# A Statistical Investigation of the Sensitivity of Ensemble-Based Kalman Filters to Covariance Filtering

MIKYOUNG JUN

*Department of Statistics, Texas A&M University, College Station, Texas*

ISTVAN SZUNYOGH

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

MARC G. GENTON

*Department of Statistics, Texas A&M University, College Station, Texas*

FUQING ZHANG

*Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania*

CRAIG H. BISHOP

*Marine Meteorology Division, Naval Research Laboratory, Monterey, California*

ABSTRACT

This paper investigates the effects of spatial filtering on the ensemble-based estimate of the background error covariance matrix in an ensemble-based Kalman filter (EnKF). In particular, a novel kernel smoothing method with variable bandwidth is introduced and its performance is compared to that of the widely used Gaspari–Cohn filter, which uses a fifth-order kernel function with a fixed localization length. Numerical experiments are carried out with the 40-variable Lorenz-96 model. The results of the experiments show that the nonparametric approach provides a more accurate estimate of the background error covariance matrix than the Gaspari–Cohn filter with any localization length. It is also shown that the Gaspari–Cohn filter tends to provide more accurate estimates of the covariance with shorter localization lengths. However, the analyses obtained by using longer localization lengths tend to be more accurate than those produced by using short localization lengths or the nonparametric approach. This seemingly paradoxical result is explained by showing that localization with longer localization lengths produces filtered estimates whose time mean is the most similar to the time mean of both the unfiltered estimate and the true covariance. This result suggests that a better metric of covariance filtering skill would be one that combined a measure of closeness to the sample covariance matrix for a very large ensemble with a measure of similarity between the climatological averages of the filtered and sample covariance.

## 1. Introduction

Atmospheric data assimilation is the process of estimating the spatiotemporally evolving state of the atmosphere based on observations. The resulting state estimate at a given time is called analysis. Modern data assimilation algorithms obtain the analysis by a statistical interpolation process: the analysis is computed by updating an a priori estimate of the state, called background, based on the observed information assuming that the background and observation errors are random variables with known statistical parameters (Daley 1991; Kalnay 2003). In particular, the data assimilation schemes, which are able to handle the large number of state variables and observations in a realistic operational or research application, assume that the probability distribution of the

*Corresponding author address:* Mikyoung Jun, 3143 TAMU, College Station, TX 77843-3143.
E-mail: mjun@stat.tamu.edu

background and observation errors is a multivariate normal distribution with a known covariance matrix. The focus of this paper is on the estimation of the covariance matrix that describes the distribution of the background error. This matrix is called the background error covariance matrix and we denote it by $\mathbf{P}^b$.

Sequential data assimilation schemes use a short-term forecast started from the analysis of the previous analysis time as background. Thus, $\mathbf{P}^b$ is an $M \times M$ matrix, where $M$ is the number of gridpoint variables in the model. The entries $p_{uv}^b(u, v = 1, \ldots, M)$ of $\mathbf{P}^b$ describe the presumed covariance between the errors $\varepsilon_u^b$ and $\varepsilon_v^b$ in the $x_u^b$ and $x_v^b$ components of the background $\mathbf{x}^b$. Each diagonal element $p_{uu}^b(u = 1, \ldots, M)$ represents the variance of the error for a scalar model variable (e.g., temperature, surface pressure, a component of the wind vector) at a given model grid point, while the off-diagonal elements represent the covariance between the errors in the estimate of the same state variable at different locations and between the errors in the estimate of the different state variables (e.g., temperature and meridional component of the wind vector) at the same or different locations. Hence, a physical distance, denoted by $|u - v|$, can be defined for each covariance $p_{uv}^b$ with the distance between the physical locations of the two state vector components for which $p_{uv}^b$ describes the statistical relationship between the state estimation errors. (For the diagonal entries and the entries that describe the covariance between variables at the same location, $|u - v| = 0$.)

An estimate of $\mathbf{P}^b$ is usually obtained in two steps; first computing a sample covariance matrix

$$\hat{\mathbf{P}}^b = \frac{1}{K-1} \sum_{k=1}^{K} \mathbf{X}^{b(k)}[\mathbf{X}^{b(k)}]^{\mathrm{T}}, \tag{1}$$

where $\{\mathbf{X}^{b(k)}: k = 1, \ldots, K\}$ is a $K$-member sample of the background error; then filtering the sampling noise by a statistical postprocessing of $\hat{\mathbf{P}}^b$ (e.g., Berre and Desroziers 2010). In this paper, we restrict our attention to the estimation of $\mathbf{P}^b$ in ensemble-based Kalman filter (EnKF) schemes. In such schemes, $\{\mathbf{X}^{b(k)}: k = 1, \ldots, K\}$ is obtained by first integrating the model from a $K$-member ensemble of analyses for the previous analysis time to obtain a background ensemble $\{\mathbf{X}^{b(k)}: k = 1, \ldots, K\}$; then, taking the difference between the members of the background ensemble and the background mean

$$\bar{\mathbf{x}}^b = \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}^{b(k)}, \tag{2}$$

to obtain $\{\mathbf{X}^{b(k)} = \mathbf{x}^{b(k)} - \bar{\mathbf{x}}^b : k = 1, \ldots, K\}$.

Because of the large computational expense associated with an ensemble of model integrations, the computationally affordable sample size is limited; for example, the typical value of $K$ in a practical application is between 20 and 100 (e.g., Houtekamer and Mitchell 2005; Hamill 2006). The small ensemble size $K$ poses two important challenges for the statistical postprocessing algorithm that provides the final estimate $\tilde{\mathbf{P}}^b$ of $\mathbf{P}^b$ based on $\hat{\mathbf{P}}^b$. First, statistical fluctuations in a small ensemble $\{\mathbf{X}^{b(k)}: k = 1, \ldots, K\}$ lead to a low signal-to-noise ratio for those entries of $\hat{\mathbf{P}}^b$ that estimate small entries of $\mathbf{P}^b$, which makes filtering the sample noise challenging. This problem typically occurs for entries $p_{uv}^b$ for which the distance $|u - v|$ is large. Second, $\hat{\mathbf{P}}^b$ is a highly rank-deficient estimate of $\mathbf{P}^b$ ($K \ll M$), because the typical dimension of the state vector in a state-of-the-art atmospheric model is $M = 10^5$–$10^8$.

Practical implementations of the EnKF address the aforementioned two problems by applying a physical-distance-dependent stationary (time independent) filter function to the sample covariances (e.g., Houtekamer and Mitchell 1998, 2001; Hamill et al. 2001; Anderson 2001; Ott et al. 2004). This approach is called *localization*, because it forces the covariance to zero beyond a prescribed distance $d$. Some localization functions do not change the sample-based estimate $\hat{p}_{uv}^b$ of the covariance when the distance $|u - v|$ associated with the pair of state vector components is smaller than $d$, but replaces $\hat{p}_{uv}^b$ with zero when $|u - v| \geq d$ (e.g., (Houtekamer and Mitchell 1998; Ott et al. 2004; Szunyogh et al. 2005, 2008; Hunt et al. 2007); other localization functions taper the covariance to zero gradually with increasing distance $|u - v|$ (e.g., Houtekamer and Mitchell 2001; Hamill et al. 2001). Such tapering functions modify all entries of the sample covariance matrix, except for the diagonal elements. Experience with the different EnKF algorithms and localization strategies suggests that tapering greatly increases the accuracy of the analyses, especially when the size of the ensemble is small (e.g., Houtekamer and Mitchell 2005; Hamill 2006) or in the presence of model error (e.g., Zhang et al. 2009b).

While there exist dynamical arguments to support localization (Yoon et al. 2010) and it may formally solve the problems it is designed to address,[1] the particular shape of the tapering function is typically selected based on intuition. Moreover, if the true covariance structure of the system is complex and does not decrease monotonically with distance, the localized sample covariance may not be a good fit to the true covariance

---

[1] Localization also makes the data assimilation algorithms more suitable for implementation on massively parallel computer architectures.

structure. The ad hoc nature of most localization algorithms used in the literature motivates us to start exploring the effects of covariance filtering on the performance of the EnKF in a more systematic way. In particular, we compare the results obtained with localization to the results obtained with an adaptive *nonparametric* method to estimate the entries of $\mathbf{P}^b$. The property that makes such a nonparametric method particularly suitable for our study is that, instead of an ad hoc choice, the shape of the localization function and the distance at which it becomes zero is determined from the sample itself. We are aware of only two earlier attempts at adaptively determining the localization distance: the hierarchical filter of Anderson (2007) estimates the localization function using an ensemble of ensembles, while the adaptive localization approach in Bishop and Hodyss (2007) and the ensemble correlations raised to a power (ECO-RAP) approach of Bishop and Hodyss (2009a,b) propagate and adapt the width of the localization by computing powers of the sample correlations. Zhang et al. (2009a), on the other hand, proposed a successive covariance localization (SCL) method that uses several different localization distances to account for different physical scales in the background error covariance.

The rest of the paper is organized as follows. Section 2 is a formal description of covariance filtering in EnKF. Section 3 explains the design of the data assimilation runs that provide the data for our quantitative investigations. This section includes a brief description of the particular implementation of the serial square root EnKF scheme of Whitaker and Hamill (2002) on the Lorenz-96 model (Lorenz 1996; Lorenz and Emanuel 1998), which we use to carry out all data assimilation experiments described in this paper. In section 4, we explain the nonparametric statistical method we use to estimate the covariance, while in section 5 we compare the performance of a standard localization method to that of the nonparametric scheme in estimating the background covariance. Then, in section 6, we compare the performance of the different filtering strategies by numerical experiments. Our conclusions are drawn in section 7.

## 2. Covariance filtering in EnKF

In this section, we illustrate the process of estimating $\mathbf{P}^b$ in an ensemble-based data assimilation system with the help of the algorithm introduced in Whitaker and Hamill (2002). We choose this particular algorithm because it allows for a straightforward implementation of distance-dependent filtering of the covariance. It also performs well for small ensemble sizes, which makes

it an ideal scheme for testing the effects of distance-dependent filtering. Later in the paper, we use an implementation of this algorithm on the Lorenz-96 model. While all matrices, vectors, and scalars described in this section are dependent on the analysis time, we do not indicate this time dependence in our notation, as all operations are carried out at the same analysis time.

### a. The EnKF algorithm of Whitaker and Hamill (2002)

In the extended Kalman filter (Jazwinski 1970), the background is updated with the equation

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}[\mathbf{y}^o - \mathcal{H}(\mathbf{x}^b)], \quad (3)$$

where the Kalman-gain matrix $\mathbf{K}$ is defined by

$$\mathbf{K} = \mathbf{P}^b \mathbf{H}^T (\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R})^{-1}. \quad (4)$$

In Eq. (3), the $M$-vector $\mathbf{x}^a$ is the analysis; $\mathcal{H}(\cdot)$ is the observation operator, that is, $\mathcal{H}(\mathbf{x}^b)$ is the prediction of the observations based on the background; $\mathbf{H}$ is the matrix that represents the linearization of $\mathcal{H}(\cdot)$ about $\mathbf{x}^b$; and $\mathbf{R}$ is the observation error covariance matrix.

In an EnKF, $\mathbf{x}^b$ is replaced with the ensemble mean $\overline{\mathbf{x}}^b$ and the result of Eq. (3) is the ensemble mean analysis $\overline{\mathbf{x}}^a$. In addition to the ensemble mean analysis, an EnKF also computes the ensemble of analysis perturbations $\{\mathbf{X}^{a(k)}: k = 1, \ldots, K\}$, which then can be added to the analysis mean to obtain the ensemble of analyses, $\{\mathbf{x}^{a(k)} = \overline{\mathbf{x}}^a + \mathbf{X}^{a(k)}: k = 1, \ldots, K\}$. EnKF schemes generate the ensemble of analysis perturbations, $\{\mathbf{X}^{a(k)}: k = 1, \ldots, K\}$, such that they satisfy the condition

$$\hat{\mathbf{P}}^a = \frac{1}{K-1} \sum_{k=1}^{K} \mathbf{X}^{a(k)} [\mathbf{X}^{a(k)}]^T, \quad \text{where} \quad (5)$$

$$\hat{\mathbf{P}}^a = (\mathbf{I} - \mathbf{KH}) \hat{\mathbf{P}}^b \quad (6)$$

is the ensemble-based estimate of the analysis error covariance matrix.[2]

The algorithm of Whitaker and Hamill (2002) is a serial algorithm, that is, the observations are assimilated one by one, serially updating the mean analysis $\overline{\mathbf{x}}^a$ and the ensemble of analysis perturbation $\{\mathbf{X}^{a(k)}: k = 1, \ldots, K\}$. To simplify the notation, we describe the algorithm for the case when each model variable is

---

[2] The set of analysis perturbations that satisfies this condition is not uniquely defined.

directly observed by a single observation[3] and the errors in the observations of the different components of the state vector are uncorrelated. We index each observation with the index of the variable it observes, that is, the observation of $x_u$ is $y_u^o$. For the particular configuration of the observing network, the component $\mathcal{H}_u(\mathbf{x}^b)$ of the observation operator for $y_u^o$ is

$$x_u^b = \mathcal{H}_u(\mathbf{x}^b). \tag{7}$$

Each observation $y_u^o$ is assimilated by the

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{k}_u(y_u^o - \overline{x}_u^b), \tag{8}$$

equivalent of Eq. (3) for a single observation, where the vector

$$\mathbf{k}_u = \mathbf{p}_u^b(p_{uu}^b + r_{uu})^{-1} \tag{9}$$

is the Kalman-gain associated with the observation of $x_u$. In Eq. (9), the components of the vector $\mathbf{p}_u^b$ are defined by the covariance between the error $\varepsilon_u^b$ and the errors $\varepsilon_v^b(v = 1, \ldots, M)$, $p_{uu}^b$ is the variance of $\varepsilon_u^b$, and $r_{uu}$ is the observation error variance for the observation $y_u^o$. The analysis perturbations are updated by

$$\mathbf{X}^{a(k)} = \mathbf{X}^{b(k)} - X_u^{b(k)}\tilde{\mathbf{k}}_u, \quad k = 1, \ldots, K, \quad \text{where} \tag{10}$$

$$\tilde{\mathbf{k}}_u = \alpha\mathbf{k}_u, \quad \alpha = \left(1 + \sqrt{\frac{r_{uu}}{p_{uu}^b + r_{uu}}}\right)^{-1}. \tag{11}$$

### b. Covariance filtering

In the computational algorithm defined by Eqs. (8)–(11), the background error covariance enters in Eq. (9) through $\mathbf{p}_u^b$. A distance-dependent filtering, for example, localization, can be implemented by filtering the $p_{uv}^b(v = 1, \ldots, M)$ components of $\mathbf{p}_u^b$ based on the distance $|u - v|$.

A potential problem with covariance filtering is that it produces an analysis ensemble, $\{\mathbf{X}^{a(k)}: k = 1, \ldots, K\}$, that is not fully consistent with the background error covariance used in the computation of the Kalman gain: while the Kalman gain is computed based on the filtered estimate $\tilde{\mathbf{P}}^b$ [see Eq. (4)], the Kalman gain is applied to the sample covariance matrix $\hat{\mathbf{P}}^b$ to compute the ensemble perturbations, $\{\mathbf{X}^{a(k)}: k = 1, \ldots, K\}$ [by Eq. (6) in a square root filter]. While the effects of this inconsistency on the accuracy of the analysis cannot be

predicted based on strictly theoretical considerations, the fact that it exists suggests that a filtering algorithm that provides a more accurate estimate of $\mathbf{P}^b$ does not necessarily produce a more accurate analysis.

### c. Covariance inflation

In a square root EnKF scheme, such as the algorithm of Hamill and Whitaker (2005), the diagonal entries of the sample covariance matrix $\hat{\mathbf{P}}^b$ tend to underestimate the diagonal entries of $\mathbf{P}^b$ (the variance). There are different ways to account for the underestimation of the variance in EnKF (e.g., Hamill and Whitaker 2005). The common feature of these techniques is that they increase the magnitude of the background ensemble perturbations $\{\mathbf{X}^{b(k)}: k = 1, \ldots, K\}$. The most popular approach, which is also the one used in this paper, is to multiply each background perturbations by a $\rho > 1$ variance inflation factor. This approach is called multiplicative variance inflation and was introduced in Anderson and Anderson (1999). It should be noted that the multiplicative variance inflation increases not only the diagonal elements of $\hat{\mathbf{P}}^b$, but also the off-diagonal ones. Therefore, the variance inflation also inflates the covariance estimates in the $\tilde{\mathbf{P}}^b$-filtered estimate of the covariance matrix.

## 3. Experiment design

In this section, we briefly introduce the Lorenz-96 model and our approach to generate the time evolving "true" states and the simulated observations of the "true" states.

### a. The Lorenz-96 model

The governing equation of the Lorenz-96 model is the

$$\frac{dx_v(t)}{dt} = [x_{v+1}(t) - x_{v-2}(t)]x_{v-1}(t) - x_v(t) + F, \tag{12}$$

system of ordinary differential equations for $M = 40$ scalar variables $x_v$, $v = 1, \ldots, M$; where $x_{M+1}(t) = x_1(t)$, $x_{-1}(t) = x_{M-1}(t)$, $x_0(t) = x_M(t)$, and $F$ is a constant forcing term. While a partial differential equation to which Eq. (12) would be a finite-dimensional approximation is not known to exist, the variables $x_v$, $v = 1, \ldots,$ 40 are usually thought of as gridpoint values of a scalar atmospheric variable along a latitude circle. Using this analog, the time evolution of the model for $F = 8$ resembles the propagation of a wavenumber 8 dispersive wave characterized by a westward (in the direction of decreasing $v$) phase speed and an eastward (in the direction of increasing $v$) group velocity. The model is chaotic: it has 13 positive Lyapunov exponents and its Lyapunov dimension is 27.1 (Lorenz and Emanuel

---

[3] This assumption is solely made to simplify the notation and has no effect on the generality of our results.

1998). Thus, in the Lorenz-96 model, similar to the situation in the storm-track regions in an NWP model, the spatiotemporal evolution of the uncertainties is governed by unstable dispersive waves and accurate estimates of the state over an extended time period can be obtained only by the frequent assimilation of observations. In spite of its skill in mimicking an important feature of the propagation of state estimation errors in realistic models of the atmosphere, the Lorenz-96 model is a highly idealized analog of a realistic atmospheric model. Most importantly, the spatial correlations between $x_v$ at the different "grid points" does not change smoothly with distance as in a realistic model.[4] We choose the Lorenz-96 model because (i) its low dimensionality allows us to test a computationally relatively expensive nonparametric scheme to filter the ensemble-based estimate of the covariance; (ii) filtering the ensemble-based covariance by localization has a well-documented positive effect on the accuracy of the analyses for this model (Whitaker and Hamill 2002); and (iii) this model has an excellent track record in providing the initial test environment for EnKF schemes (e.g., Whitaker and Hamill 2002; Ott et al. 2004; Zhang et al. 2009b), which later proved competitive with the state-of-the-art data assimilation schemes of the operational centers.

### b. Generation of the time series of "true" states and the observations

We solve Eq. (12) using a fourth-order Runge–Kutta time integration scheme and a time step of 0.05 dimensionless time unit. This time unit is defined by the *e*-folding time of the dissipation in the model (e.g., Lorenz and Emanuel 1998). Assuming that the *e*-folding time of dissipation in the atmosphere is about 5 days, the time of 0.05 dimensionless time unit is equivalent to about 6 h in real time. We carry out all numerical experiments under the perfect model hypothesis, generating the "true" state space trajectory with a long time integration of the model. The initial condition for this integration is obtained by adding small-magnitude random noise to the unstable steady-state solution $x_u = 0$, $u = 1, \ldots, M$; then, discarding the initial transient part of the trajectory (first 1000 time steps) and defining the true states $\mathbf{x}^t(t)$ with the states along the remaining

portion of the trajectory. Simulated observations are generated for each time step by adding normally distributed random noise with expectation zero and standard deviation 1 to each variable $x_u$, $u = 1, \ldots, M$. That is, the observation error covariance matrix $\mathbf{R}$ is the identity $\mathbf{I}$.[5] We estimate the state at each observation time by assimilating the simulated observations with the algorithm described by Eqs. (8)–(11).

## 4. Statistical methodology

We now introduce a nonparametric statistical method to estimate the covariance matrix, which is computationally feasible but alleviates some of the potential problems with the sample and the localized sample covariance matrices. We first review some terminologies in spatial statistics, then introduce the nonparametric scheme, gradually relaxing the assumptions we make about the covariances. We illustrate our main points with quantitative results for the Lorenz-96 model. In these calculations, we use a sample covariance matrix based on a $K = 5000$ member ensemble, which we denote by $\hat{\mathbf{P}}^b_{5000}$, as a proxy for the true covariance matrix $\mathbf{P}^b$. We obtain $\hat{\mathbf{P}}^b_{5000}$ by running the EnKF data assimilation system for 200 analysis steps using a set of initial ensemble members, which was generated by adding Gaussian random noise, with mean zero and standard deviation $F/10$, to the unstable steady-state solution $x_u = 0$, $u = 1, \ldots, M$; then, choosing $\hat{\mathbf{P}}^b$ from the last analysis step to be $\hat{\mathbf{P}}^b_{5000}$. For this calculation, we use a weak covariance inflation factor, $\rho = 1.005$, because the primary role of variance inflation for such a large ensemble under the perfect model scenario is to compensate for the effects of nonlinearities in the model dynamics.

### a. Terminology

We introduce the notation $\varepsilon^b = (\varepsilon^b_1, \varepsilon^b_2, \ldots, \varepsilon^b_M)$ for the random vector, whose components are the background errors for the different state variables. In spatial statistics, we say that the process is *nonstationary* in space, when $p^b_{uv} = \text{Cov}\{\varepsilon^b_u, \varepsilon^b_v\}$ depends on the two locations $u$ and $v$. When the mean is constant and the covariance depends on the locations only through the difference of the two locations, that is, $C_1(u - v) = \text{Cov}\{\varepsilon^b_u, \varepsilon^b_v\}$ for an appropriate (i.e., positive definite) function $C_1$, then we call the process *stationary*.

---

[4] This shortcoming of the model was successfully corrected by a modification of Eq. (12) in Lorenz (2005). Unfortunately, this improvement of the model was achieved at the expense of increasing the number of variables, which limits the appeal of the improved model as a low computational cost alternative to a more realistic atmospheric model.

[5] We also carried out experiments in which we observed every third location, but because the results were qualitatively similar to those for observing all locations, we do not report the results for that setting.

Finally, when the mean is constant and the covariance depends on the locations only through the distance between the two locations, that is, $C_2(|u - v|) = \text{Cov}\{\varepsilon_u^b, \varepsilon_v^b\}$ for an appropriate function $C_2$, we call the process *isotropic*.

One option to estimate the covariance function is to use a *parametric* covariance model. There are various parametric covariance models that are isotropic, for example, the exponential function $C_2(x) = \alpha e^{-x/\beta}$, the Gaussian function $C_2(x) = \alpha e^{-(x/\beta)^2}$, or the Matérn function $C_2(x) = \alpha(x/\beta)^\nu \mathcal{K}_\nu(x/\beta)$, where $\alpha$, $\beta$, $\nu$ are positive parameters and $\mathcal{K}_\nu$ is the modified Bessel function ($x > 0$). Parametric covariance models are also available for certain nonstationary processes (Paciorek and Schervish 2006; Jun and Stein 2007). In some situations, we may model a nonstationary process as a sum of several independent, locally stationary processes with simple parametric stationary (or isotropic) covariance functions (Fuentes 2001). Parametric methods, however, in general, are not as flexible as *nonparametric* methods, which do not make presumptions about the general shape of the covariance function. For instance, the nonstationarity of the background covariance structure for the Lorenz-96 model is complex and there is no obvious flexible parametric covariance model to fit the covariance structure. This complexity is illustrated with Fig. 1, which shows the sample covariance function for various ensemble sizes in the Lorenz-96 model: the sample covariance structure is not symmetric around the center and has a strong dependence on the location. These are clear signs of strong anisotropy and nonstationarity of the background error. This result motivates us to search for an appropriate nonparametric method.

We use a *kernel smoothing* approach as our nonparametric estimation method, which requires the selection of a kernel function and a bandwidth. A kernel is a nonnegative function, which is symmetric around zero, and decreases monotonically as $|u - v|$ goes to $\infty$; for example, $G(u - v) = \exp[-(u - v)^2]$, is a Gaussian kernel function. There are various statistical methods to choose optimal kernels and bandwidths, and in many applications the effect of the selection of the kernel is insignificant compared to the selection of the bandwidth.[6] In what follows, we first develop a nonparametric method to estimate the covariance function for the stationary case and then we extend the method to the nonstationary case.

### b. Stationary estimate

Suppose a spatial process is stationary. Then, we estimate $p_{uv}^b = C_1(u - v)$ by the kernel estimator given in Hall and Patil (1994):

$$\tilde{p}_{uv}^b = \frac{\dfrac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{M}\sum_{j=1}^{M} G\left[\dfrac{|(u - v) - (i - j)|}{h}\right][X_i^{b(k)} - \overline{X}^{b(k)}][X_j^{b(k)} - \overline{X}^{b(k)}]}{\sum_{i=1}^{M}\sum_{j=1}^{M} G\left[\dfrac{|(u - v) - (i - j)|}{h}\right]}. \tag{13}$$

Here, $h$ is a bandwidth, $G(\cdot)$ is a kernel function, and $\overline{X}^{b(k)}$ is the spatial average, $\overline{X}^{b(k)} = (1/M)\sum_{i=1}^{M} X_i^{b(k)}$, of the $k$th ensemble perturbation $\mathbf{X}^{b(x)}$. The estimator in Eq. (13) is actually isotropic in the sense that $\tilde{p}_{uv}^b = \tilde{p}_{vu}^b$ for all $u$ and $v$. As Hall and Patil (1994) note, Eq. (13) may not give a positive definite function and they suggest a modified version of Eq. (13) to ensure positive definiteness and to achieve nice mathematical properties of the estimator. Here we do not worry about the positive definiteness of the estimate provided by Eq. (13), as it is not the final scheme we intend to use. Equation (13) can be easily adapted for $d$-dimensional case ($d > 1$) by simply letting $i$ and $j$ inside of the parentheses be the coordinates of the $i$th and $j$th locations on the $d$-dimensional domain.

Figure 2 shows the estimated covariance function obtained with Eq. (13) for $u = 20$ and $v = 1, \ldots, 40$.

Because of the stationarity assumption, the shape of the estimated covariance function is the same for all $u = 1, \ldots, 40$. Moreover, the estimated covariance function is symmetric around the center. Each row shows the result for a given bandwidth. The curves in Fig. 2 do not match the sample covariances in Fig. 1, further suggesting that the assumption of stationarity for the background error in the Lorenz-96 model is not appropriate.

### c. Nonstationary estimate

We now extend Eq. (13) for the nonstationary case, estimating $p_{uv}^b$ with

---

[6] For more details on kernel estimation methods, see Fan and Yao (2003).

## Sample covariance between u=3 and v



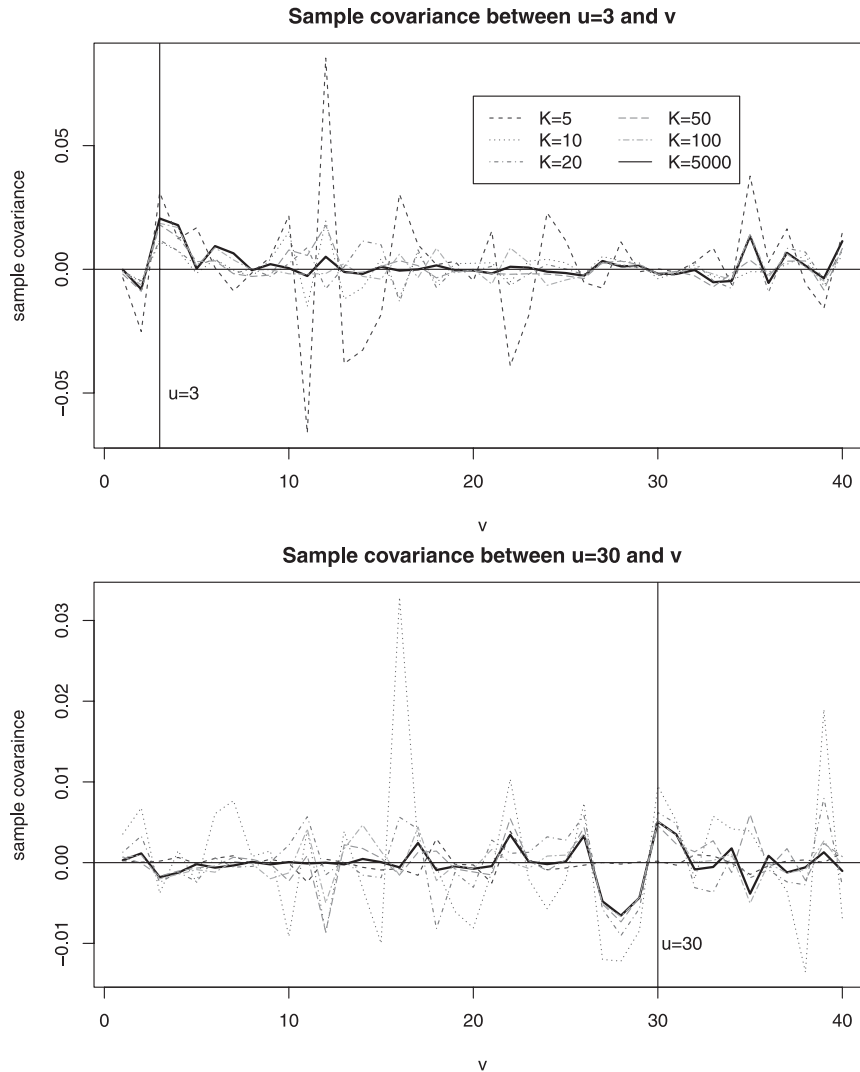## Sample covariance between u=30 and v



FIG. 1. Sample covariances between the two locations $u$ and $v$ using $K$ ensemble members for $u = 3, 30$ and $v = 1, \ldots, 40$.

$$\tilde{p}_{uv}^b = \frac{\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{M}\sum_{j=1}^{M}G_i(u;h)G_j(v;h)[X_i^{b(k)} - \overline{X}^{b(k)}][X_j^{b(k)} - \overline{X}^{b(k)}]}{\sum_{i=1}^{M}\sum_{j=1}^{M}G_i(u;h)G_j(v;h)}, \tag{14}$$

where $G_i(u; h)$ and $G_j(v; h)$ are kernel functions with a fixed bandwidth $h$. See the appendix for the sketch of proof for the positive definiteness of Eq. (14). Since the kernel functions depend on the two locations, $u$ and $v$, Eq. (14) provides a flow-dependent estimate of the covariance. We may, for instance, use the Gaussian kernel functions $G_i(u;h) = \exp[-(|i-u|/h)^2]$ and $G_j(v;h) = \exp[-(|j-v|/h)^2]$. We note that the computational cost of Eq. (14) can be greatly reduced by using a compactly supported kernel for $G$, since it reduces the number of terms in the double summation over $i$ and $j$. Figure 3 shows the estimate of the covariance function $\tilde{p}_{uv}^b$ using
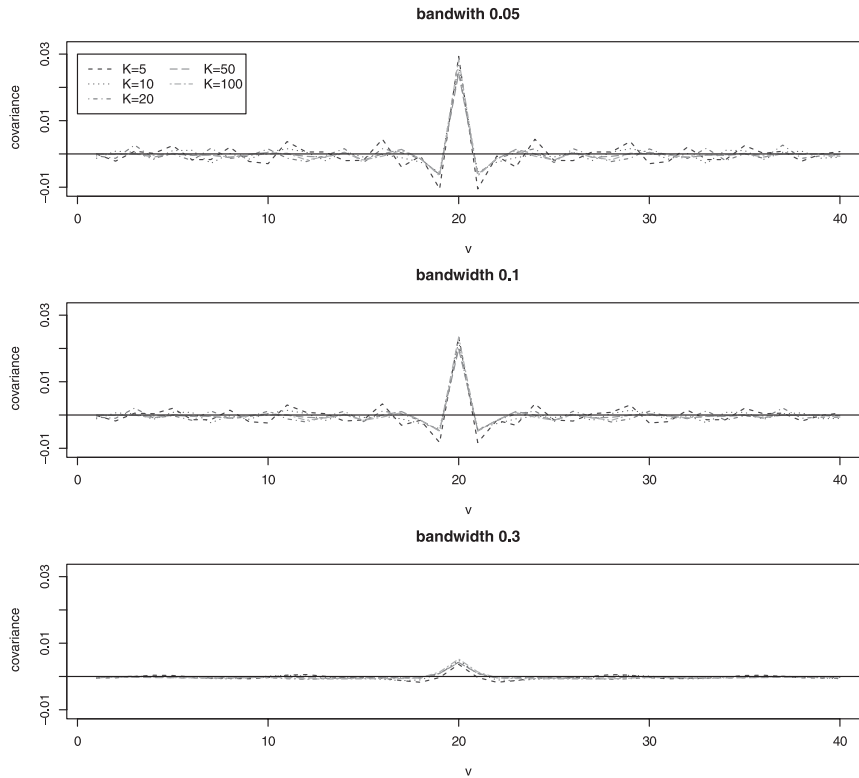
FIG. 2. Fitted covariance between the locations $u = 20$ and $v = 1, \ldots, 40$ using the nonparametric approach under stationarity assumption.

Eq. (14). The covariance curve from $\hat{\mathbf{P}}^b_{5000}$ is displayed (black curve) for comparison. The estimated covariance function changes with the location $u$ and provides a better fit to the large scale structure of the sample covariance (Fig. 1) than the stationary estimate (Fig. 2).

#### d. Adaptive bandwidth

As seen before, both the stationary and nonstationary estimates of the covariance require the selection of a bandwidth $h$: a larger value of $h$ results in a smoother estimate of the covariance, because it involves a stronger averaging of the sample errors over the neighboring locations. On the one hand, when $h$ goes to zero, the estimate provided by Eq. (14) becomes similar to the estimate provided by the sample covariance estimate, except that in Eq. (14) we divide the sums by $K$ instead of $K - 1$ and filter the background error perturbation $\mathbf{X}^{b(k)}$ with its mean $\overline{X}^{b(k)}$ over the locations. On the other hand, when $h$ is large, the estimated covariance is smooth and, in the extreme case, it becomes constant. This is because when $h$ is large, the kernel values in Eqs. (13) and (14) are almost constant for all $i, j, u,$ and $v$. See the bottom row in Fig. 3 for an example.

Because we expect the signal-to-noise ratio to be lower at larger distances $|u - v|$, we introduce a bandwidth $h = h(|u - v|)$ that increases with the distance.[7] In particular, to make $h$ smoothly varying with the distance, we let $h = h_1 \exp\{(|u - v|/h_2)^2\}$. Thus, we make the bandwidth adaptive at the price of replacing the single bandwidth parameter $h$ with a pair of parameters, $h_1$ and $h_2$. Figure 4 shows examples of the dependence of the adaptive bandwidth on the two parameters (Fig. 4, top panel) and a comparison of the corresponding covariance estimates with the sample covariance using 20 ensemble members (Fig. 4, bottom panel). The two figures use the same color scale. For the bottom panel, we also display sample covariances using 20 and 5000 ensemble members for comparison. When $h_1 = 10^{-7}$ and $h_2 = 1$, the adaptive bandwidth is about $10^{-7}$ at all distances; thus, the nonparametric estimate becomes similar to the sample covariance. The results for $h_1 = 0.01$ and $h_2 = 2$ are fairly similar. For the other choices of $h_1$ and $h_2$ the bandwidth increases with the distance and the

---

[7] Localization of the sample covariance is equivalent to using an $h$, which is nearly zero within the localization length and large outside the localization length.
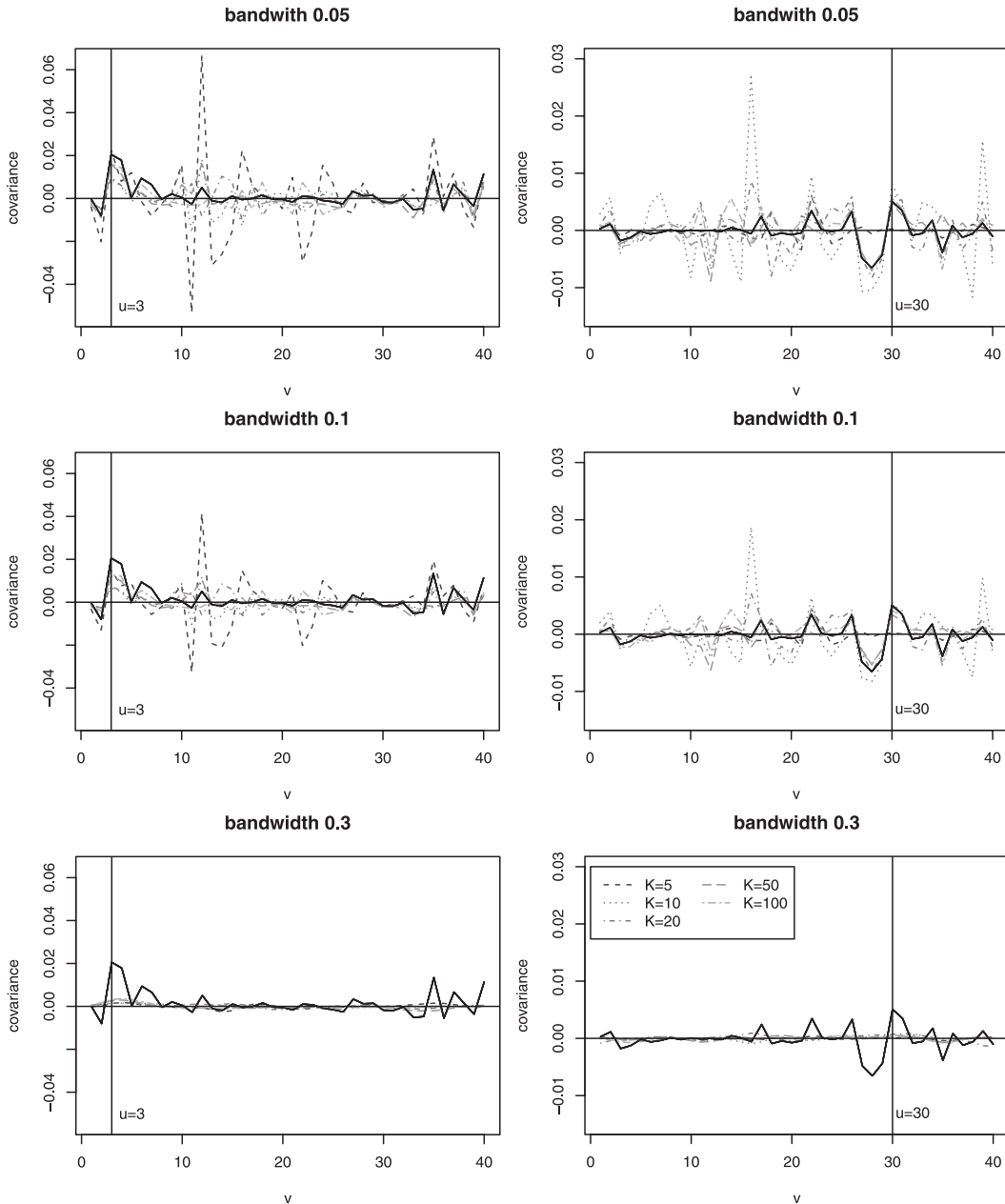
FIG. 3. Fitted covariance between the locations $u$ and $v$ using the nonparametric approach for the nonstationary case. The two chosen $u$ values are the same as in Fig. 1. The black curve gives $\hat{\mathbf{P}}_{5000}^b$.

corresponding fitted covariance values are close to zero after a short distance (Fig. 4).

## 5. Quantitative comparison

Next, we compare the covariance estimates provided by the nonparametric scheme with those based on the sample and the localized sample covariance at a fixed time step. For this comparison, we again use $\hat{\mathbf{P}}_{5000}^b$ as

a proxy for the true $\mathbf{P}^b$. We measure the accuracy of all our estimates to the true $\mathbf{P}^b$ by computing the Frobenius norm of the difference between the estimates and $\hat{\mathbf{P}}_b^{5000}$[8]. For a given ensemble size $K$ we randomly select 200 sets of $K$ ensemble members from the 5000-member

---

[8] For a matrix $\mathbf{A} = \{a_{ij}\}$, the Frobenius norm is defined as $\|\mathbf{A}\|_F = \sqrt{\sum_i \sum_j a_{ij}^2}$.
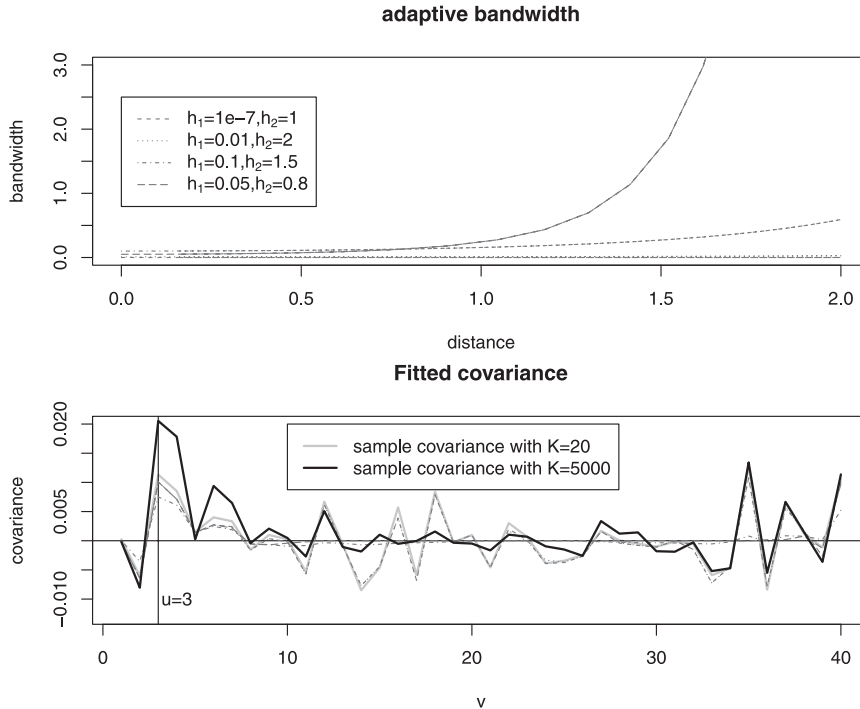
FIG. 4. A few adaptive bandwidth examples and the corresponding covariance fits with 20 ensemble members. (top) The shape of four adaptive bandwidths against the distance for the four combinations of $h_1$ and $h_2$ values. (bottom) The corresponding fitted covariance values between the locations $u = 3$ and $v = 1, \ldots, 40$ with 20 ensemble members. Same scale is used for the two figures for the adaptive bandwidth case. For the bottom panel, we have two additional curves displaying sample covariances with 20 and 5000 ensemble members for comparison.

ensemble from which $\hat{\mathbf{P}}^b_{5000}$ is derived. We compute the Frobenius norm of the difference between the estimates and $\hat{\mathbf{P}}^b_{5000}$ for each of the 200 sets of ensembles and characterize the accuracy of the estimates from the distribution of the 200 values of the difference.

Results are shown for various values of the parameters $h_1$ and $h_2$ in Fig. 5. The ''optimal'' pair of values is $h_1 = 0.05$ and $h_2 = 0.8$ in terms of the median, and $h_1 = 0.1$ and $h_2 = 1.5$ in terms of the mean. For localization, we use the fifth-order kernel function defined by Eq. (4.10) of Gaspari and Cohn (1999), which is the most widely used localization function in the EnKF literature. This function has a single parameter $c$, which controls the localization length: while the function formally becomes zero at distance $2c$, the filtered covariances are already nearly zero at distance $c$. Figure 6 shows the median and mean of the Frobenius norm for different values of $c$. For a 20-member ensemble, the ''optimal'' localization length, with respect to the Frobenius norm, is smaller than $c = 24$, the value that was reported to be optimal with respect to the analysis accuracy for a 10-member ensemble in Whitaker and Hamill (2002). Also, there is a noticeable difference between the ''optimal'' localization length

with respect to the median, $c = 7.5$, and to the mean, $c = 5$.

The covariance estimates provided by the nonparametric scheme and the localization are compared for the ''optimal'' values of the parameters of the two schemes (Fig. 7). Overall, with respect to the mean, the nonparametric scheme provides the most accurate estimate except when the ensemble size is 5. The advantage of this scheme over the sample covariance and the localization with $c = 24$ is particularly large for the small ensembles ($K \ll 50$). The advantage of the nonparametric scheme over localization with $c = 5$ is much more modest. In addition, with respect to the median, the localization with $c = 5$ outperforms the nonparametric scheme when $K < 50$. The system used here gives a highly nonstationary covariance structure for the background errors. The nonparametric scheme would be superior if the true covariance structure was stationary or isotropic [Eq. (13)].

## 6. Analysis experiments with filtered covariances

In section 5 we compared the accuracy of the filtered covariance estimates for a single analysis time. In this
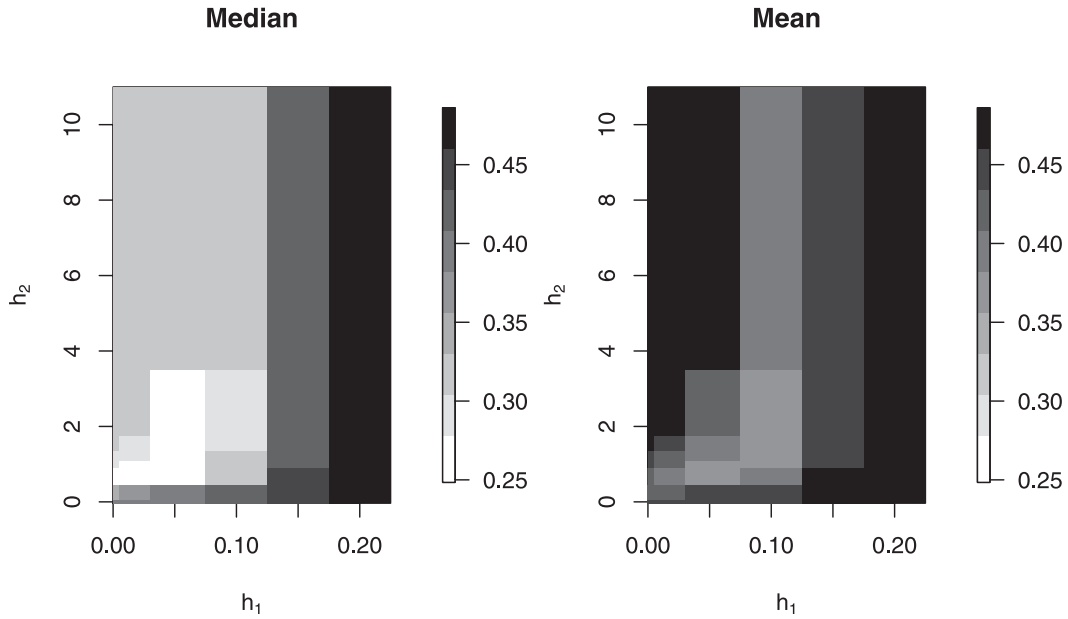
FIG. 5. Frobenius norm of the (left) median and (right) mean of the differences between the fitted covariances from the nonparametric scheme and $\hat{\mathbf{P}}^{b}_{5000}$ for various $h_1$ and $h_2$ values. The median and mean are computed from 200 independent trials.

section, we carry out analysis experiments with a $K = 20$ member ensemble using the filtered estimates of the background covariance in the computation of the Kalman gain. The filter is run for 1500 time steps and statistics are computed based on the last 1000 time steps.

*a. Verification methods*

We measure the accuracy of the analysis with

$$\Delta = \frac{1}{1000} \sum_{n=1}^{1000} \left\{ \frac{1}{M} \sum_{m=1}^{M} [\varepsilon_m^a(t_n)]^2 \right\}^{1/2}, \qquad (15)$$

the time-mean of the root-mean-square error over all model variables. Here, $\varepsilon_m^a(t_n)$ is the analysis error at grid point $m$ at time $t_n$, where $\varepsilon_m^a(t_n) = \overline{x}_m^a(t_n) - x_m^t(t_n)$. Using the mean of a time series of root-mean-square
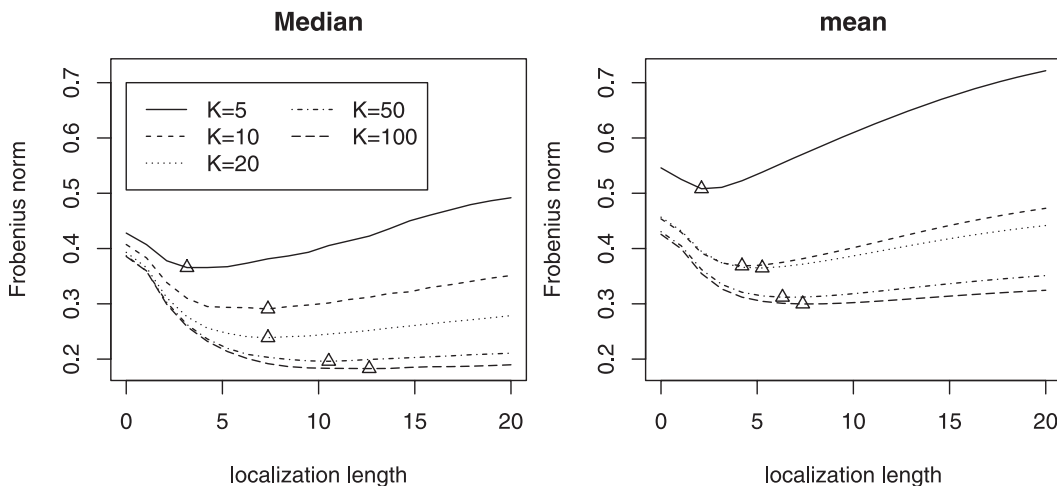


FIG. 6. Frobenius norm of the differences between the fitted covariances from localization and $\hat{\mathbf{P}}^{b}_{5000}$ for various localization lengths. We use the same ensemble members for each trial at a given ensemble size as in Fig. 5. (left) The median from the 200 independent trials; (right) the mean. The minimum values for each ensemble size are given by the triangles.
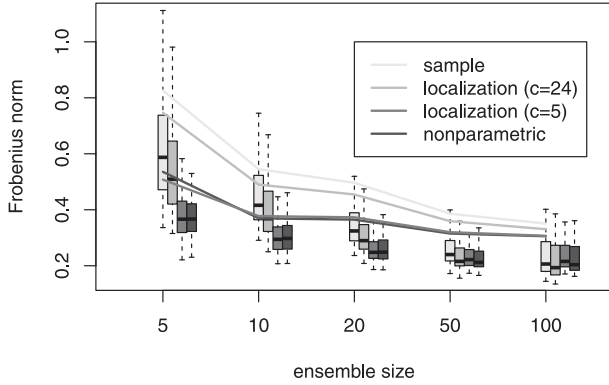
FIG. 7. Frobenius norm of the differences between the fitted covariances and $\hat{\mathbf{P}}^b_{5000}$ for various ensemble sizes. We use the same ensemble members for each trial at a given ensemble size as in Fig. 5. The box plots indicate the distribution of differences found in the 200 test trials in the following way: The lower and upper bounds of the box respectively give the 25th and 75th percentile of differences (as measured by the Frobenius norm) found in the 200 trials. The thick line going across the interior of the box gives the median of the differences found in the 200 trials. The whiskers or thin lines depend on the interquartile range (IQR) that is precisely equal to the vertical length of the box. The whiskers extend to the most extreme values (differences obtained in the 200 trials) which are no more than 1.5 IQR from the box. The solid lines connect the means of the differences in the 200 trials at each ensemble size.

errors is a standard practice in numerical weather prediction to measure the quality of an analysis/forecast system: the root-mean-square error computed over all grid points characterizes the accuracy of the state estimate at a given time by a single number and the time mean over an extended time period can be easily computed retrospectively based on the archived root-mean-square values. (Computing the root-mean-square over time and space would require processing all grid-point values at all times at the end of the verification period.)

To characterize the estimates of the background error, we examine the estimates of $\mathbf{p}^b_u$ for a fixed location $u$. Since $\mathbf{p}^b_u$ is a probabilistic variable, its estimates cannot be evaluated for a single analysis time $t_n$, for which only a single realization, $\varepsilon^b(t_n) = \overline{\mathbf{x}}^b(t_n) - \mathbf{x}^t(t_n)$, of the background error is available. {Here we use the notation $\varepsilon^b(t_n) = [\varepsilon^b_1(t_n), \varepsilon^b_2(t_n), \ldots, \varepsilon^b_M(t_n)]$}. Thus, we use the

$$\mathbf{P}^{\mathrm{btc}} = \frac{1}{1000} \sum_{n=1}^{1000} \varepsilon^b(t_n)[\varepsilon^b(t_n)]^{\mathrm{T}} \qquad (16)$$

estimate of the "climatological" value of the true background error covariance matrix: $E\{\varepsilon^b(t_n)[\varepsilon^b(t_n)]^{\mathrm{T}}\}$. We compare $\mathbf{P}^{\mathrm{btc}}$ with

$$\hat{\mathbf{P}}^{\mathrm{bc}} = \frac{1}{1000} \sum_{n=1}^{1000} \hat{\mathbf{P}}^b(t_n) = \frac{1}{1000} \sum_{n=1}^{1000} \frac{1}{K-1} \sum_{k=1}^{K-1} \mathbf{X}^{b(k)}[\mathbf{X}^{b(k)}]^{\mathrm{T}}$$

$$= \frac{1}{K-1} \sum_{k=1}^{K-1} \frac{1}{1000} \sum_{n=1}^{1000} \mathbf{X}^{b(k)}[\mathbf{X}^{b(k)}]^{\mathrm{T}}, \qquad (17)$$

the time mean of the sample covariance matrix $\hat{\mathbf{P}}^b(t_n)$. For a properly constructed ensemble, $\varepsilon^b(t_n)$ and the ensemble perturbations $\{\mathbf{X}^{b(k)}: k = 1, \ldots, K\}$ are from the same distribution; thus

$$\hat{\mathbf{P}}^{\mathrm{bc}} \approx \frac{1}{K-1} \sum_{k=1}^{K-1} \mathbf{P}^{\mathrm{btc}} \approx \mathbf{P}^{\mathrm{btc}}, \quad \text{and} \qquad (18)$$

$$\hat{\mathbf{p}}^{\mathrm{bc}}_u \approx \mathbf{p}^{\mathrm{btc}}_u, \qquad (19)$$

where $\hat{\mathbf{p}}^{\mathrm{bc}}_u$ and $\mathbf{p}^{\mathrm{btc}}_u$ are the $u$th column of $\hat{\mathbf{P}}^{\mathrm{bc}}_u$ and $\mathbf{P}^{\mathrm{btc}}_u$, respectively. Similar arguments can be made to show that the filtered estimate should also satisfy

$$\tilde{\mathbf{p}}^{\mathrm{bc}}_u \approx \mathbf{p}^{\mathrm{btc}}_u, \qquad (20)$$

where $\tilde{\mathbf{p}}^{\mathrm{bc}}_u$ is the $u$th column of $\tilde{\mathbf{P}}^{\mathrm{bc}}_u$.

We emphasize that Eqs. (19) and (20) do not provide information about the accuracy of the estimates $\hat{\mathbf{p}}^b(t_n)$ and $\tilde{\mathbf{p}}^b(t_n)$ at a single analysis time $t_n$; instead, these relations can be used to verify the consistency between a time series of the estimated background error covariance matrix and a time series of the true background error. Such consistency is a necessary, but not sufficient, condition for $\hat{\mathbf{P}}^b(t_n)$ or $\tilde{\mathbf{P}}^b(t_n)$ to be an accurate representation of $\mathbf{P}^b(t_n)$ at the different analysis times.

### b. Dependence of the results on the filtering parameters

To establish a baseline of the analysis error, against which we can measure the effectiveness of covariance filtering, we first carry out an analysis experiment using the unfiltered sample covariance matrix $\hat{\mathbf{P}}$. We find that $\Delta$ takes its minimum value of 0.23 when $\rho = 1.06$; using smaller values of $\rho$ makes the filter unstable, while increasing $\rho$ leads to increasingly larger values of $\Delta$.

First, we filter the sample covariance estimates using localization. Whitaker and Hamill (2002) studied, in detail, the sensitivity of the analysis error to the localization parameter $c$ and the variance inflation coefficient $\rho$ for $K = 10$. They found that the root-mean-square error in the analysis mean $\overline{\mathbf{x}}^a$ took its smallest value of 0.2 for $c = 24$ and $\rho = 1.03$.[9] Using the same values of

--------

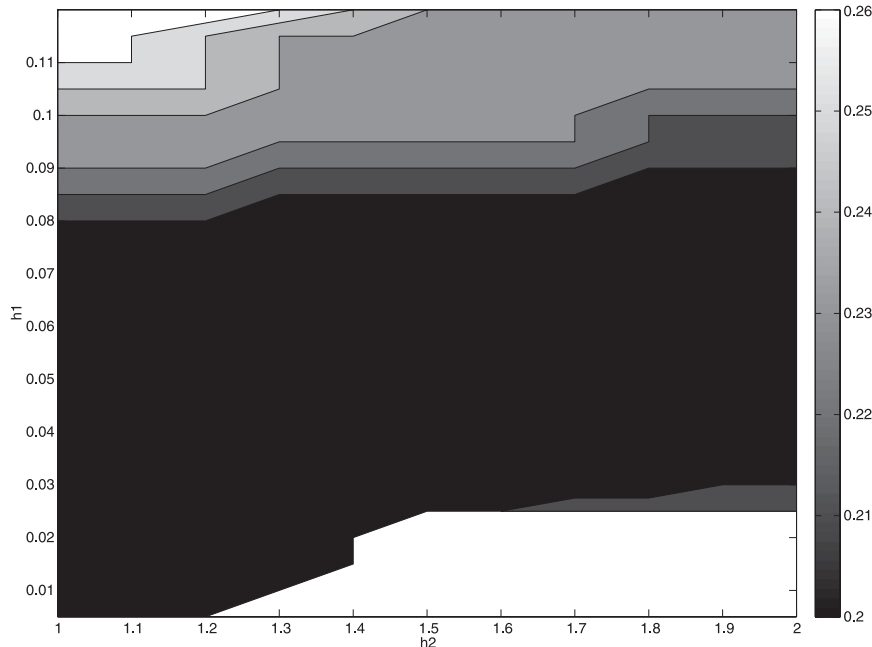[9] The correct figure to support their result can be found in the corrigendum of Whitaker and Hamill (2002).

FIG. 8. The grayscale shades show the value of $\Delta$ for $\rho = 1.025$ as a function of the parameters $h_1$ and $h_2$.

$c$ and $\rho$ for $K = 20$, we obtain a root-mean-square error of $\Delta = 0.19$. This value indicates an error reduction of 0.04 (about 17%) compared to the case when no covariance filtering is used. For $c = 5$, the value that we found optimal for a single analysis time in section 5, the minimum value of $\Delta$ is 0.22, which can be obtained using $\rho > 1.025$. That is, the performance of the EnKF for $c = 5$ is clearly inferior to that for $c = 24$, despite our earlier result that $c = 5$ provides a more accurate estimate at a single analysis time.

Finally, we investigate the sensitivity of the analysis error to the parameters $h_1$ and $h_2$ using the nonparametric scheme with adaptive bandwidth to filter the covariance. We show results for $\rho = 1.025$, the value of the variance inflation we found to provide the smallest value of $\Delta$. The results are summarized in Fig. 8. In this figure, the white area indicates the parameter range where the filter fails, indicated by a value of $\Delta$, which is larger than one, the root-mean-square of the observation error. The typical value of $\Delta$ in this region is between 3.0 and 4.0, which is similar to the standard deviation of the temporal changes in the model variables. An interesting feature in Fig. 8 is the sharp boundary between the parameter ranges where the filter fails and where the analysis error is the smallest. The parameter range where the performance of the EnKF is nearly optimal (marked by blue shades) is wide: a large value of $h_2$ cannot be used with a small value of $h_1$, but once $h_1$ is larger than about 0.03, the

analysis error becomes insensitive to the choice of $h_2$. In essence, an increase of the bandwidth with distance is important only when the bandwidth at zero distance is small. In the blue region, the root-mean-square error is about $\Delta = 0.2$, lower than that for localization with $c = 5$, but slightly worse than that for localization with $c = 24$.

## c. Consistency between the estimated and the true errors

Figure 9 illustrates the level of consistency between the sample covariance and the covariance for the true background error. While the general shape of the covariance function is captured well by the sample covariance, at short distances, $|u - v| \leq 2$ ($18 \leq v \leq 22$), the absolute value of the covariance is somewhat overestimated. This overestimation can be easily corrected by reducing the variance inflation, but reducing the variance inflation quickly leads to a loss of the stability of the EnKF. At longer distances, beyond $|u - v| \geq 5$, the time mean of the sample covariance is about zero, while the time mean of the true error is small, but not zero at most distances. We note that the near-zero values at the larger distances for the sample covariance are due to the filtering effect of time averaging, as we observe relatively large instantaneous estimates of the covariance (results not shown) at those distances. This result suggests that the relatively large instantaneous values at the larger distances are dominated by statistical fluctuations.
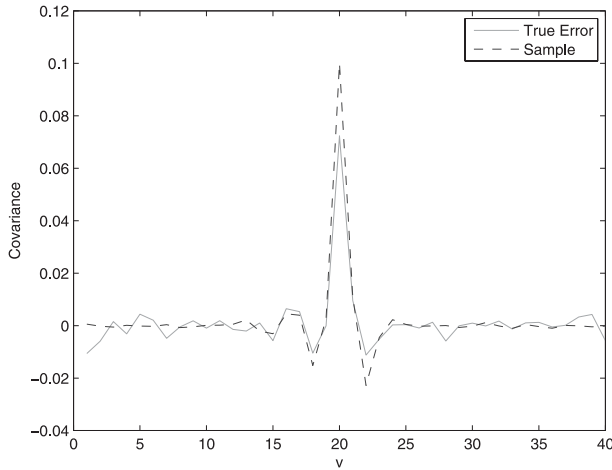
FIG. 9. The components $\hat{p}_{uv}^{bc}, v = 1, \ldots, 40$, of $\hat{\mathbf{p}}_u^{bc}$ (gray dashed) and $\hat{p}_{uv}^{btc}$, $v = 1, \ldots, 40$ of $\mathbf{p}_u^{btc}$ (light gray solid) for $u = 20$.
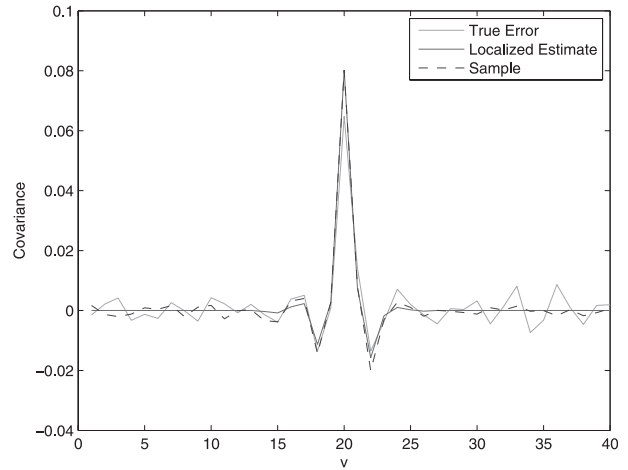


FIG. 10. The components $\hat{p}_{uv}^{bc}$, $v = 1, \ldots, 40$ of $\hat{\mathbf{p}}_u^{bc}$ (gray dashed) and $p_{uv}^{btc}, v = 1, \ldots, 40$, of $\mathbf{p}_u^{btc}$ (light gray solid) and $\tilde{p}_{uv}^{bc}, v = 1, \ldots, 40$, of $\tilde{\mathbf{p}}_u^{bc}$ for localization with $c = 5$ for $u = 20$.

We show results on the consistency of the estimates of the covariance for $c = 5$ (Fig. 10) and $c = 24$ (Fig. 11). In these figures, both the sample and the localized estimates slightly overestimate the variance. As in the case of the results shown in Fig. 9, this overestimation can be corrected by reducing the variance inflation, but reducing the variance inflation makes the filter unstable. At distances $|u - v| \leq 2$, the consistency is clearly better for $c = 24$ than for $c = 5$, as for the latter choice of $c$, the magnitude of the negative covariance at $|u - v| = 2$ is underestimated. The difference in the accuracy of the estimates at $|u - v| = 2$ may explain the better performance of the EnKF for $c = 24$ than for $c = 5$. For $c = 5$, at distances $|u - v| > 10$, the filtered estimate is zero, while the sample covariance shows some distance-dependent variation. For $c = 24$, the sample and the filtered covariance shows a high level of consistency with each other. This is an indication that the localized covariance with $c = 24$ leads to a better consistency between the localized covariance estimate and the ensemble perturbations.

We show results for two different choices of the bandwidth parameters of the nonparametric estimation scheme: $h_1 = 0.05, h_2 = 1.5$ (Fig. 12) and $h_1 = 0.11$, $h_2 = 1.5$ (Fig. 13). A common feature of these figures is that the consistency between the sample and the filtered estimates is lower than for the localization-based filtering for $c = 24$. (We recall that localization with $c = 24$ provides the most accurate analysis among all filtering schemes tested here.) Thus, we conclude that the scheme that provides the most accurate analysis, on average, is the scheme for which the sample and the filtered estimates are most consistent with each other (localization with $c = 24$). This is also the scheme that provides the best consistency with the

true errors for short distances. The results of this section suggest that a better metric of covariance filtering skill would be one that combined a measure of closeness to the sample covariance matrix for a very large ensemble with a measure of similarity between the climatological averages of the filtered and sample covariance.

## 7. Conclusions

In this paper, we investigated the effects of filtering the sample covariances used in the computation of the Kalman gain in an EnKF on the accuracy of the estimates of the background error covariance. We considered two particular approaches to filter the sample
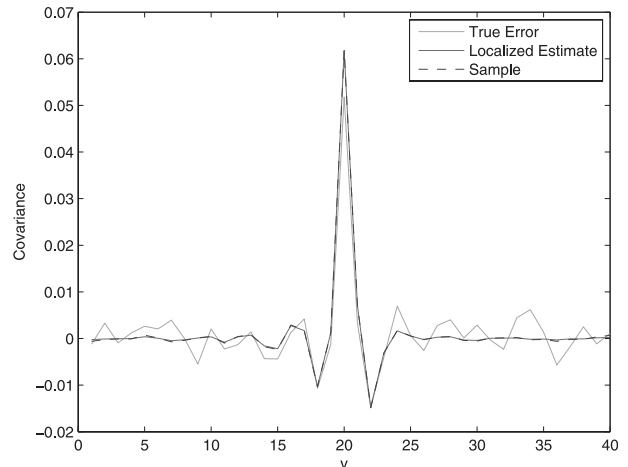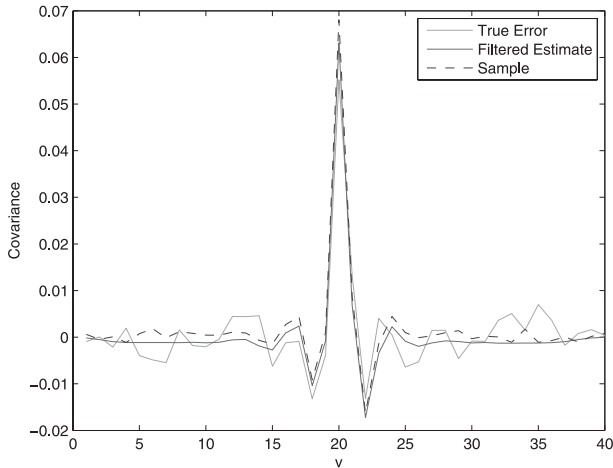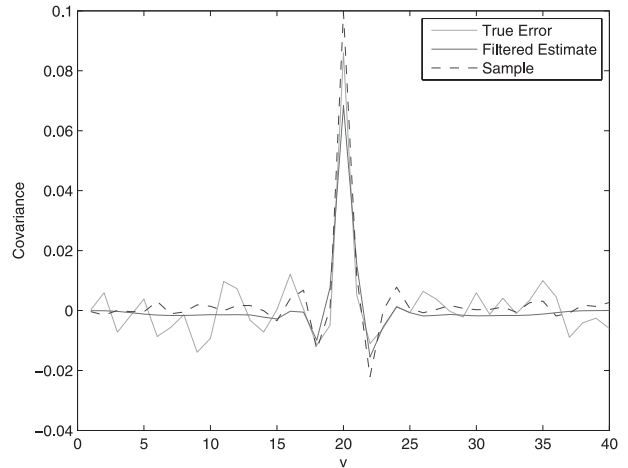


FIG. 11. The components $\hat{p}_{uv}^{bc}$, $v = 1, \ldots, 40$, of $\hat{\mathbf{p}}_u^{bc}$ (gray dashed) and $p_{uv}^{btc}, v = 1, \ldots, 40$ of $\mathbf{p}_u^{btc}$ (light gray solid) and $\tilde{p}_{uv}^{bc}, v = 1, \ldots, 40$ of $\tilde{\mathbf{p}}_u^{bc}$ for localization with $c = 24$ for $u = 20$.

FIG. 12. The components $\hat{p}^{bc}_{uv}$, $v = 1, \ldots, 40$ of $\hat{\mathbf{p}}^{bc}_u$ (gray dashed) and $p^{btc}_{uv}$, $v = 1, \ldots, 40$ of $\mathbf{p}^{btc}_u$ (light gray solid) and $\tilde{p}^{bc}_{uv}$, $v = 1, \ldots, 40$ of $\tilde{\mathbf{p}}^{bc}_u$ for the nonparametric scheme with $h_1 = 0.05$ and $h_2 = 1.5$ for $u = 20$.

FIG. 13. The components $\hat{p}^{bc}_{uv}$, $v = 1, \ldots, 40$ of $\hat{\mathbf{p}}^{bc}_u$ (gray dashed) and $p^{btc}_{uv}$, $v = 1, \ldots, 40$ of $\mathbf{p}^{btc}_u$ (light gray solid) and $\tilde{p}^{bc}_{uv}$, $v = 1, \ldots, 40$ of $\tilde{\mathbf{p}}^{bc}_u$ for the nonparametric scheme with $h_1 = 0.11$ and $h_2 = 1.5$ for $u = 20$.

covariance: the more traditional approach of covariance localization and a new kernel smoothing method with variable bandwidth to obtain nonparametric estimates of the covariances.

We found that for a single analysis time, the nonparametric scheme provided the overall most accurate estimate of the covariances and that localization provided more accurate estimates with short localization length than with long localization length. We also found, however, that when the analysis is cycled, localization with long localization length provided more accurate analysis in the root-mean-square sense than the nonparametric scheme or localization with short localization length. We explained this result with the better consistency between the covariance estimates, the true background error covariance, and the ensemble perturbations that represent the background uncertainty. Our results suggest that preserving such consistency is important.

Finally, we note that the Lorenz-96 model, with its rapidly changing background covariance between locations, poses a considerable challenge to the covariance estimation methods. It is plausible that some of our findings would not hold for a more realistic model in which the background error covariance changes in a smoother way. In such a model, distinguishing between the spatially rapidly changing sampling noise and the smoother changing true covariance should be easier, which we expect to benefit the new kernel smoothing method more than localization.

## APPENDIX

### Proof of the Positive Definiteness of Eq. (14)

We give a brief sketch of the proof for the positive definiteness of the covariance function in Eq. (14).

It is sufficient to show the positive definiteness of

$$\sum_{i=1}^{M}\sum_{j=1}^{M} G_i(u;j)G_j(v;h)[X_i^{b(1)} - \overline{X}^{b(1)}][X_j^{b(1)} - \overline{X}^{b(1)}] \left/ \sum_{i=1}^{M}\sum_{j=1}^{M} G_i(u;j)G_j(v;h). \right.$$

The idea is similar to the proof given by Eq. (4) of Paciorek and Schervish (2006).

For any $a_1, \ldots, a_n \in \mathbb{R}$ and any $n \in \mathbb{N}$,

$$\sum_{p=1}^{n} \sum_{q=1}^{n} a_p a_q \sum_{i=1}^{M} \sum_{j=1}^{M} G_i(u_p; h) G_j(u_q; h) [X_i^{b(1)} - \overline{X}^{b(1)}][X_j^{b(1)} - \overline{X}^{b(1)}] \bigg/ \sum_{i=1}^{M} \sum_{j=1}^{M} G_i(u_p; h) G_j(u_q; h)$$

$$= \sum_{p=1}^{n} a_p \left\{ \sum_{i=1}^{M} G_i(u_p; h) [X_i^{b(1)} - \overline{X}^{b(1)}] \bigg/ \sum_{i=1}^{M} G_i(u_p; h) \right\} \times \sum_{q=1}^{n} a_q \left\{ \sum_{j=1}^{M} G_j(u_q; h) [X_j^{b(1)} - \overline{X}^{b(1)}] \bigg/ \sum_{j=1}^{M} G_j(u_q; h) \right\}$$

$$= \left( \sum_{p=1}^{n} a_p \left\{ \sum_{i=1}^{M} G_i(u_p; h) [X_i^{b(1)} - \overline{X}^{b(1)}] \bigg/ \sum_{i=1}^{M} G_i(u_p; h) \right\} \right)^2 \geq 0.$$

## REFERENCES

Anderson, J. L., 2001: An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.,* **129,** 2884–2903.

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

——, and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.,* **127,** 2741–2758.

Berre, L., and G. Desroziers, 2010: Filtering of background error variance and correlations by local spatial averaging: A review. *Mon. Wea. Rev.,* **138,** 3693–3720.

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.

——, and ——, 2009a: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus,* **61,** 84–96.

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

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

Fan, J., and Q. Yao, 2003: *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer-Verlag, 552 pp.

Fuentes, M., 2001: A high frequency kriging approach for nonstationary environmental processes. *Environmetrics,* **12,** 469–483.

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

Hall, P., and P. Patil, 1994: Properties of nonparametric estimators of autocovariance for stationary random fields. *Probab. Theory Relat. Fields,* **99,** 399–424.

Hamill, T. M., 2006: Ensemble-based data assimilation. *Predictability of Weather and Climate,* T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 124–156.

——, and J. S. Whitaker, 2005: Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Mon. Wea. Rev.,* **133,** 3132–3147.

——, ——, 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.

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.,* **129,** 123–137.

——, and ——, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.,* **131,** 3269–3289.

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.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory.* Academic Press, 376 pp.

Jun, M., and M. L. Stein, 2007: An approach to producing space–time covariance functions on spheres. *Technometrics,* **49,** 468–479.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, 341 pp.

Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability,* Shinfield Park, Reading, Berkshire, United Kingdom, European Centre for Medium-Range Weather Forecasts, 1–18.

——, 2005: Designing chaotic models. *J. Atmos. Sci.,* **62,** 1574–1587.

——, and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulations with a small model. *J. Atmos. Sci.,* **55,** 399–414.

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

Paciorek, C., and M. Schervish, 2006: Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics,* **17,** 483–506.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott, and J. A. Yorke, 2005: Assessing a local ensemble Kalman filter: perfect model experiments with the National Centers for Environmental Prediction global model. *Tellus,* **57A,** 528–545.

——, ——, ——, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. Yorke, 2008: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus,* **60A,** 113–130.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.,* **130,** 1913–1923; Corrigendum, **134,** 1722.

Yoon, Y.-N., E. Ott, and I. Szunyogh, 2010: On the propagation of information and the use of localization in ensemble Kalman filtering. *J. Atmos. Sci.,* **67,** 3823–3834.

Zhang, F., Y. Weng, J. Sippel, and C. Bishop, 2009a: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.,* **137,** 2105–2125.

——, M. Zhang, and J. Hansen, 2009b: Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Adv. Atmos. Sci.,* **26,** 1–8.