

Comparison of Local Ensemble Transform Kalman Filter, 3DVAR, and 4DVAR in a Quasigeostrophic Model

SHU-CHIH YANG*

Earth System Science Interdisciplinary Center, Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, and Global Modeling Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

MATTEO CORAZZA

ARPAL-CFMI-PC, Genoa, Italy

ALBERTO CARRASSI

Institut Royal Meteorologique de Belgique, Brussels, Belgium

EUGENIA KALNAY

Earth System Science Interdisciplinary Center, Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland

TAKEMASA MIYOSHI

Numerical Prediction Division, Japan Meteorological Agency, Tokyo, Japan

(Manuscript received 28 September 2007, in final form 3 June 2008)

ABSTRACT

Local ensemble transform Kalman filter (LETKF) data assimilation, three-dimensional variational data assimilation (3DVAR), and four-dimensional variational data assimilation (4DVAR) schemes are implemented in a quasigeostrophic channel model. Their advantages and disadvantages are compared to assess their use in practical applications. LETKF and 4DVAR, which take into account the flow-dependent errors, outperform 3DVAR under a perfect model scenario. Given the same observations, LETKF produces more accurate analyses than 4DVAR with a 12-h window by effectively correcting the fast-growing errors with the flow-dependent background error covariance. Even though 4DVAR performance benefits substantially from using a longer assimilation window, LETKF is also able to achieve a satisfactory accuracy compared to the 24-h 4DVAR analyses. It is shown that the advantage of the LETKF over 3DVAR is a result of both the ensemble averaging and the information about the “errors of the day” provided by the ensemble. The analysis corrections at the end of the 12-h assimilation window are similar for LETKF and the 12-h window 4DVAR, and they both resemble bred vectors. At the beginning of the assimilation window, LETKF analysis corrections obtained using a no-cost smoother also resemble the corresponding bred vectors, whereas the 4DVAR corrections are significantly different with much larger horizontal scales.

1. Introduction

The procedure that combines all available information on the state of a physical system to obtain the best estimate of its state is known as data assimilation (DA); such a best estimate is usually referred to as an *analysis*. In geophysics and specifically in numerical weather prediction (NWP), DA algorithms are designed to provide optimal initial conditions for prediction models. By extracting information from the observations and

* Current affiliation: Department of Atmospheric Sciences, National Central University, Jhongli, Taiwan.

Corresponding author address: Shu-Chih Yang, Department of Atmospheric Sciences, National Central University, Jhongli 32001, Taiwan.
E-mail: shuchih.yang@atm.ncu.edu.tw

system physics, a good DA algorithm should improve the analysis and consequently the forecast skill. Due to the chaotic nature of nonlinear systems such as the atmosphere and the oceans, accurate initial conditions are crucial to making reliable forecasts. Most current operational DA schemes statistically combine the observations and a short-range forecast.

Three-dimensional variational data assimilation (3DVAR) is considered an economically and statistically reliable DA method and is implemented in many operational centers. It assumes that the two sources of information, forecast and observations, have errors that are adequately described by static error covariances (Talagrand 1997). Although these assumptions make solving realistic NWP problems computationally tractable (Parrish and Derber 1992), 3DVAR misses the time-dependent and nonnormal error dynamics common in nonlinear chaotic systems, that is, the “errors of the day.”

Four-dimensional variational data assimilation (4DVAR) is an advanced DA technique that computes the model trajectory that best fits the observations distributed within a given time interval with the dynamical constraints of the model equations (Talagrand and Courtier 1987; Courtier et al. 1994; Rabier et al. 2000). The 4DVAR employs optimal control theory (Le Dimet and Talagrand 1986) to minimize the cost function defined over the time interval by using an adjoint model to determine its gradient. Given its high computational cost of minimizing the cost function, 4DVAR has been implemented operationally only in its simplified incremental form (Courtier et al. 1994), and the tangent linear and the adjoints models with the simplified physics are used in the inner loop to speed up the minimization. The forecast covariance matrix in 4DVAR is implicitly evolved within the assimilation window from a constant initial background error covariance such that an updated analysis covariance is nontrivial to obtain for the next assimilation cycle (e.g., Kalnay 2003). Although studies show that 4DVAR analyses could be improved by using a longer assimilation window (Pires et al. 1996), this extension comes with computational costs that limit its application for dynamically complex models and also the limitation from model errors. Operational centers have not yet used an assimilation window of more than one day for a 6-h analysis cycle.

In contrast to variational methods, the ensemble Kalman filter (EnKF) uses a sequential method based on the Kalman filter (KF; Kalman 1960; Kalman and Bucy 1961). A KF uses an evolving full-rank error covariance, which is computationally extremely expensive. A KF produces equivalent results to 4DVAR at the end of the assimilation window given a linear model, the same Gaussian error statistics and linear observa-

tion operators (Lorenc 1986). During the past decade, experience with ensemble forecasting has suggested that an ensemble approach could address some of the computational costs of KF with large and complex dynamics. The first and most celebrated ensemble-based scheme is the ensemble Kalman filter (Evensen 1994). It applies estimation theory with a Monte Carlo statistical approach to the conceptual and mathematical framework of KF. Most of the current ensemble-based data assimilation schemes have been designed to offer a nonlinear extension of the KF approach while reducing its computational cost. The use of the full nonlinear model may be beneficial for analyses for situations in which nonlinearity is strong and statistics exhibit some nonnormality (Hamill 2006; Yang 2005). Several types of ensemble-based KF algorithms have been developed using either stochastic (perturbed observations) or deterministic (square root) filters. In each ensemble-based scheme, a nonlinearly evolved ensemble of trajectories samples the unknown flow-dependent error distribution using an ensemble number several orders of magnitude smaller than the dimensions of the system’s state vector. Given the same limited ensemble size, the square root filters have proven to be more accurate than the stochastic methods because random errors are introduced through the perturbed observations (Whitaker and Hamill 2002). Discussions and reviews of the ensemble-based data assimilations method can be found in Hamill (2006) and Evensen (2003). The results obtained so far indicate that they represent a feasible alternative to 4DVAR (Houtekamer and Mitchell 1998; Anderson 2001; Whitaker and Hamill 2002; Evensen, 2003; Ott et al. 2004; Kalnay et al. 2007a; Caya et al. 2005; Miyoshi and Yamane 2007).

Compared to 4DVAR, an ensemble-based KF scheme is easy to implement and maintain, because it does not require the development and maintenance of the tangent linear and adjoint models. Its analysis ensemble also provides a set of dynamically consistent states to initialize an ensemble prediction system while the 3DVAR/4DVAR schemes require an additional procedure to start a probabilistic forecast. In ensemble KF methods, flow-state information, such as uncertainties associated with flow instabilities, is propagated through the DA cycle, unlike variational-based methods. Lorenc (2003) and Kalnay et al. (2007a) provide more detailed discussions of the pros and cons of ensemble Kalman filters and 4DVAR. Comparisons between variational-based and ensemble-based DA schemes have mostly used simple models (e.g., Anderson 2001; Fertig et al. 2007). Recently, Caya et al. (2005) compared 4DVAR and EnKF with a realistic atmospheric model for the convective-scale assimilation and found

that 4DVAR was more accurate in the initial stages and that EnKF became more accurate later in the development of a storm. This suggests that EnKF has a longer spinup, a problem addressed by E. Kalnay and S.-C. Yang (2008, manuscript submitted to *Quart. J. Roy. Meteor. Soc.*, hereinafter KY08). The impact of assuming a constant background error covariance used in 4DVAR, given the short assimilation window affordable in operational centers, still remains to be clearly assessed.

This work compares variational and ensemble-based DA schemes based on the quality of their analyses as well as their computational costs. One ensemble-based and two variational schemes are applied to a quasigeostrophic channel model using the same “noisy” observations. The two variational schemes are 3DVAR (developed by Morss 1999 and Morss et al. 2001, following Parrish and Derber 1992) and 4DVAR (newly implemented for this study). The ensemble scheme is the local ensemble transform Kalman filter (LETKF; based on Hunt et al. 2007), an efficient ensemble square root filter in a parallel computational setup (Whitaker et al. 2008). In this study, the perfect model assumption is made for all the experiments to focus on the ability of the DA schemes to control and reduce errors coming from an incorrect estimate of the initial conditions. This article explores the differences between the variational-based and the LETKF methods and discusses considerations applicable to using these methods operationally.

The paper is organized as follows: section 2 outlines the model and observation network setup, section 3 describes the DA schemes used in this study, and section 4 presents the results. Finally, the findings are summarized and discussed in Section 5.

2. Model and observing system design

a. The quasigeostrophic, tangent linear, and adjoint models

All the data assimilation schemes are implemented in the quasigeostrophic (QG) model developed by Rotunno and Bao (1996). It is a periodic channel model on a beta plane. At the resolution used in this study, it has 64 grid points in the zonal direction, 33 grid points in the meridional direction, and 7 vertical levels. Physical processes include advection, diffusion, relaxation, and Ekman pumping at the bottom level. The model variables are nondimensional potential temperature at the bottom and top levels, and nondimensional potential vorticity at the five inner levels. Note that the model variables are also the analysis variables in all the following assimilation schemes. The integration time step is 30 min. The numerical schemes used for advecting and

inverting PV are described in Rotunno and Bao (1996). The forcing and dissipation included in the model are specified in Snyder et al. (2003), where some characteristics of the statistically steady turbulent flow are also discussed. This model has been widely used for testing data assimilation, error characteristics, and adaptive observations methods (e.g., Morss 1999; Hamill and Snyder 2000; Snyder et al. 2003; Corazza et al. 2003, 2007; Kim et al. 2004; Carrassi et al. 2007). In this study, the model is assumed to be perfect and the true state, from which observations are extracted, is represented by a reference trajectory integrated by the QG model.

The implementation of the 4DVAR requires the development of the tangent linear and adjoint models. To this end, the Tangent Linear and Adjoint Model Compiler (TAMC; Giering and Kaminski 1998) was used to generate a preliminary version of the codes, which did not fulfill the boundary conditions automatically. Several very subtle corrections (i.e., a long debugging effort) were required for the TAMC-generated linear and adjoint models to eliminate a spurious accumulation of extreme values at the meridional and vertical walls and at zonal periodic boundaries. Verification checks for tangent linear and adjoint codes, following Navon et al. (1992), indicate that the linear regime is valid for forecasts up to 5 days.

b. The observing system configuration

The simulated “rawinsonde observations” consist of the velocity components and temperature at all levels. They are generated from the true state through a linear observation operator, \mathbf{H} , mapping from model variables into observation variables (Morss 1999). In \mathbf{H} , the wind and temperature are calculated through finite differences of the streamfunction, which is obtained from the model variables (potential vorticity and temperature). Sixty-four rawinsonde observations are used and their locations are randomly chosen and remain fixed afterward. The observation locations are on the model grid points and cover about 3% of the domain. Observations are available every 12 h and the analysis cycle is also performed every 12 h.

Observation errors are generated by adding white random noise sampled by a Gaussian distribution consistent with the observational error covariance matrix (Morss 1999; Morss et al. 2001). The observation error covariance matrix is constructed following Dey and Morone (1985): the observation error is assumed to be uncorrelated between observations and between different variables. Only vertical correlations for the same variable are considered. The wind and temperature observation error variances are adapted from Parrish and Derber (1992) and the corresponding values are provided in Morss (1999).

3 Data assimilation schemes

a. 3DVAR

The 3DVAR system implemented for the QG model was developed by Morss (1999). The time-independent background error covariance matrix \mathbf{B}_{3DVAR} (in the model gridpoint coordinates is truncated and saved in the form of spectral coordinates as in Parrish and Derber (1992). The background error covariance in spectral coordinates (\mathbf{C}) is assumed to have separable horizontal and vertical structures. It is built by calculating the horizontal error covariances at each level and linking them through a vertical correlation matrix as

$$\mathbf{C} = \hat{\mathbf{C}}^{1/2} \mathbf{V} \hat{\mathbf{C}}^{1/2}. \quad (1)$$

In (1) $\hat{\mathbf{C}}$, the matrix of horizontal background error covariance at each level, is diagonal because the errors corresponding to different wavelengths are assumed to be uncorrelated in horizontal spectral coordinates. The \mathbf{V} is a 7×7 matrix of background error correlations between different levels averaged from all the horizontal spectral components. The gridpoint space matrix \mathbf{B}_{3DVAR} is then obtained using an operator (\mathbf{S}) to transform from gridpoint to spectral space. Thus, $\mathbf{B}_{3DVAR} = \mathbf{S} \mathbf{C} \mathbf{S}^T$. This structure of the error covariance in spectral space is also used for 4DVAR (section 3b).

The matrices $\hat{\mathbf{C}}$ and \mathbf{V} given in Morss (1999) were generated with 32 observations. Because baroclinic instability is the dominant dynamical process in this model, changes in the number of observations have little impact on the structure \mathbf{C} . The performance of 3DVAR is not sensitive to the assumption that \mathbf{C} is the same for all observation densities. The amplitude of \mathbf{C} is reduced here by a factor of 0.4 to optimize the 3DVAR results for the present observation configuration.

3DVAR is implemented and solved in the incremental cost-function form as follows:

$$J(\delta \mathbf{x}_a) = \frac{1}{2} \{ \delta \mathbf{x}_a^T \mathbf{B}_{3DVAR}^{-1} \delta \mathbf{x}_a + [\mathbf{y}^o - H(\mathbf{x}_b)]^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x}_b)] \}, \quad (2)$$

where \mathbf{y}^o is the observational vector, \mathbf{R} is the observation error covariance matrix, and H is the observation operator. The control variable for the minimization of the cost function is the analysis increment $\delta \mathbf{x}_a = \mathbf{x}_a - \mathbf{x}_b$, where \mathbf{x}_a is the analysis and \mathbf{x}_b is the background state vector. The minimum in the cost function (2) is obtained by solving for the model state \mathbf{x}_a , which has a cost-function gradient equal to zero:

$$(\mathbf{I} + \mathbf{B}_{3DVAR} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})(\mathbf{x}_a - \mathbf{x}_b) = \mathbf{B}_{3DVAR} \mathbf{H}^T \mathbf{R}^{-1} \times (\mathbf{y}^o - \mathbf{x}_b). \quad (3)$$

In (3), \mathbf{H} is the linearized observation operator. In all the applications described in this work, the observation operator is inherently linear. The analysis increment is solved with conjugate residual minimization.

In this study, the 3DVAR provides the benchmark for comparison with other assimilation schemes. Further details on the configuration setup of 3DVAR are discussed in Morss (1999).

b. 4DVAR

The 4DVAR cost function (4) generalizes the 3DVAR cost function in that it considers observations distributed within a time interval. Given the background state at the beginning of the assimilation window, \mathbf{x}_b , and the observations \mathbf{y}_i^o at time t_i , the minimization of the cost function (4) provides the initial condition (at the beginning of the time interval) leading to the forecast trajectory that best fits all the observations within the assimilation window (Courtier et al. 1994):

$$J[\mathbf{x}(t_0)] = \frac{1}{2} [\mathbf{x}(t_0) - \mathbf{x}_b]^T \mathbf{B}^{-1} [\mathbf{x}(t_0) - \mathbf{x}_b] + \frac{1}{2} \sum_{i=0}^{i=n} \{ H[\mathbf{x}(t_i)] - \mathbf{y}_i^o \}^T \mathbf{R}^{-1} \{ H[\mathbf{x}(t_i)] - \mathbf{y}_i^o \}. \quad (4)$$

In (4), H is the observation operator. The incremental approach (Courtier et al. 1994) is used to seek the minimum of the 4DVAR cost function with the tangent linear approximation. Defining the analysis increment as the difference between the minimum in (4) and the background state $[\mathbf{x}(t_0) - \mathbf{x}_b = \delta \mathbf{x}(t_0)]$, the incremental form of the cost function is shown in (5), and the corresponding gradient of (5) with respect to $\delta \mathbf{x}(t_0)$ is shown in (6):

$$J[\delta \mathbf{x}(t_0)] = \frac{1}{2} [\delta \mathbf{x}(t_0)]^T \mathbf{B}^{-1} [\delta \mathbf{x}(t_0)] + \frac{1}{2} \sum_{i=0}^n [\mathbf{H}_i \mathbf{L}(t_0, t_i) \delta \mathbf{x}(t_0) - \mathbf{d}(t_i)]^T \mathbf{R}^{-1} \times [\mathbf{H}_i \mathbf{L}(t_0, t_i) \delta \mathbf{x}(t_0) - \mathbf{d}(t_i)] \quad \text{and} \quad (5)$$

$$\nabla J[\delta \mathbf{x}(t_0)] = \mathbf{B}^{-1} [\delta \mathbf{x}(t_0)] + \sum_{i=0}^K \mathbf{L}^T(t_i, t_0) \mathbf{H}_i^T \mathbf{R}^{-1} \times [\mathbf{H}_i \mathbf{L}(t_0, t_i) \delta \mathbf{x}(t_0) - \mathbf{d}(t_i)]. \quad (6)$$

In these equations, $\mathbf{d}(t_i)$ is the innovation (the difference between the background state and observations) at observing time t_i , $\mathbf{L}(t_0, t_i)$ is the tangent linear

(forward) model advancing a perturbation from t_0 to t_i , and $\mathbf{L}^T(t_i, t_0)$ is the adjoint (backward) operator. In (4) or (5) and (6), the background error covariance \mathbf{B} at the beginning of the window t_0 is important in initially distributing the correction. In Courtier et al. (1994), the process of minimization of (5) is referred to as the “inner loop,” because only the linear operators are involved and the nonlinearity of the trajectory is not considered. We also note that in the operational framework, these linear operators typically use simplified physics and/or a low-resolution grid. To account for nonlinearity, an outer loop is applied so that the increment is used to update the background and its distance $\mathbf{d}(t_i)$ to the observations. Then, this improved background and the innovations are used for the next cycle of the inner loop. In this study, the nonlinearity has been considered by applying the full nonlinear model to compute $[\mathbf{H}\mathbf{L}(t_0, t_i)\delta\mathbf{x}(t_0) - \mathbf{d}(t_i) = \mathbf{H}M_{t_0 \rightarrow t_i}[\mathbf{x}(t_0)] - \mathbf{y}_i^o]$, so that the outer loop is unnecessary.

The incremental cost function (5) is preconditioned using a variable transformation that avoids inverting \mathbf{B} to compute the cost function directly and make the minimization more efficient. The transformation operator \mathbf{U} is chosen to be the square root of the inverse of \mathbf{B} , that is, $\mathbf{U} = \mathbf{B}^{-1/2}$. With the preconditioned variable $\delta\mathbf{v}$, the analysis increment is expressed as $\delta\mathbf{x} = \mathbf{U}^{-1}\delta\mathbf{v}$ and the cost function is reformulated as

$$J[\delta\mathbf{v}(t_0)] = \frac{1}{2}[\delta\mathbf{v}(t_0)]^T \delta\mathbf{v}(t_0) + \frac{1}{2} \sum_{i=0}^n [\mathbf{H}\mathbf{L}(t_0, t_i)\mathbf{U}^{-1}\delta\mathbf{v}(t_0) - \mathbf{d}(t_i)]^T \mathbf{R}^{-1} [\mathbf{H}\mathbf{L}(t_0, t_i)\mathbf{U}^{-1}\delta\mathbf{v}(t_0) - \mathbf{d}(t_i)]. \quad (7)$$

The gradient of (7) for $\delta\mathbf{v}(t_0)$ becomes

$$\nabla J[\delta\mathbf{v}(t_0)] = \delta\mathbf{v}(t_0) + \sum_{i=0}^K (\mathbf{U}^{-1})^T \mathbf{L}^T(t_i, t_0) \mathbf{H}^T \mathbf{R}^{-1} \times [\mathbf{H}\mathbf{L}(t_0, t_i)\mathbf{U}^{-1}\delta\mathbf{v}(t_0) - \mathbf{d}(t_i)]. \quad (8)$$

This allows an iterative minimization with respect to the variable $\delta\mathbf{v}$, which is initially set to zero.

In this study, \mathbf{B} is assumed to have the same structure as $\mathbf{B}_{3\text{DVAR}}$, a reasonable assumption for the perfect model experiments, and the amplitude of the \mathbf{B} is optimized by tuning a constant factor b_0 :

$$\mathbf{B} = b_0 \times \mathbf{B}_{3\text{DVAR}} = \mathbf{S}(\sqrt{b_0}\hat{\mathbf{C}}^{1/2})\mathbf{V}(\sqrt{b_0}\hat{\mathbf{C}}^{1/2})\mathbf{S}^T. \quad (9)$$

For the 12-h assimilation window, $b_0 = 0.05$ was determined to be optimal, indicating that the analysis error variance from 4DVAR is much smaller than 3DVAR.

TABLE 1. Mean RMS of analysis errors (in terms of generalized potential vorticity) of 4DVAR with different assimilation window lengths averaged over 80 days.

	3DVAR	4DVAR					
		12 h	24 h	48 h	72 h	96 h	120 h
RMSE ($\times 10^2$)	6.24	2.63	1.81	1.50	1.37	1.29	1.19

As the assimilation window becomes longer, the accuracy is less sensitive to the amplitude of \mathbf{B} . In the following results, this optimally tuned \mathbf{B} is used as the initial background error covariance for all the 4DVAR experiments. From (1), \mathbf{U}^{-1} is defined as

$$\mathbf{U}^{-1} = \mathbf{S}(\sqrt{b_0}\hat{\mathbf{C}}^{1/2})\mathbf{V}^{1/2}. \quad (10)$$

Because $\hat{\mathbf{C}}^{1/2}$ and \mathbf{V} are the same as in 3DVAR, only the square root of the vertical correlation matrix ($\mathbf{V}^{1/2}$) is required, a trivial computation for a 7×7 matrix.

Given an initial guess for $\delta\mathbf{v}(t_0)$ defined in spectral coordinates, the cost function and its gradient are computed using (6) and (7). A Liu and Nocedal (1989) update of Broyden (1969), Fletcher (1970), Goldfarb (1970), and Shanno (1970) (L-BFGS) quasi-Newton minimizer is used to determine how $\delta\mathbf{v}(t_0)$ should be modified to reduce the value of (7) with respect to its gradient. After each iteration, $\delta\mathbf{v}(t_0)$ is converted back to grid coordinates to derive a new initial increment $[\delta\mathbf{x}(t_0)]$, which is then added to the initial background state to generate a new initial state $\mathbf{x}(t_0)$. The full nonlinear model is used to forward integrate $\mathbf{x}(t_0)$ to the end of the assimilation window. The process is repeated until the minimization criterion for the L_2 norm of the cost-function gradient is within a chosen threshold tolerance value of 10^{-3} and terminated if the maximum number of 30 iterations is reached. As mentioned before, the innovation vector is also approximated using the full nonlinear model and assuming that the analysis state is close to the background state, as follows:

$$\mathbf{H}\mathbf{L}(t_0, t_i)\mathbf{U}^{-1}\delta\mathbf{v} - \mathbf{d}(t_i) \approx \mathbf{H}M_{t_0 \rightarrow t_i}[\mathbf{x}(t_0)] - \mathbf{y}_i^o. \quad (11)$$

By doing so, we also take into account the nonlinearity of the model trajectory. To determine a practical assimilation window length, 4DVAR runs were made for lengths ranging from 12 h to 5 days. Table 1 shows the time and domain root-mean-square (RMS) analysis errors, in terms of the generalized potential vorticity norm,¹ for

¹ The generalized potential vorticity (\bar{q}) is the same as the model potential vorticity (q), except for the first and fifth levels where the potential temperatures (θ) are incorporated. For these levels, it is defined as $q_1 = q_1 + (N_b/\delta z)\theta_b$; $q_5 = q_5 - (N_t/\delta z)\theta_t$, where N is the nondimensional static instability and subscript b is for the bottom level and t is for the top level.

different assimilation windows averaged over 80 days. These results confirm that longer windows, although computationally costlier, improve the 4DVAR analyses (Pires et al. 1996; Kalnay et al. 2007a). Most of the improvement is gained when the assimilation window increases from 12 h to one day, while beyond one day the improvement is small. By comparing the experiment on a Linux PC machine with a Pentium 2.66-GHz processor and a 1GB of memory, the computational time needed for a 5-day window is 9 times larger than that needed for a 12-h window. As a consequence, it may become impractical to use a very long assimilation window with the 4DVAR scheme in an operational framework (in addition to the fact that the perturbations dynamics would become very nonlinear in a more unstable system than this QG model).

Here, 12- and 24-h assimilation windows are used as proxies of short and long windows that are operationally affordable. Comparing the error doubling time of about 2.5–4 days in this QG model (Morss 1999) to a realistic atmospheric NWP model with a doubling rate of 1.5–2 days (Toth and Kalnay 1993; Simmons et al. 1995), the 4DVAR performance with a 12-h assimilation window in this QG model could be a proxy for the performance of the 6-h 4DVAR with an NWP model. Thus, we can refer to a 12-h window as a short window and use it to compare with the LETKF performance. Also, the result of 4DVAR with a 24-h window is used to represent the performance of a longer window. Despite not being very long, most of the advantages of longer windows were attained at 24 h.

For the 12-h assimilation window, observations are available at the end of the window; for the 24-h assimilation window, observations are available at the middle and the end of the window. To perform an analysis every 12 h and avoid observations being used twice, two successive assimilation windows in the 24-h 4DVAR are overlapped but initialized independently from each other, as done operationally (A. Lorenc 2008, personal communication). These two experiments will be referred to as 4DV12H and 4DV24H, respectively. Note that only the 4DV12H window uses the same information as 3DVAR and LETKF.

c. Local ensemble transform Kalman filter

The local ensemble Kalman filter (LEKF) was first proposed by Ott et al. (2004), who solved the ensemble Kalman filter equations in local patches exploiting the low-dimensional properties of atmospheric instabilities (Patil et al. 2001). Szunyogh et al. (2005) successfully tested this scheme in a large realistic atmospheric primitive equation model with complicated physics [National Centers for Environmental

Prediction (NCEP) GFS]. Corazza et al. (2007) compared LEKF with 3DVAR with the same QG model used here. The LEKF scheme was modified by Hunt et al. (2007) into LETKF, an equivalent but more efficient approach. This scheme provides essentially identical results to the ensemble square root filter of Whitaker and Hamill (2002) but is computationally more efficient as the number of processors increases (Whitaker et al. 2008). The LETKF scheme assimilates observations in local domains, allowing both the DA step and the construction of new ensemble vectors to be performed locally and in parallel, in contrast to the variational methods (3DVAR/4DVAR). In LETKF, a transform matrix (Bishop et al. 2001) is used to map from the K -dimensional space back to the local physical space.

The local analysis is performed sequentially in a local domain (local volume) with total $N = (2l + 1)^2 \times (2l_z + 1)$ grid points, where l and l_z are the chosen horizontal and vertical number of grid points surrounding the central analysis grid point. The horizontal projection of the local volume is referred to as the local “patch.” In a single-processor computer, LETKF computes local analyses sequentially. The local background ensemble mean $\bar{\mathbf{x}}_f$ is an $N \times 1$ vector, and the ensemble perturbations \mathbf{X}_f are arranged by column in an $N \times K$ matrix. A similar notation is adopted for the ensemble of analysis states, with the mean denoted by $\bar{\mathbf{x}}_a$ and the matrix of deviations denoted by \mathbf{X}_a . The standard KF formula (Kalman 1960; Kalman and Bucy 1961) is used to solve the analysis with the corresponding “local” error statistics (Hunt et al. 2007). The local analysis ensemble mean is represented by

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_f + \mathbf{P}_f \mathbf{H}^T (\mathbf{H} \mathbf{P}_f \mathbf{H}^T + \mathbf{R})^{-1} [\mathbf{y}_o - H(\bar{\mathbf{x}}_f)], \quad (12)$$

where \mathbf{y}_o , \mathbf{H} , and \mathbf{R} have the same definitions as in sections 3a,b but now are defined in the local domain. The \mathbf{P}_f is the local forecast error covariance in the model grid coordinate (physical space), estimated from \mathbf{X}_f , the matrix whose columns are K local ensemble forecast perturbations. Thus,

$$\mathbf{P}_f = \frac{1}{K-1} \mathbf{X}_f \mathbf{X}_f^T. \quad (13)$$

LETKF transforms the ensemble background perturbations into the analysis ensemble perturbations in the K -dimension ensemble space so that

$$\mathbf{P}_a = \frac{1}{K-1} \mathbf{X}_a \mathbf{X}_a^T = \mathbf{X}_f \tilde{\mathbf{P}}_a \mathbf{X}_f^T, \quad (14)$$

where $\tilde{\mathbf{P}}_a$ denotes the analysis error covariance in the ensemble space, computed as

TABLE 2. Settings adopted for the LETKF system for the simulations described in the text.

Horizontal dimension of the local domain	19×19 ($l = 9$)
Number of ensemble members	$K = 40$
Method to update the global field	Center point
Amplitude of the random perturbations	2% vectors amplitude
Multiplicative variance inflation	1.08 1.06 1.07 1.08 1.075 1.07 1.14
Decorrelation length	$r_d = 7$

$$\tilde{\mathbf{P}}_a = [(\mathbf{H}\mathbf{X}_f)^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{X}_f) + (K - 1)\mathbf{I}]^{-1}. \quad (15)$$

The analysis ensemble perturbations are

$$\mathbf{X}_a = \mathbf{X}_f [(K - 1)\tilde{\mathbf{P}}_a]^{1/2}. \quad (16)$$

Note that a symmetric square root of $\tilde{\mathbf{P}}_a$ is used to obtain the analysis ensemble perturbations. This ensures that there is a zero mean for ensemble perturbations, the matrix $[(K - 1)\tilde{\mathbf{P}}_a]^{1/2}$ is closest to the identity matrix, given the constraint of the analysis error covariance, and the analysis ensemble perturbations depend continuously on $\tilde{\mathbf{P}}_a$ so as to be consistent with the background ensemble perturbations (Ott et al. 2004). Using (15) and (16), (12) can be rewritten as

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_f + \mathbf{X}_f \tilde{\mathbf{P}}_a (\mathbf{H}\mathbf{X}_f)^T \mathbf{R}^{-1} [\mathbf{y}_o - H(\bar{\mathbf{x}}_f)]. \quad (17)$$

Equations (16) and (17) provide the analysis ensemble perturbations and its mean. [See Hunt et al. (2007) for more details and discussion on LETKF.]

Three procedures are used in this study to improve the performance of the LETKF. First, a multiplicative variance inflation is applied to the background ensemble perturbations. The inflation magnitude is vertically dependent (Table 2) to reflect error characteristic (see Fig. 4; section 4b). This reduces the RMS analysis error by about 4% for the potential vorticity at interior levels and 6.5% for the potential temperature at the bottom level compared to a vertically constant inflation variance. Second, additive variance inflation (Corazza et al. 2007) is used to improve the performance of the system by “refreshing” the ensemble vectors. This prevents the ensemble from collapsing into a space that is too small and avoids the rank problems characteristic of evolving perturbations converging to the leading Lyapunov vector (Wang and Bishop 2003; Etherton and Bishop 2004). We note that the procedure of localization in LETKF to use observations locally would also address some of the rank problems (Hamill et al. 2001). In this study, the refreshing is computed by generating random

perturbations with a size of about 2% of the field variability at the observation locations, and then converting them back to model coordinates by applying the transpose of the observation operator.

Third, an observation localization method that multiplies the observation error covariance by the inverse of a Gaussian localization operator is used to weight the observations located farther away from the center of the volume (Miyoshi 2005). This has a significant impact when the observations are sparse and large local patches are required. All seven vertical levels are included in the local volume without vertical localization ($2l_z + 1 = 7$). The size of the local patch needs to be optimized depending on the ensemble size and observational density. Enlarging local patches improves performance, but this effect saturates beyond a certain size [see discussion and Table 4 in Corazza et al. (2007)]. The optimal local (horizontal) domain for 64 observations and 40 ensemble members is 19×19 ($l = 9$), which allows the assimilation of about three to four observations per local patch. The experimental configuration of LETKF follows closely the LEKF configuration in Corazza et al. (2007). Table 2 summarizes the experimental configuration of the LETKF used in this study.

4. Results

a. The flow-dependent background error covariance in LETKF

In this section, we focus on whether the explicit computation of the flow-dependent background–analysis error covariance by LETKF represents a significant advantage over 3DVAR. Figure 1 compares the RMS analysis errors to show that the analysis derived from LETKF (blue line) outperforms the 3DVAR analysis (black line).

To explore this question, the ensemble mean background state from the LETKF system was first provided as a background to 3DVAR. The 3DVAR analysis from the modified background states is much more accurate (red line in Fig. 1) than the original 3DVAR analysis and only slightly worse than the LETKF analysis. This result might be interpreted as evidence that the effect of ensemble averaging in LETKF is more important than the information on so-called errors of the day. However, this interpretation would be incorrect because the latter allows LETKF to minimize the errors of the day and to provide a better first guess than 3DVAR. When 3DVAR is continued without further information from the LETKF background, the errors of the day that were suppressed in the ensemble mean immediately grow and the 3DVAR analysis error increases over a few days to its normal value (the green line in Fig. 1). By

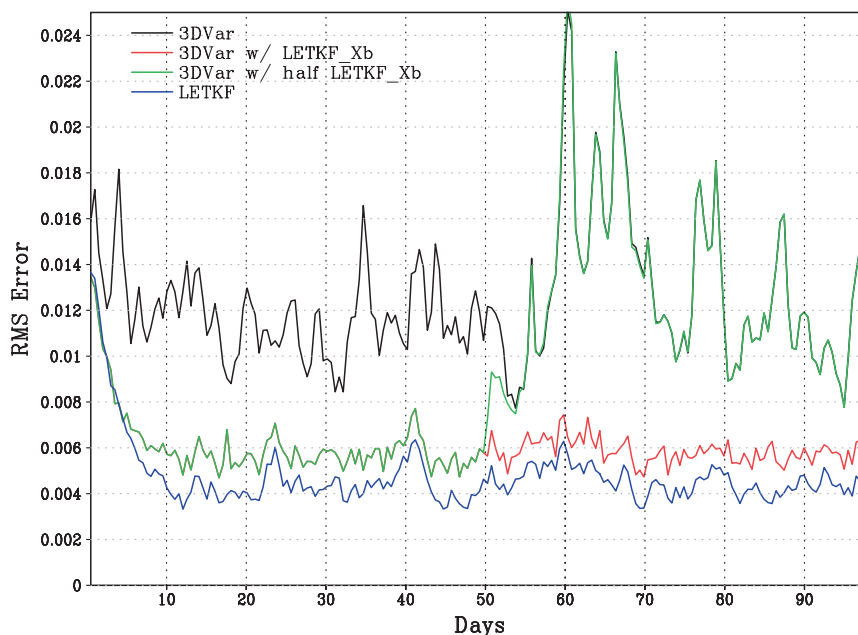


FIG. 1. Time series of RMS analysis errors in potential temperature at the bottom level from 3DVAR (black line), 3DVAR with the background replaced with the LETKF background (ensemble mean) state (red line), 3DVAR with the background replaced with the LETKF background state only during the first 50 days (green line), and LETKF (blue line).

projecting the corrections onto the local dynamical instabilities estimated by the ensemble, the LETKF background error covariance is able to properly correct the background state (ensemble mean) with the available observations. At the same time, the accuracy of the mean state also determines the effectiveness of ensemble perturbations (see more discussion in section 4b). This accumulated information makes LETKF perform better than 3DVAR.

In a complementary experiment, the LETKF background ensemble perturbations are replaced with Gaussian random perturbations drawn from the 3DVAR background error covariance \mathbf{B}_{3DVAR} , that is, generating K ensemble perturbations $\mathbf{B}_{3DVAR}^{1/2} \eta_k$, where η_k are Gaussian random perturbations with unit variance and model dimension. Figure 2 shows that the quality of local LETKF analysis with isotropic ensemble perturbations from \mathbf{B}_{3DVAR} not representing the errors of the day quickly degrades to a level even worse than 3DVAR (red line). From Figs. 1 and 2, we conclude that the improvement from LETKF is the combined effect of the ensemble averaging to filter out the unpredictable uncertainties, and of the information that the ensemble brings on the errors of the day.

b. Comparisons of 3DVAR, 4DVAR, and LETKF

Table 3 shows the temporally and spatially averaged RMS analysis errors of 3DVAR, 4DV12H, 4DV24H,

and LETKF with different-sized local domains. LETKF, with a small local patch of 7×7 grid points ($l = 3$) and only 20 ensemble members, already outperforms the 3DVAR. The LETKF analysis computed from a local patch of 11×11 grid points ($l = 5$) and 20 ensemble members has comparable accuracy to 4DV12H and is more accurate with a 40-member ensemble. Results for local patches larger than 11×11 are slightly worse than the results for 4DV24H. The remaining results are based on LETKF, with a local patch of 19×19 ($l = 9$) and 40 ensemble members.

Table 3 also lists the computation time required to perform one analysis cycle with 3DVAR, 4DV12H, 4DV24H, and LETKF using the same (serial) computer system described in section 3b. The measurement of the computing time does not include the ensemble forecasts because in most operational centers the ensemble forecasts are required in either type of the assimilation scheme. On a serial computer, the LETKF computational cost increases with the size of the local patch, without including the efficiency gains due to the intrinsically parallel characteristic of LETKF (Hunt et al. 2007) compared to 4DVAR.

Figure 3 shows the RMS analysis error for the time series of the potential temperature for the bottom level from the experiments listed in Table 3. Note that both 4DVAR and LETKF successfully avoid the large error spikes that occur in 3DVAR (e.g., day 60 in Fig. 3). The

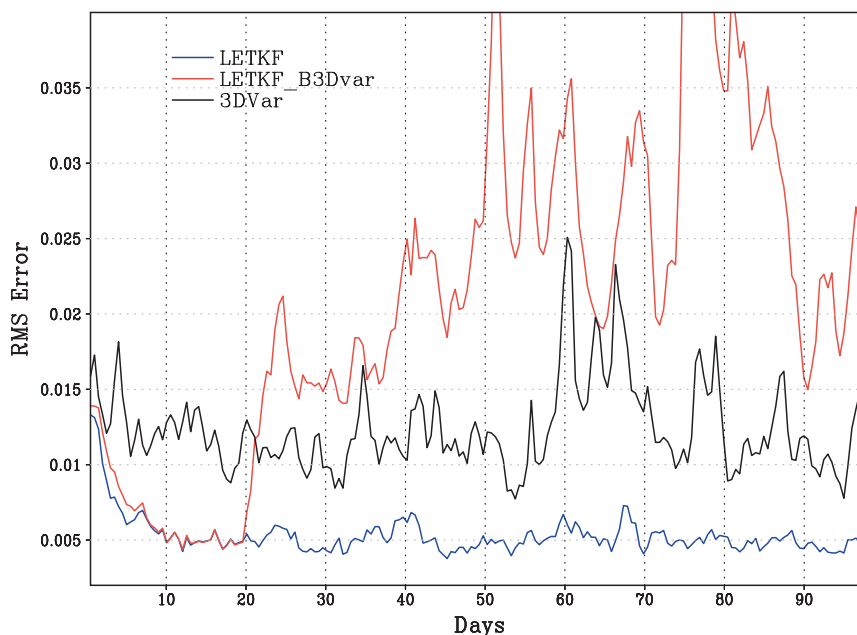


FIG. 2. Time series of RMS analysis errors in potential temperature at the bottom level from LETKF (blue line), LETKF with the background ensemble perturbations replaced with Gaussian random perturbations drawn from $\mathbf{B}_{3\text{dvar}}$ (red line), and 3DVAR (black line).

accuracy of the LETKF analysis is between 4DV12H and 4DV24H. However, 4DV24H shows more variability in the analysis error between successive 12-h analyses, reflecting the fact that they are computed independently.

The spinup time for LETKF depends on the accuracy of the initial background state in the first analysis cycle, while this has little influence in 4DVAR. The spinup time is much longer if LETKF is initialized with the climatology mean state (not shown). Caya et al. (2005) also showed that for a realistic storm-scale forecast using radar data, the advantage of EnKF over 4DVAR only became apparent after several assimilation cycles. As discussed in section 4a, the LETKF corrections require that the ensemble perturbations be representative of the relevant background growing errors and capable of depicting the structures of the flow-dependent

background dynamical instabilities. The long spinup required for LETKF can be avoided if the 3DVAR analysis is used as the initial condition, because it is sufficiently close to the true state. Then the spinup times for LETKF and 4DV24H (10 days) are comparable and faster than for 4DV12H (15 days). Alternatively, it is possible to use an algorithm based on the LETKF no-cost smoother (appendix A) to accelerate the spinup (KY08).

The flow-dependent structures carried in both 4DVAR and LETKF substantially reduce the large vertical dependence in 3DVAR analysis errors (Fig. 4), and the LETKF analysis accuracy is in between 4DV12H and 4DV24H for all levels. Such improvement is most evident in the variables with large error growth rates (e.g., the potential temperature at the bottom and the top levels). Snyder et al. (2003) demonstrated that

TABLE 3. Mean RMS analysis errors of generalized potential vorticity for the 3DVAR, 4DVAR, and LETKF schemes averaged over 80 days (160 analysis cycles) and the corresponding computational times required for one analysis cycle in a single processor. For 4DVAR, the average number of iterations required for the minimization process is included.

	3DVAR	4DVAR		LETKF				
		12H	24H*	$K = 20$	$K = 40$			
				$l = 5$	$l = 3$	$l = 5$	$l = 7$	$l = 9$
RMSE ($\times 10^2$)	6.24	2.63	1.75	2.54	3.41	2.39	2.14	2.12
Time (min)	0.007	0.67 (14 iterations)	1.4 (19 iterations)	0.12	0.21	0.36	0.51	0.67

* The value is slightly different from the 24-h 4DVAR in Table 1 because of the use of an overlapped window to derive analysis every 12 h.

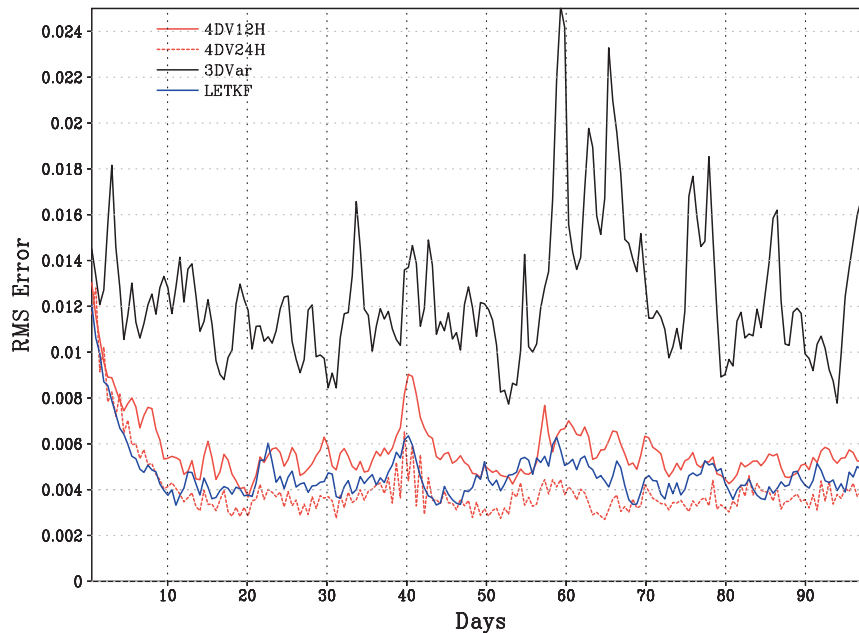


FIG. 3. Time series of RMS analysis errors in potential temperature at the bottom level from the 3DVAR and 4DVAR systems (12- and 24-h window time) and the LETKF schemes.

there is a strong relationship of the perturbations to the gradient of the reference flow in this QG model, and with the large gradient of the reference flow concentrating at the bottom and top levels, this dominates the growth of the perturbations. In our study, the accuracy in the midlevel is also greatly improved with respect to 3DVAR. This can be attributed to the fact that Morss (1999) used a small vertical correlation in the 3DVAR errors, whereas the advantage of being flow dependent allows 4DVAR and LETKF to have vertical correlations that better reflect local instabilities.

Figure 5 shows the analysis and forecast RMSE for the potential temperature at the bottom level as a function of forecast lead time up to 5 days for all the DA schemes. With the setup of the perfect model, the forecast errors are dominated by the dynamically growing errors. The forecast skill with LETKF is again between that of 4DV12H and 4DV24H, and much better than the 3DVAR forecast error throughout the integration. We note that during the first 12 h, the 3DVAR forecast error grows more slowly than in the other schemes, indicating that the 3DVAR initial analysis error includes not only growing but also nongrowing errors.

To further explain the results, we compare the corrections and errors from 3DVAR, LETKF, and 4DV12H because they have the same length of analysis cycle using the same observations. Figure 6 presents 12-h forecast (background) (left) errors and (right)

analysis errors in colors, superimposed with the analysis corrections (contours) at an arbitrarily chosen time for the three analysis schemes. Figure 6a shows that the slower 12-h forecast error growth of 3DVAR can be attributed to the fact that the analysis corrections derived from 3DVAR have isotropic shapes and do not project well on the structures of the dynamically stretched errors of the day that dominate the 12-h forecast errors in LETKF (Fig. 6c) and 4DVAR (Fig. 6e). The importance of these growing errors of the day, similar to bred vectors or locally leading Lyapunov vectors (section 4c), has been recognized in several previous studies (e.g., Corazza et al. 2002; Snyder et al. 2003).

Figure 6b shows that in 3DVAR the background errors are partially corrected in the analysis but the isotropic corrections also introduce errors with shapes that grow less than the dynamically stretched growing errors. As a result, the 3DVAR analysis error contains large amounts of uncorrected growing and nongrowing errors, and therefore the initial growth of errors in 3DVAR is slower, as observed in Fig. 5. After 12 h, the stretched, evolving, growing errors dominate the forecast errors (not shown), and the 3DVAR error growth becomes exponential and similar to the other schemes. In contrast, the analysis corrections from the LETKF and the 4DVAR are flow dependent with shapes closely related to the dynamically growing forecast errors (Figs. 6c,e) and able to remove more (but not all) of the

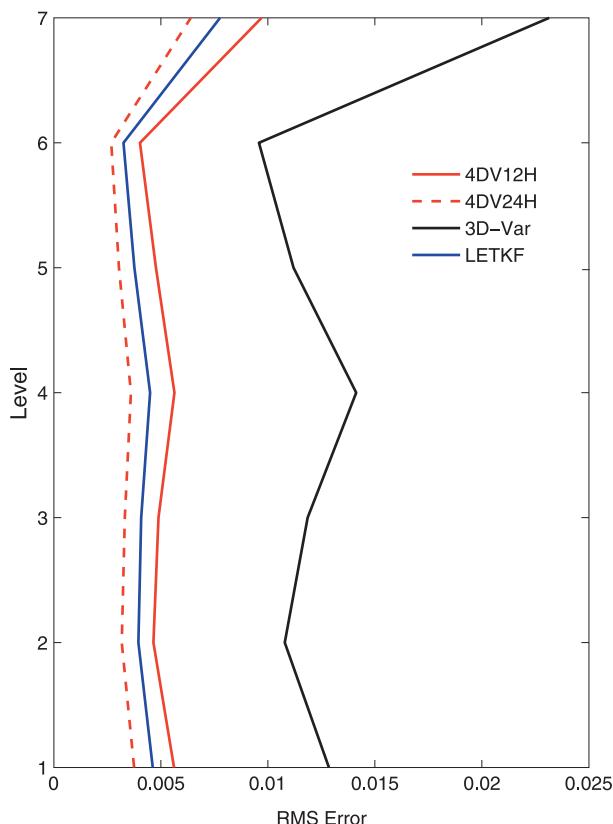


FIG. 4. Mean RMS analysis error in vertical levels from all assimilation schemes.

background errors than 3DVAR. The errors left in the analysis still have shapes related to uncorrected dynamically growing directions (Figs. 6d,f), which dominate the error growth (section 4c). Therefore, the forecast errors from LETKF and 4DVAR show a more consistent exponential error growth (Fig. 5), related to the dynamical instabilities.

To quantify the relationship between analysis corrections and errors at the analysis time shown in Fig. 6, we computed the temporally and spatially averaged local-explained variance of the analysis error for the background error, the correction (analysis increment) for the background error, and the correction for the analysis error from LETKF, 4DV12H, and 3DVAR (Table 4). The local-explained variance is computed as the inner product between two local vectors divided by the square of the norm of the projected vector. This essentially measures how much the shapes of the fields shown in Fig. 6 “locally” agree.

The first column of Table 4 shows that after one analysis cycle, the analysis errors have shapes that still agree substantially with the background errors, from 71.5% for the LETKF to 80% in the 3DVAR (a result

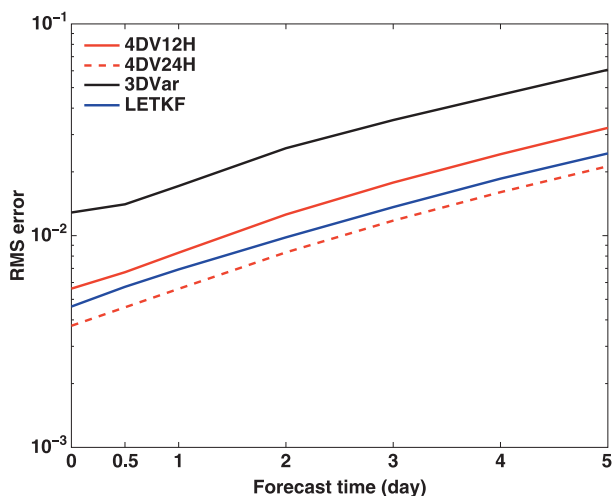


FIG. 5. Analysis and forecast errors of the potential temperature at the bottom level as a function of forecast length (days) from different assimilation schemes.

that is very apparent comparing the colored fields on the left- and the right-hand sides in Fig. 6). The second column indicates that on average, the analysis corrections capture more than 30% of the background error variance in the LETKF, a larger percentage than 4DV12H or 3DVAR. This confirms that the corrections from LETKF are able to remove more background errors (e.g., from Figs. 6c,d, the LETKF corrections effectively remove or reduce the large background errors).

That the analysis errors also project on the analysis increments (third column in Table 4) suggests that although the shapes of the errors are well captured, the local amplitudes are not always optimal due to the lack of observations. Because the analysis errors contain the incompletely removed background errors, the corrections from LETKF also agree more closely with the analysis errors than 4DV12H or 3DVAR. We note that this is also valid for 4DV24H, which gives the most accurate analysis. With a longer 4DVAR assimilation window, the analysis corrections from 4DV24H project more on both the background and the analysis errors (60% and 53%, respectively). Therefore, to obtain accurate analyses, the analysis corrections from assimilation schemes should contain the structures that dominate the growth of the background errors at the analysis time.

c. Initial and final analysis corrections in LETKF and 4DV12H

Swanson and Vautard (1998) showed that the structures of 4DVAR analysis errors strongly project on the subspace of the global leading Lyapunov vectors (the unstable manifold of the system). The strength of these

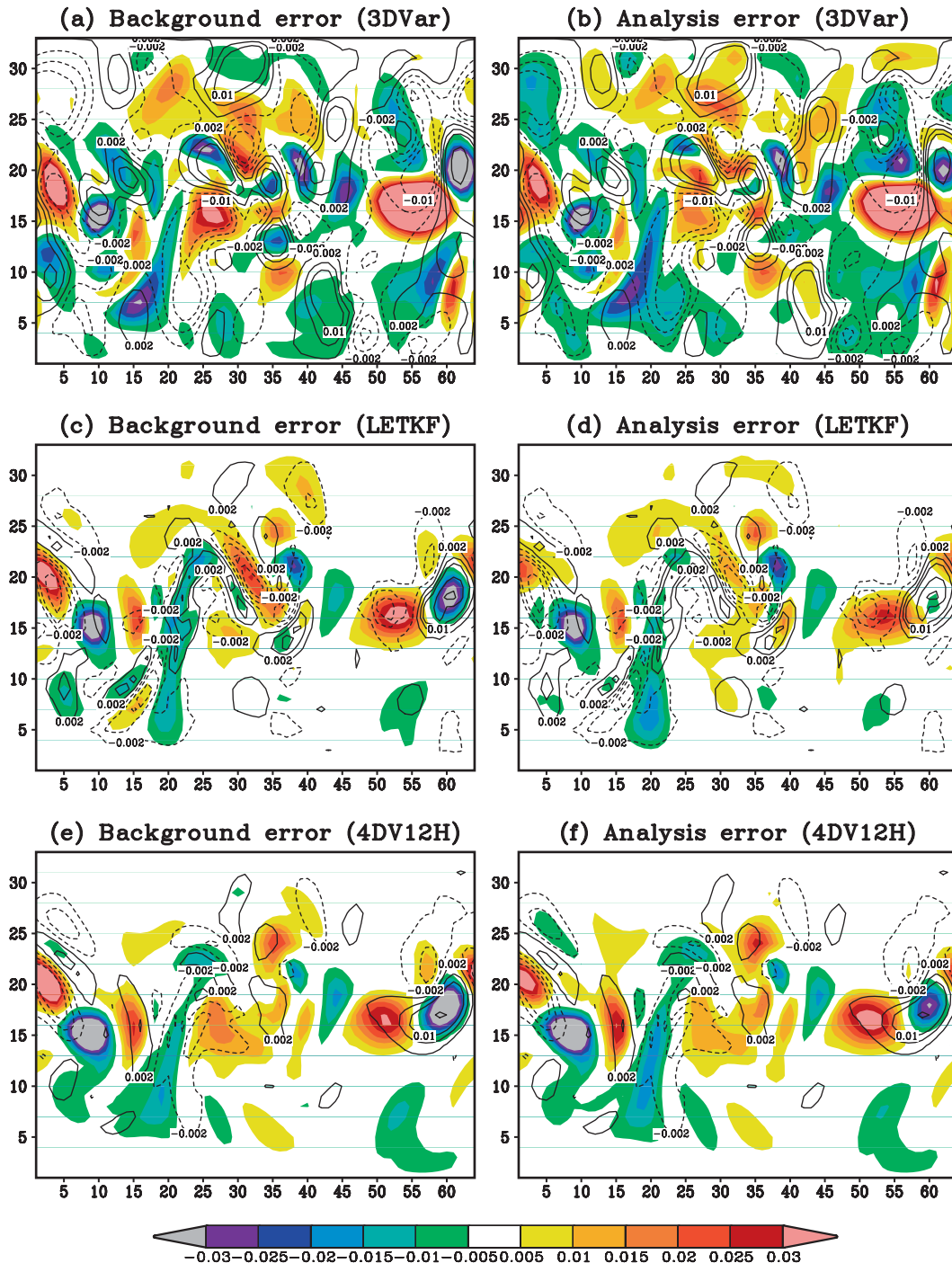


FIG. 6. (a) Background error (color shades) and analysis corrections (contours) of the potential temperature at the bottom level from 3DVAR at day 41 1200 UTC; (b) as in (a), except that the color shades are the 3DVAR analysis errors; (c),(d) Same variables as in (a),(b), but from LETKF; (e),(f) Same variables as in (a),(b), but from the 12-h 4DVAR. The intervals for the contours are -0.03 , -0.02 , -0.01 , -0.006 , -0.002 , 0.002 , 0.006 , 0.01 , 0.02 , and 0.03 .

projections increases as the assimilation window lengthens, especially for small scales. Considering the local properties of the low-dimensional dynamical errors (Patil et al. 2001) and the results from Table 4, we ex-

amine the local structures of the assimilation corrections in the unstable space for both schemes. We use bred vectors (BVs; a finite amplitude equivalent to the leading Lyapunov vectors; Toth and Kalnay 1993, 1997)

TABLE 4. Local explained variance (%) of analysis error for background error, analysis increments for background errors, and analysis increments for analysis errors. The local explained variance is computed in local domains of 11×11 grid points and the averaging is both temporal and spatial.

	Analysis error vs background error	Analysis increments vs background errors	Analysis increment vs analysis error
LETKF	71.5	31.3	24.5
4DV12H	77.4	24.9	20.1
3DVAR	80.3	21.7	16.8

to represent the fast-growing dynamical instabilities of the evolving flow and to compare with the 4DVAR and LETKF analysis increments both at the beginning and at the end of an assimilation window. The BV shown in Fig. 7 is randomly chosen from an ensemble of 20 BVs (appendix B).

To obtain the initial and final increments for a given assimilation window, we compute the LETKF initial increment by using a no-cost smoother (Kalnay et al. 2007b). Because the LETKF selects the ensemble trajectory that best fits the data throughout the assimilation window, we use the same weight obtained at the end of the assimilation window for the analysis increment [Eq. (17)] to obtain the “smoothed” analysis increments at the beginning of the window (see appendix A for more details). This smoothed analysis increment is equivalent to the 4DVAR analysis increment at the beginning of the window, so that, as in 4DVAR, evolving the smoothed LETKF analysis forward to the end of the window approximates the same analysis computed by LETKF.

Figures 7a,b show the initial and final analysis increments (color shading) from the LETKF superimposed on the corresponding BVs (contours). Similarly, the initial and final 4DV12H analysis increments (color shading) are superimposed to the corresponding BVs (contours) in Figs. 7c,d. Both the smoothed initial increment and final analysis increments from the LETKF have local structures related to the BVs’ structures at the corresponding times. This indicates that the shapes of the corrections and their evolutions are strongly influenced by the local dynamical instabilities. The final analysis increments from 4DV12H (valid at the end of the window) also show a similarity with the corresponding BV (Fig. 7d), in agreement with Swanson and Vautard (1998). However, the initial time analysis increments from 4DV12H have larger scales and are more isotropic than the BVs due to their stronger dependence on the initial 3DVAR-like isotropic background error covariance. Thus, the initial 4DVAR larger-scale analysis corrections quickly stretch into the dynamical un-

stable structures given by the BVs. As a result, at the end of the assimilation window, the 4DVAR analysis makes useful corrections to the background errors (Fig. 7d) that are similar to BVs.

5. Summary and discussion

In this study, data assimilation schemes related to variational and ensemble methods were implemented in a quasigeostrophic model. Three different schemes were compared: 3DVAR, 4DVAR, and LETKF. Experiments were conducted to compare individual performance and to understand their differences. This information should be useful to operational centers that face the choice of continuing with 3DVAR or 4DVAR, or testing ensemble Kalman filters as the next-phase data assimilation system. Under perfect model conditions, the results focus on the error structure in different data assimilation schemes, using 3DVAR as the benchmark. We compared the performance of LETKF with 4DVAR using the same observations with a 12-h assimilation window or using more time-dependent observations with a longer window of 24 h.

This study confirms that 4DVAR and LETKF are superior to 3DVAR because they are able to lower the overall errors and eliminate large analysis errors spikes seen in 3DVAR. The corrections made by 4DVAR and LETKF agree well with the corresponding local instabilities at the analysis time as depicted by bred vectors (a finite time equivalent of leading Lyapunov vectors). The results also confirm the Pires et al. (1996) result that using a longer assimilation window with 4DVAR improves the analysis, although the improvement becomes small for windows longer than about 2 days. With real observations, longer windows also require accounting for model errors and nonlinearity of the perturbations.

Our results indicate that LETKF, requiring low computer resources (20 ensemble members and small local domains of 11×11 grid points), gives results comparable to 4DVAR with a 12-h assimilation window. With 40 ensemble members, LETKF performs significantly better than the 12-h 4DVAR. With a larger local domain of 19×19 , LETKF provides a result comparable to 4DVAR with a 24-h assimilation window. When the initial condition for the first analysis cycle is close enough to the true state (such as when it is started from a 3DVAR analysis), the LETKF has a similar spinup time as the 4DVAR with a 24-h window and is faster than the 12-h 4DVAR.

The superior performance of LETKF compared to 3DVAR is due to the combined effect of having flow-dependent perturbation structures related to errors of the day and to ensemble averaging. By strongly projecting the

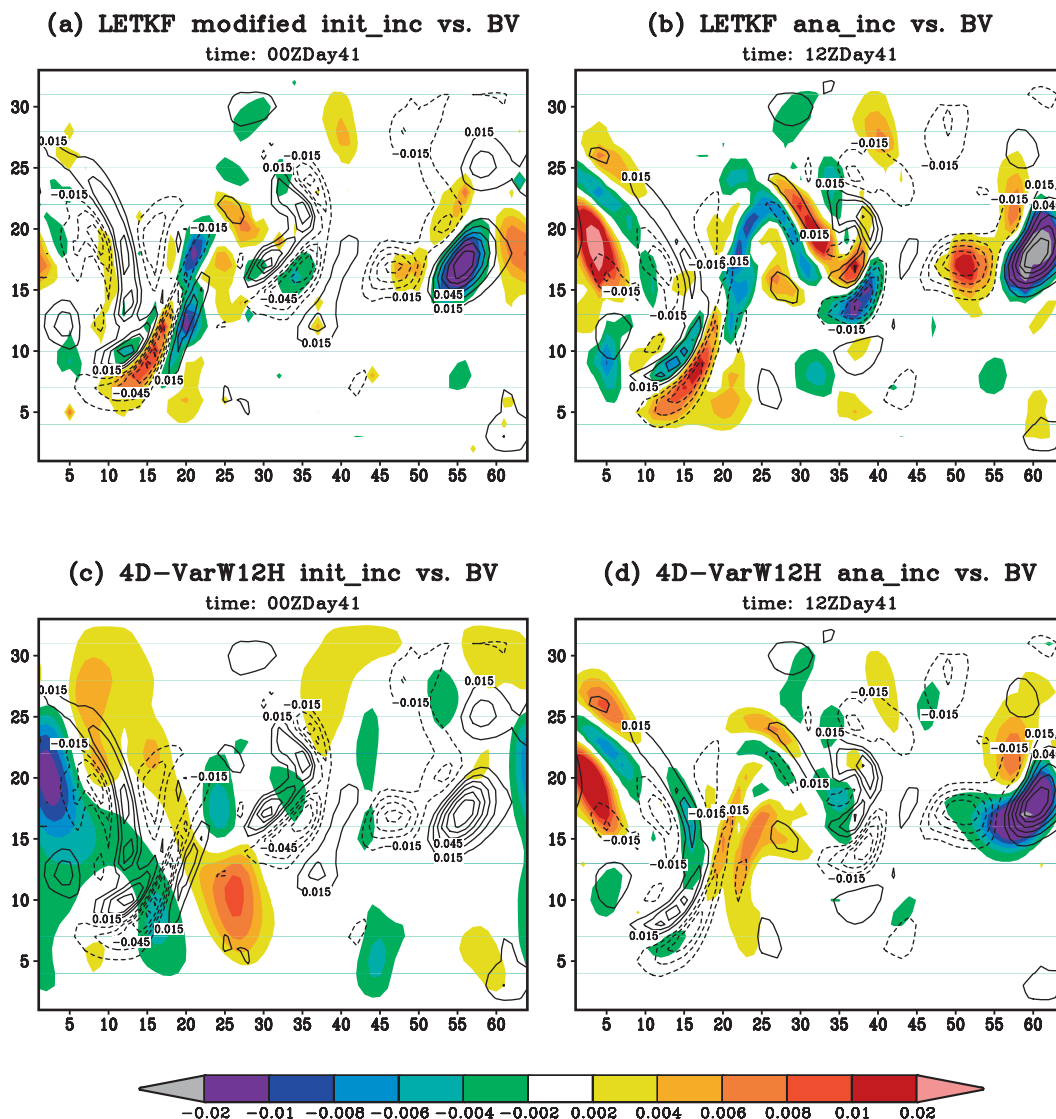


FIG. 7. (a) Modified (smoothed, see appendix A) initial increments of LETKF (color shades) and BV (contours) of the potential temperature at the bottom level at day 41 0000 UTC; (b) as in (a), but for the LETKF analysis increments and BVs at day 41 1200 UTC; (c) initial increments of the 12-h 4DVAR and BV (contours) at day 41 0000 UTC; (d) analysis increment of the 12-h 4DVAR and BV at day 41 1200 UTC. The contour interval is 0.015.

corrections on the local dynamical instabilities, LETKF is better able to correct the background state (ensemble mean) with the available observations.

The structures of the analysis increments (corrections) from these DA schemes were examined to understand the performance differences between the LETKF and 4DVAR with a 12-h assimilation window. Both the analysis increments and smoothed initial increments from the LETKF, obtained using a no-cost smoother, have local structures characterized by the errors of the day. Those structures strongly match the BVs valid at their corresponding times. The initial increments from 4DVAR are larger scale and show weak projections on

the dominant initial BVs, but the 4DVAR analysis increments at the end of the window exhibit strong similarities to the corresponding BVs. This suggests that the corrections needed in the background state (at the analysis time) are strongly related to the structures of BVs (and the final SVs, not shown), that is, to the fast-growing errors corresponding to the analysis time.

This study considered the performance of 4DVAR and LETKF under the “perfect model” assumption, so that in the presence of model errors the performance could be significantly worse. An important related area of research is that of hybrid schemes that combine the ensemble-based and variational assimilation techniques.

Initial studies suggest that this combination can efficiently overcome the limitations of time-independent background error covariance in the 3DVAR and 4DVAR schemes (Hamill and Snyder 2000; Corazza et al. 2002; Etherton and Bishop 2004; Wang et al. 2007).

Acknowledgments. We express our deep gratitude to Rebecca Morss for providing the QG model and the 3DVAR system, to Debra Baker for substantially improving a draft of this article, and to Joaquim Ballabrera for helping to correct the adjoint model; C. Snyder and an anonymous reviewer gave valuable suggestions that also improved the manuscript. We are very grateful to Profs. Istvan Szunyogh, Ed Ott, and Brian Hunt and the other members of the Chaos and Weather Group at the University of Maryland, as well as Jeff Whitaker from NOAA/OAR/PSD, for helpful interactions. Author S.-C. Yang was supported by NASA Grants NNG004GK78A and NNG06GB77G; A. Carrassi was supported by the Belgian Federal Science Policy Program under Contract MO/34/017.

APPENDIX A

No-Cost LETKF Smoother

From Eq. (17), the mean analysis computed from the LETKF at time i is

$$\bar{\mathbf{x}}_a^i = \bar{\mathbf{x}}_f^i + \mathbf{X}_f^i \mathbf{W}^i \quad (\text{A1})$$

where \mathbf{W}^i is a matrix of weights [see (A2) following] and \mathbf{X}_f^i is the matrix whose columns are the ensemble background perturbations at time i :

$$\mathbf{W}^i = \tilde{\mathbf{P}}_a^i (\mathbf{H} \mathbf{X}_f^i)^T \mathbf{R}^{-1} [\mathbf{y}_o^i - H(\bar{\mathbf{x}}_f^i)]. \quad (\text{A2})$$

Because the weight matrix provides the linear combination of background perturbations that best fit the observations within the assimilation window, we can improve the analysis at the beginning of the time window (at time $i - 1$) by applying a smoother with the same weights, \mathbf{W}^i , to the ensemble perturbations at time $i - 1$:

$$\tilde{\mathbf{x}}_a^{i-1} = \bar{\mathbf{x}}_a^{i-1} + \mathbf{X}_a^{i-1} \mathbf{W}^i, \quad (\text{A3})$$

where $\bar{\mathbf{x}}_a^{i-1}$ and \mathbf{X}_a^{i-1} are the mean analysis and the ensemble analysis perturbations at time $i - 1$, and $\tilde{\mathbf{x}}_a^{i-1}$ is the modified analysis at the same time. Although this smoothed analysis $\tilde{\mathbf{x}}_a^{i-1}$ at the beginning of the window does improve the original analysis $\bar{\mathbf{x}}_a^{i-1}$, because it uses the “future” observations available at time i , it approximates the analysis at time i , as indicated in (A4). In addition to the effect of the nonlinear model, the ap-

proximation in (A4) considers that the weight, \mathbf{W}^i , varies with locations. Such no-cost smoothing, proposed by Kalnay et al. (2007b), would be particularly useful in the context of reanalysis and can be used for accelerating the LETKF spinup by “running in place” (KY08). Note that in (A4), M is the nonlinear model advancing the state from time $i - 1$ to i and \mathbf{L} is the tangent linear model defined in section 3b:

$$\begin{aligned} M(\tilde{\mathbf{x}}_a^{i-1}) &= M(\bar{\mathbf{x}}_a^{i-1} + \mathbf{X}_a^{i-1} \mathbf{W}^i) \approx M(\bar{\mathbf{x}}_a^{i-1}) \\ &+ \mathbf{L} \mathbf{X}_a^{i-1} \mathbf{W}^i = \bar{\mathbf{x}}_f^i + \mathbf{X}_f^i \mathbf{W}^i = \bar{\mathbf{x}}_a^i. \end{aligned} \quad (\text{A4})$$

APPENDIX B

The Dynamically Fast-Growing Perturbations

Bred vectors, defined as the differences between perturbed and nonperturbed nonlinear runs, represent the fast-growing dynamical instabilities of the evolving flow and naturally carry information on errors of the day (Toth and Kalnay 1993, 1997). They are the local Lyapunov vectors bred through nonlinear integration. In this study, 20 breeding cycles, initialized from different random perturbations, are bred upon the 3DVAR analyses with a 12-h rescaling interval. The bred perturbations are rescaled uniformly according to the mean squared 3DVAR analysis error at the midlevel. Also, a small amount of Gaussian random perturbations are added to BVs at every breeding cycle to “refresh” BVs and avoid the tendency of BVs to converge to a too small dimensional subspace (Wang and Bishop 2003). The process of refreshing plays an important role in accelerating the BV convergence to the subspace of growing instability during the transition period and also increases the BVs’ ability to capture the subspace of the evolving background errors. Details of the characteristics of the BVs from this QG model and their relationship to the errors of the day are discussed in Corazza et al. (2002).

REFERENCES

- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.
- Broyden, C. G., 1969: A new double-rank minimization algorithm. *Not. Amer. Math. Soc.*, **16**, 670.
- Carrassi, A., A. Trevisan, and F. Uboldi, 2007: Adaptive observations and assimilation in the unstable subspace by breeding on the data-assimilation system. *Tellus*, **59A**, 101–113.
- Caya, A., J. Sun, and C. Snyder, 2005: A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.*, **133**, 3081–3094.

- Courtier, P., J. N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-VAR, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387.
- Corazza, M., E. Kalnay, D. J. Patil, E. Ott, J. Yorke, I. Szunyogh, and M. Cai, 2002: Use of the breeding technique in the estimation of the background error covariance matrix for a quasigeostrophic model. Preprints, *Symp. on Observations, Data Assimilation, and Probabilistic Prediction*, Orlando, FL, Amer. Meteor. Soc., 154–157.
- , and Coauthors, 2003: Use of the breeding technique to estimate the structure of the analysis “error of the day.” *Nonlinear Proc. Geophys.*, **10**, 233–243.
- , E. Kalnay, and S.-C. Yang, 2007: An implementation of the local ensemble Kalman filter for a simple quasi-geostrophic model: Results and comparison with a 3D-Var data assimilation system. *Nonlinear Proc. Geophys.*, **14**, 89–101.
- Dey, C., and L. L. Morone, 1985: Evolution of the national meteorological center global data assimilation system: January 1982–December 1983. *Mon. Wea. Rev.*, **113**, 304–318.
- Etherton, B. J., and C. H. Bishop, 2004: Resilience of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error. *Mon. Wea. Rev.*, **132**, 1065–1080.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162.
- , 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367.
- Fertig, E., J. Harlim, and B. R. Hunt, 2007: A comparative study of 4D-VAR and a 4D ensemble Kalman filter: Perfect model simulations with Lorenz-96. *Tellus*, **59A**, 96–101.
- Fletcher, R., 1970: A new approach to variable metric methods. *Comput. J.*, **13**, 317–322.
- Giering, R., and T. Kaminski, 1998: Recipes for adjoint code construction. *ACM Trans. Math. Software*, **24**, 437–474.
- Goldfarb, D., 1970: A family of variable-metric methods derived by variational means. *Math. Comp.*, **24**, 23–26.
- Hamill, T. M., 2006: Ensemble-based data assimilation. *Predictability of Weather and Climate*, T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 124–156.
- , and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919.
- , J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- Hunt, B. R., E. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35–45.
- , and R. S. Bucy, 1961: New results in linear filtering and prediction Theory. *J. Basic Eng.*, **83**, 95–107.
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 340 pp.
- , H. Li, T. Miyoshi, S.-C. Yang, and J. Ballabrera-Poy, 2007a: 4D-Var or ensemble Kalman filter? *Tellus*, **59A**, 758–773.
- , —, —, and —, 2007b: Response to the discussion on “4-D-Var or EnKF?” by Nils Gustafsson. *Tellus*, **59A**, 778–780.
- Kim, H. M., M. C. Morgan, and R. E. Morss, 2004: Evolution of analysis error and adjoint-based sensitivities: Implications for adaptive observations. *J. Atmos. Sci.*, **61**, 795–812.
- Le Dimet, F. X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97–110.
- Liu, D. C., and J. Nocedal, 1989: On the limited memory BFGS method for large scale optimization. *Math. Program.*, **45**, 503–528.
- Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194.
- , 2003: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203.
- Miyoshi, T., 2005: Ensemble Kalman filter experiments with a primitive-equation global model. Ph.D. dissertation, University of Maryland, College Park, 197 pp. [Available online at <http://hdl.handle.net/1903/3046>.]
- , and S. Yamane, 2007: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. *Mon. Wea. Rev.*, **135**, 3841–3861.
- Morss, R. E., 1999: Adaptive observations: Idealized sampling strategies for improving numerical weather prediction. Ph.D. thesis, Massachusetts Institute of Technology, 225 pp.
- , K. A. Emanuel, and C. Snyder, 2001: Idealized adaptive observation strategies for improving numerical weather prediction. *J. Atmos. Sci.*, **58**, 210–232.
- Navon, I. M., X. Zou, J. Derber, and J. Sela, 1992: Variational data assimilation with an adiabatic version of the NMC spectral model. *Mon. Wea. Rev.*, **120**, 1433–1446.
- Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428.
- Parrish, D., and J. Derber, 1992: The National Meteorology Center’s spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- Patil, D., B. R. Hunt, E. Kalnay, J. A. Yorke, and E. Ott, 2001: Local low dimensionality at atmospheric dynamics. *Phys. Rev. Lett.*, **86**, 5878–5881.
- Pires, C., R. Vautard, and O. Talagrand, 1996: On extending the limits of variational assimilation in chaotic systems. *Tellus*, **48A**, 96–121.
- Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143–1170.
- Rotunno, R., and J. W. Bao, 1996: A case study of cyclogenesis using a model hierarchy. *Mon. Wea. Rev.*, **124**, 1051–1066.
- Shanno, D. F., 1970: Conditioning of quasi-Newton methods for function minimization. *Math. Comp.*, **24**, 647–657.
- Simmons, A. J., R. Mureau, and T. Petroligis, 1995: Error growth estimates of predictability from the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.*, **121**, 1739–1771.
- Snyder, C., T. M. Hamill, and S. B. Trier, 2003: Linear evolution of error covariances in a quasigeostrophic model. *Mon. Wea. Rev.*, **131**, 189–205.
- Swanson, K., and R. Vautard, 1998: Four-dimensional variational assimilation and predictability in a quasi-geostrophic model. *Tellus*, **50A**, 369–390.
- Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, E. Kalnay, E. Ott, and J. A. Yorke, 2005: Assessing a local ensemble Kalman filter: Perfect model experiments with the National Centers for Environmental Prediction global model. *Tellus*, **57A**, 528–545.
- Talagrand, O., 1997: Assimilation of observations, an introduction. *J. Meteor. Soc. Japan*, **75**, 191–209.

- , and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteor. Soc.*, **113**, 1311–1328.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- , and —, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158.
- , C. Snyder, and T. M. Hamill, 2007: On the theoretical equivalence of differently proposed ensemble/3DVAR hybrid analysis schemes. *Mon. Wea. Rev.*, **135**, 222–227.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- , —, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482.
- Yang, S.-C., 2005: Appendix B: Errors of the day, bred vectors and singular vectors in a QG atmospheric model: Implications for ensemble forecasting and data assimilation. Bred vectors in the NASA NSIPP Global Coupled Model and their application to coupled ensemble predictions and data assimilation. Ph.D. thesis, University of Maryland, College Park, 174 pp. [Available online at <http://hdl.handle.net/1903/2477>.]