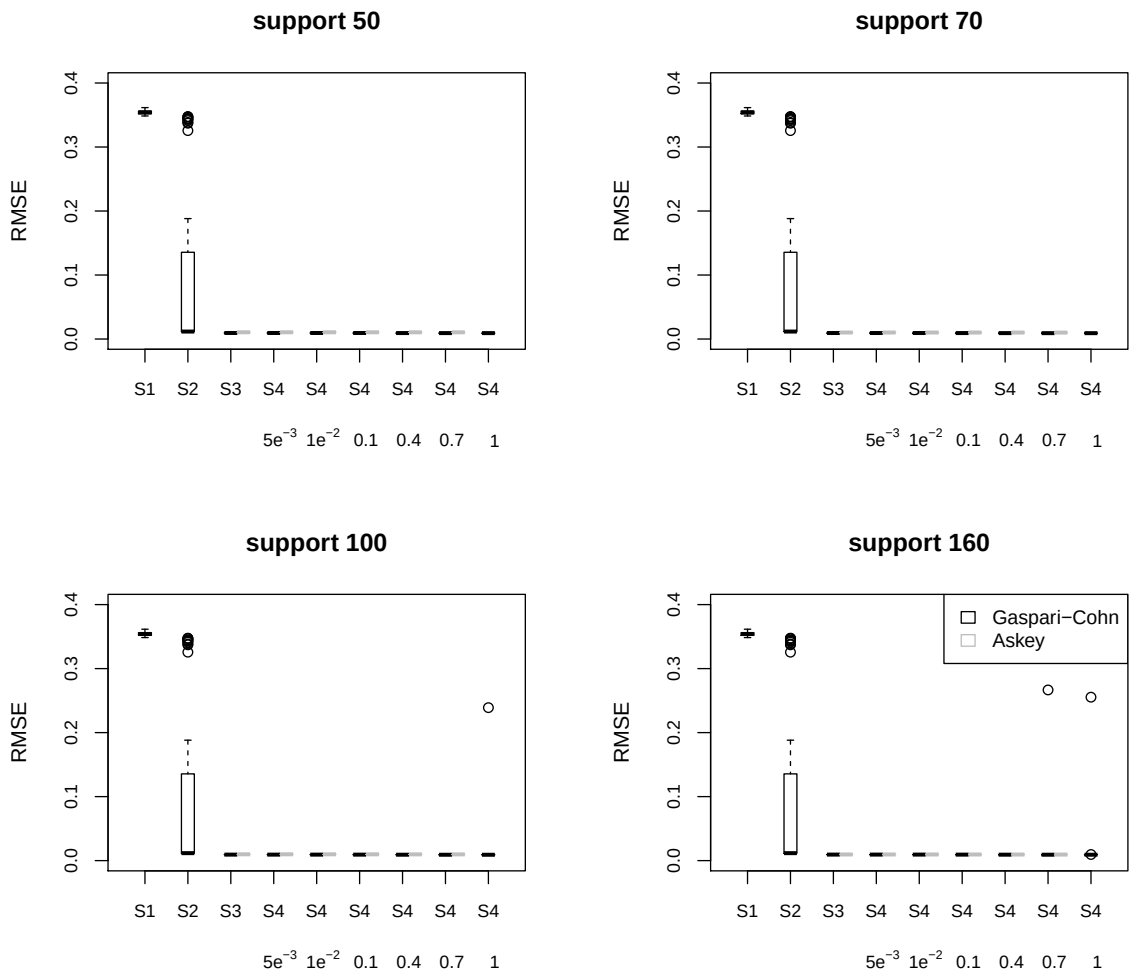

Figure 11: Same as 7, except for 500 ensemble members.