Tellus (2005), 57A, 528–545
Printed in Singapore. All rights reserved

TELLUS

Assessing a local ensemble Kalman filter: perfect model experiments with the National Centers for Environmental Prediction global model

By ISTVAN SZUNYOGH[1]*, ERIC J. KOSTELICH[2], G. GYARMATI[3], D. J. PATIL[3], BRIAN R. HUNT[4], EUGENIA KALNAY[1], EDWARD OTT[5] and JAMES A. YORKE[6], [1]*Department of Meteorology and Institute for Physical Science and Technology, University of Maryland, College Park, MD, USA;* [2]*Department of Mathematics and Statistics, Arizona State University, AZ, USA;* [3]*Institute for Physical Science and Technology, University of Maryland, College Park, MD, USA;* [4]*Institute for Physical Science and Technology and Department of Mathematics, University of Maryland, College Park, MD, USA;* [5]*Institute for Research in Electronics and Applied Physics, Department of Electrical and Computer Engineering and Department of Physics, University of Maryland, College Park, MD, USA;* [6]*Institute for Physical Science and Technology, Department of Mathematics and Department of Physics, University of Maryland, College Park, MD, USA*

(Manuscript received 29 June 2004; in final form 10 January 2005)

ABSTRACT

The accuracy and computational efficiency of the recently proposed local ensemble Kalman filter (LEKF) data assimilation scheme is investigated on a state-of-the-art operational numerical weather prediction model using simulated observations. The model selected for this purpose is the T62 horizontal- and 28-level vertical-resolution version of the Global Forecast System (GFS) of the National Center for Environmental Prediction. The performance of the data assimilation system is assessed for different configurations of the LEKF scheme. It is shown that a modest size (40-member) ensemble is sufficient to track the evolution of the atmospheric state with high accuracy. For this ensemble size, the computational time per analysis is less than 9 min on a cluster of PCs. The analyses are extremely accurate in the mid-latitude storm track regions. The largest analysis errors, which are typically much smaller than the observational errors, occur where parametrized physical processes play important roles. Because these are also the regions where model errors are expected to be the largest, limitations of a real-data implementation of the ensemble-based Kalman filter may be easily mistaken for model errors. In light of these results, the importance of testing the ensemble-based Kalman filter data assimilation systems on simulated observations is stressed.

1. Introduction

The time has come when ensemble-based Kalman filter data assimilation schemes can be considered for implementation on operational weather forecast systems in the foreseeable future. For the first time, an ensemble Kalman filter has been reported to break even with a sophisticated operational 3D-Var system (Houtekamer et al., 2005), to outperform the National Centers for Environmental Prediction (NCEP) 3D-Var in reconstructing the state of the mid-troposphere from surface pressure observations (Whitaker et al., 2004), and to be efficient in assimilating simulated and real Doppler-radar observations of convective systems (Snyder and Zhang, 2003; Dowell et al., 2004; Zhang et al., 2004). The main goal of the present paper is to demonstrate the potential of one particular ensemble-based Kalman filter scheme; we explain the implementation of the local ensemble Kalman filter (LEKF), introduced in Ott et al. (2002, 2004), using the Global Forecast System (GFS) of the NCEP.

The LEKF scheme is an ensemble square-root filter (e.g. Tippett et al., 2002). In an ensemble square-root filter, one first obtains an estimate of the most likely state of the atmosphere and an analysis error covariance matrix that describes the uncertainty in the best estimate. Then, an ensemble of analyses is generated centered on the most likely state and representative of the uncertainty reflected by the analysis error covariance matrix. A distinguishing feature of the LEKF is that it solves the Kalman filter equations locally in model grid space; other square-root filters solve the Kalman filter equation locally in observation space (Anderson, 2001; Bishop et al., 2001; Whitaker and

*Corresponding author.
e-mail: szunyogh@ipst.umd.edu

Hamill, 2002). More precisely, the LEKF obtains the analysis at the different grid points independently, using all observations that are thought to improve the analysis at the individual grid points. In this scheme, the same observation is used to obtain the analysis at multiple grid points. By contrast, sequential schemes assimilate the observations one by one (or by small groups when the errors between the observations are correlated), iteratively updating the state estimate at those grid points where the accuracy of the analysis is thought be positively affected by a given observation (or group of observations).

In Ott et al. (2004) we speculated that solving the Kalman filter equations locally in grid space may be computationally advantageous and, at the same time, may not noticeably degrade the accuracy of the assimilation. The local analyses can be processed in parallel, involve relatively small matrices, and treat all data simultaneously. The numerical experiments presented here are designed to test the accuracy and computational efficiency of the LEKF on an operational model.

In what follows, we first provide a short summary of the LEKF algorithm (Section 2), and then we explain the implementation of the scheme on the NCEP GFS (Section 3). This implementation is tested under the 'perfect model scenario', i.e. by assuming that a model run provides a perfect representation of the true evolution of the atmosphere, making possible the generation of simulated observations (with known error statistics) and the exact computation of the analysis errors (Section 4). Experiments are carried out for ensembles of different size and for varying observational data coverage (Section 5). The accuracy of the analysis scheme is measured by the root-mean-square (rms) distance between the true states and the analyses. The computational efficiency is measured by the wall-clock time needed to complete the analysis. The results of this experiment indicate the potential operational feasibility of the LEKF scheme (Section 6).

2. Local ensemble Kalman filter

A detailed description and mathematical justification of the different components of the LEKF scheme can be found in Ott et al. (2002, 2004). Here we provide only a brief algorithmic summary needed to understand the implementation of the scheme on the NCEP GFS. The version of the scheme that we describe assumes that the rank of the background and analysis covariance matrices is k when the ensemble has $k+1$ members. Ott et al. (2004) describe a more general formulation that allows for a reduction of the rank. Finally, we consider the case where all observations collected for the current analysis are taken at the same time. [A simple technique to extend the scheme to the assimilation of asynchronous observations is presented in Hunt et al. (2004).]

2.1. Global and local background vectors

A $(k+1)$-member ensemble ($k \geq 1$) of global background state vectors, $\mathbf{x}_{\mathrm{g}}^{\mathrm{b}(i)}$, $i = 1, 2, \ldots, k+1$, is obtained by integrating the forecast model started from a $(k+1)$-member ensemble of analysis fields created in the previous analysis cycle.

For each grid point $\mathbf{m}$ of the three-dimensional model grid, we define a corresponding local volume that consists of all grid points within a suitably prescribed neighborhood of $\mathbf{m}$. Let $\mathbf{x}(\mathbf{m})$ be the d-dimensional local vector representing the model state within the local region centered at the grid point $\mathbf{m}$. The dimension d depends on both the number of model grid points in the local volume and the number of variables defined at each grid points. For instance, if the model has v variables (e.g. two components of wind, temperature, and specific humidity, etc.) defined at all $n(\mathbf{m})$ grid points within the local volume, $d = vn(\mathbf{m})$. The construction of this local vector is a linear mapping $\mathbf{L}(\mathbf{m})$ of the D-dimensional vector $\mathbf{x}_{\mathrm{g}}$ that represents the state of the model in the space defined by the global three-dimensional grid of N grid points. (For the aforementioned example, $D = vN$.) Formally, the mapping is performed by a multiplication of the $d \times D$ matrix that represents $\mathbf{L}(\mathbf{m})$ and the D-dimensional vector $\mathbf{x}_{\mathrm{g}}$ that represents the global state. In practice, this operation is executed by simply copying the d relevant components of $\mathbf{x}_{\mathrm{g}}$ into $\mathbf{x}(\mathbf{m})$. Because all the analysis operations take place at a fixed time t and are repeated for all local regions, henceforth we suppress the dependence of all vectors and matrices on t and $\mathbf{m}$. The local background error covariance matrix $\mathbf{P}^{\mathrm{b}}$ and the most probable local background state $\bar{\mathbf{x}}^{\mathrm{b}}$ are derived from the $(k+1)$-member ensemble of global state field vectors $\mathbf{x}_{\mathrm{g}}^{\mathrm{b}(i)}$, $i = 1, 2, \ldots, k+1$. The most probable local state is estimated by

$$\bar{\mathbf{x}}^{\mathrm{b}} = \mathbf{L}\left[(k+1)^{-1}\sum_{i=1}^{k+1}\mathbf{x}_{\mathrm{g}}^{\mathrm{b}(i)}\right], \tag{1}$$

while the $d \times d$ local background error covariance matrix $\mathbf{P}^{\mathrm{b}}$ is estimated by

$$\mathbf{P}^{\mathrm{b}} = k^{-1}\sum_{i=1}^{k+1}\delta\mathbf{x}^{\mathrm{b}(i)}\left[\delta\mathbf{x}^{\mathrm{b}(i)}\right]^{\mathrm{T}}, \tag{2}$$

where the superscript 'T' denotes transpose and

$$\delta\mathbf{x}^{\mathrm{b}(i)} = \mathbf{L}\mathbf{x}_{\mathrm{g}}^{\mathrm{b}(i)} - \bar{\mathbf{x}}^{\mathrm{b}}. \tag{3}$$

We can express $\mathbf{P}^{\mathrm{b}}$ in terms of the $d \times (k+1)$ matrix,

$$\mathbf{X}^{\mathrm{b}} = k^{-1/2}\left[\delta\mathbf{x}^{\mathrm{b}(1)} \mid \delta\mathbf{x}^{\mathrm{b}(2)} \mid \ldots \mid \delta\mathbf{x}^{\mathrm{b}(k+1)}\right], \tag{4}$$

as

$$\mathbf{P}^{\mathrm{b}} = \mathbf{X}^{\mathrm{b}}\mathbf{X}^{\mathrm{bT}}. \tag{5}$$

2.2. Projection onto the k-dimensional analysis space

By using a $(k+1)$-member ensemble, we assume that an estimate of the background covariance matrix of rank k is sufficient to obtain accurate analyses. Experience accumulated by others (Houtekamer and Mitchell, 2001; Keppenne and Rienecker, 2002; Whitaker et al., 2004) suggests that $k+1$ may be reasonably small. For the purpose of subsequent computations, we

consider the coordinate system of the k-dimensional space determined by the k orthonormal eigenvectors $\{\mathbf{u}^{(j)}\}$ of $\mathbf{P}^{\mathrm{b}}$, which we use to form the internal coordinate system for the k-dimensional local analysis space. Because $\mathbf{P}^{\mathrm{b}}$ has rank k, it has k positive eigenvalues

$$\lambda^{(1)} \geq \lambda^{(2)} \geq \ldots \geq \lambda^{(r)} \geq \ldots \geq \lambda^{(k)} > 0. \tag{6}$$

Thus,

$$\mathbf{P}^{\mathrm{b}} = \sum_{j=1}^{k} \lambda^{(j)} \mathbf{u}^{(j)} \left[\mathbf{u}^{(j)}\right]^{\mathrm{T}}. \tag{7}$$

Because the size of the ensemble $(k+1)$ is envisioned to be much smaller than the dimension d of $\mathbf{x}^{\mathrm{b}}$, the computation of the basis vectors $\{\mathbf{u}^{(j)}\}$ is most efficiently done in the basis of the ensemble vectors. That is, we consider the eigenvalue problem for the $(k+1)\times(k+1)$ matrix $\mathbf{X}^{\mathrm{bT}}\mathbf{X}^{\mathrm{b}}$, whose non-zero eigenvalues are those of $\mathbf{P}^{\mathrm{b}}$ and whose corresponding eigenvectors left-multiplied by $\mathbf{X}^{\mathrm{b}}$ (and then normalized) are the k eigenvectors $\mathbf{u}^{(j)}$ of $\mathbf{P}^{\mathrm{b}}$.

We denote the projection of vectors into the k-dimensional space and the restriction of matrices to the same space by a superscribed circumflex (hat). The matrix representation of the corresponding projection operator is

$$\mathbf{Q} = \left\{\mathbf{u}^{(1)} \mid \mathbf{u}^{(2)} \mid \ldots \mid \mathbf{u}^{(k)}\right\}. \tag{8}$$

For instance, for the d-dimensional local background vector $\mathbf{x}^{\mathrm{b}}$, the vector $\hat{\mathbf{x}}^{\mathrm{b}}$ is a k-dimensional column vector given by

$$\hat{\mathbf{x}}^{\mathrm{b}} = \mathbf{Q}^{\mathrm{T}}\mathbf{x}^{\mathrm{b}}. \tag{9}$$

Similarly, for a $d \times d$ matrix, such as the local background covariance matrix $\mathbf{P}^{\mathrm{b}}$, the matrix $\hat{\mathbf{P}}^{\mathrm{b}}$ is $k \times k$ and is given by

$$\hat{\mathbf{P}}^{\mathrm{b}} = \mathbf{Q}^{\mathrm{T}}\mathbf{P}^{\mathrm{b}}\mathbf{Q}. \tag{10}$$

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

$$\hat{\mathbf{P}}^{\mathrm{b}} = \mathrm{diag}\left[\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(k)}\right], \tag{11}$$

and so is trivial to invert.

2.3. Local analysis

We solve the Kalman filter equation in the local low-dimensional subspaces. Let $\mathbf{y}^{\mathrm{o}}$ be the vector of current observations within the local region, and let $\mathbf{x}^{\mathrm{a}}$ denote a prospective local analysis field. Let

$$\Delta\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{a}} - \bar{\mathbf{x}}^{\mathrm{b}}. \tag{12}$$

We show in Ott et al. (2004) that the most probable value of the analysis increment $\Delta\hat{\mathbf{x}}^{\mathrm{a}} = \mathbf{Q}^{\mathrm{T}}\Delta\mathbf{x}^{\mathrm{a}}$ in the local analysis space is

$$\Delta\hat{\bar{\mathbf{x}}}^{\mathrm{a}} = \hat{\mathbf{P}}^{\mathrm{a}}\hat{\mathbf{H}}^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y}^{\mathrm{o}} - \mathcal{H}(\bar{\mathbf{x}}^{\mathrm{b}})]. \tag{13}$$

Here $\mathbf{R}$ is the observational error covariance matrix for the observations in the local analysis space, $\hat{\mathbf{P}}^{\mathrm{a}}$ is the analysis error covariance matrix in the local analysis space, and $\mathbf{H}$ is the Jacobian matrix of partial derivatives of the observation operator $\mathcal{H}$ (evaluated at $\bar{\mathbf{x}}^{\mathrm{b}}$); thus, $\hat{\mathbf{H}} = \mathbf{H}\mathbf{Q}$ maps variables from the k-dimensional representation of the analysis to the space of observations within local region l. In eq. (13), $\hat{\mathbf{P}}^{\mathrm{a}}$ is determined from the usual Kalman filter equations (e.g. Kalnay, 2003, p. 156), but restricted to the k-dimensional internal coordinate system:

$$\hat{\mathbf{P}}^{\mathrm{a}} = \hat{\mathbf{P}}^{\mathrm{b}}[\mathbf{I} + \hat{\mathbf{H}}^{\mathrm{T}}\mathbf{R}^{-1}\hat{\mathbf{H}}\hat{\mathbf{P}}^{\mathrm{b}}]^{-1}. \tag{14}$$

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

$$\bar{\mathbf{x}}^{\mathrm{a}} = \mathbf{Q}\Delta\hat{\bar{\mathbf{x}}}^{\mathrm{a}} + \bar{\mathbf{x}}^{\mathrm{b}}. \tag{15}$$

2.4. Ensemble of local analyses

The ensemble of local analysis fields $\{\mathbf{x}^{\mathrm{a}(i)}\}$, $i = 1, 2, \ldots, k+1$ is obtained by first finding the $k+1$ local analysis perturbations $\delta\mathbf{x}^{\mathrm{a}(i)}$,

$$\delta\mathbf{x}^{\mathrm{a}(i)} = \mathbf{Q}\delta\hat{\mathbf{x}}^{\mathrm{a}(i)}, \tag{16}$$

then forming the local analysis ensemble

$$\mathbf{x}^{\mathrm{a}(i)} = \bar{\mathbf{x}}^{\mathrm{a}} + \delta\mathbf{x}^{\mathrm{a}(i)}. \tag{17}$$

The local analysis perturbations $\delta\hat{\mathbf{x}}^{\mathrm{a}(i)}$ are a linear combination of the local background perturbations in the k-dimensional analysis space

$$\hat{\mathbf{X}}^{\mathrm{a}} = \hat{\mathbf{X}}^{\mathrm{b}}\mathbf{Y}, \tag{18}$$

where

$$\hat{\mathbf{X}}^{\mathrm{a,b}} = k^{-1/2}\left\{\delta\hat{\mathbf{x}}^{\mathrm{a,b}(1)} \mid \delta\hat{\mathbf{x}}^{\mathrm{a,b}(2)} \mid \ldots \mid \delta\hat{\mathbf{x}}^{\mathrm{a,b}(k+1)}\right\} \tag{19}$$

and

$$\mathbf{Y} = [\mathbf{I} + \hat{\mathbf{X}}^{\mathrm{bT}}(\hat{\mathbf{P}}^{\mathrm{b}})^{-1}(\hat{\mathbf{P}}^{\mathrm{a}} - \hat{\mathbf{P}}^{\mathrm{b}})(\hat{\mathbf{P}}^{\mathrm{b}})^{-1}\hat{\mathbf{X}}^{\mathrm{b}}]^{1/2}. \tag{20}$$

The matrix square root of the positive definite matrix in eq. (20) is chosen to be the positive definite square root. This construction of the local analysis perturbations has the desirable properties that it does not distort the mean of the analysis ensemble, it correctly represents the analysis uncertainty, and it preserves the smoothness of the background ensemble fields as closely as possible (see Ott et al., 2004, for details). For discussion of other square-root choices, see Tippett et al. (2002) and Ott et al. (2004).

2.5. Ensemble of global analyses

The components of the most probable global analysis field $\bar{\mathbf{x}}^{\mathrm{a}}_{\mathrm{g}}$ at the grid point $\mathbf{m}$ are obtained by first selecting the local analysis vector $\bar{\mathbf{x}}^{\mathrm{a}}(\mathbf{m})$ associated with the local region centered at $\mathbf{m}$, then copying the components of $\bar{\mathbf{x}}^{\mathrm{a}}(\mathbf{m})$ at its central grid point. The same strategy is used to obtain the members of the global analysis ensemble $\{\mathbf{x}^{\mathrm{a}(i)}_{\mathrm{g}}\}$, $i = 1, 2, \ldots, k+1$.

Table 1. The computational complexity of the main steps in the LEKF algorithm for each local region: k denotes the size of the ensemble; d, the dimension of the local atmospheric vector; n, the number of observations in the local region. By 'here' we mean for the simulated observations we used in our numerical experiments, which are uncorrelated and located at model grid points. In practice, computing $\hat{\mathbf{H}}$ and $(\hat{\mathbf{H}})^{\mathrm{T}}\mathbf{R}^{-1}$ will be more expensive, but still significantly less than the upper estimates given

$\bar{\mathbf{x}}^{\mathrm{b}}$; eq. (1)	$O(kd)$
$\mathbf{X}^{\mathrm{b}}$; eqs. (3) and (4)	$O(kd)$
$(\mathbf{X}^{\mathrm{b}})^{\mathrm{T}}\,\mathbf{X}_l^{\mathrm{b}}$	$O(k^2d)$
$\hat{\mathbf{P}}^{\mathrm{b}}$; eqs. (11)	$O(k^3)$
$\mathbf{Q}$, eq. (8)	$O(k^2d)$
$\mathbf{H}$	Depends on $\mathcal{H}$; trivial here
$\hat{\mathbf{H}} = \mathbf{HQ}$	Up to $O(knd)$, but trivial here
$\hat{\mathbf{H}}^{\mathrm{T}}\mathbf{R}^{-1}$	Up to $O(kn^2 + n^3)$, but $O(kn)$ here
$\hat{\mathbf{P}}^{\mathrm{a}}$; eq. (14)	$O(k^2n + k^3)$
$\mathbf{y}^{\mathrm{o}} - \mathcal{H}(\bar{\mathbf{x}}^{\mathrm{b}})$	Depends on $\mathcal{H}$; $O(n)$ here
$\delta\bar{\mathbf{x}}^{\mathrm{a}}$; eq. (13)	$O(k^2n)$
$\hat{\mathbf{x}}_l^{\mathrm{a}}$; eq. (15)	$O(kd)$
$\hat{\mathbf{X}}^{\mathrm{b}} = \mathbf{Q}^{\mathrm{T}}\mathbf{X}^{\mathrm{b}}$	$O(k^2d)$
$\mathbf{Y}$; eq. (20)	$O(k^3)$
$\hat{\mathbf{X}}^{\mathrm{a}}$; eq. (18)	$O(k^2)$
$\mathbf{x}^{\mathrm{a}(i)}$; eq. (17)	$O(k^2d)$

2.6. Algorithmic complexity

An algorithm is $O[f(n)]$ if, given input data of length n, the required number of machine instructions is bounded by $Cf(n)$ for some constant C when n is sufficiently large. This so-called 'order notation' describes the computational complexity of a given algorithm and provides a rough measure of the computing time on a single processor. For instance, multiplying an n-vector by a scalar is an $O(n)$ procedure. The product of an $n \times k$ matrix with a k-vector is $O(nk)$, in so far as one must form a linear combination of k vectors, each of length n; the classical algorithm for computing the product of an $n \times k$ matrix with a $k \times m$ matrix is $O(nkm)$.

Table 1 summarizes the computational complexity of each step in the LEKF algorithm for a single local region containing n observations together with d dynamical variables in each of k ensemble solutions. In typical applications, we expect d and n to be much larger than k. Therefore, with the possible exception of the computation of $\hat{\mathbf{H}}$ and $\hat{\mathbf{H}}^{\mathrm{T}}R_l^{-1}$, which we discuss in the next paragraph, the most expensive steps are those with complexity $O(k^2d)$ and $O(k^2n)$. The $O(k^2d)$ terms are related to changing coordinates between the d-dimensional local model space and the k-dimensional local analysis space, which the $O(k^2n)$ terms are related to assimilating n observations in the local analysis space. In typical applications, d is expected to be one to two orders of magnitude larger than k. Indeed, our initial timing results, described in Section 5.9, suggest that the overall run time is approximately quadratic in the size of the ensemble.

In the idealized experiments described here, the observation operator $\mathcal{H}$ is linear and equivalent to multiplication by $\mathbf{H}$, which is simply a gather–scatter observation. In practice, of course, $\mathcal{H}$ and $\mathbf{H}$ are more complex, depending on the nature of the data. For conventional observations, such as radiosonde data, $\mathcal{H}$ may involve only simple linear interpolation between model grid points, $\mathbf{R}$ may be banded or even diagonal, and the evaluation of $\hat{\mathbf{H}}^{\mathrm{T}}\mathbf{R}^{-1}$ may still be $O(kn)$. Remotely sensed data, such as satellite measurements, may involve the evaluation of a non-local observation operator and, hence, $\hat{\mathbf{H}}$ and $\hat{\mathbf{H}}^{\mathrm{T}}\mathbf{R}^{-1}$ may be more expensive to compute. However, in either case, if $\mathbf{R}$ is approximately constant, then $\mathbf{R}^{-1}$ may be pre-computed, thus reducing the cost of evaluating $\hat{\mathbf{H}}^{\mathrm{T}}\mathbf{R}^{-1}$. (The more general case, involving a non-linear $\mathcal{H}$, is discussed in Appendix A.)

The total time required to complete the LEKF is proportional to the total number of grid points (i.e. the number of local regions). However, because the assimilation is performed on each region independently, the algorithm is amenable to efficient implementation on parallel computer architectures, which substantially reduces the wall-clock time.

3. Implementation on the NCEP GFS

3.1. Forecast model

The NCEP GFS is a spectral model, which means that the model state variables are height-dependent coefficients of a spherical harmonic expansion on the globe. We use a version of the NCEP GFS that was in operational use at the beginning of 2001. In this version, the model variables are spectral coefficients of the two-dimensional vorticity and divergence, virtual temperature, logarithm of the surface pressure, specific humidity, and ozone mixing ratio. The only difference between our version and the operational one is in the resolution, which we have reduced to T62 in the horizontal direction and to 28 levels in the vertical direction. This resolution is well tested in so far as it has been used for many operational forecast products of NCEP for more than a decade. In addition, with this resolution we were able to perform a large number of experiments within the computational resources available to us.

3.2. Definition of the local volume

The non-linear and physical parametrization terms in the T62 resolution NCEP GFS are computed on a 192 × 94 Gaussian longitude–latitude grid. We utilize this grid to implement the LEKF, which is formulated in model grid space. This means that our three-dimensional global grid has 194 × 92 × 28 points. Because all variables are defined on the global grid, except for the logarithm of the surface pressure, the total number of grid-point variables is 2 544 768. The number of model variables, 1 137 024 spectral coefficients, is about 44% of the number of grid point variables.

In our experiments, the observed variables are the two horizontal components of the wind, the virtual temperature, and the surface pressure. We use this information to obtain an analysis of all grid-point variables. Then the spectral transform is applied to the grid-point variables of different types to obtain an analysis in spectral space at each level. Finally, the spectral coefficients of the two-dimensional vorticity and divergence are computed from the spectral coefficients of the two horizontal components of the wind vector.

Our implementation of the LEKF scheme involves both horizontal and vertical localizations of the $192 \times 94 \times 28$ model grid. In all experiments reported here, the numbers of grid points in the local volume in the zonal and meridional directions are equal, while the number of vertical levels in the local volume varies with the height of the center point (see Section 5). For center points near the poles, the model grid artificially truncates the local volumes. We augment these local volumes to have the same number of horizontal grid points by adding grid points from across the pole. For example, consider a center point near the north pole at 0° longitude. If the number of grid latitudes available to the north of this point is two fewer than the number required for the local volume, then we include grid points near 180° longitude at the two highest grid latitudes. Accommodating the vertical localization requires special treatment of the surface pressure variable. The surface pressure is included in the definition of the local vectors in each vertical layer, but the analyzed value of the surface pressure is solely determined by the analyzed value in the lowest atmospheric layer. This localization strategy ensures that the background and observed surface pressure pose a constraint on the analysis of the other variables in all local regions (layers). Our numerical experiments show that this is essential for maintaining a proper balance of the analyzed fields.

4. Experimental design

We assume that the NCEP GFS provides a perfect representation of the true atmosphere, an approach frequently called the 'perfect model scenario'. Under this assumption, forecast errors arise and grow exclusively due to uncertainties in the initial conditions and the sensitivity of the model solutions to these uncertainties. In other words, the model is a chaotic system, in the sense that uncertainties in the initial conditions are more frequently amplified than damped during the forecast phase of the analysis cycle. The role of the data assimilation system, on the one hand, is to use the information contained in the observations to remove the (typically exponentially) growing component of the errors from the background. (If the data assimilation was not able to extract useful information from the observations, the state estimation by the NCEP GFS would completely lose its usefulness within two to three weeks. At that time, the rms distance between the estimated state and the true state would become equal to the rms distance between two randomly chosen states from a climatological archive (see, for example, fig. 5 in Szunyogh and Toth, 2002). On the other hand, the data assimilation must use the information contained in the background to filter the observational noise and to spread the information to unobserved locations. We have designed a series of experiments that measure the efficiency of the LEKF in all of these three areas (removing growing errors, reducing noise, and spreading information).

4.1. Observations

First, a time series of 'true' states, $\mathbf{x}_g^t(t)$, was generated by a 60-day integration of the T62 GFS model, started from the operational NCEP analysis at 0000 UTC on 1 January 2000. Then, simulated observations were prepared at each grid point by adding zero-mean Gaussian random noise (simulated observational error) to the true states every 6 h (at 0000, 0600, 1200 and 1800 UTC). The standard deviations of the assumed observational errors are 1 K, 1.1 m s^{-1}, and 1 hPa for the virtual temperature, horizontal wind components, and surface pressure, respectively. The humidity and ozone variables and the physical parameters describing the conditions of the underlying surface (e.g. sea surface temperature, albedo, snow and ice coverage, soil type, etc.) are not observed. In most of the experiments, to simulate reduced observational networks, only subsets of the observations are assimilated. These subsets are created in a systematic manner, gradually removing observational locations at randomly selected locations to obtain sparser observational data sets. Results are shown for four different observational networks: observing vertical profiles at all (i.e. 17 848), 2000, 1000, and 500 locations around the globe

4.2. Measures of accuracy

In what follows, the accuracy of an analysis field will be assessed using three different types of diagnostics.

(i) A single number reflecting the overall error at a given height level and within a specified horizontal verification region will be obtained by first computing the spatial rms distance between the analyzed and the 'true' meteorological field at each analysis time, and then averaging these numbers over time.

(ii) When latitude versus height distributions of error are shown, the error at a given point on the plot is obtained by first computing at each time the rms average over longitude at fixed latitudes and height, and then averaging over time. [We note that (i) and (ii) follow the convention of numerical weather prediction in that the rms is never taken over both space and time.]

(iii) When horizontal distributions at a fixed height are shown, the error fields at a given point are obtained by averaging the absolute value of the error field at each grid point.

4.3. Explained variance and E-dimension

While the rms-based error measures can efficiently characterize the accuracy of the analysis scheme, diagnostics measuring the

accuracy of the assumptions made in the scheme can help find ways to further reduce the rms errors. In particular, the main motivation behind a Kalman filter scheme is the hope that the scheme can capture and efficiently remove a large portion of the actual background error, $\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}}$. The explained variance, $\|\mathbf{Q}\,\mathbf{Q}^{\mathrm{T}}(\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}})\|^2\,\|\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}}\|^{-2}$, measures the portion of the background error that is captured by the background ensemble in the local region centered at a given grid point. By mapping the explained variance, we can distinguish between regions where the background ensemble provides good and poor representations of background uncertainties. The optimal value of the explained variance is one, indicating that the ensemble fully captures the space in which uncertainties evolve. It is important to note that a nearly optimal value of the explained variance is a necessary, but not a sufficient, condition to obtain an accurate analysis. Large analysis errors can occur when the amount of available observed information is insufficient to reduce the background error in some of the well-captured directions of uncertainties. Our numerical results, presented in Section 5, show that relatively large analysis error can indeed occur for cases of high explained variance on some rare occasions.

We note that the efficiency of the Kalman filter can be measured by the explained variance only under the perfect model scenario employed here. What makes this case special is that (i) the actual background error can be determined, (ii) the observational error statistics are exactly known due to the lack of representativeness errors, and (iii) the quality of the estimated background error is fully determined by the quality of the estimated analysis error at the previous analysis time. Combining (ii) and (iii) we can see that the quality of the estimated background covariance matrix is fully determined by the quality of the estimated background covariance matrix at earlier times. When real observations are assimilated with a non-perfect model, the explained variance diagnostic has to be replaced with innovation statistics, such as the χ^2 diagnostic (e.g. Dee, 1995; Menard et al., 2000).

The explained variance can be small, either because the ensemble fails to capture important background error direction(s) due to a flaw in the ensemble design or because the background error is spread over many phase space directions and a limited-size ensemble cannot capture a sufficient portion of the background error. We use the ensemble dimension (E-dimension)

$$E = \frac{\left[\sum_{j=1}^{k} \lambda^{(j)^2}\right]}{\sum_{j=1}^{k} \lambda^{(j)}}, \tag{21}$$

where $\lambda^{(j)}$ is defined in eq. (6), to quantify the number of directions spanned by the background ensemble. The E-dimension, introduced in Patil et al. (2001) and studied in detail in Oczkowski et al. (2005), is a positive, typically non-integer number. It takes its maximum value, k, when the uncertainty is evenly distributed among k directions. The less even the distribution of uncertainty between the k directions, the smaller the E-dimension. The minimum value, $E = 1$, occurs in the extreme case where the uncertainty is confined to a single direction. At grid points where the analysis error is large and the E-dimension is close to k, there is a high probability that the ensemble size is insufficient to represent the background uncertainty. [Formally, eq. (21) is equivalent to the effective number of degrees of freedom (ES-DOF) defined by Bretherton et al. (1999), although the latter notion is applied to global state vectors taken at different times rather than an ensemble of local model vectors taken at the same time. We emphasize that here we are assessing the dimensionality of the background uncertainty locally in time and space, rather than the overall dimensionality of the dynamics.]

5. Numerical experiments

To validate and tune our implementation of the LEKF on the NCEP GFS, we first designed a base experiment, then explored changes in the behavior of the data assimilation system under gradual changes to selected parameters of the scheme. In the base experiment, the surface pressure, horizontal wind components at all height levels, and virtual temperature are observed at 2000 randomly selected geographical grid points on the globe, and the parameters of the LEKF are the following: the number of ensemble members is 40, and the size of the local region is $7 \times 7 \times v$ (the depth v of the local layers varies with altitude according to Fig. 1). The scheme also applies a uniform 4% multiplicative variance inflation (Anderson and Anderson, 1999) to the background ensemble to compensate for the variance lost to non-linearities and limited sample size. According to the results of the numerical experiments described below, the base configuration is a reasonable, although not optimal, configuration of the LEKF.

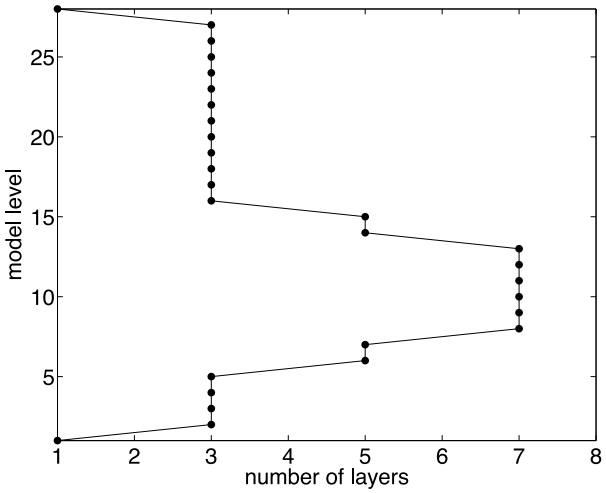

Fig. 1. Number of model levels in the local volumes centered at the different model levels.

5.1. Temporal evolution of errors

To start the LEKF cycle, we first chose the operational NCEP background forecast at 0000 UTC on 1 January 2000 to be the background mean $\bar{\mathbf{x}}_g^b$, then we added zero-mean Gaussian random noise to $\bar{\mathbf{x}}^b$ to generate the $(k+1)$-member ensemble of global background state vectors $\mathbf{x}_g^{b(i)}$, $i = 1, 2, \ldots, k+1$. That is, the initial estimate of the local background error covariance matrix $\mathbf{P}^b$ was a random matrix for each local region. While the standard deviation of the assumed initial background uncertainty was the same as for the observational error, the random samples were independently generated for the different ensemble members and the simulated observations. Because the initial 'true state' was defined by the operational NCEP analysis at 0000 UTC on 1 January 2000, the initial error in the background mean was equal to the analysis increment (the difference between analysis and background) in the NCEP system. While choosing the operational background to be the initial estimate of the background mean seems to be a reasonable strategy to start the first analysis cycle, another analysis experiment, choosing the operational NCEP analysis at 1800 UTC on 29 March 2002 to be initial estimate of the background mean, was also carried out. Because this time is much further apart from the 'true' date (0000 UTC on 1 January 2000) than the predictability time limit for the atmosphere, this second choice for $\bar{\mathbf{x}}_g^b$ can be viewed as a randomly selected possible states of the atmosphere (as represented by the operational NCEP analysis).

The global rms analysis error rapidly (in a few steps) tends to a level that is much smaller than the rms error of the observations regardless of the selection of the initial estimate of the background mean. We have found this to be a robust property; it is characteristic of all experiments reported in this paper. While the speed of convergence is rapid, it is slightly different for the different variables: fastest for the temperature and slowest for the wind components. Figure 2 shows an example of the time evolution of the rms error for the surface pressure of the base experiment. In this example, the rms error tends to a level that is about 40% of the observational error within a couple of days. While some slow temporal fluctuations of the error can be observed, the efficient filtering of the observational noise by the data assimilation scheme is evident. Because observations are taken at only 2000 locations (or about 11% of all the grid points), the results also indicate that the LEKF efficiently propagates information to unobserved locations.

Because the temporal fluctuation of the errors is modest, the different configurations of the LEKF can be meaningfully compared using time-mean results. To ensure that the time means are not affected by the initial transient, the time averages are computed for the last 45 d (the last 180 analysis cycles) of each experiment, i.e. the first 15 d (60 analysis cycles) are ignored.

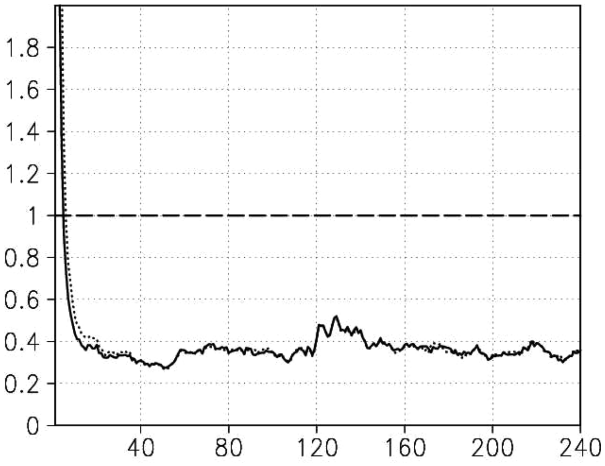

Fig. 2. Time evolution of the analysis surface pressure rms error for two different estimates of the initial background mean. The solid line represents the estimation error for the case where the operational background is used as initial estimate, while the dotted line represents the estimation error for the case of a 'randomly' selected initial estimate. The rms error of the observations is shown by dashes. The abscissa is the number of analysis cycles performed every 6 h.

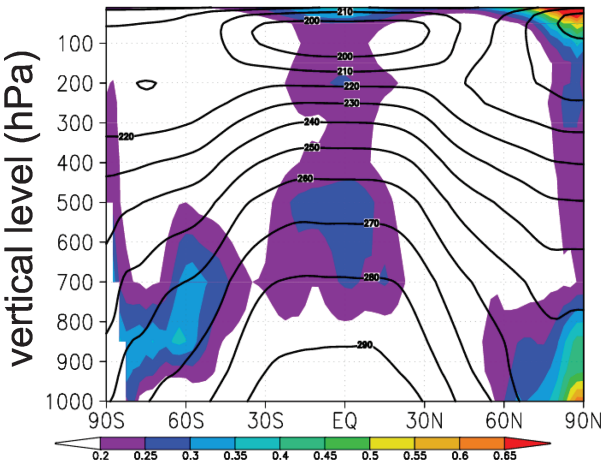

Fig. 3. Zonal average of the time-mean rms error in the temperature analysis (shades). The zonal average of the 'true' time-mean temperature is also shown (contours).

5.2. Spatial distribution of errors

The analysis errors have a strong zonal dependence (Figs. 3 and 4): the largest errors are in the tropics and over the polar regions; the errors in the mid-latitudes are the smallest. The time-mean errors are small compared to the rms observational errors (1 K and 1.1 m s^{-1}). The most important virtue of the data assimilation system is the high accuracy of the analyses in the mid-latitudes. In these regions, the errors are practically negligible in the lower troposphere (below the 500-hPa level). This is most remarkable, because these are the regions where baroclinic instabilities – the most energetic instabilities in the Earth's atmosphere – convert

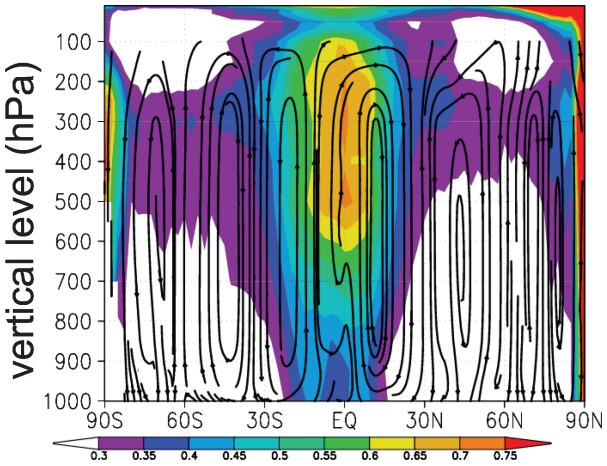

Fig. 4. Zonal average of the time-mean rms error in the analysis of the zonal component of the wind (shades). The zonal average of the 'true' time-mean circulation is shown by streamlines (contours with small arrowheads showing the direction of the flow).

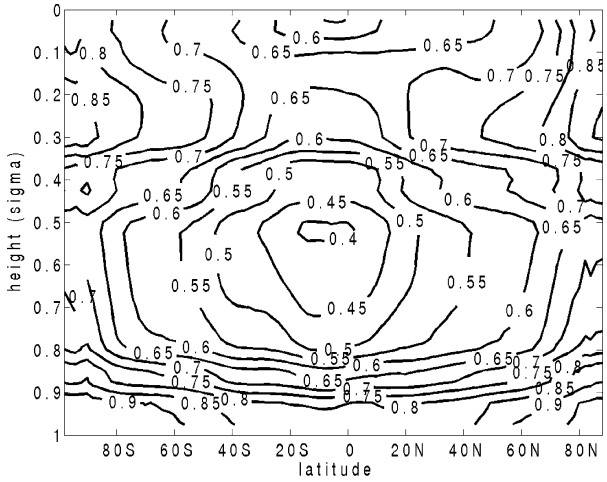

Fig. 5. Zonal average of the time-mean explained variance.

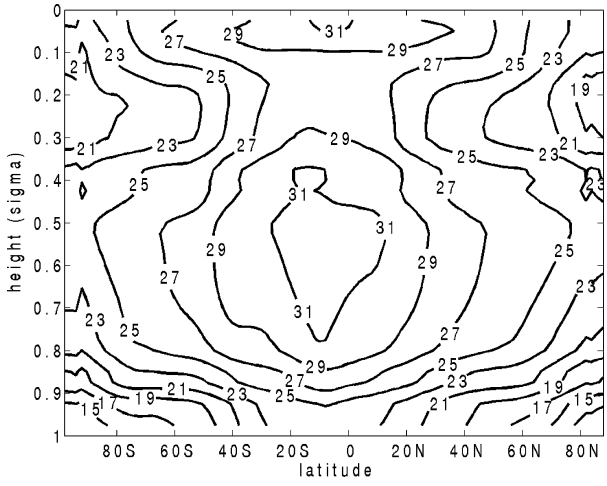

Fig. 6. Zonal average of the time-mean E-dimension.

the available potential energy to kinetic energy. (The main zones of baroclinic energy conversion can be recognized in Fig. 4 by the on-average upward motions that they generate in the mid-latitudes.) This finding shows the efficiency of the Kalman filter in correcting fast-growing background errors associated with processes initiated by baroclinic instabilities. The largest, but still modest, analysis errors in the mid-latitudes occur in and immediately below the jet layer (between the 200- and 500-hPa levels).

Estimation of the atmospheric state is the most difficult (i) in the region of ascending motions in the Hadley cells, (ii) in the stratosphere aloft, (iii) in the stratosphere over the two poles, especially at the top of the model atmosphere, and (iv) in the lower troposphere in the Northern Hemisphere (NH) polar region. In our quest for a better understanding of the mechanisms leading to the development of relatively large errors in these regions, we first examine the geographical distribution of the explained variance and the E-dimension.

5.3. *Explained variance and E-dimension*

The explained variance is the lowest in the tropics, especially in the mid-troposphere and the stratosphere aloft (Fig. 5). The zonal distribution of the explained variance is remarkably symmetric, characterized by a center of low values (somewhat shifted to the south of the equator) and by increasing values toward the poles. In the vertical, the explained variance is the highest at the surface, quickly decreasing in the lower troposphere, becoming minimum in the mid-troposphere, and then increasing again in the upper troposphere. There is also a secondary minimum in the tropical stratosphere above the absolute minimum in the mid-troposphere.

The explained variance and the E-dimension are strongly anti-correlated; the space of background uncertainties is always well captured in regions of low dimensionality, and always poorly captured in regions of high dimensionality (Figs. 6 and 7). Had the explained variance not saturated at low E-dimensions, the anticorrelation between the explained variance and the E-dimension would be even higher. The strong anticorrelation at the higher E-dimension indicates that the limited size ensemble can capture only a decreasing portion of the background error. We note that these results do not change qualitatively when the correlation is computed for grid-point values instead of zonal means (the anticorrelation for the grid-point values 0.91).

Because the observational coverage is uniform over the globe, the dimensionality of the analysis (background) errors is a property of the model dynamics. Our results demonstrate that, as conjectured by Ott et al. (2004), the local dimensionality of the dynamics plays a central role in the efficiency of the LEKF. The

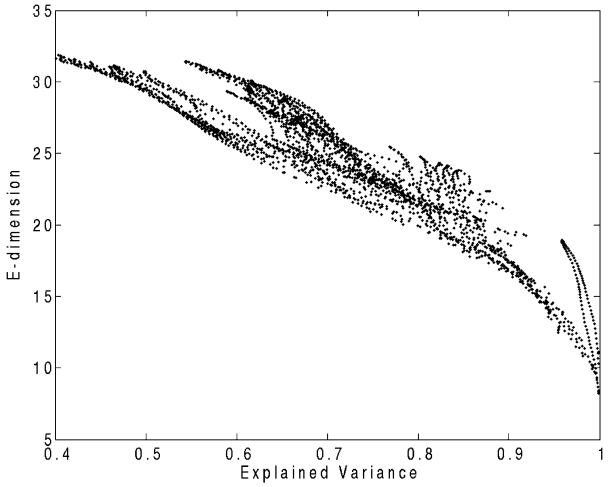

Fig. 7. Zonal average of the explained variance versus zonal average of the E-dimension. The sample is collected over all latitudes and analysis cycles. The correlation between the explained variance and the E-dimension is −0.93.

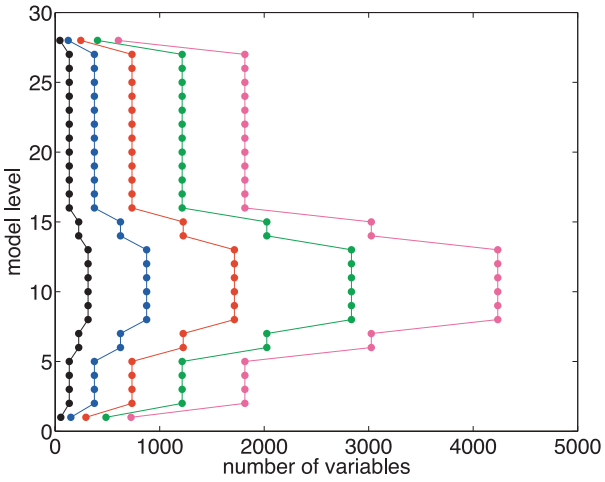

Fig. 8. Number of model variables in the local regions centered at the different model levels. The number of variables is shown for local regions consisting of 3 × 3 (black), 5 × 5 (blue), 7 × 7 (base experiment, bright red), 9 × 9 (green) and 11 × 11 (cyan) horizontal grid points at each level.

larger errors in the tropics are clearly due to the higher local dimensionality of the model dynamics. (The only regions where the explained variance is high and the E-dimension is low, but the analysis error is still large, are over the poles. As noted earlier, such regions can exist where there is no sufficient number of observations to remove the background errors in certain phase space directions.) This gives us hope that the analysis accuracy can be improved in the tropics by either decreasing the size of the local regions (dimensionality of the local dynamics) or by increasing the number of ensemble members, thus increasing the local dimensionality that can be accurately represented by the LEKF. Care should be taken though, as each of these strategies has its potential drawbacks. On the one hand, reducing the dimension d of the local vectors inevitably leads to using fewer observations in obtaining the local analyses. The resulting degradations may offset the beneficial effects of having a better representation of the local dynamics. On the other hand, increasing the ensemble size $k + 1$ leads to a significant increase of the computational cost (see Table 1). In what follows, experiments are carried out to find an ideal range of d and $k + 1$.

5.4. *Sensitivity to the horizontal size of the local regions*

First we vary the horizontal size of the local regions, using $3 \times 3 \times v$, $5 \times 5 \times v$, $7 \times 7 \times v$, $9 \times 9 \times v$, and $11 \times 11 \times v$ patches of grid points, and keeping the ensemble size fixed at 40 members. The dimension d of the local vectors for the different-sized regions is shown in Fig. 8. While the rms of the surface pressure analysis error is essentially the same for all experiments (including the base experiment shown in Fig. 2) the patch size has a noticeable impact on the accuracy of the temperature and wind analyses. The results are summarized in Figs. 9 and 10.

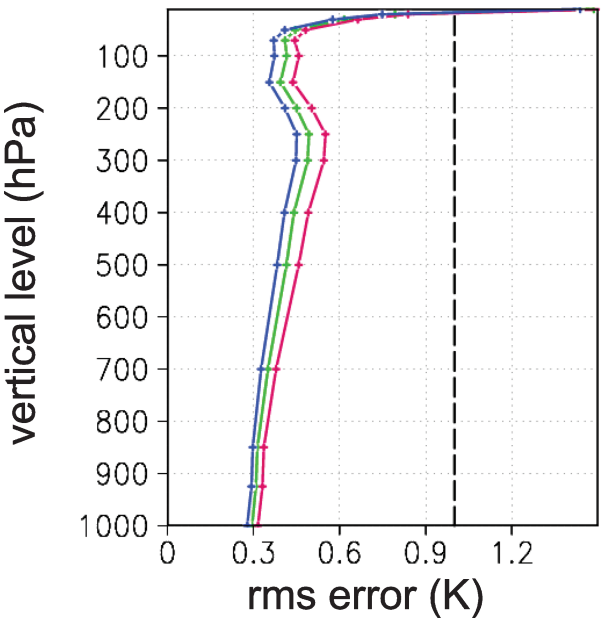

Fig. 9. The time mean of the rms error of the temperature analysis over the globe. The color scheme is the same as in Fig. 8. The curves are very similar for the 5 × 5 and 7 × 7 regions, and for the 9 × 9 and 3 × 3 regions; thus, results for the 7 × 7 regions and for the 3 × 3 regions are not shown for the sake of transparency. The rms error of the observations is shown by black dashes.

Increasing the region size to $9 \times 9 \times v$ or $11 \times 11 \times v$ leads to an increase of errors in both the tropics and extratropics, with the degradation being much more substantial in the tropics. An inspection of the explained variance and E-dimension in figures similar to Figs. 5 and 6 (not shown)

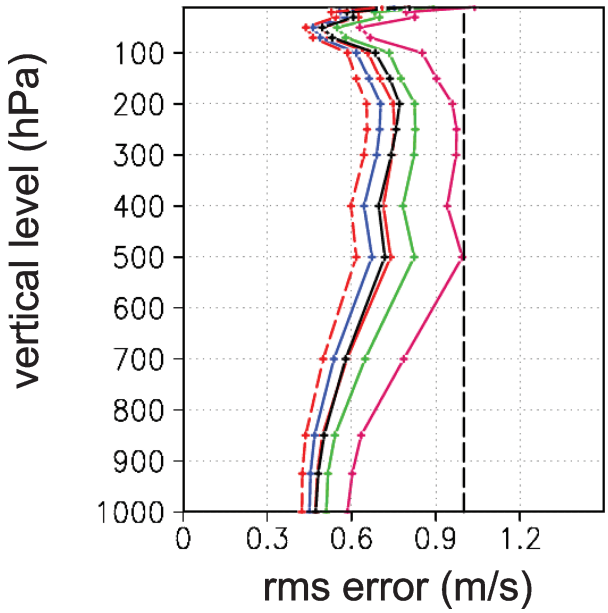

Fig. 10. The time mean of the rms error of the zonal wind component analysis in the tropics, which is defined by the latitude belt between 30°S and 30°N. The color scheme is the same as in Fig. 8. The solid lines indicate performance for the 40-member ensemble, while the dashed (red) line shows results for the 80-member ensemble using 7 × 7 horizontal regions. For the 80-member ensemble, the performance is very similar for the 5 × 5 (not shown) and 7 × 7 horizontal regions, while the accuracy for the 3 × 3 regions (not shown) is about the same as for the 5 × 5 regions using a 40-member ensemble (solid blue curve). The rms error of the observations is shown by dashes.

suggested that the degradation of the analysis accuracy is associated with a poorer representation of the space of uncertainties. For instance, for the 9 × 9 local regions the highest E-dimension is 33 (compared to 31 in the base experiment), while the lowest explained variance is 0.3 (compared to 0.45 in the base experiment). Also, the anticorrelation between the two quantities increases to −0.94 indicating a narrower saturation region than observed for the 7 × 7 × 7 local regions. Most importantly, these results demonstrate that increasing the amount of local observed information by increasing the local region size becomes counterproductive when a given ensemble size becomes insufficient to capture the important background error patterns in the larger regions.

Decreasing the size of the local regions to 5 × 5 × v, on the other hand, reduces the highest E-dimension to 29, increases the lowest explained variance to 0.55, and reduces the anticorrelation to −0.89 (figures not shown). The associated improvement of the analysis accuracy is the largest in the tropics (Fig. 10). However, when the horizontal size of the local regions is further reduced to 3 × 3 × v the analysis accuracy degrades. We can conclude that the optimal horizontal region size for a 40-member ensemble and our choice of v (Fig. 1) is 5 × 5.

5.5. *Sensitivity to the vertical size of the local regions*

We carried out experiments with three different choices for the layer depth in the vertical localization (results are not shown). These include the vertical localization depicted in Fig. 1, the case when localization in the vertical is not applied, and the case in which each local layer contains three levels (except at the top and the bottom boundaries of the atmosphere, where the local layers consist of only one level.)

Decreasing the local layer depth has a noticeable influence on the analysis accuracy only in the tropics, where it leads to improvements similar to those obtained by reducing the horizontal region size to 5 × 5. This is a plausible result considering that both changes lead to a 50% reduction in the dimension d of the local regions, resulting in a very similar distribution of the explained variance and E-dimension (for the thin layers the minimum E-dimension in 29, while the lowest explained variance is 0.6).

Not applying vertical localization at all, on the other hand, drastically degrades the analysis quality. The errors in the temperature increase by about 50%, while the rms error in the wind doubles in both the jet layer in the extratropics and at most levels in the tropics. In the latter region, the wind analysis errors are even larger than the observational errors. This degradation is due to a severe reduction in the capability of the ensemble to capture the space of uncertainty. The explained variance varies between 0.15 and 0.45, while the E-dimension varies between 25 and 35.

The vertical localization has an obviously large beneficial effect on the analysis accuracy, although our results also indicate that beyond a certain level of localization not much can be gained by reducing the depth of the local layers. Fine tuning the vertical localization will become an important practical issue once the assimilation of satellite radiance observations is considered. These observations can be affected by the atmospheric temperature and humidity at distant vertical levels. This may severely limit the maximum possible degree of vertical localization, thus limiting the smallest d that can be considered in an implementation of the LEKF. [We note that some types of radiance observations, such AMSU-A radiances, can be assimilated with a Kalman filter using vertical localization (Houtekamer et al., 2005).]

5.6. *Sensitivity to the ensemble size*

Now we turn our attention to the improvements that can be achieved by increasing the number of ensemble members. Because a larger region size allows for a better utilization of more observed information, we can expect that increasing the number of ensemble members leads to a reduction of the analysis errors in two ways: (i) a larger ensemble increases the explained variance for a given d, and (ii) it also allows for a larger d leading to a better utilization of more observed information. We find

(results are not shown) that doubling the ensemble size (to $k = 80$) increases the maximum E-dimension to 55, increases the minimum explained variance to 0.65, and reduces the anti-correlation to -0.86.

In terms of analysis accuracy, we find that increasing the ensemble size has the largest positive effect when the configuration uses $7 \times 7 \times v$ grid points (Fig. 10). (Because the overall improvement obtained by increasing the ensemble size from 40 to 80 is relatively modest in the extratropics, the results are illustrated only for the tropics.) This configuration now breaks even, in all regions and for all variables, with the configuration using $5 \times 5 \times v$ grid points and a 80-member ensemble. This result suggests that more than 80 ensemble members would be required to improve the performance of the scheme for patch sizes larger than 7×7.

5.7. *Sensitivity to the density of observations*

Experiments with simple models show that Kalman filters have a distinct advantage, as the observational density decreases, over schemes that use static estimates of the background covariance matrix (e.g. Hamill et al., 2001; Ott et al., 2004). More precisely, while the performance of the static schemes, such as Optimal Interpolation (OI) and 3D-Var, degrades dramatically with decreasing observational density, the accuracy of the Kalman filter is only slightly affected until a critically low density is reached. To test whether our implementation of the LEKF on the NCEP GFS retains this important property, further experiments with differing numbers of observations were also carried out. While the observational density has only a modest impact on the quality of the temperature analysis (Fig. 11), the accuracy of the wind analysis, especially in the tropics, degrades more dramatically as the density of the observations is reduced (Fig. 12). This figure also shows that increasing the number of observations is a very efficient way to reduce the analysis errors in the tropics. A more detailed inspection of the zonal mean errors (figures are not shown) indicates that when all model grid points are observed (i) the temperature and wind errors are greatly reduced in the tropics in both the troposphere and the lower stratosphere, (ii) the temperature errors in the arctic mid-troposphere are also greatly reduced, but (iii) the large errors in the NH polar stratosphere and boundary layers are essentially unchanged.

The changes in the E-dimension are practically negligible, indicating that they are mainly determined by the ensemble size, the dimension of the local regions and the model dynamics, and not by the observation density. Increasing the number of observations slightly reduces the explained variance (e.g. when all locations are observed, the lowest explained variance decreases from 0.45 to 0.35). This shows that increasing the number of observations improves the analysis by completely removing the background errors along some of the well-captured directions.

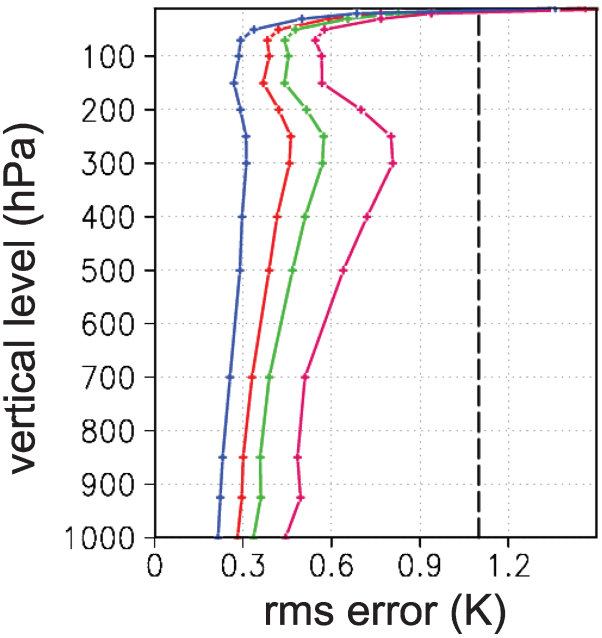

Fig. 11. The time mean of the rms error of the temperature analysis over the globe. The analysis error is shown for observing networks consisting of 18 048 (blue), 2000 (base experiment, bright red), 1000 (green) and 500 (purple) observational locations. This corresponds to observing 100% points (blue), 11% (bright red), 5.5% (green), and 2% (cyan) of all grid points. The rms error of the observations is shown by dashes.

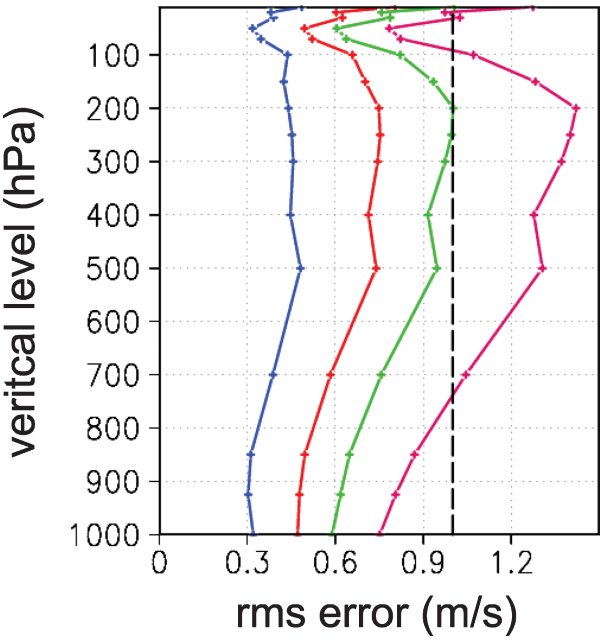

Fig. 12. The time mean of the rms error of the zonal wind component analysis in the tropics. The color scheme is the same as in Fig. 11.

5.8. *Gravity waves*

Wave solutions of the atmospheric primitive equations admit both shallow and deep gravity wave motions. The importance of correctly analyzing these motions is one of the most challenging aspects of atmospheric data assimilation. On the one hand, the goal is to ensure that the initial conditions do not lead to the excitation of spurious deep gravity waves in the extratropics where high-frequency oscillation of the surface pressure, characteristic of these waves, is not observed in nature. On the other hand, the goal is to retain those shallow gravity waves in the entire atmosphere and those deep gravity waves in the tropics that have important effects on synoptic scale motions in the atmosphere. Current operational systems use a variety of techniques to eliminate spurious gravity waves (e.g. Daley, 1991; Kalnay, 2003). Some of these techniques are integral parts of the data assimilation schemes: 3D-Var schemes define cross-correlations between the different variables by some variant of the balance equation and/or by imposing an extra constraint on the divergence of the horizontal wind in the statistical optimization problem. The so-called 'initialization' methods, on the other hand, are applied after assimilation of the observations have been completed. Currently, the most popular initialization technique is the digital filter, which empirically (by numerical integration of the model) identifies and eliminates all features of the analyzed field that excite motions with frequencies larger than a given threshold (Lynch and Huang, 1992). The more traditional non-linear normal mode initialization is based on the somewhat 'heavier handed' approach of first transforming the state vector into a form where the low-frequency 'slow' and the high-frequency 'fast' motions are represented by different components of the transformed state vector, and then setting the time derivative of the 'fast' components to zero (Machenauer, 1977). [See Lynch (2002) for an illustration of the problem of separating and initializing the 'fast' component of the atmospheric model state.]

The aforementioned techniques alone or in some combinations can efficiently filter deep gravity waves in the extratropics. Nevertheless, they are based on formal assumptions that are valid for the real atmospheric flow only to some low-order approximations. These approximations generally lead to overly aggressive filtering, or in some other cases insufficient filtering, of the initial conditions. One potential advantage of ensemble Kalman filter schemes is that they may eliminate the need for inherently inaccurate explicit balance constraints or initialization. To support this view, we first note that deep inertia-gravity waves in the extratropics are typically transient features of numerical weather predictions, because they are dispersive, and the diffusion and drag terms and the time integration schemes rapidly dissipate the energy of these waves. This means that the forecast phase of the analysis cycle is unlikely to introduce deep gravity waves in the extratropics. In addition, in ensemble Kalman filters the analysis is a linear combination of the background ensemble members. Combining these two factors, we can expect that the analysis should contain only those types of gravity waves that can occur during longer-term (longer than an initial transient) integration of the model, unless generation of the analysis perturbations leads to excitation of gravity waves. To avoid excitation of such unwanted gravity waves, the LEKF use a spatially smoothly varying local transformation of the background ensemble to obtain an analysis ensemble, also ensuring that the transformation introduces the smallest possible changes into the background ensemble members with respect to a suitable norm (Ott et al., 2004).

For the perfect model scenario, investigated in this paper, gravity waves should ideally occur at the same locations and times in the analysis and the 'true' state. To verify whether our system satisfies this requirement, we first identified gravity waves by searching for cases of high-frequency oscillations in the surface pressure in both the analyses and the true states. We found that our system performed as we had expected; high-frequency oscillation of the surface pressure occurred only at the times and locations where high-frequency oscillations also occurred in the true state. The most interesting case we have found is a rare event of gravity wave propagation from the tropics deep into the subtropical region. The unusually large surface pressure errors between cycles 120 and 150 in Fig. 2 are associated with this event. Animations have revealed that these errors (which are still smaller than the observational errors) are associated with deep gravity waves propagating from the South Pacific convergence zone toward the extratropics, where they eventually die out between the latitudes 30°N and 35°N. Figure 13 shows that high-frequency oscillation is present in both the analysis and the 'true' states and that the analysis accurately tracks these waves. (Although the mismatch between the analyses and the 'true' states is somewhat larger deeper in the tropics, the analyses always

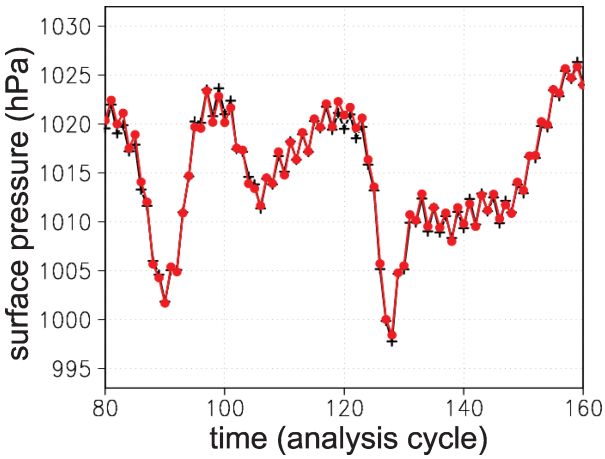

Fig. 13. The 'true' (black crosses) and analyzed (red dots) time evolution of the surface pressure at 30°N 150°E. The analyses are based on 500 soundings.

correctly capture the frequency of the changes.) We can conclude that the LEKF can correctly capture gravity waves, without exciting spurious gravity waves where they are not present in the 'true' state. This finding demonstrates that solving the analysis problem locally in physical space does not necessarily lead to artificial gravity waves in the analysis. This is in contrast to the general wisdom mainly based on earlier results with the OI. It has yet to be seen whether this advantageous behavior of the LEKF carries over to the case where the model provides an imperfect representation of the true atmospheric dynamics. However, the fact that the LEKF analysis is a combination of balanced forecasts should ensure that the spurious generation of gravity waves is avoided.

5.9. *Physical interpretation*

In what follows, we focus our attention on identifying the atmospheric processes that affect the accuracy of the LEKF analyses. While this part is more speculative in nature than the discussion presented so far, it may prove useful in drawing more general conclusions, applicable to atmospheric models other than the NCEP GFS.

We start our investigation in the tropical region. Figure 4 indicates that the analysis of the wind is the most difficult in the region of ascending motions in the Hadley cells. The error quickly decreases toward higher latitudes in the neighboring regions of descending motions. A comparison of Fig. 3 and fig. 14.3 of Emanuel (1994) suggests that the errors are the largest in the layers where the convective available potential energy (CAPE) is the largest. A picture emerges in which the largest errors in the tropical wind analysis are associated with deep convective processes. This conclusion is also well supported by Fig. 14, which depicts the geographical distribution of the errors at the 300-hPa level. The regions of largest errors in the tropics are

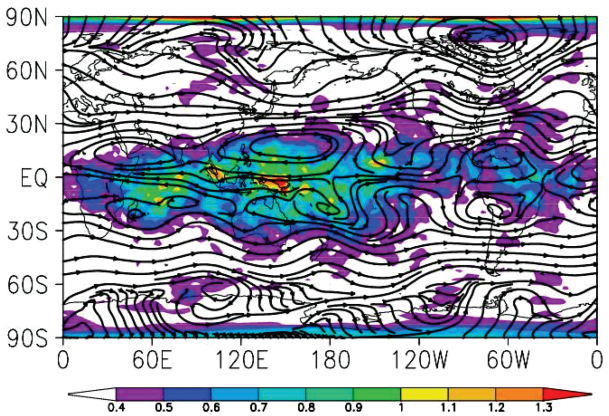

Fig. 14. Time-mean of the rms error in the analysis of the zonal wind component at the 300-hPa pressure level (shades). The 'true' time-mean horizontal circulation at the same level is shown by streamlines (contours with small arrowheads showing the direction of the flow).

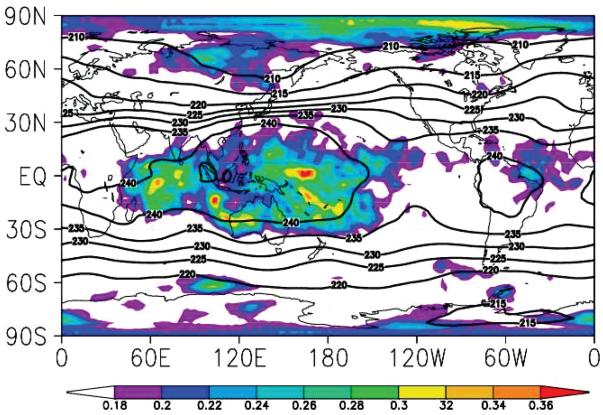

Fig. 15. Time mean of the rms error in the temperature analysis at the 600-hPa pressure level (shades). The 'true' time-mean temperature is also shown (contours).

sandwiched between the regions of easterly trade winds to the north and south. These regions are associated with pools of warm air located in the south-west Pacific/north-east Indian and mid-Atlantic equatorial regions (Fig. 15).

The sources of CAPE are air parcels lifted from the surface. Because the analysis is accurate near the surface, the temperature flux associated with the ascending warm air parcels may be relatively well analyzed in the subcloud layer. This suggests that the relatively poor analysis in the deep convective clouds originates from a poor analysis of processes within the clouds themselves. Also, because the wind analyses are of poorer quality than the temperature analysis, it may be considerably more difficult to analyze the momentum fluxes in deep convective clouds than the associated temperature fluxes. These difficulties, encountered in the regions of deep convection, cannot be explained by the inadequate parametrization of convection because, in our experiments, the true state has also been generated by the same model.

The version of the NCEP GFS that we use employs a modified Arakawa–Schubert scheme (Pan and Wu, 1995) for the parametrization of deep convection. The scheme is simpler than the original Arakawa–Schubert scheme (Arakawa and Schubert, 1974): it assumes that the deep convection is associated with one type of cloud (the deepest cloud) instead of a spectrum of clouds. Nevertheless, the scheme corrects one important deficiency of the original scheme: it allows for the transport of momentum by downdrafts. Otherwise, the scheme retains the central hypothesis of the Arakawa–Schubert scheme: convection is essentially a rapid-response mechanism to neutralize the destabilizing effects of such large-scale processes as surface fluxes and radiation. More precisely, the scheme assumes that the consumption of CAPE, by an ever present ensemble of deep convective clouds, is in a statistical equilibrium with the CAPE generated by the large-scale processes. (We note that, since 2001, NCEP has made several important upgrades to the parametrization of deep convection in the operational GFS. We would not be surprised if an

implementation of the LEKF on the current operational model behaved somewhat differently than reported here.)

One may ask whether our results have any significance in the case where real observations, collected in real deep convective clouds, are assimilated. As Emanuel (1994) pointed out, the observational evidence to support the central hypothesis of Arakawa and Schubert (1974) is striking. He also pointed out, however, that the entraining plume model, on which the Arakawa–Schubert scheme is based, is a poor representation of individual clouds. Thus, in regions of deep convection, we can expect large model errors to occur, which makes a good analysis even more difficult to obtain than we have found here.

This example shows the importance of carrying out experiments in the perfect model scenario. Had we started assimilating real observations first, we might have attributed all difficulties to the inadequate parametrization of deep convection in the model. Our results show that preparing the analysis for regions of deep convection can be challenging even in the absence of model errors. This occurs due to the complex nature of the dynamics (reflected by high E-dimensions) that leads to complex background (and analysis) error structures. Our results show that in the perfect model setup, the analysis error can be efficiently reduced in regions of deep convection by increasing the ensemble size, finding the size of the local regions that is optimal for the given ensemble size, and increasing the number of observations.

Animations of the error propagation reveal that in the tropics, errors and error reduction patterns can episodically propagate from the troposphere to the lower stratosphere. This explains the close relationship between the errors in the regions of ascending motions in the Hadley cells and in the lower stratosphere aloft. We speculate that these episodes are associated with 'convective overshoots', in which the convective updrafts gain enough kinetic energy to break into the stable lower stratospheric air above the cloud tops. ['Convective overshoots' are thought to be responsible for transporting most of the air and water vapor into the stratosphere (e.g. James, 1994, pp. 331–334).]

The origin of the relatively large stratospheric errors is not absolutely clear to us, and we suspect that minor flaws in our implementation of the LEKF may play some role. (For instance, we hope that using a reduced grid near the poles would take into account the decreasing distance between grid points and would help to reduce the errors. Similarly, we may find a better strategy for localization near the upper boundary of the model atmosphere.) Nevertheless, we believe that model dynamics plays some role in the development of the unusually large errors in the polar regions. Animations show that, in the NH polar region, the typical direction of error propagation is from the lower stratosphere to the troposphere, indicating that the sources of errors are instabilities in the lower stratosphere. More precisely, we observe rapidly developing errors in the polar stratosphere (north of 60°N). These errors first propagate toward the pole and then downward in a narrow channel over the pole. These errors are typically much larger than those in the tropical stratosphere, but after a sudden burst they are efficiently removed by the Kalman filter (in contrast to the more persistent nature of the tropical errors). Similar error bursts were observed in our earlier data assimilation experiments with the low-order Lorenz-96 model (Lorenz, 1996; Lorenz and Emanuel, 1998) for both the LEKF scheme (Ott et al., 2004) and a less advanced 3D-Var scheme (Baek et al., 2004). This phenomenon can occur when the data assimilation scheme underestimates the background error (e.g. due to the presence of non-linear error growth), when there are no observations that can help remove the critical errors. The high explained variance in the polar regions suggests that in our case insufficient observational coverage is more likely to be responsible for the large errors in the polar region.

5.10. Timing results

The timing results described here are from our initial implementation of the LEKF, which is on a relatively modest Beowulf cluster consisting of 25 dual-processor nodes, each with 2 GB of random access memory and connected by a 1-gigabit Ethernet; each processor is a 2.8-GHz Intel Pentium Xeon with hyperthreading disabled. Most runs use 40 processors on the cluster. (See Appendix B for more details on the implementation of the code.)

Typically, one complete cycle of the algorithm takes 15 min of wall-clock time for an ensemble of 40 solutions when observations are available at every model grid point using the GFS at T62/L28 resolution and vertical localization, as described below. This computation involves $192 \times 94 \times 28 = 505\,344$ local regions and slightly more than 1.5 million observations; each local region is a cube of $7 \times 7 \times v$ model grid points, where $v = 1$, 3, 5 or 7, depending on altitude (Fig. 1). The time includes: (i) that spent computing the transforms from spectral space to physical space and back (which could be parallelized, but which we have implemented only on a single processor); (ii) the i/o and network overhead to transport the appropriate model grid and observation data to each processor; (iii) one step of the LEKF algorithm to each local region; (iv) 40 6-h forecasts from the resulting analysis. Excluding the spectral transforms and forecasts, the wall-clock time is about 505 s for the above parameters. The average time needed to process one local region, where the ensemble size is 40 and the patch size is 1078, is about 31 ms. Table 2 shows timing results for local regions that consist of $7 \times 7 \times v$ cubes ($v = 1$, 3) as described above, with either 40 or 80 ensemble members. The observing network consists of observations of temperature and wind speed at all 28 vertical levels, plus the surface pressure, at each of N points. The notation, $N =$ all, refers to the case where observations are available at each model grid point (1.5 million observations in total); $N = 2000$ refers to an observing network that consists of 2000 randomly chosen model grid points at which the observations are assumed to exist at each vertical level (i.e. a total of 170 000 observations). The wall-clock time includes the total

Table 2. Time needed to run the LEKF algorithm

Parameters $k+1, N$	Wall-clock time (min)	LEKF step only (s)	Patch mean (10^{-3} s)
40, all	15	505	31
40, 2000	14	447	28
80, all	45	1973	122
80, 2000	42	1682	104

time needed to perform all spectral transforms, the LEKF algorithm, and $k+1$ 6-h forecasts from the resulting analysis. The column labeled 'LEKF only' refers to the maximum amount of time that any given processor spends performing only the data assimilation step after the model grid is distributed more or less evenly across the cluster; the last column shows the average time spent processing a single local region.

These results suggest that, for a given observing network and patch size, the overall time required to perform the LEKF assimilation step grows roughly quadratically with the number of ensemble solutions and is relatively insensitive to the amount of data to be assimilated. In practice, of course, the work required to decode observations from data files and to evaluate the observation operators may be substantially greater than what is needed here. Nevertheless, we are optimistic that, with further tuning and a somewhat larger computer, the LEKF data assimilation algorithm can be performed within the time constraints of a typical operational forecast center. It should be noted that the 6-h forecasts do not represent an additional cost in the operational centers that have ensemble forecasting systems, and that the LEKF would provide initial perturbations representative of the analysis uncertainty.

6. Conclusions

Our conclusions fall into three categories: (i) conclusions that apply generally to ensemble Kalman filter methods that employ localization; (ii) conclusions that, although most directly applicable to the LEKF scheme, may also provide some guidance to other Kalman-filter based schemes; (iii) conclusions that are specific to the LEKF scheme.

6.1. General conclusions

(i) A modest size (40-member) ensemble is sufficient to track the 'true' evolution of the atmospheric state when the model provides a perfect representation of the atmospheric dynamics.

(ii) The more complex the local model dynamics (the higher the E-dimension), the smaller the portion of the background error that the Kalman filter can capture. This leads to relatively large errors in the regions where the model dynamics is complex. The most important regions of complex model dynamics are the tropical convergence zones, where parametrized deep convection plays an important role. We note that unlike the explained variance, the E-dimension of the ensemble can be computed without knowing the 'true' state. It may be possible to improve filter performance by adjusting filter parameters according to the E-dimension.

(iii) While relatively large errors are typically associated with locally high-dimensional behavior of the background forecast, on rare occasions, the analysis error can also be relatively large in regions of low local dimensionality. This anomalous behavior is observed in the polar regions, especially during the polar night. We speculate that this may be because in our implementation in this paper the surface variables were not analyzed. We conjecture that the surface analysis would be best prepared by including the variables representing the time-dependent state of the surface in the state vector analyzed by the LEKF.

(iv) In a real-data implementation, some analysis errors due the limitations of the ensemble Kalman filter approach or to formulation errors in a particular implementation may be easily mistaken for the effects of model errors. For instance, one would expect important model errors in the regions where parametrized deep convection plays an important role. Our results show that part of the large errors in these regions would be due to the high local complexity of the dynamics. (We note that the high complexity of the dynamics and the difficulties with accurately modeling the atmospheric processes have closely related roots.) This example shows the importance of a rigorous initial testing of ensemble-based Kalman filter data assimilation systems on simulated observations.

6.2. Conclusions with limited generality

Some of our other findings are valid only for the LEKF in a strict sense, but they may also be useful as a guide in tuning other ensemble based Kalman filters.

(i) Increasing the number of ensemble members provides a better representation of the space of uncertainties in the background and analysis. The number of observations has little influence on the capability of the scheme to capture the space of uncertainties. Instead, a better observational coverage more efficiently removes background errors in the space identified by the ensemble. There exists an optimal region size for a given number of ensemble members. For our system the optimal horizontal size of the local region in the mid-latitude is about 800×800 km^2.

(ii) The analysis accuracy is most sensitive to the number of ensemble members, dimension of the local regions, and the observational density in the area of the most complex dynamics (in the tropics). In the polar regions, where the analysis errors are large but the local dimensionality of the dynamics is low, the analysis accuracy is mainly sensitive to the observational density.

6.3. LEKF-specific conclusions

The most important LEKF specific finding is the very reasonable computational efficiency of the scheme. We achieved a less than 9 min per analysis performance for the T62 and 28-level version of an operational model, using a cluster of 40 2.8-GHz processors communicating through a 1-Gbit ethernet. We anticipate that the extra computational cost that would arise at the higher resolution of current operational systems could be compensated by the much higher performance of the computers used by the operational centers. Encouraged by the positive results of our attempt at assimilating simulated observations, we have started working on the assimilation of real observations. We are cautiously optimistic about the outcome of these efforts, expecting major but not insurmountable challenges stemming from the presence of model errors.

7. Acknowledgments

The authors would like to thank Jean-Francois Estrade (NCEP, currently ECMWF) Joseph Sela (NCEP) and Yannick Tremolet (NCEP, currently ECMWF) for their advice with implementing the NCEP GFS on our computers, and Zoltan Toth (NCEP) and Wan-Shu Wu (NCEP) for helpful discussions on ensemble forecasting and data assimilation. The two anonymous reviewers provided helpful comments on an earlier version of the manuscript. This work was supported by the Army Research Office, by a James S. McDonnell 21st Century Research Award, by a National Oceanic and Atmospheric Administration THORPEX grant, by the NPOESS Integrated Program Office (IPO), by the Office of Naval Research (Physics), and by the National Science Foundation (Grants 0104087 and PHYS 0098632). EJK gratefully acknowledges partial support from the Arizona State University College of Liberal Arts and Sciences.

8. Appendix A: Non-linear observation operators

If the observation operator $\mathcal{H}$ is non-linear, a slightly different formulation than that of Section 2.3 may be advantageous. By rewriting eqs. (13) and (14), we can place each instance of the linearized operator $\mathbf{H}$ adjacent to the matrix $\mathbf{X}^{\mathrm{b}}$ of ensemble perturbations defined in eq. (4), and then replace $\mathbf{HX}^{\mathrm{b}}$ with the matrix of non-linear observation increments

$$\mathbf{E}^{\mathrm{b}} = \left\{\delta\mathcal{H}\left[\mathbf{x}^{\mathrm{b}(1)}\right] | \, \delta\mathcal{H}\left[\mathbf{x}^{\mathrm{b}(2)}\right] | \, \ldots \, | \, \delta\mathcal{H}\left[\mathbf{x}^{\mathrm{b}(k+1)}\right]\right\}, \tag{A1}$$

where

$$\delta\mathcal{H}[\mathbf{x}^{\mathrm{b}(i)}] = \mathcal{H}[\mathbf{x}^{\mathrm{b}(i)}] - \mathcal{H}(\bar{\mathbf{x}}^{\mathrm{b}}). \tag{A2}$$

Alternatively, one can define $\delta\mathcal{H}[\mathbf{x}^{\mathrm{b}(i)}]$ to be $\mathcal{H}[\mathbf{x}^{\mathrm{b}(i)}]$ minus the mean over i of this expression (in other words, take the mean after applying $\mathcal{H}$ rather than before). This is the approach described by Houtekamer and Mitchell (2001), and is mathematically equivalent to the joint state-observation space approach described by Anderson (2001). With either definition, this approach avoids the potentially non-trivial implementation of $\mathbf{H}$ and $\mathbf{H}^{\mathrm{T}}$, and may produce better analyses by avoiding the unnecessary linearization.

Notice that $\hat{\mathbf{X}}^{\mathrm{b}} = \mathbf{Q}^{\mathrm{T}}\mathbf{X}^{\mathrm{b}}$ and, because $\mathbf{X}^{\mathrm{b}}$ lies in the space spanned by the columns of $\mathbf{Q}$, conversely $\mathbf{X}^{\mathrm{b}} = \mathbf{Q}\hat{\mathbf{X}}^{\mathrm{b}}$. Thus, using eqs. (5) and (10), $\hat{\mathbf{P}}^{\mathrm{b}} = \hat{\mathbf{X}}^{\mathrm{b}}\hat{\mathbf{X}}^{\mathrm{bT}}$. Also, $\mathbf{HX}^{\mathrm{b}} = \mathbf{HQ}\hat{\mathbf{X}}^{\mathrm{b}} = \hat{\mathbf{H}}\hat{\mathbf{X}}^{\mathrm{b}}$, so based on the preceding paragraph we will replace $\hat{\mathbf{H}}\hat{\mathbf{X}}^{\mathrm{b}}$ by $\mathbf{E}^{\mathrm{b}}$. One can show that eq. (14) is equivalent to

$$\hat{\mathbf{P}}^{\mathrm{a}} = \hat{\mathbf{X}}^{\mathrm{b}}\left[\mathbf{I} + \hat{\mathbf{X}}^{\mathrm{bT}}\hat{\mathbf{H}}^{\mathrm{T}}\mathbf{R}^{-1}\hat{\mathbf{H}}\hat{\mathbf{X}}^{\mathrm{b}}_{l}\right]^{-1}\hat{\mathbf{X}}^{\mathrm{bT}}. \tag{A3}$$

Substituting this expression into eq. (13) and substituting $\mathbf{E}^{\mathrm{b}}$ for $\hat{\mathbf{H}}\hat{\mathbf{X}}^{\mathrm{b}}$ by $\mathbf{E}^{\mathrm{b}}$ yields

$$\Delta\hat{\mathbf{x}}^{\mathrm{a}} = \hat{\mathbf{X}}^{\mathrm{b}}[\mathbf{I} + \mathbf{E}^{\mathrm{bT}}\mathbf{R}^{-1}\mathbf{E}^{\mathrm{b}}]^{-1}\mathbf{E}^{\mathrm{bT}}\mathbf{R}^{-1}[\mathbf{y}^{\mathrm{o}} - \mathcal{H}(\bar{\mathbf{X}}^{\mathrm{b}})]. \tag{A4}$$

Also, substituting $\mathbf{E}^{\mathrm{b}}$ into eq. (A3), we have

$$\hat{\mathbf{P}}^{\mathrm{a}} = \hat{\mathbf{X}}^{\mathrm{b}}[\mathbf{I} + \mathbf{E}^{\mathrm{bT}}\mathbf{R}^{-1}\mathbf{E}^{\mathrm{b}}]^{-1}\hat{\mathbf{X}}^{\mathrm{bT}}. \tag{A5}$$

We can then use eqs. (A1), (A4), and (A5) in place of eqs. (13) and (14).

The algorithmic complexity described in Section 2.6 is affected very little by this alternative formulation. We no longer need to compute $\mathbf{H}$ and $\hat{\mathbf{H}}$, but must instead compute $\mathbf{E}^{\mathrm{b}}$ at a cost of roughly k times the cost of computing $\mathcal{H}(\bar{\mathbf{x}}^{\mathrm{b}})$. Because $\mathbf{E}^{\mathrm{b}}$ has $k+1$ columns, while $\hat{\mathbf{H}}$ has k, some operations will take slightly longer accordingly.

9. Appendix B: Computational implementation

Our initial implementation of the LEKF is in FORTRAN 95, which provides a simple, portable, and efficient notation for handling dense matrices. We have used version 3 of the LAPACK library (Anderson et al., 1999) to compute matrix inverses, eigenvalues, and eigenvectors, because it is numerically robust, thoroughly tested, and widely available; most computer vendors provide optimized implementations of the Basic Linear Algebra Subroutines (BLAS) (Lawson et al., 1979; Dongarra et al., 1988, 1990) upon which the LAPACK library is built. The LAPACK routine DSYEVR implements the algorithm of choice for finding all eigenvalues and eigenvectors of a symmetric $k \times k$ matrix. However, for maximum efficiency, DSYEVR requires IEEE-754 (IEEE, 1985) infinity arithmetic to be implemented without trapping (Anderson et al., 1999, p. 146). This requirement can be problematic, depending on the processor and FORTRAN compiler, although we have had no difficulty with version 6.1 of the Lahey FORTRAN compiler (Lahey Computer Systems, see http://www.lahey.com) on Intel Pentium Xeon processors running Red Hat Linux (Red Hat, Inc., see http://www.redhat.com). Because the underlying algorithm is iterative, the operation count associated with DSYEVR is not fixed: the convergence rate depends on the data, but in general is

expected to be $O(k^3)$. The overall efficiency of the LEKF algorithm also is influenced by the quality of the FORTRAN intrinsic function MATMUL, which we use heavily to multiply matrices. The level-3 BLAS routines DGEMM and DGEMV (Dongarra et al., 1990) may be substituted to yield faster code, depending on the FORTRAN implementation.

References

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

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

Anderson, E., Bai, Z., Bischof, C., Blackford, L. S., Demmel, J. and co-authors. 1999. LAPACK *Users' Guide*, 3rd Edition. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Arakawa, A. and Schubert, W. H. 1974. Interaction of a cumulus cloud ensemble with the large-scale environment.Part I. *J. Atmos. Sci.* **31**, 674–701.

Baek, S-J., Hunt, B. R., Szunyogh, I., Zimin. A. and Ott, E. 2004. Localized error bursts in estimating the state of spatiotemporal chaos. *Chaos* **14**, 1042–1049.

Bishop, C. H., Etherton, B. J. and Majumdar, S. 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.* **129**, 420–436.

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M. and Blade, I. 1999. The effective number of spatial degrees of freedom of a time-varying field. *J. Climate* **12**, 1990–2009.

Daley, R. 1991. *Atmospheric Data Analysis*. Cambridge Univ. Press, Cambridge.

Dee, D. P. 1995. Testing the perfect-model assumption in variational data assimilation. *Proc. 2nd Int. Symp. on Assimilation of Observations in Meteorology and Oceanography*. World Meteorological Organization, Tokyo, Japan, 225–228.

Dongarra, J. J., Du Croz, J., Hammarling, S. and Hanson, R. J. 1988. An extended set of FORTRAN Basic Linear Algebra Subprograms. *ACM Trans. Math. Soft.* **14**, 1–17.

Dongarra, J. J., Du Croz, J., Hammarling, S. and Duff, I. S. 1990. A set of Level 3 Basic Linear Algebra Subprograms. *ACM Trans. Math. Soft.* **16**, 1–17.

Dowell, D. C., Zhang, F., Wicker, L. J., Snyder, C. and Crook, N. A. 2004. Wind and thermodynamic retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: ensemble Kalman filter experiments. *Mon. Wea. Rev.* **132**, 1982–2005.

Emanuel, K. A. 1994. *Atmospheric Convection*. Oxford Univ. Press, New York.

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

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

Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M. and co-authors. 2005. Atmospheric data assimilation with the ensemble Kalman filter: results with real observations. *Mon. Wea. Rev.* **133**, 604–620.

Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E. and Patil, D. J. and co-authors. 2004. Four-dimensional ensemble Kalman filtering. *Tellus* **56A**, 273–277.

IEEE, 1985. IEEE Standard for Binary Floating-Point Arithmetic. *ANSI/IEEE Std. 754-1985*. American National Standards Institute.

James, I. N. 1994 *Introduction to Circulating Atmospheres*. Cambridge Univ. Press, Cambridge.

Kalnay, E. 2003. *Atmospheric Modeling, Data Assimilation, and Predictability*. Cambridge Univ. Press, Cambridge.

Keppenne, C. and Rienecker, H. 2002. Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.* **130**, 2951–2965.

Lawson, C. L., Hanson, R. J., Kincaid, D. R. and Krogh F. T. 1979. Basic Linear Algebra Subprograms for FORTRAN usage. *ACM Trans. Math. Soft.* **5,** 308–323.

Lorenz, E. N. 1996. Predictability: a problem partly solved. In *Proc. Seminar on Predictability, Vol. 1*. European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire, RG2 9AX, UK.

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

Lynch, P. 2002. The swinging spring: a simple model of atmospheric balance. In: *Large-scale Atmosphere–Ocean Dynamics II* (eds. J. Norbury, and I. Roulstone). Cambridge Univ. Press, Cambridge.

Lynch, P. and Huang, P. M. 1992. Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.* **120**, 1019–1034.

Machenauer, B. 1997. On the dynamics of gravity oscillation in a shallow-water model, with application to normal mode initialization. *Beitr. Phys. Atmos.* **50**, 253–251.

Menard, R., Cohn, S. E., Chang, L-P. and Lyster, P. M. 2000. Assimilation of stratospheric chemical tracer observations using a Kalman Filter. Part I: Formulation. *Mon. Wea. Rev.* **128**, 2654–2670.

Oczkowski, M., Szunyogh, I., Patil, D. J., and Zimin, A. V. 2005. Mechanisms for the development of locally low dimensional atmospheric dynamics. *J. Atmos. Sci.* **65**, 1135–1156.

Ott, E., Hunt, B. H., Szunyogh, I., Corazza, M., Kalnay, E. and co-authors. 2002. Exploiting local low dimensionality of the atmospheric dynamics for efficient Kalman filtering. Preprint (physics/0203058).

Ott, E., Hunt, B. H., Szunyogh, I., Zimin, A. V., Kostelich, E. J. and co-authors. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* **56A**, 415–428.

Pan, H-L. and Wu, W-S. 1995. Implementing a Mass Flux Convection Package for the NMC Medium-Range Forecast model. *NMC Office Note* **409**. Available from NCEP, 5200 Auth Road, Washington, 20233.

Patil, D. J., Hunt, B. R., Kalnay, E., Yorke, J. A. and Ott, E. 2001. Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.* **86**, 5878–5881.

Snyder, C. and Zhang, F. 2003. Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.* **131**, 1663–1677.

Szunyogh, I., and Toth, Z. 2002. The effect of increased horizontal resolution on the NCEP global ensemble mean forecasts. *Mon. Wea. Rev.* **130**, 1125–1143.

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

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

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

Zhang, F., Snyder, C. and Sun, J. 2004. Impacts of initial estimate and observation availability on convective-scale dataassimilation with an ensemble Kalman filter. *Mon. Wea. Rev.* **132**, 1238–1253.