J2.4 A LOCAL ENSEMBLE KALMAN FILTER FOR THE NCEP GFS MODEL

Istvan Szunyogh *
Eric J. Kostelich†, Gyorgyi Gyarmati, Brian R. Hunt, Edward Ott, Aleksey V. Zimin, Eugenia Kalnay,
Dhanurjay Patil, and James A. Yorke
University of Maryland, College Park, MD 20742

Abstract

This paper outlines the basic concept and mathematical formulation of the Local Ensemble Kalman Filter (LEKF, Ott et al. 2002 and 2003) data assimilation scheme. Some important properties of the scheme are illustrated by numerical experiments with the Lorenz-96 model. Initial implementations of the LEKF on the National Centers for Environmental Prediction Global Forecast System (NCEP GFS) are discussed. Tests of these implementations under the perfect model hypothesis are conducted.

1. INTRODUCTION

The accuracy of a data assimilation scheme strongly depends on the accuracy of the assumed background error statistics. Since the true background errors are not known, the implementation of a data assimilation system requires the development of statistical models that can provide an estimate of the background error covariance matrix.

In the case of linear dynamics, the mathematically consistent technique to define a background error covariance matrix is the Kalman Filter (Kalman 1960; Kalman and Bucy 1961) which utilizes the dynamical equations to evolve the most probable state and the error covariance matrix in time. In the case of linear systems with unbiased normally distributed errors the Kalman Filter provides estimates of the system state that are optimal in the mean square sense. The method has also been adapted to nonlinear systems, but, in this case, optimality no longer applies. Although the Kalman Filter approach has been successfully implemented for a wide range of applications and has been considered for atmospheric data assimilation for a long while (e.g., Ghil et al. 1981) the computational cost involved does not allow for an operational implementation in the foreseeable future (see Daley 1991 for details).

Currently the most popular approach toward reducing the cost of the Kalman Filter is to use a relatively small (10-100 member) ensemble of background forecasts to estimate the background error covariances (e.g. Evensen 1994; Houtekamer and Mitchell 1998, 2001; Anderson 2001; Bishop et al. 2001; Hamill et al. 2001; Whitaker and Hamill 2002; Keppenne and Rienecker 2002). In ensemble-based data assimilation schemes the ensemble of background forecasts is generated by using initial conditions distributed according to the result of the previous analysis.

The main difference between the existing ensemble-based schemes is in the generation of the analysis ensemble. One family of schemes is based on *perturbed observations* (Evensen and van Leeuwen 1996; Houtekamer and Mitchell 1998, 2001; Hamill and Snyder 2000, 2001, Keppenne and Rienecker 2002). In this approach, the analysis ensemble is obtained by assimilating a different set of observations to each member of the background ensemble. The different sets of observations are created by adding random noise to the real observations, where the random noise component is generated according to the observational error covariance matrix. Such a system has been developed at the Canadian Meteorological Service and was recently reported to break even with the operational 3D-Var scheme (Houtekamer et al., 2003).

The other family of schemes, the Kalman square-root filters, uses a different approach to reduce the size of the ensemble. These techniques do the analysis only once, to obtain both the mean analysis and the analysis error covariance matrix. Then the analysis ensemble perturbations (to the mean analysis) are generated by linearly transforming the background ensemble perturbations to a set of vectors that can be used to represent the analysis error covariance matrix. Thus, the analysis is confined to the subspace of the ensemble. Since there is an infinite set of analysis perturbations that can be used to represent the analysis error covariance matrix, many different schemes can be derived following this approach (Tippett et al., 2002). Existing examples of the square root filter approach are the Ensemble Transform Kalman Filter (Bishop et al. 2001), the Ensemble Adjustment Filter (An-

* *Corresponding author address:* University of Maryland, IPST, CSS Building, College Park, MD 20742-2431, USA
† *Current address:* Arizona State University, Tempe, AZ 85287, USA

derson 2001), and the Ensemble Square-root Filter (Whitaker and Hamill 2001).

2. THE BASIC CONCEPT

The scheme we propose is a Kalman square-root filter (Ott et al., 2002, 2003). In a sense, our technique is related to previous work that attempted to construct a simplified Kalman filter by explicitly taking into account the dominant unstable directions of the state space (Kalnay and Toth 1994; Fisher 1998). The most important difference between our scheme and the other Kalman square-root filters is that the LEKF is not a sequential data assimilation algorithm. This difference is illustrated by a pair of schematic figures (figures 1 and 2).

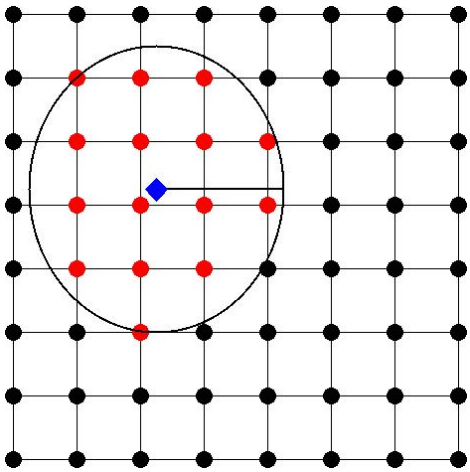

FIG. 1: Illustration of the process of updating the state estimate in a sequential data assimilation scheme. Observations (blue diamond) are sequentially used to update the state estimate is updated at all grid points (red grid points) within the correlation distance (black circle) from the observations. Background information is used from the grid points at which the state is updated. At grid points beyond the correlation distance (black grid points) the state estimate is not updated.

On the one hand, a sequential scheme updates the model variables at all grid points within a predefined correlation length from the observations. This means that the analysis at a given grid-point is updated iteratively until all observations within the correlation length from the grid point have been assimilated. On the other hand, the LEKF is based on the construction of local regions about each grid point. The basic idea is that we do the analysis at each grid point simultaneously using the state variables and all observations in the local region centered at that point. (A similar computational model

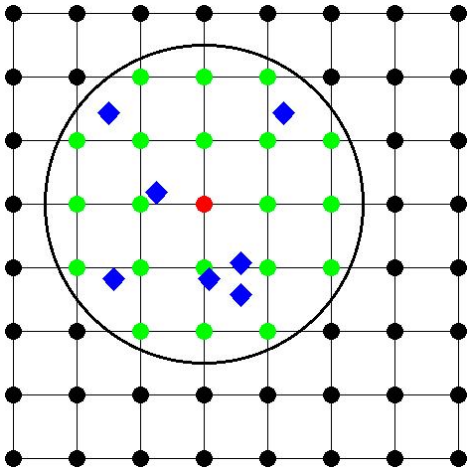

FIG. 2: Illustration of the process of updating the state estimate in the LEKF. The knowledge of all observations within the local region (blue diamonds) is simultaneously used to update the state is at the center of the local region (red grid point) using background information from all grid points within the local region (green grid points). The shape of the local region is somewhat arbitrary. Here we have indicated it as circular, but in section 5, for convenience of implementation, we will use a square shape.

was outlined in Anderson (2001), where the the equivalent of the local region was called compute domain.)

In the LEKF, the assimilations in each local grid point are independent, thus facilitating a massively parallel update approach using a straightforward static load balancing strategy between the computing units. We believe that this is an advantage over the sequential schemes, which can be implemented in a parallel computational environment, but for which load balancing is problematic. This is due to the feature of the sequential schemes that the main computational cost is proportional to the number of observations. In contrast, the computational cost of the LEKF is proportional to the number of ensemble members. The suitable number of ensemble members depends on the complexity (dimensionality) of the dynamics of the model, which does not change for a given model and is independent of the number of observations.

Another potential disadvantage of the sequential schemes arises when the neighboring observations have correlated errors throughout a large area. Distributing such observations between the computing unit would introduce artificial zero correlations between the distributed subsets of data. The LEKF does not have this problem, since it only assumes that correlations between observations in-

side and outside of the local regions have no effect on the analysis at the center of the local regions. These correlations are taken into account in other local regions.

While the sequential assimilation of observations has been shown to be a powerful approach for the case when the number of observations is much lower than the number of grid points (Houtekamer et al. 2003; Whitaker et al. 2003); our scheme will be of particular interest when the dimensionality of the model dynamics in the local regions is typically much lower than the number of observations. In the future this condition will become progressively better satisfied as ever-increasing amounts of remotely-sensed data become available. The ultimate test of our scheme is whether it can be used to build an efficient data assimilation scheme. Later in this paper, we show some preliminary evidence that this is indeed the case.

3. MATHEMATICAL FORMULATION

For each grid point, we define a corresponding *local region* that consists of all grid points within a suitably prescribed neighborhood of that grid point. While the size and shape of the local regions are somewhat arbitrary, we desire that the local regions be sufficiently large that on average, include all grid points whose state field values have significant correlation with the state fields at the corresponding central grid point. In general, the model state within a local region is specified by all the model field variables evaluated at all grid points within the local region.

Let $\mathbf{x}_l(m, n, o)$ the d-dimensional local vector representing the model state within the local region centered at the grid point (m, n, o). The construction of this local vector can be represented by a linear localization operator$\mathbf{L}$,

$$\mathbf{x}_l(m, n, o, t) = \mathbf{L}\mathbf{x}_g(t). \tag{1}$$

where the vector $\mathbf{x}_g$ represents the state of the model on the global three-dimensional grid. Since all the analysis operations take place at a fixed time t and are repeated for all local regions, we will suppress the t, m, n, and o dependence of all vectors and matrices introduced henceforth.

We solve the Kalman-filter equations for the local vectors. The most probable state of the local background vector $\mathbf{x}_l^b$ is $\overline{\mathbf{x}}_l^b$ and the local background vector is $\mathbf{P}_l^b$. As explained subsequently, the rank of the d by d covariance matrix $\mathbf{P}_l^b$ for our approximate probability distribution function is taken to be much less than d.

In the data assimilation procedure we describe in this paper, the background error covariance matrix $\mathbf{P}_l^b$ and the most probable background state $\overline{\mathbf{x}}_l^b$ are derived from a $k+1$ member ensemble of global state field vectors $\mathbf{x}_g^{b(i)}$, $i = 1, 2, \cdots, k+1$; $k \geq r \geq 1$. The most probable local state is given by

$$\overline{\mathbf{x}}_l^b = \mathbf{L}\left[(k+1)^{-1}\sum_{i=1}^{k+1}\mathbf{x}_g^{b(i)}\right]. \tag{2}$$

The local background error covariance matrix $\mathbf{P}^b$ is estimated by $\mathbf{P}_l^b$ given by

$$\mathbf{P}_l^b = k^{-1}\sum_{i=1}^{k+1}\delta\mathbf{x}_l^{b(i)}\left(\delta\mathbf{x}_l^{b(i)}\right)^T, \tag{3}$$

where the superscribed T denotes transpose, and

$$\delta\mathbf{x}_l^{b(i)} = \mathbf{L}\mathbf{x}_g^{b(i)} - \overline{\mathbf{x}}_l^b. \tag{4}$$

It is also useful to introduce the notation

$$\mathbf{X}_l^b = (k)^{-1/2}\left[\delta\mathbf{x}_l^{b(1)} \mid \delta\mathbf{x}_l^{b(2)} \mid \cdots \mid \delta\mathbf{x}_l^{b(k+1)}\right], \tag{5}$$

in terms of which (3) can be rewritten,

$$\mathbf{P}_l^b = \mathbf{X}_l^b\mathbf{X}_l^{bT}. \tag{6}$$

We anticipate that we can approximate the background error covariance matrix by one of lower rank than d, and this motivates our assumption that an ensemble of size of $k+1$, where $k+1$ may be substantially less than d, will be sufficient to yield a good approximate representation of the background covariance matrix. [In the version of the LEKF method reported here we take the rank r of our approximate $\mathbf{P}_l^b$ to be $r = k$, but we note that the theory for the more general choice $r \leq k$ is also available (Ott et al., 2003).] Since $\mathbf{P}_l^b$ has rank k, it has k positive eigenvalues. Let the eigenvalues of the matrix $\mathbf{P}_l^b$ be denoted by $\lambda^{(j)}$, where the labeling convention for the index j is

$$\lambda^{(1)} \geq \lambda^{(2)} \geq \ldots \geq \lambda^{(r)} \geq \cdots \geq \lambda^{(k)}. \tag{7}$$

Since $\mathbf{P}_l^b$ is a symmetric matrix, it has k orthonormal eigenvectors $\left\{\mathbf{u}^{(j)}\right\}$ corresponding to the k eigenvalues (7). Thus

$$\mathbf{P}_l^b = \sum_{j=1}^{k}\lambda^{(j)}\mathbf{u}^{(j)}(\mathbf{u}^{(j)})^T. \tag{8}$$

Since the size of the ensemble ($k+1$) is envisioned to be much less than the dimension of $\mathbf{x}_l^b$ (d), the computation of the eigenvalues and eigenvectors of $\mathbf{P}_l^b$ is most effectively done in the basis of the ensemble vectors. That is, we consider the

eigenvalue problem for the $(k+1)\times(k+1)$ matrix $\mathbf{X}^{bT}\mathbf{X}^{b}$, whose nonzero eigenvalues are those of $\mathbf{P}_l^b$ [7] and whose corresponding eigenvectors left-multiplied by $\mathbf{X}^b$ are the k eigenvectors $\mathbf{u}^{(j)}$ of $\mathbf{P}_l^b$.

For the purpose of subsequent computation, we consider the coordinate system for the k dimensional space determined by the basis vectors $\{\mathbf{u}^{(j)}\}$. We call this the *internal coordinate system* for the k-dimensional *analysis space*. We denote the projection of vectors into the k-dimensional space and the restriction of matrices to the same space by suppressing the subscript l. The operator of this projection is

$$\mathbf{Q} = \left\{\mathbf{u}^{(1)}|\mathbf{u}^{(2)}|\cdots|\mathbf{u}^{(k)}\right\}. \tag{9}$$

For instance, for the d dimensional local background vector $\mathbf{x}_l^b$, the vector $\mathbf{x}_b$ is an k dimensional column vector given by

$$\mathbf{x}^b = \mathbf{Q}^T\mathbf{x}_l^b. \tag{10}$$

Note that this operation consists of both projecting $\mathbf{x}_l^b$ into the r-dimensional space and changing to the internal coordinate system. Similarly, for a d by d matrix, such as the local background covariance matrix $\mathbf{P}_l^b$, the matrix $\mathbf{P}^b$ is k by k and given by

$$\mathbf{P}^b = \mathbf{Q}^T\mathbf{P}_l^b\mathbf{Q}. \tag{11}$$

We also note, that in the internal coordinate system $\mathbf{P}^b$ is diagonal,

$$\mathbf{P}^b = \mathrm{diag}\left(\lambda^{(1)}, \lambda^{(2)}, ..., \lambda^{(k)}\right), \tag{12}$$

and thus it is trivial to invert.

Let $\mathbf{x}_l^a$ be the random variable at the current analysis time t representing the local vector after knowledge of the observations and background mean are taken into account. For simplicity, we assume that all observations collected for the current analysis were taken at the same time t. [A simple technique to extend the scheme to the assimilation of of asyncronous observations is presented in Sauer et al. (2004).] Let $\mathbf{y}_l^o$ be the vector of current observations within the local region, and $\mathbf{R}$ be the observational error covariance matrix of those observations. An ideal (i.e., noiseless) measurement is a function of the true atmospheric state. Considering measurements within the local region, we denote this function $\mathcal{H}_l(\cdot)$. That is, if the true local state is $\mathbf{x}_l^a$, then the error in the observation is $\mathbf{y}_l^o - \mathcal{H}_l(\mathbf{x}_l^a)$. Assuming that the true state is near the mean background state $\overline{\mathbf{x}}_l^b$, we approximate $\mathcal{H}_l(\mathbf{x}_l^a)$ by linearizing about $\overline{\mathbf{x}}_l^b$,

$$\mathcal{H}_l(\mathbf{x}_l^a) \approx \mathcal{H}_l(\overline{\mathbf{x}}_l^b) + \mathbf{H}_l\Delta\mathbf{x}_l^a, \tag{13}$$

where

$$\Delta\mathbf{x}_l^a = \mathbf{x}_l^a - \overline{\mathbf{x}}_l^b, \tag{14}$$

and the matrix $\mathbf{H}_l$ is the Jacobian matrix of partial derivatives of $\mathcal{H}_l$ evaluated at $\overline{\mathbf{x}}_l^b$. (If there are s scalar observations in the local region at analysis time t, then $\overline{\mathbf{y}}_l^o$ is s dimensional and the rectangular matrix $\mathbf{H}_l$ is s by d). The data assimilation step determines $\overline{\mathbf{x}}_l^a$ (the *local analysis*) and $\mathbf{P}_l^a$ (the *local analysis covariance matrix*).

The most probable value of $\Delta\mathbf{x}^a$ is

$$\Delta\overline{\mathbf{x}}^a = \mathbf{P}^a\mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{y}^o - \mathcal{H}(\overline{\mathbf{x}}^b)\right), \tag{15}$$

Here $\mathbf{H} = \mathbf{H}_l\mathbf{Q}$ maps variables from the analysis space to the observation space, $\Delta\mathbf{x}^a = \mathbf{Q}^T\Delta\mathbf{x}_l^a$ is the analysis increment in the k-dimensional analysis space, and $\mathbf{P}^a$ is the analysis error covariance matrix in the same k-dimensional space. In (15), $\mathbf{P}^a$ is determined from (e.g., Kalnay 2003)

$$\mathbf{P}^a = \mathbf{P}^b\left[\mathbf{I} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{P}^b\right]^{-1}, \tag{16}$$

Finally, going back to the local space representation, we have

$$\overline{\mathbf{x}}_l^a = \mathbf{Q}\Delta\overline{\mathbf{x}}^a + \overline{\mathbf{x}}^b. \tag{17}$$

The components of the most probable global analysis field $\overline{\mathbf{x}}_g^a$ at the grid point (m, n, o) are obtained by first selecting the local analysis vector $\overline{\mathbf{x}}_l^a(m, n, o)$ associated with the local region centered at $(m.n, o)$, and then copying the components of $\overline{\mathbf{x}}_l^a(m, n, o)$ at the central grid point.

Our only remaining task is to obtain an ensemble of global analysis fields $\left\{\mathbf{x}_g^{a(i)}\right\}$; $i = 1, 2, \cdots, k+1$ based on the local analysis information, $\mathbf{P}^a$ and $\overline{\mathbf{x}}^a$. Once these fields are determined, they can be used as initial conditions for the atmospheric model. Integrating these global fields forward in time to the next analysis time $t + \Delta t$, we obtain the background ensemble $\left\{\mathbf{x}_g^{b(i)}(t + \Delta t)\right\}$. This task is accomplished by first finding the $(k+1)$ local analysis perturbations $\delta\mathbf{x}_l^{a(i)}$,

$$\delta\mathbf{x}_l^{a(i)} = \mathbf{Q}\delta\mathbf{x}^{a(i)}, \tag{18}$$

then forming the local analysis ensemble

$$\mathbf{x}_l^{a(i)} = \overline{\mathbf{x}}_l^a + \delta\mathbf{x}_l^{a(i)}. \tag{19}$$

and finally obtaining the members of the global analysis ensemble $\left\{\mathbf{x}_g^{a(i)}\right\}$; $i = 1, 2, \cdots, k+1$, following the same strategy we used to create the global mean analysis based on the local information. Integrating these global fields forward in time to the next analysis time $t + \Delta t$, we obtain the background ensemble $\left\{\mathbf{x}_g^{b(i)}(t + \Delta t)\right\}$.

The local analysis perturbations $\delta\mathbf{x}^{a(i)}$ are found by linearly combining the local background perturbations in the k-dimensional analysis space

$$\mathbf{X}^a = \mathbf{X}^b\mathbf{Y}. \tag{20}$$

where

$$\mathbf{X}^{a,b} = (k)^{-1/2}\left\{\delta\mathbf{x}^{a,b(1)}|\delta\mathbf{x}^{a,b(2)}|\cdots|\delta\mathbf{x}^{a,b(k+1)}\right\}. \tag{21}$$

and

$$\mathbf{Y} = \left[\mathbf{I} + \mathbf{X}^{bT}\left(\mathbf{P}^b\right)^{-1}\left(\mathbf{P}^a - \mathbf{P}^b\right)\left(\mathbf{P}^b\right)^{-1}\mathbf{X}^b\right]^{1/2}. \tag{22}$$

This construction of the local analysis perturbations has the desirable properties that *(i)* it does not distort the mean of the analysis ensemble

$$\sum_{i=1}^{k+1}\delta\mathbf{x}^{a(i)} = \sum_{i=1}^{k+1}\delta\mathbf{x}_l^{a(i)} = \mathbf{0}, \tag{23}$$

(ii) it correctly represents the analysis uncertainty

$$\mathbf{P}^a = \mathbf{X}^a\mathbf{X}^{aT}. \tag{24}$$

and *(iii)* it preserves the smoothness of the background ensemble fields as closely as possible, by satisfying the constraint that the functional

$$\mathcal{F}(\delta\hat{\mathbf{x}}_{mn}^{a(i)}) = \sum_{i=1}^{k'+1}\left\|\delta\hat{\mathbf{x}}_{mn}^{a(i)} - \delta\hat{\mathbf{x}}_{mn}^{b(i)}\right\|^2 \tag{25}$$

$$= \sum_{i=1}^{k+1}\left[\delta\hat{\mathbf{x}}_{mn}^{a(i)} - \delta\hat{\mathbf{x}}_{mn}^{b(i)}\right]^T\mathbf{D}^{-1}\left[\delta\hat{\mathbf{x}}_{mn}^{a(i)} - \delta\hat{\mathbf{x}}_{mn}^{b(i)}\right], \tag{26}$$

is minimal when the metric $\mathbf{D}^{-1}$ is chosen to be $\mathbf{D}^{-1} = \mathbf{P}^a$ or $\mathbf{D}^{-1} = \mathbf{P}^b$. Constraint (25) ensures that changes to the physical fields, introduced by transforming the background perturbations to ensemble perturbations, are minimal with respect to the metric $\mathbf{D}^{-1}$. Without imposing constraint *(iii)* the transformation matrix $\mathbf{Y}$ would not be uniquely defined (see Ott 2003 for details). In fact, past work (Tippett et al., 2003) provide no basis for choosing among the solutions of (23) and (25).

4. TESTS WITH THE LORENZ-96 MODEL

Initial testing of the LEKF scheme was carried out with Lorenz-96 (L96) model (Lorenz 1996; Lorenz and Emanuel 1998),

$$\frac{dx_m}{dt} = (x_{m+1} - x_{m-2})x_{m-1} - x_m + F. \tag{27}$$

Here, $m = 1,\cdots,M$, where $x_{-1} = x_{M-1}$, $x_0 = x_M$, and $x_{M+1} = x_1$. This model mimics the time evolution of an unspecified scalar meteorological quantity, x, at M equidistant grid points along a latitude circle. We solve (27) with a fourth-order Runge-Kutta time integration scheme with a time step of 0.05 nondimensional unit (which may be thought of as nominally equivalent to 6-h in real world time assuming that the characteristic time scale of dissipation in the atmosphere is 5-days; see Lorenz 1996 for details).

For our chosen forcing, $F = 8$, the steady state solution, $x_m = F$ for $m = 1,\cdots,M$, is linearly unstable. This instability is associated with unstable dispersive waves characterized by westward (i.e., in the direction of decreasing m) phase velocities and eastward group velocities. Lorenz and Emanuel (1998) demonstrated by numerical experiments for $F = 8$ and $M = 40$ that the x field is dominated by a wave number 8 structure, and that the system is chaotic; it has 13 positive Lyapunov exponents, and its Lyapunov dimension (Kaplan and Yorke 1979) is 27.1. We carried out experiments with three different size systems ($M = i \times 40$, $i = 1, 2, 3$) and found that increasing the number of variables did not change the wavelength, i.e. the x fields were dominated by wavenumber $i \times 8$ structures.

The 40-variable version of the L96 model was also used by Whitaker and Hamill (2002) to validate their ensemble square root filter (EnSRF) approach. In designing our numerical experiments we follow their approach of first generating the 'true state', $x_m^t(t)$, $m = 1,\cdots,M$, by a long (40,000 time-step) model integration; then creating 'observations' of all model variables at each time step by adding uncorrelated normally distributed random noise with unit variance to the 'true state' (i.e., $\mathbf{R}_m = \mathbf{I}$). (The rms random observational noise variance of 1.00 is to be compared with the value 3.61 of the time mean rms deviation of solutions, $x_m(t)$, of (27) from their mean.) We found that our results were the same for Gaussian noise and for truncated Gaussian noise (we truncated at three standard deviations). The effect of reduced observational networks is studied by removing observations one by one, starting from the full network, at randomly selected locations. The reduced observational networks are fixed for all experiments. That is, the difference between a network with O observations and another with $O + 1$ observations is that there is a fixed location at which only the latter takes observations.

The observations are assimilated at each time step, and the accuracy of the analysis is measured by the time mean of the rms error,

$$E = \left(\frac{1}{M}\sum_{m=1}^{M}(\overline{x}_m^a - x_m^t)^2\right)^{1/2}. \tag{28}$$

In order to the assess the skill of our data assimilation scheme in shadowing the true state, we considered three alternative schemes for comparison.

Full Kalman filter. For the sake of comparison with our local ensemble Kalman filter results, we first establish a standard that can be regarded as the best achievable ensemble Kalman filter result that could be obtained given that computer resources placed no constraint on computations of the analysis. (In contrast with operational weather prediction, for our simple M-variable Lorenz model, this is indeed the case.) For this purpose, we considered the state $\mathbf{x}(t) = \big(x_1(t), x_2(t), \cdots, x_M(t)\big)$ on the entire domain rather than on a local patch.

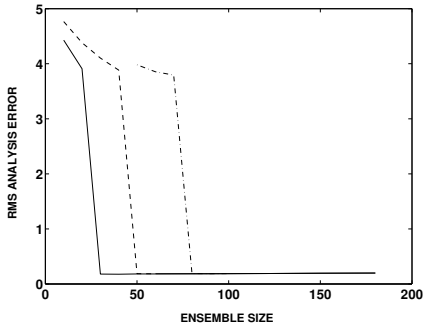

FIG. 3: The rms error of the full Kalman filter as function of the number of ensemble members. Shown are the results for M=40 (solid line), M=80 (dashed line), and M=120 (dotted dashed line).

Figure 3 shows, for the case in which all $M = 40$ variables are observed, that by increasing the number of ensemble members the time mean of E converges to 0.20 regardless of M. The only difference between the different size systems (characterized by different values of M) is that more ensemble members are required to reach the minimum value for the larger systems. We refer to 0.2 as the *"optimal" error*, and we regard it as a comparison standard for our local Kalman filter method. (However, we note that it is not truly optimal since Kalman filters are rigorously optimal only for linear dynamics.)

Conventional method. We designed another comparison scheme that we call the *conventional method*, to obtain an estimate of the analysis error that can be expected from a procedure analogous to a 3D-Var scheme adapted to the L96 model. In this scheme, only the best estimate of the true state is sought (not an ensemble of analyses) using a constant estimate of the background error covariance matrix that does not change with time or position. This background error covariance matrix was determined by an iterative process (see Ott et al. 2003) and was recomputed for each observational networks. The estimate obtained this way is not necessarily optimal in the sense of providing the smallest possible analysis error of any constant background error matrix, but it has the desirable feature that the background error statistics are correctly estimated by the analysis scheme. This is a big advantage compared to the operational schemes, for which the estimate of the background error covariance matrix has to be computed by rather ad hoc techniques, since the true state, and therefore the true background error statistics, are not known. Thus, it might be assumed that our "conventional method" provides an estimate of the analysis error that is of good accuracy as compared to analogous operational schemes.

Direct insertion. We now give a third standard designed to decide whether the data assimilation schemes provide any useful information compared to an inexpensive and simple scheme, not requiring matrix operations. This scheme updates the state estimate by replacing the background with the observations, where observations are available, and leaving the background unchanged, where there are no observations.

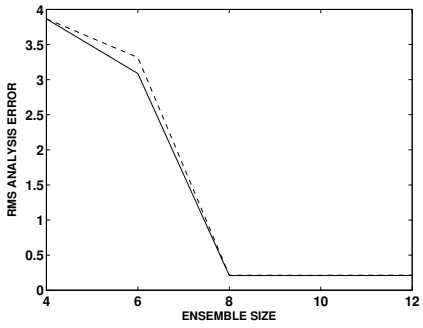

FIG. 4: The rms error of the local ensemble Kalman filter as function of the number of ensemble members. Shown are the results for M=40 (solid line), and for M=80 and M=120 which coincide (dashed line).

Local ensemble Kalman filter Figure 4 shows, for the case when all M=40 variables are observed, that when the ensemble has at least eight members, the analysis error settles at the level (0.2) of the "optimal" scheme, independently of the number of variables. This is roughly consistent with the supposition of an effective correlation length for the dynamics that is less than M. Thus our method appears to be effective on large systems of this type. Moreover, the (non-parallelized) analysis computa-

tional time scales linearly with the number of local regions (i.e., with M). This favorable scaling is to be expected, since the analysis computation size in each local region is independent of M.

We note, that the aforementioned scaling property of the local Kalman filter is in contrast to the behavior of the full Kalman filter, which requires many more members, and also an increasing number of members for an increasing number of variables, to achieve the "optimal" precision. This demonstrates the potential superiority of the local Kalman filter (and presumably other methods that use covariance localization) in terms of computational efficiency when applied to large systems. On the other hand, it also means that, since the minimum error was independent of M, it suffices to use the smallest, 40-variable, system for further experimentation.

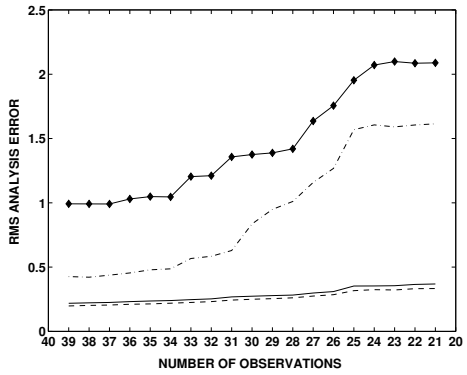

FIG. 5: The rms error of the different analysis schemes as function of the number of observations. Shown are the results for the full Kalman filter [4% variance inflation] (dashed line), conventional scheme (dashed-dotted line), direct insertion (solid line with diamonds), and the local ensemble Kalman filter [3% variance inflation] (solid line).

The four data assimilation schemes (local ensemble Kalman Filter, full Kalman filter, conventional method, and direct insertion) were compared for different numbers of observations (figure 5). The two Kalman filter schemes give almost identical error results, although the full Kalman filter has a very small advantage. The two Kalman filter schemes and the conventional data assimilation scheme are always more accurate than direct insertion, indicating that they are always able to retrieve nontrivial, useful information about the true state. The two Kalman filter schemes, in addition, have a growing advantage over the conventional scheme as the number of observations is decreased. This shows that, as the observational network and the background error become more inhomogeneous, the adaptive nature of the background error covariance matrix in the Kalman filters leads to a growing advantage over the static schemes.

	vertical localization	NO	PG
blue	no	80	100%
green	no	80	10%
red	no	40	10%
magenta	yes	40	10%

Table 1: Parameters of the different implementations of the LEKF scheme on the NCEP GFS.

5. INITIAL IMPLEMENTATION ON THE NCEP GFS

Tests have been carried out with four different implementations of the LEKF algorithm on the T62 horizontal resolution, 28-level, 2001 version of the NCEP GFS. In three of these implementations the localization is done only in the horizontal direction, while in the fourth implementation vertical localization is also included. The number of ensemble members (NO) and portion of grid points observed (PG) also varies between the implementations. In figures 6-8, the different implementations are distinguished by colors and the parameters associated with the different colors are listed in table 1 (the black line always represents the rms observational error).

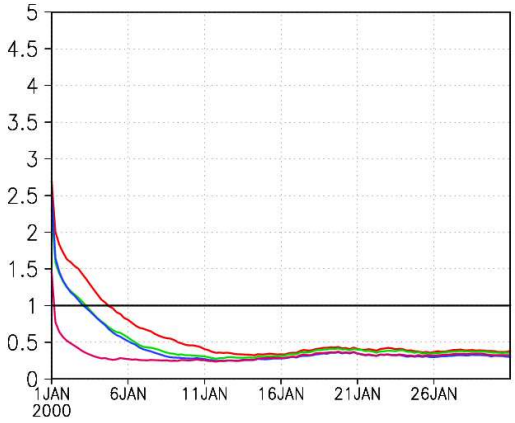

FIG. 6: The global rms error of the temperature analysis at the 50 kPa pressure level (y-axis) as function of time (x-axis.

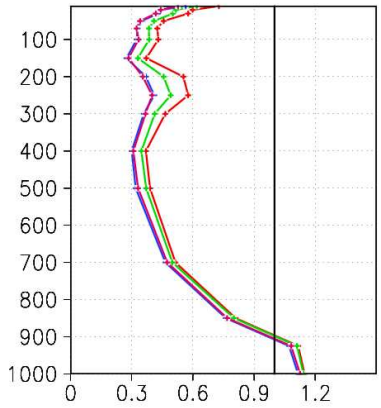

FIG. 7: The global rms error in the analysis of the temperature (x-axis) as function of altitude (y-axis) [hPa].

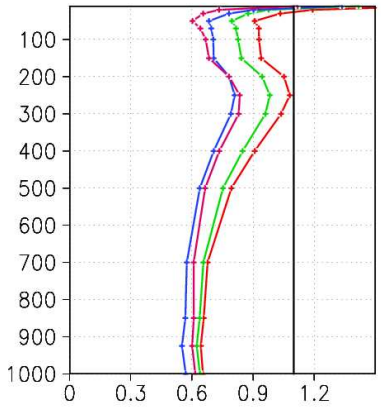

FIG. 8: The rms error in the analysis of the zonal wind component in the NH extratropics ($30^0N - 70^0N$, x-axis) as function of altitude (y-axis) [hPa].

The horizontal localization is done on the 144-by-73 grid points of the Gaussian model grid, and, at the poles, in order to avoid introducing artificial singularities extra latitudes were also added, making the proper definition of the local regions possible on the entire globe. Visually, one may think of our approach as defining local regions on polar stereographic map projections near the poles. In the experiments reported here, the local regions are uniformly defined by 9-by-9 horizontal grid points. When vertical localization is used, the number of model levels in the local region varies with the altitude. In layers, where the vertical gradient of the atmospheric fields are bigger (e.g., near the surface and the jet layer) the local regions consist of fewer layers.

In our experiments, the "true" state was generated by a 30-day integration of the T62 GFS model, started from the operational NCEP analysis at 0000 UTC on 1 January 2000. The observed data were generated by adding zero-mean, Gaussian random noise (observational error) to the true state. The variances of the errors are 1 K for the temperature, 1.1 m/s for the two horizontal components of the wind, and 1 hPa for the surface pressure. The humidity and ozone variables were not observed. Neither were surface analyses prepared; the lower boundary conditions were substituted from the nature run.

As we expected, the scheme proved to be highly efficient. The assimilation of 1.5×10^6 simulated observations (wind, temperature, and surface pressure), using a 40-member ensemble and horizontal localization, takes about 6 minutes of wall-clock time on a Beowulf cluster of forty 2.8-GHz Intel Xeon processors, a reasonably fast result for a Kalman filter of this resolution. When the number of ensemble members was increased to 80, the wall-clock time became about 12 minutes. When the ensemble had 40 members and localization in the vertical direction was also included, the wall-clock time was about 10 minutes.

We found that the analysis error converges to its asymptotic value for all variables within a few days (an example is shown in figures 6). The asymptotic value of the analysis error is typically smaller than the observational uncertainty. The only exceptions are the temperature in the boundary layer and the wind components in the jet layer. These are the vertical layers, where the variables have the largest gradient. Having more observations and ensemble members, and/or using vertical localization all increase the speed of convergence. The effect of vertical localization is especially striking, considering that it leads to a much faster convergence than the combined effect of doubling the ensemble size and increasing the number of observations by a factor of ten. This property is not totally unexpected, since the vertical localization decreases the dimension of the local regions (d), allowing for a better local representation of the background structures by the relatively small ensembles. The effect of changing the parameters on the asymptotic values of the analysis errors is less dramatic, though the magenta and the blue im-

plementations are clearly superior to the other two. The difference between the two runs (magenta and blue) is practially negligible, except near to the top of the model atmosphere, where the the vertical localization leads to a superior performance.

The accuracy of the analysis in the jet layer clearly improves with increasing the number of ensemble members and observation and/or introducing vertical localization. This leaves us with only one major concern regarding the performance of our initial implementations of the LEKF scheme: the large temperature error at the surface. A detailed analysis of this problem showed the following interesting features: the large global error is due to large errors over land; the large rms error is due to a large bias (mean error) component; the error is essentially identical for the different implementations, except for the first 5-10 days, when the error is much smaller than the observational uncertainty and the better (blue and magenta) implementations have a clear advantage. If our experiments were not conducted under the perfect model hypothesis, we would almost certainly conclude that the large temperature errors near the surface are due to a model bias. Having a perfect model experimental design, however, we have no other choice than to accept that the errors at the surface are due to a formulation problem in our scheme. We suspect that this problem is due to the strong influence of the Earth's surface in shaping the background error covariance near the surface. Since this influence is more closely related to the boundary conditions than to the dynamical evolution of the flow, the Kalman filter has difficulties with tracking the errors near the surface. This picture is also supported by our forecast experiments (not presented in this paper), which indicate that the temperature errors near the surface are not growing, i.e., they are not associated with instabilities. Nevertheless, the fact that the LEKF was able to efficiently correct the temperature errors near the surface in the first few days, makes us believe that these errors can be reduced by a randomized additive variance inflation of the temperature component of the background ensemble perturbations (Corazza et al. 2002). Finally, we note that this example illustrates the importance of performing the initial testing under the perfect model assumption.

6. CONCLUSIONS

Our ultimate goal is to develop the present implementation of the LEKF method on the NCEP GFS into an operationally attainable Kalman filter, capable of assimilating a wide range of meteorological observations, including those collected by remote-sensing instruments. Our experiments so far indicate that the the algorithm has the computationally efficiency required to accomplish this ambitious goal. Our plan is to proceed through completing the following tasks:

1. Elimination of the relatively large temperature analysis errors near the surface, by introducing and tuning an additive random variance inflation scheme.

2. Assimilation of real wind, temperature and surface pressure observations taken by radiosondes. Major challenges, associated with adapting the scheme to an imperfect model environment, are expected.

3. Assimilation of simulated humidity and ozone observations within the LEKF procedure.

4. Assimilation of a gradually expanding set of observations including a growing number of remotely-sensed observations.

7. Acknowledgments

This work was supported by the W. M. Keck Foundation, a James S. McDonnell 21st Century Research Award, by the Office of Naval Research (Physics), by the Army Research Office, and by the National Science Foundation (Grants #0104087 and PHYS 0098632).

8. References

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

Anderson, J. L., 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.

Bishop, C. H., B. J. Etherton, and S. Majumdar, 2001: Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical aspects. *Mon. Wea. Rev.,* **129,** 420-436.

Corazza, M., E. Kalnay, DJ Patil, E. Ott, J. Yorke, I Szunyogh and M. Cai, 2001: Use of the breeding technique in the estimation of the background error covariance matrix for a quasigeostrophic model.

AMS Symposium on Observations, Data Assimilation and Probabilistic Prediction. 13-17 January 2002, Orlando, Florida, pp 154-157.

Daley, R., 1991: *Atmospheric data analysis.* Cambridge University Press, New York.

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.

Evensen, G., and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman Filter with a quasi-geostrophic model. *Mon. Wea. Rev.,* **124,** 85-96.

Fisher, M., 1998: Development of a simplified Kalman Filter. ECMWF Research Department Tech. Memo. 260, 16 pp. [Available from European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire, RG2 9AX, United Kingdom.]

Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, l98l: Applications of estimation theory to numerical weather prediction, Dynamic Meteorology:Data Assimilation Methods, L. Bengtsson, M. Ghil and E. KŁlln (Eds.), Springer Verlag, pp. l39224.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman Filter-3D variational analysis scheme. *Mon. Wea. Rev.,* **128,** 2905-2919.

Hamill, T. M., J. 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, 2001: A sequential ensemble Kalman Filter for atmospheric data assimilation. *Mon. Wea. Rev.,* **129,** 796-811.

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

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2003: Atmospheric data assimilation with the ensemble Kalman filter: Results with real observations. Submitted. Avaibale from peter.houtekamer@ec.gc.ca.

Kalman, R., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D, J. Basic Eng.,* **82,** 35-45.

Kalman, R., R. Bucy, 1961: New results in linear filtering and prediction theory. *Trans. ASME, Ser. D, J. Basic Eng.,* **83,** 95-108.

Kalnay, E., 2003: *Atmospheric modeling, data assimilation, and predictability.* Cambridge University Press, Cambridge, 341 pp.

Kalnay, E., and Z. Toth, 1994: Removing growing errors in the analysis. Preprints, *10th AMS Conference on Numerical Weather Prediction,* Portland, OR, 212-215.

Keppenne, C., and H. Rienecker, 2002: Initial testing of a massively parallel ensemble Kalman Filter with the Poseidon Isopycnal Ocean General Circulation Model. *Mon. Wea. Rev.,* **130,** 2951-2965.

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

Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability,* Vol. 1, ECMWF, Reading, Berkshire, UK, 1-18.

Ott E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, J. A. Yorke, 2003: A Local Ensemble Kalman Filter for atmospheric data assimilation. http://arxiv.org/abs/physics/0203058.

Ott, E., B. R. Hunt, I. Szunyogh, M. Corazza, E. Kalnay, D. J. Patil, J. A. Yorke, A. V. Zimin, and E. Kostelich, 2002: Exploiting local low dimensionality of the atmospheric dynamics for efficient Kalman filtering. http://arxiv.org/abs/physics/0203058.

Sauer, T., B. R. Hunt, J. A. Yorke, A. V. Zimin, E. Ott, E. J. Kostelich, I. Szunyogh, G. Gyarmati, E. Kalnay, D. J. Patil, 2004: 4D Ensemble Kalman filtering for assimilation of asynchronous observations. Abstract J5.4.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2002: Ensemble square-root filters. *Mon. Wea. Rev.,* **131,** 1485-1490.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteorol. Soc.,* **74,** 2317-2330.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.,* **127,** 1374-1377.

Wang, X., and C. H. Bishop, 2002: A comparison of breeding and Ensemble Transform Kalman Filter ensemble forecast schemes. Preprints, *Symposium on Observations, Data Assimilation, and Probabilistic Prediction*, Orlando, FL., Amer. Meteor. Soc., J28-J31.

Whitaker, J. S., and T. H. Hamill, 2002: Ensemble Data Assimilation without perturbed observations. *Mon. Wea. Rev.,* **130,** 1913-1924.

Whitaker, J. S., G. P. Compo, X. Wei, and T. M. Hamill, 2003: Reanalysis without radiosondes using ensemble data assimilation. Mon. Wea. Rev., submitted. Available from jeffrey.s.whitaker@noaa.gov.