Empirical Localization of Observation Impact in Ensemble Kalman Filters

JEFFREY ANDERSON

Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research,* Boulder, Colorado

LILI LEI

Advanced Study Program, National Center for Atmospheric Research,* Boulder, Colorado

ABSTRACT

Localization is a method for reducing the impact of sampling errors in ensemble Kalman filters. Here, the regression coefficient, or gain, relating ensemble increments for observed quantity y to increments for state variable x is multiplied by a real number α defined as a localization. Localization of the impact of observations on model state variables is required for good performance when applying ensemble data assimilation to large atmospheric and oceanic problems. Localization also improves performance in idealized low-order ensemble assimilation applications. An algorithm that computes localization from the output of an ensemble observing system simulation experiment (OSSE) is described. The algorithm produces localizations for sets of pairs of observations and state variables: for instance, all state variables that are between 300- and 400-km horizontal distance from an observation. The algorithm is applied in a low-order model to produce localizations from the output of an OSSE and the computed localizations are then used in a new OSSE. Results are compared to assimilations using tuned localizations that are approximately Gaussian functions of the distance between an observation and a state variable. In most cases, the empirically computed localizations produce the lowest root-mean-square errors in subsequent OSSEs. Localizations derived from OSSE output can provide guidance for localization in real assimilation experiments. Applying the algorithm in large geophysical applications may help to tune localization for improved ensemble filter performance.

1. Introduction

Ensemble Kalman filters are used for data assimilation (DA) in geophysical applications including the atmosphere, ocean, and land surface. Basic ensemble Kalman filters (Evensen 1994, Burgers et al. 1998) without localization work well for some applications but often diverge from the true state or fail for large atmospheric problems with affordable ensemble sizes (Houtekamer and Mitchell 1998).

The optimal solution to a DA problem with a linear forecast model, linear observation operators, and normal observation error is given by the Kalman filter (Kalman 1960). For sufficiently large ensembles, some deterministic ensemble Kalman filters like the ensemble adjustment Kalman filter (EAKF; Anderson 2001) are just methods for computing the Kalman filter solution (Anderson 2009b). However, the ensemble sample covariance of an observed variable and an unobserved variable can be suboptimal when conditions for Kalman filter optimality are violated, in particular when the model is nonlinear, when the ensemble size is too small (always the case for large geophysical applications), or for stochastic ensemble filters (Burgers et al. 1998).

Houtekamer and Mitchell (1998) showed that ensemble filters for large atmospheric applications worked better when an observation only impacted a state variable if the two were separated by less than a specified physical distance. Limiting the impact of an observation on a state variable is called localization or tapering (Hamill et al. 2001; Furrer and Bengtsson 2007). It is often implemented by multiplying the regression coefficient, or gain, for observation y and state variable x by a factor between 0 and 1. Some filters like the local ensemble

transform Kalman filter (Ott et al. 2004; Miyoshi et al. 2007) implicitly localize by implementing the assimilation algorithm on local regions. Localization is an essential part of large geophysical ensemble DA systems. Ensemble filters with fewer than 100 members work for the largest atmosphere and ocean models, but choosing appropriate localizations normally requires expensive tuning. The common localization method for ensemble filters is closely related to the alpha control variable method that is used to model flow-dependent covariances in some variational assimilation systems (Lorenc 2003).

The most common form for localization uses the approximately Gaussian compactly supported form of Gaspari and Cohn (1999), referred to as the Gaspari–Cohn (GC) function. The GC function is a piecewise continuous fifth-order polynomial approximation to a Gaussian. When a single GC function is used in an application, localization is a function of only the horizontal distance between an observation and a state variable and a real parameter that is equal to one-half the distance at which the GC function goes to zero. Tuning the localization only requires finding an appropriate value for the parameter that defines the half-width, but even this can be costly for large models.

Localization that is also a function of the vertical distance between an observation and a state variable may be required and a number of functions have been used in addition to the GC (Whitaker et al. 2004; Houtekamer and Mitchell 2005). Using different localizations for the impact of different observation types (Houtekamer and Mitchell 2005; Tong and Xue 2005; Kang et al. 2011) can also be helpful. Anderson (2007) shows that the type of the state variable may also lead to different localization. For instance, collocated surface pressure and temperature state variables might require different localizations for the same observation (Anderson 2012).

Forward operators for some observations, like Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) radio occultations (Liu et al. 2007) or satellite radiances (Houtekamer and Mitchell 2005; Campbell et al. 2010) are functions of many state variables and have no obvious spatial location. Effective localization for such observations may be quite different from a GC function (Anderson 2007). Additional localization may also be required for an observation and a state variable that are separated in time (Anderson 2007). Comprehensive tuning of localization for large applications with a variety of observation types and state variable types is clearly very challenging.

One way to address this challenge is to develop methods that dynamically compute localization. For a petroleum reservoir application, Chen and Oliver (2009)

used the prior ensemble covariance between an observation and a state variable to compute localization while Emerick and Reynolds (2010) based localization on observation sensitivity matrices. Zhang and Oliver (2010) present a bootstrap method that directly estimates errors in ensemble covariances and can be used to develop a localization. Anderson (2012) developed a similar algorithm that computed localization as a function of correlation and ensemble size. Bishop and Hodyss (2007, 2009a,b) compute localization as a power of the sample correlation of a smoothed ensemble.

Anderson (2007) proposed running a group of ensemble filters that differ only in the initial ensemble. A localization that minimizes the expected ensemble mean root-mean-square error for a state variable, given the set of regression coefficients from the group is computed. Results were promising, but the group filter algorithm may be too expensive for large applications. Anderson (2007) suggested that the time mean localization from a short group filter assimilation could be used as a static localization for a traditional ensemble filter.

This manuscript proposes a method for using the output of an observing system simulation experiment (OSSE) to generate empirical localizations. These empirical localizations can then be applied in another OSSE and compared to experiments with carefully tuned GC localization. Of particular interest is whether localizations that are distinctly different in shape from the GC function can produce enhanced assimilation performance. Section 2 develops the algorithm, section 3 describes low-order model experiments to explore its behavior, and section 4 presents results of the experiments. Section 5 discusses the challenges of computing empirical localization outside of the OSSE context, and sections 6 and 7 present additional discussion and conclusions.

2. Empirical localization functions

Localization is a method for reducing the impact of errors in ensemble Kalman filters (Houtekamer and Mitchell 1998; Hamill et al. 2001; Furrer and Bengtsson 2007). The regression coefficient, or gain, relating ensemble increments for observed quantity y to increments for state variable x is multiplied by a real number α called a localization,

$$\Delta x_n = \alpha \hat{b} \Delta y_n, \quad n = 1, \ldots, N, \quad (1)$$

where n indexes the N ensemble members, Δy_n is the increment for the nth ensemble estimate of the observed quantity, Δx_n is the corresponding increment for the nth ensemble estimate of the state variable component, and $\hat{b}$ is the prior sample regression coefficient. The increments

Δy_n for the observed quantity can be computed by a variety of algorithms (Anderson 2003), but the details are not relevant to the localization. Results here use the deterministic ensemble adjustment Kalman filter (Anderson 2001) with a sequential parallel implementation (Anderson and Collins 2007). All subsequent results assume that the observational errors of any pair of observations are uncorrelated; the method can be extended to the case of correlated observational errors by using a singular value decomposition (SVD) to rotate to a frame in which observational errors are uncorrelated during the assimilation (Anderson 2003).

a. Empirical localization from an OSSE

An algorithm to compute localizations given the results of an ensemble filter OSSE is described here. Let **X** be the set of all model state variable instances archived from an OSSE. A state variable instance is defined as a model state variable at a particular assimilation time. For instance, in a climate model OSSE **X** would include gridded values of temperature, wind components and surface pressure every 6 h. The size M of set **X** is the product of the number of gridded model state variables and the number of assimilation times. Let $x_{m,n}$ be the nth ensemble sample of the mth state variable instance, $m \in \{1, \ldots, M\}$ and $n \in \{1, \ldots, N\}$, where N is the ensemble size.

Let **Y** be the set of all observations for which localization is to be computed. It is important to note that elements of **Y** need not be the same as observations that were used in the OSSE and can be any function of the model state variable instances. Let

$$y_{j,n} = h_j(x_{1,n}, x_{2,n}, \ldots, x_{M,n}) \qquad (2)$$

be the nth ensemble sample of the jth observation with $j \in \{1, \ldots, J\}$, where h_j is the jth observation operator.

Note that in all equations that follow, x refers to the prior value of a state variable.

A localization function L is defined as a real-valued function whose domain is the set of all pairs of a member of set **Y** and a member of set **X**, $L = L(y, x)$, where $y \in \mathbf{Y}$ and $x \in \mathbf{X}$. For instance, L could give the localization for a radiosonde temperature observation at (43.6°N, 128.2°W) at 452 hPa at 1313 UTC and a model gridpoint value of zonal wind component at (39.3°N, 126.4°W) at 5300 m at 1200 UTC. In most ensemble DA applications, the range of localization has been confined to [0, 1] by definition but this is not required in the definition here.

Localization functions are frequently defined as a composite function, $L(y, x) = G[F(y, x)]$. For example, $F(y, x)$ could be defined as

$$F(y,x) = \infty \quad \text{if} \quad D_t(y,x) > \Delta t_{\text{crit}},$$

$$F(y,x) = D_s(y,x) \quad \text{if} \quad D_t(y,x) \leq \Delta t_{\text{crit}},$$

where $D_t(y, x)$ is the difference in time between y and x and $D_s(y, x)$ is the horizontal spatial distance between y and x. In this case, the most common choice for G has been the Gaspari–Cohn function (Gaspari and Cohn 1999) with a specified half-width.

An empirical localization function (ELF) is defined as a function $L(y, x)$ constructed using the output from an ensemble OSSE. Here, an empirical localization is computed for subsets of the pairs (y, x) of observations and state variable instances that comprise the domain of L. The empirical localization minimizes the root of the mean square difference between the posterior ensemble mean of the state variable instance and its known true value over the subset. It is the value of α that minimizes

$$\sqrt{\sum_{k=1}^{K} (\overline{x}_k + \Delta \overline{x}_k - x_{k,t})^2 \Big/ K} = \sqrt{\sum_{k=1}^{K} (\overline{x}_k + \alpha \hat{b}_k \Delta \overline{y}_k - x_{k,t})^2 \Big/ K}, \qquad (3)$$

where k indexes the elements of the subset, K is the number of (y, x) pairs in the subset, an overbar is an ensemble mean, and $x_{k,t}$ is the true value. The subscript k refers to the state variable instance and observation that are associated with the kth pair in the subset, not to the kth state variable instance or kth observation. The minimum occurs when the derivative with respect to α of the sum in (3) is 0,

$$\sum_{k=1}^{K} 2(\overline{x}_k + \alpha \hat{b}_k \Delta \overline{y}_k - x_{k,t}) \hat{b}_k \Delta \overline{y}_k = 0. \qquad (4)$$

Solving for α gives

$$\alpha = \sum_{k=1}^{K} (x_{k,t} - \overline{x}_k) \hat{b}_k \Delta \overline{y}_k \Big/ \sum_{k=1}^{K} (\hat{b}_k \Delta \overline{y}_k)^2. \qquad (5)$$

The hypothetical observation y_j^o is assumed to be drawn from the normal distribution $N(y_{j,t}, \sigma_{o,j}^2)$, where

$$y_{j,t} = h_j(x_{1,t}, x_{2,t}, \ldots, x_{M,t}) \qquad (6)$$

is the true value and $\sigma_{o,j}^2$ is the specified observation error variance.

If y_j^o is observed, the Kalman filter increment to the observed quantity is a weighted sum of the prior ensemble mean $\overline{y}_{p,j}$ and y_j^o,

$$\Delta \overline{y}_j = A_j + B_j y_j^o, \tag{7}$$

where

$$A_j = (\sigma_{u,j}^2/\sigma_{p,j}^2 - 1)\overline{y}_{p,j} \tag{8}$$

$$B_j = \sigma_{u,j}^2/\sigma_{o,j}^2. \tag{9}$$

In (8) and (9), $\sigma_{p,j}^2$ is the prior ensemble sample variance for observation y_j and

$$\sigma_{u,j}^2 = [(\sigma_{p,j}^2)^{-1} + (\sigma_{o,j}^2)^{-1}]^{-1} \tag{10}$$

is the posterior variance.

Equation (7) can be substituted in (5), and one can integrate over the probability distribution of the observation, giving

$$\alpha = \sum_{k=1}^K \left[(x_{k,t} - \overline{x}_k)\hat{b}_k \int_{-\infty}^\infty (A_k + B_k y_k^o) N(y_{k,t}, \sigma_{o,k}^2)\, dy_k^o \right] \Bigg/ \sum_{k=1}^K \int_{-\infty}^\infty [\hat{b}_k(A_k + B y_k^o)]^2 N(y_{k,t}, \sigma_{o,k}^2)\, dy_k^o. \tag{11}$$

Integrating over the distribution of the observation helps to reduce sampling noise that would occur from just using the noisy scalar observed value. Again, subscript k refers to the observation associated with the kth pair in the subset. Evaluating the integrals in (11) gives

$$\alpha = \sum_{k=1}^K \hat{b}_k (x_{k,t} - \overline{x}_k)(A_k + B_k y_{k,t}) \Bigg/ \sum_{k=1}^K \hat{b}_k^2 [A_k^2 + 2A_k B_k y_{k,t} + B_k^2(\sigma_{o,k}^2 + y_{k,t}^2)]. \tag{12}$$

For example, in a global climate model, one might define $S = 11$ subsets of the (y, x) pairs. Subset s, $s = 1, \ldots, 10$ contains all (y, x) pairs where the observation y and state variable instance x differ in time by less than 3 h and where $100(s - 1) \le D_x(y, x) < 100(s)$ km; subset 11 contains all other (y, x) pairs. For instance, the second subset would contain only (y, x) pairs that are between 100 and 200 km apart and separated by less than 3 h. The value of $L(y, x)$ might be set to 0 for all elements of subset 11 while a single value of $L(y, x)$ is computed for each of the first 10 subsets from the archived output of an ensemble OSSE as follows:

1) For each pair (y_k, x_k) in the subset,
   (i) compute all ensemble members for observation y_k using (2) and compute the ensemble mean;
   (ii) compute the true value of the observation using (6);
   (iii) compute the sample regression coefficient $\hat{b}$ between y_k and x_k;
   (iv) compute A and B for the observation using (8)–(10); and
   (v) compute the kth term of the numerator and denominator of (12).
2) Compute the localization α by summing the terms in (12).

Recall that observations in $\mathbf{Y}$ need not be the same as those that were assimilated in the OSSE.

b. Empirical localization for parallel algorithms

For large atmospheric applications of ensemble data assimilation, efficient parallel computation is essential. The parallel implementation of the sequential filter algorithm described in Anderson and Collins (2007) as implemented in the Data Assimilation Research Testbed (Anderson et al. 2009) is used for all results here. This algorithm uses a joint state space (Anderson 2001; Jazwinski 1970) where the joint state/observation set is $\mathbf{Z} = \mathbf{X} \cup \mathbf{Y}$. Assimilation using the parallel algorithm requires a localization for the impact of one observation on another. The domain of a localization function L is redefined as the set of all pairs of a member of set $\mathbf{Y}$ and a member of set $\mathbf{Z}$, $L = L(y, z)$, where $y \in \mathbf{Y}$ and $z \in \mathbf{Z}$. The prior ensemble members for all potential observations y at a given time are computed at the start of the algorithm using (2). The algorithm in section 2a is modified replacing all instances of $\mathbf{X}$ and x with $\mathbf{Z}$ and z.

c. Empirical localization without knowing the truth

A modified procedure can be applied to find empirical localizations using the output from a real ensemble assimilation where the truth is unknown. In this case, a localization function L is defined as a real-valued function whose domain is the set of all pairs of members

of the set of observations **Y** that were assimilated, $L = L(y, u)$, where $y, u \in \mathbf{Y}$ and y and u are unique (although they could be two independent observations of the same quantity at the same time).

An empirical localization is computed for subsets of the pairs of observations to minimize the root-mean-square difference between the posterior ensemble mean of u_k and the observed value u_k^o over the subset,

$$\sqrt{\sum_{k=1}^{K} (\overline{u}_k + \alpha \hat{b}_k \Delta \overline{y}_k - u_k^o)^2 / K}. \tag{13}$$

The mean increment for a given observation $\Delta \overline{y}_j$ can be computed from (7) to (10). In this case, there is a single observed value and no need to integrate over an observational error distribution, so the localization can be computed from

$$\alpha = \sum_{k=1}^{K} (u_k^o - \overline{u}_k) \hat{b}_k \Delta \overline{y}_k \bigg/ \sum_{k=1}^{K} (\hat{b}_k \Delta \overline{y}_k)^2. \tag{14}$$

Localization can only be computed between quantities that are observed. Since it is unusual to have observations of all state variables, it is generally impossible to compute a complete ELF that can be used in a subsequent assimilation; however, partial ELFs could be used in conjunction with other methods of defining localization.

3. Design of low-order model experiments

a. Empirical localization for Lorenz96 40-variable model

The 40-variable configuration of the Lorenz96 model (Lorenz and Emanuel 1998) is used with standard parameter settings discussed in their section 3: forcing $F = 8$, time step of 0.05 units, and fourth-order Runge–Kutta time differencing scheme. The 40 state variables are defined to be equally spaced in a periodic one-dimensional domain of length 40. All observations have a location that is the same as one of the state variables. Then 40 subsets are defined by the displacement of the state variable x from the observation y for pairs that are at the same time; the displacement has values from the set $\{-20, -19, \ldots, 18, 19\}$. For plotting purposes, values for displacement -20 are repeated for 20. All (y, x) pairs where the observation and state variable are at different times are assigned to a 41st subset with localization 0; observations only impact state at the assimilation time.

b. Experimental design

A 100 000-step integration of the model is performed starting from a state with the first state vector element
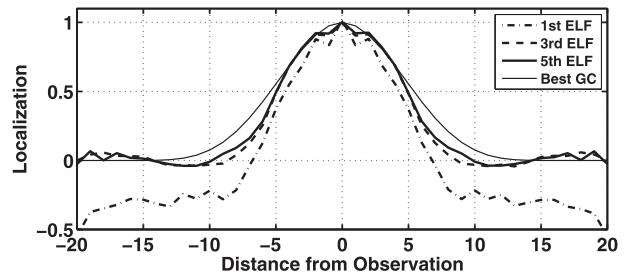

FIG. 1. First, third, and fifth empirical localization functions for observations with error variance 16 assimilated every model time step with a 20-member ensemble and a Gaspari–Cohn function with half-width of 20% of the domain. The units for the horizontal axis are the number of grid intervals separating the state variable from the observation.

8.01 and the 39 other elements 8.0. The resulting model state is perturbed to create 320 additional states by adding random draws from $N(0, 0.01)$ to each of the state vector elements for each of the perturbed states. These 321 states are advanced in time for another 10 000 steps to produce an initial truth state and 320 initial ensemble members that are samples of the model's climatological distribution. The truth is then advanced for an additional 110 000 assimilation time intervals (experiments here assimilate either every model time step or every 12 model time steps) and synthetic observations are generated at each assimilation time.

A sequence of five ELFs is constructed for a given set of observations and ensemble size N. First, a 6000-step assimilation with no localization is done starting from the first N of the 320 ensemble initial conditions. The last 5000 steps of output from this assimilation are used to construct empirical localizations for each of the 40 displacement subsets described in section 3a. The resulting ELF is used as localization for a second 6000-step assimilation starting from the ensemble initial conditions and the output of this experiment is used to construct a second ELF. This is repeated five times, resulting in five successive ELFs.

The computed range of ELFs is not limited to [0, 1] (Fig. 1). However, when ELFs are used as localization in subsequent assimilation experiments, negative values are assumed to be the results of noise and are replaced with 0. Using negative values often led to larger ensemble mean RMSE error while using values larger than 1 was not found to have detrimental effects. The use of negative localizations needs further exploration (see section 6).

The five ELFs are used as the localization for long assimilations and compared to results that use a variety of GC function localization half-widths. Each N-member ensemble assimilation experiment starts with the first N

of the 320 initial ensemble members and proceeds for 110 000 assimilation times; data from the first 10 000 steps are discarded, leaving 100 000 assimilation times for diagnostic evaluation. Since only assimilation steps 1000–6000 are used to compute ELFs, this provides an independent evaluation. All assimilation experiments (including those used to compute the ELFs) use the spatially and temporally varying adaptive inflation algorithm of Anderson (2009a) with a fixed value of 0.1 for the inflation standard deviation and initial inflation of 1. The sensitivity of results to the inflation parameters was not explored since the ratio of time mean RMS error to spread was close to one for all experiments (see also section 6). All results used the ensemble adjustment Kalman filter (Anderson 2001) with the sequential least squares algorithm of Anderson (2003) in the parallel implementation of Anderson and Collins (2007). Some cases were repeated with a perturbed observation ensemble Kalman filter (Burgers et al. 1998) and with the non-Gaussian rank histogram filter (Anderson 2010). There were no obvious qualitative differences for the different filter variants. All computations used the Data Assimilation Research Testbed software (Anderson et al. 2009), which is freely available online (from http://www.image.ucar.edu/DAReS/DART).

c. Evaluation metrics

The time mean value of the root-mean-square error (RMSE) of the ensemble mean from the truth,

$$\text{RMSE} = \sqrt{\sum_{m=1}^{40} (\overline{x}_m - x_{m,t})^2 / M}, \qquad (15)$$

is used to evaluate results where $\overline{x}_m$ is the ensemble mean, $x_{m,t}$ is the true value, and the subscript m indexes the model state variable. Results are shown for the RMSE of the prior (forecast), but no qualitative differences were found between the prior and posterior RMSE. The value of the adaptive inflation averaged over all assimilation times and all state variables is also examined.

Mitchell et al. (2002), Kepert (2009), and Oke et al. (2007) have shown that localization can lead to increasingly unbalanced posterior model states and consequent transient nonequilibrium oscillations such as gravity waves. The mean absolute value of the time tendency of the state variables in all ensembles immediately after each assimilation averaged over all times and state variables

$$\sum_{k=1}^{K} \sum_{n=1}^{N} \sum_{m=1}^{40} \left| \frac{dx_{m,n,k}}{dt} \right| \bigg/ 40KN \qquad (16)$$

is used as a measure of imbalance in the analyzed states. In (16), K is the number of assimilation times, N is the ensemble size, and m indexes the 40 variables in the model state vector. Large average time tendencies indicate that the analyzed states have increased spatial gradients that are not consistent with those found in free integrations of the model (Greybush et al. 2011).

4. Results

a. Frequent identity observations with large error variance

For this case, every state variable is observed every model time step with an observational error variance of 16. This error variance is not small compared to the model's climatological variance, but it produces an assimilation where a fairly tight GC localization is optimal for an initial test. The 20-member ensemble case is a baseline for comparison for other experiments. Figure 1 shows the first, third, and fifth ELF computed as described in section 3b for this case along with a GC function with half-width of 20% of the domain. Figure 2a shows that the 20% GC and the second through fifth ELFs produced similar RMSE; the fourth ELF actually produced the smallest RMSE but differences are certainly not significant. Figure 1 shows that the third and fifth ELFs are qualitatively similar to the 20% GC localization for small displacements but are smaller than the GC for intermediate displacements and spatially noisy for large displacements where values are close to zero. The first ELF has significant negative values for larger displacements that are discussed in section 6.

Figure 2b shows the prior spread for the five ELFs and a number of GC half-widths. Comparing these to the RMSE in Fig. 2a shows that the adaptive inflation keeps the RMSE and spread roughly comparable. The spread for the ELF cases is slightly larger than the spread for the 20% GC case.

Figure 2c shows that the adaptive inflation required for the GC cases increases monotonically as the half-width increases. The inflation required for the ELF cases is slightly smaller than that for the 20% half-width GC case. Apparently the central portion of the ELFs is slightly less broad than the 20% GC (Fig. 1) and leads to less erosion of spread and a reduced need for inflation.

Figure 2d shows the time and spatial mean of the absolute value of initial time tendencies for analyses from the GC cases and the five ELF cases. The time tendency is smallest for GC half-width 15% and results are noisy. All but one of the ELFs have smaller time tendencies than the 20% GC case that gave the minimum RMSE. This suggests the ELFs produce analyses
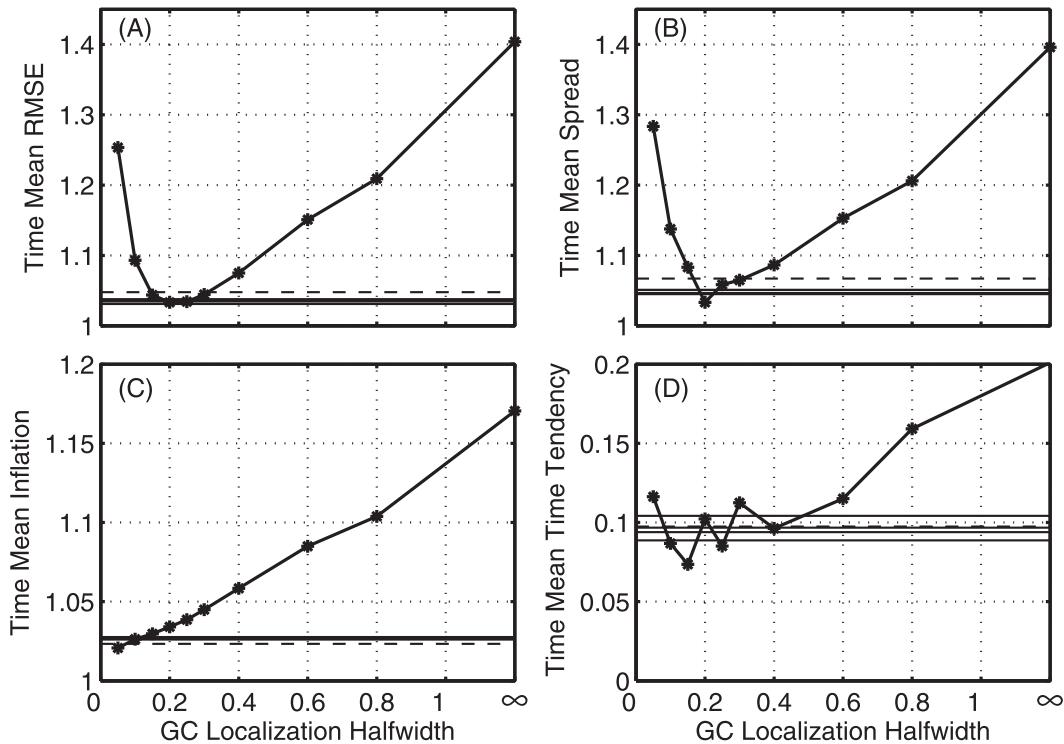
FIG. 2. Comparison between assimilations with Gaspari–Cohn localization and the five empirical localization functions for observations with error variance 16 assimilated every model time step with 20-member ensembles. (a) The time mean root-mean-square error of the prior ensemble mean, (b) the prior ensemble spread, (c) the time mean of the spatial mean of the adaptive inflation, and (d) the time mean of the spatial mean of the absolute value of the model time tendency after assimilation. Results for the empirical localization functions are independent of the horizontal axis. The second through fifth empirical localization functions are solid while the first empirical localization function is marked by a dashed line and is not visible in (d) because it is coincident with one of the other empirical localization functions. The close spacing of the ELF results in most panels indicates that they vary little compared to the variation as a function of GC localization half-width.

that are about as well balanced as for the very smooth GC function.

Figure 3 shows the RMSE for 10- and 40-member ensembles for a set of GC localizations and for the third ELF for each. For $N = 10$, the first two ELFs (not shown) produce larger RMSEs than the last three, which have RMSEs that are slightly larger than the best GC case. For $N = 40$ the first four ELF cases have slightly smaller RMSE than the best GC case while the fifth ELF has slightly larger RMSE. It is not clear why the fifth ELF has higher RMSE in this case. Exploring a sequence of five additional ELFs (ELFs 6–10) showed that their RMSEs were all between those of the fourth and fifth ELF.

Figure 4 shows half of the fifth ELF for ensemble sizes of $N = 20$, 80, 160, and 320. The ELFs get broader as the ensemble size increases but are close to zero at larger displacements. This is consistent with localization estimates produced using the hierarchical filter method in Anderson (2007). The spatial noise in the ELFs far from

the observation increases as the ensemble size grows. This may be because larger ensembles better represent very small correlations that would require larger numbers of OSSE assimilation steps to accurately estimate the empirical localization.

Figure 5 shows the five ELFs for ensemble size $N = 10$. In this case, it takes three iterations for the ELFs to approximately converge. The first ELF, computed from an assimilation that used no localization, has large areas of negative localization indicating that the impact of a distant observation is expected to degrade the estimate of the true state. The third, fourth, and fifth ELFs are similar to the best GC localization function except for the two points immediately adjacent to the observation (only the one to the right is shown in the figure) where the ELFs are smaller than the GC. There is a hint of this feature in the ELFs for $N = 20$, but it is not apparent for larger ensemble sizes (see Fig. 4). In all cases, however, the shape of the ELFs is approximately Gaussian, making it difficult for assimilations using the ELFs to
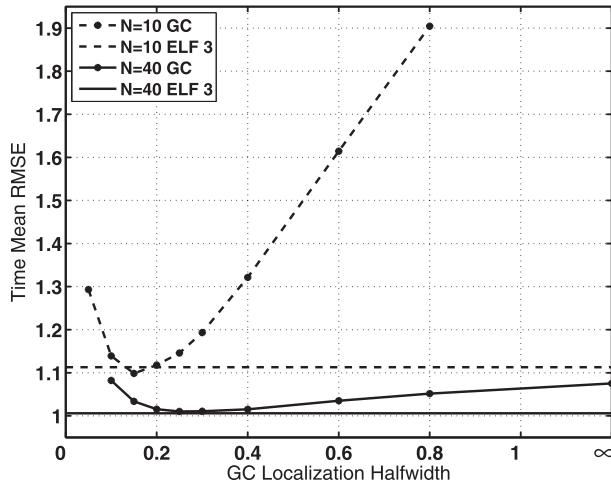
FIG. 3. Time mean root-mean-square error for the third empirical localization function and a variety of Gaspari–Cohn half-widths for observations with error variance 16 assimilated every model time step with 10- and 40-member ensembles. Results for the empirical localization functions are independent of the horizontal axis.

produce significantly lower RMSE than obtained with the best GC function.

The ELFs shown here were all constructed using 5000 assimilation steps to compute the sums in (12). For each assimilation time, there are 40 (y, x) pairs for each displacement so that a total of 200 000 terms are summed for each subset. A total of 5000 steps were chosen because more did not produce much reduction in RMSE when the ELFs are used to localize in assimilations. However, the RMSE for ELFs obtained from assimilations using only 250 steps were nearly as small. As a baseline for comparison, an ELF was computed from a 20 000-step OSSE (after discarding the first 1000 steps) $N = 20$ case that used the fourth ELF from the standard experiments as its localization. In Fig. 6, the ELF computed using all 20 000 steps is compared to ELFs that use only every 4th, every 32nd, or every 256th assimilation time. Figure 6 shows the absolute value of the difference of the ELFs computed with reduced data from the one computed with all 20 000 steps. The largest differences are found for displacements greater than four grid intervals and differences are larger when fewer time steps were used in computing the ELFs. Differences are less than 0.01 for the every 4th case and less than 0.04 for the every 32nd case while differences larger than 0.1 are found for the case using the fewest data. Further research is needed to understand how large a sample size is required to produce good ELFs for various subsets of (y, x) pairs in large geophysical prediction models.
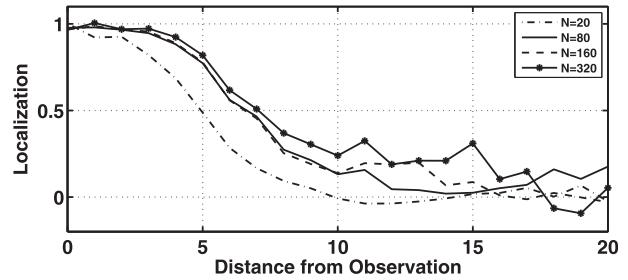

FIG. 4. Fifth empirical localization function for observations with error variance 16 assimilated every model time step with ensemble sizes of 20, 80, 160, and 320. Only positive displacements are shown since the functions are approximately symmetric around the observation point. The units for the horizontal axis are the number of grid intervals separating the state variable from the observation.

b. Infrequent identity observations with small error variance

The next set of results is for observational error variance of 1 with observations of each state variable assimilated every 12 model time steps. In this case, the prior RMSE is about twice that for experiments in section 4a, but the posterior RMSE is much smaller than the posterior RMSE in section 4a. Figure 7 shows the RMSE for $N = 10$, 20, and 40 for a variety of GC half-widths and the five ELFs in each case. For $N = 10$ and 20, the RMSE for all but the first ELF is smaller than the best GC case, while all five $N = 40$ ELFs give smaller RMSE than the best GC. The best GC localizations in this case have considerably smaller half-widths than for the section 4a experiments.

Figure 8 shows the fifth ELF and the best GC localization for the $N = 10$ and 40 cases. Both ELFs have a local minimum for the grid point adjacent to the observation, similar to the $N = 10$ case in section 4a. The
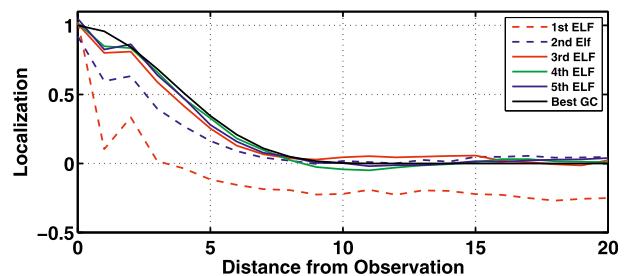

FIG. 5. The five empirical localization functions for observations with error variance 16 assimilated every model time step with a 10-member ensemble and a Gaspari–Cohn function with half-width of 15% of the domain. The units for the horizontal axis are the number of grid intervals separating the state variable from the observation.
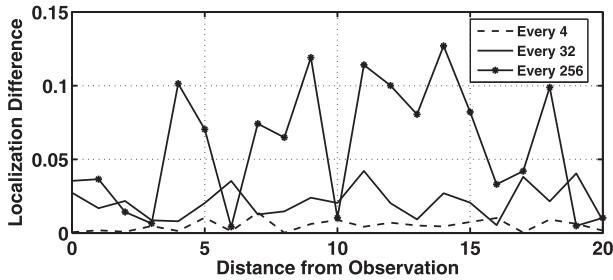
FIG. 6. The absolute value of the difference between empirical localization functions computed from every 4th, every 32nd, and every 256th assimilation step of a 20 000-step assimilation and an empirical localization function computed from all 20 000 assimilation steps as a function of displacement between the observation and the state variable. The assimilation has 20 ensemble members for the error variance 16 case with observations every time and used the fourth ELF. The units for the horizontal axis are the number of grid intervals separating the state variable from the observation.

best GC localization functions are a compromise between the sharp decrease adjacent to the observation and the broad tails at greater displacements in the ELFs. In this case, the ELFs are less Gaussian and the resulting RMSEs are smaller than for the best GC half-widths.

c. Observations of the sum of state variables

In this case, each observation is the sum of 17 consecutive state variables with an observational error variance of 1. The location of the middle (9th out of 17) of the summed variables is the location assigned to the observation. For instance, an observation located at grid point 20 is drawn from

$$N\left(\sum_{j=12}^{28} x_{j,t}, 1\right).$$

Eight observations at each location are taken every model time step. These observations are analogous to satellite radiance observations of the atmosphere where the forward operator is a weighted average of state variables in some volume of the atmosphere. Also, like radiances each individual observation has a relatively small amount of information: hence the need for 320 observations at each time to keep the prior RMSE comparable to the previous cases. Finally, because the observations are weakly correlated with a large number of state variables, they are expected to be particularly susceptible to sampling error (Anderson 2012).

Figure 9 shows the first, third and fifth ELFs for $N = 20$ and the GC function with half-width 40% of the domain that gave the smallest RMSE. The ELFs are distinctly non-Gaussian with a local minimum close to the observation and a pair of maxima for the grid points
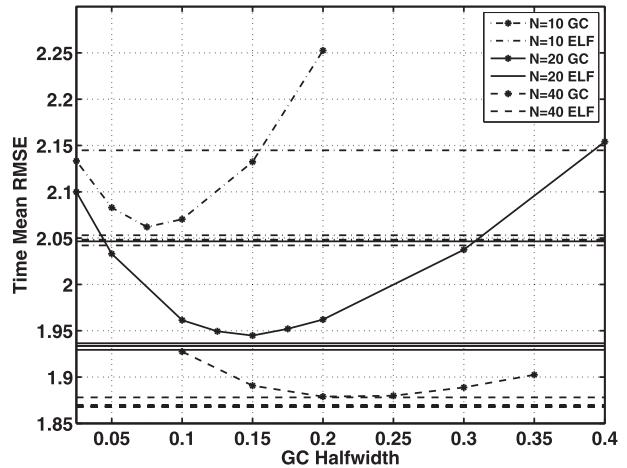

FIG. 7. Time mean root-mean-square error for the five empirical localization functions and a variety of Gaspari–Cohn half-widths for observations with error variance 1 assimilated every 12th model time step for 10-, 20-, and 40-member ensembles. The values for all five empirical localization functions for a given ensemble size are very similar, except for the first ones for the 10- and 20-member ensembles, which are much larger than the subsequent four. Results for the empirical localization functions are independent of the horizontal axis.

at displacement ±7. The maxima of the ELFs are about 0.8. This is similar to localization patterns found with a hierarchical filter for observations that are sums of state variables (see Fig. 12 in Anderson 2007). It is also similar in shape to the time mean absolute values of the correlation between the observation and the individual state variables as a function of displacement. Anderson (2012) showed that localization values for an observation often are closely related to the absolute value of the correlation.

Figure 10 shows the RMSE for the five ELFs and for a sample of GC half-widths. All five ELFs produce significantly smaller RMSE than the best GC case as expected from the non-Gaussian shape of the ELFs. The dependence of the RMSE on the GC half-width in this case has multiple minima.

The inflation for the ELF cases varies between 1.12 and 1.14 while the inflation for the best GC case is about 1.2. This is consistent with the GC reducing ensemble variance too much by not localizing the impact of an observation on nearby state variables enough. The average initial time tendency for the best GC case is about 0.22, while the time tendencies for all but the first ELF are less than 0.175, indicating that the ELFs give more balanced analyses. These results show that using an ELF can produce lower RMSE with less noisy analyses for cases where good localizations are non-Gaussian.

The parallel sequential algorithm for the EAKF that is used here computes the impact of observations on the
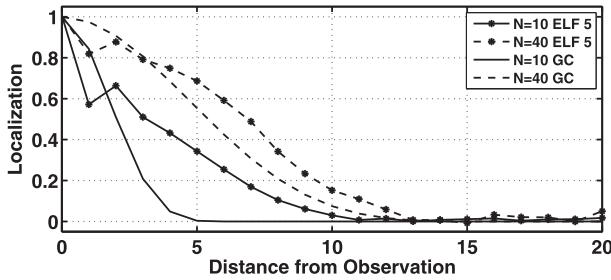
FIG. 8. The fifth empirical localization function and the Gaspari–Cohn function with the smallest RMSE for observations with error variance 1 assimilated every 12th model time step for 10- and 40-member ensembles. The units for the horizontal axis are the number of grid intervals separating the state variable from the observation.
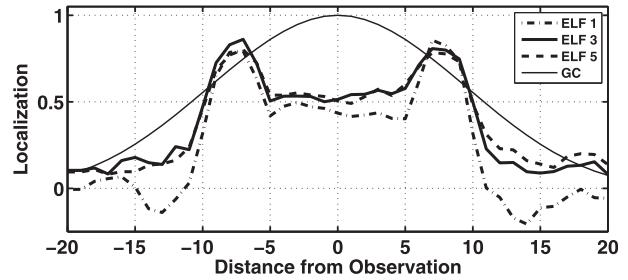

FIG. 9. The first, third, and fifth empirical localization functions and the Gaspari–Cohn function with the smallest associated RMSE for 320 observations of the sum of 17 state variables with observation error variance 1 assimilated every time step with 20-member ensembles. The units for the horizontal axis are the number of grid intervals separating the state variable from the observation.

prior ensemble for other contemporaneous observations that have not yet been assimilated. It is possible that appropriate localizations for the impact of an averaged observation on another averaged observation could be different from the localization for impact on a state variable. In these experiments, this was not explored and the localization optimized for state variables was also used for the impact on other observations.

Another consequence of the sequential assimilation of observations at the same assimilation time is that the appropriate localization is surely related to the order in which the observations are assimilated. The first observation is impacting prior state estimates with relatively large uncertainty and might be expected to require a moderately compact localization. Subsequent observations impact a progressively more certain state estimate and might require progressively less compact localization. In the algorithm used here, the optimization in (12) is applied only to the prior ensemble at the start of each assimilation step and this sequential assimilation issue is ignored. It is possible to compute an approximate upper bound on the impact that would be expected by doing sequential assimilation. The ELF computation can be repeated using the posterior ensemble distribution from an OSSE rather than the prior. The last observation assimilated sequentially would see a state estimate very similar to the final posterior. ELFs computed using the posterior did not differ qualitatively from those shown here and generally differed quantitatively by less than the differences between ELF4 and ELF5 computed with the prior ensemble.

d. Imperfect model: Frequent identity observations with large error variance

The observations used here are the same as in section 4a. The forcing parameter F of the Lorenz96 model was changed from the default value of 8 to 5 for the

assimilating model to simulate model error. With forcing of 5, the Lorenz96 model is no longer chaotic, instead undergoing repeated vacillations. Figure 11 shows the $N = 20$ fifth ELF and the best GC localization function for this case along with the corresponding functions for the perfect model case described in section 4a. Localizations for the imperfect case are more compact. Some previous methods for estimating localization like the hierarchical filter (Anderson 2007) or sampling error correction (Anderson 2012) are unable to explore the impact of model error on localization. The RMSE for the second through fifth ELF assimilation experiments in this case (ranging from 1.605 to 1.607) are just slightly worse than the RMSE for the best GC localization (1.603) while the first ELF had larger RMSE (1.619).

e. Frequent identity observations with small error variance

This case assimilates identity observations with an error variance of 1 every model time step with a 20-member ensemble. The RMSE for the five ELF localizations in this case were all significantly larger than the RMSE for the best GC case. Figure 12 shows the fifth ELF and GC localization functions with half-widths of 25%, 35%, and 45% of the domain. The GC with 45% half-width produced the smallest RMSE. The ELF in Fig. 12 is closest to the GC 25% half-width and produced similar RMSE. The dependence of RMSE on GC half-width is fairly small in this case with RMSE of about 0.201 for the 45% case and 0.217 for no localization; the ELFs produce RMSE between 0.204 and 0.209. A number of additional localization functions were explored to understand why the ELFs did not do as well as the best GC cases. Spatially smoothing the noisy tails of the ELF localizations did not significantly change the RMSE. Using much longer assimilations (up to 20 000 steps instead of 5000 steps after discarding the first 1000
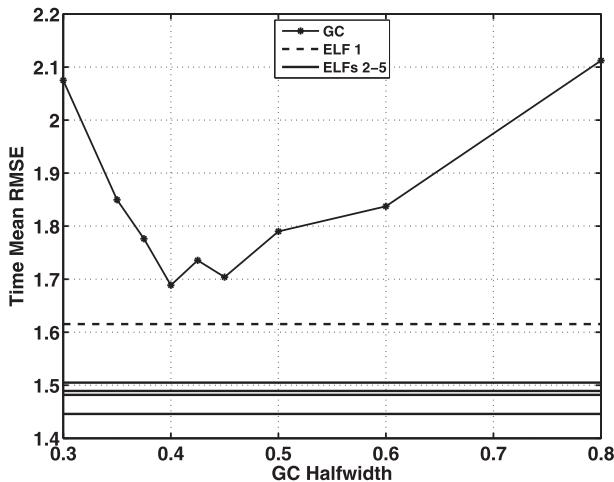
FIG. 10. The RMSE for five empirical localization functions and a variety of Gaspari–Cohn functions for 320 observations of the sum of 17 state variables with observation error variance 1 assimilated every time step with 20-member ensembles. Results for the empirical localization functions are independent of the horizontal axis.

steps) produced smoother ELFs that were just slightly broader than 5000-step ELFs, and this led to a slight reduction in RMSE. Artificially broadening the tails of the ELFs so that they retained their structure near the observation but were closer to the 45% GC function for distant states variables did result in RMSEs that were comparable to the GC results. This suggests that the ELFs in this case are too narrow for reasons that are not clear. It is likely that it becomes increasingly difficult to create ELFs when the impacts of localization are relatively small.

5. ELFs without knowing the truth

Results shown so far are based on (12) that requires knowing the true values of the state variable instances from an OSSE. If only observations are available, (14) can be used instead to compute a localization between subsets that contain pairs of observations. Because the observations are noisy, the estimate of the localization from a given assimilation length is noisier than when the truth is known.

Figure 13 shows the fifth ELF for the same observations as in section 4a using only observations of the state variable instances with $N = 20$ and the corresponding ELF and best GC functions using the truth. The ELFs are roughly similar in shape, but the observation only computation results in more spatial noise, especially far from the observation.

Here, all state variables are observed. In real assimilation experiments, observations of all state variables

are rarely available so this extension would have reduced value. However, it is possible to compute ELFs for pairs of observations for an OSSE as discussed in section 2b. Comparing results from an OSSE to results for the same subsets from a real assimilation might be useful in calibrating OSSE ELFs for use in real assimilations.

6. Discussion

Several aspects of the ELFs shown in previous sections require additional discussion. First, ELFs for identity observations do not have a value of 1 for zero displacement (when the observation and the state variable are identical). This is particularly evident in Fig. 12 but can be seen in most ELF figures. For zero displacement, the ELF is actually correcting for suboptimal estimates of the prior spread. If the prior spread is systematically too small in the OSSE generating the ELF, then observations receive too little weight. The resulting ELF will compensate for this by being larger than one; if
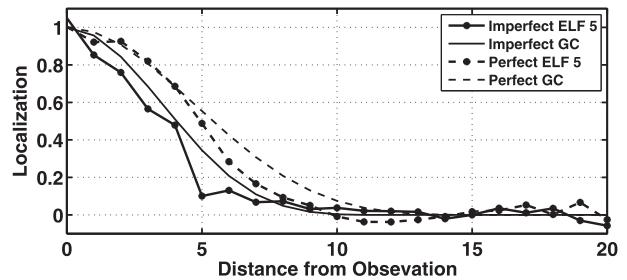

FIG. 11. The fifth empirical localization function and the Gaspari–Cohn function with the smallest associated RMSE for observations with error variance 16 assimilated every model time step with 20-member ensembles for a case with a perfect model and for a case with an imperfect model with forcing $F = 5$. The units for the horizontal axis are the number of grid intervals separating the state variable from the observation.
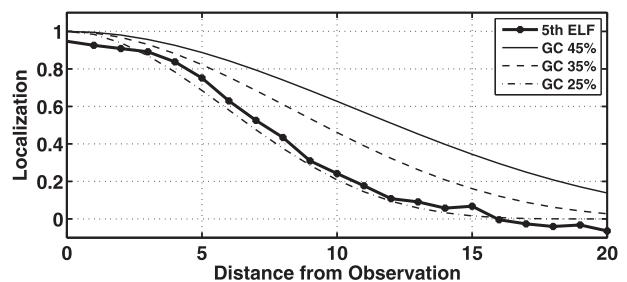

FIG. 12. The fifth empirical localization function and Gaspari–Cohn functions with half-widths of 0.45, 0.35, and 0.25 for observations with error variance 1 assimilated every model time step with 20-member ensembles. The units for the horizontal axis are the number of grid intervals separating the state variable from the observation.
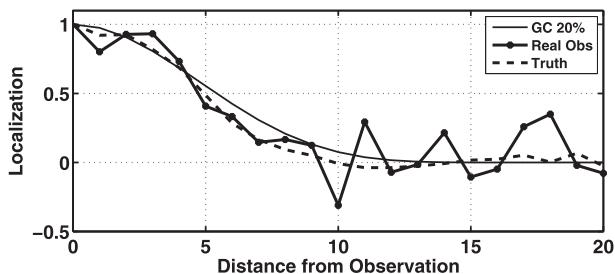
FIG. 13. The fifth empirical localization functions computed using knowledge of the true state and computed using only observations along with the Gaspari–Cohn function with the smallest associated RMSE for observations with error variance 16 assimilated every time step with 20-member ensembles. The perfect model and GC cases are as in Fig. 1. The units for the horizontal axis are the number of grid intervals separating the state variable from the observation.

the spread is too large, the ELF will be less than 1. The fact that most ELFs are close to 1 at zero distance indicates that the adaptive inflation scheme is doing a good job at producing appropriate prior spread.

Second, the first ELFs have significantly negative values of localization for all of the experiments that assimilate observations every model time step. This can be seen in Figs. 1, 5, and 9. However, significantly negative values were not found in the first ELF for experiments that assimilate observations every 12 time steps (section 4b). An explanation for the negative localizations can be found by comparing the 20-member OSSEs with no localization (used to compute the first ELF) for a case with frequent observations with high error variance (section 4a) and a case with infrequent observations with low error variance (section 4b). Examining the time series of the prior ensemble mean and the truth shows that the error changes sign on average every 5.3 (1.9) assimilation steps for the frequent (infrequent) observations. One can also look at the time series of the prior ensemble correlations between state variables and observations with displacements greater than 10. The correlations change sign on average every 6.0 (1.9) assimilation steps for the frequent (infrequent) observations. In the frequent observation case, there is a significant probability that spurious correlations for widely separated observations and state variables maintain the same sign for several consecutive assimilation times. Similarly, there is an enhanced probability that the increments to the observation will have the same sign for several consecutive assimilation times. These two facts combined lead to an enhanced probability that the erroneous increment induced in a remote state variable will have the same sign for several consecutive assimilation times. In other words, the observation is likely to push the state variable consistently away from the truth for a set of consecutive

steps. This leads to a localization that corrects this error by reversing the sign of the increment. In the low-frequency observation case, 12 model time steps are sufficient for the model to mostly eliminate consistent errors over consecutive assimilation times, so significant negative localizations are not found in the first ELF.

As noted earlier, negative localization values are set to 0 when an ELF is used as the localization in an OSSE. If the negative values from ELF1 are not set to 0, the time mean RMS error was found to be significantly higher for all high-frequency observation cases. For instance, ELF1 of the 20-member ensemble for high-frequency observations with large error variance (Fig. 1) produced RMS errors of 1.047 without negative values and 1.1696 with negative values. In addition, if one retains negative values, the subsequent ELF2, ELF3, etc., also have significantly negative values and produce larger RMS errors in OSSEs.

7. Conclusions

A method for producing localization functions for ensemble assimilation using the output of an OSSE has been described. In a low-order model, the method is able to produce localization functions that are competitive with the best Gaspari–Cohn localizations in most cases. In a case where the derived localization is notably non-Gaussian, results that are much better than those produced with the best Gaspari–Cohn are obtained.

There are challenges in extending this methodology to large atmospheric applications. First, large numbers of pairs of observations and state variables are needed to compute accurate localizations. As appropriate localizations get smaller, progressively larger numbers of pairs are needed to get accurate estimates. Ways of reducing or detecting noisy localization estimates will be essential for application to large models. Instead of computing localization for a large number of a priori selected subsets of observation and state variable pairs, one could also use a similar method to estimate optimal parameters of some functional form for localization. For instance, one could attempt to find an optimal half-width for a Gaspari–Cohn function. Also, the localizations in the simple model explored here are expected to be spatially homogenous, but this is definitely not the case for realistic atmospheric models.

A second challenge is extrapolating localizations computed with an OSSE to real observation assimilation. Results here suggest that the presence of model error will lead to narrower localization functions. Directly estimating an empirical localization when the truth is not known is likely to require very long assimilations and is probably not viable for large models.

Here, good localizations were estimated by running a sequence of OSSEs, each using the localization estimated from the previous OSSE. It may be possible to dynamically estimate the localization by finding an optimal localization for subsets over some moving window in a single OSSE. Initial experiments with Lorenz96 where the localization optimized for the last 1000 assimilation steps is used to localize the next assimilation step converge to localizations that are comparable to those discussed here. This dynamic estimation method might be useful for large model applications where an OSSE without some initial localization will not run without diverging or failing.

Future work will explore the application of the empirical localization to large global and regional atmospheric models. Of particular interest are cases where good localizations are expected to be significantly non-Gaussian. For instance, one might anticipate that low-level observations in mountainous regions have localizations that are partly determined by the topography. Good localizations for observations like satellite radiances or radio occultations that have forward operators that depend on many state variables may also be non-Gaussian. In these cases, even a relatively noisy empirical localization estimated from a short OSSE might provide improved assimilation results. Comparing the empirical localization results to group filter results in large models is also interesting. The group filter only accounts for localization due to ensemble sampling error, so differences from empirical localization should give insight into the importance of other error sources in the assimilation.

REFERENCES

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

——, 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.,* **131,** 634–642.

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

——, 2009a: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus,* **61A,** 72–83.

——, 2009b: Ensemble Kalman filters for large geophysical applications. *IEEE Contr. Syst. Mag.,* **29,** 66–82.

——, 2010: A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.,* **138,** 4186–4198.

——, 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.,* **140,** 2359–2371.

——, and N. Collins, 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.,* **24,** 1452–1463.

——, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.,* **90,** 1283–1296.

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,* **61A,** 84–96.

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

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.,* **126,** 1719–1724.

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.

Chen, Y., and D. S. Oliver, 2009: Cross-covariances and localization for EnKF in multiphase flow data assimilation. *Comput. Geosci.,* **14,** 579–601, doi:10.1007/s10596-009-9174-6.

Emerick, A., and A. Reynolds, 2010: Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. *Comput. Geosci.,* **15,** 251–269, doi:10.1007/s10596-010-9198-y.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

Furrer, R., and T. Bengtsson, 2007: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivariate Anal.,* **98,** 227–255.

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

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.,* **139,** 511–522.

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.

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

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

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.,* **82D,** 35–45.

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,** D09110, doi:10.1029/2010JD014673.

Kepert, J. D., 2009: Covariance localisation and balance in an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.,* **135,** 1157–1176.

Liu, H., J. L. Anderson, Y.-H. Kuo, and K. Raeder, 2007: Importance of forecast error multivariate correlations in idealized assimilations of GPS radio occultation data with the ensemble adjustment filter. *Mon. Wea. Rev.,* **135,** 173–185.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP-A comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.,* **129,** 3183–3204.

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

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.,* **130,** 2791–2808.

Miyoshi, T., S. Yamane, and T. Enomoto, 2007: Localizing the error covariance by physical distance within a local ensemble transform Kalman filter (LETKF). *SOLA,* **3,** 89–92.

Oke, P. R., P. Sakov, and S. P. Corney, 2007: Impacts of localization in the EnKF and EnOI: Experiments with a small model. *Ocean Dyn.,* **57,** 32–45.

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

Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.,* **133,** 1789–1807.

Whitaker, J. S., G. P. Compo, X. Wei, and T. M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.,* **132,** 1190–1200.

Zhang, Y., and D. S. Oliver, 2010: Improving the ensemble estimate of the Kalman gain by bootstrap sampling. *Math. Geosci.,* **42,** 327–345.