Four-dimensional local ensemble transform Kalman filter: numerical experiments with a global circulation model

By JOHN HARLIM* and BRIAN R. HUNT, *Department of Mathematics and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA*

ABSTRACT

We present a four-dimensional ensemble Kalman filter (4D-LETKF) that approximately and efficiently solves a variational problem similar to that solved by 4D-VAR, and report numerical results with the Simplified-Parametrized primitive Equation Dynamics model, a simplified global atmospheric model. We discuss the relationship of 4D-LETKF to other ensemble Kalman filters and, in our simulations, compare it with two simpler approaches to assimilating asynchronous observations.

We find that 4D-LETKF significantly improves on the approach of treating asynchronous observations as if they occur at the analysis time. For a sufficiently short analysis time interval, the approach of computing innovations from the background state at the observation times and treating those innovations as if they occur at the analysis time is comparable to 4D-LETKF, but for longer analysis intervals, we find that 4D-LETKF is superior to this approach.

1. Introduction

This paper presents an efficient method of implementing a four-dimensional ensemble Kalman filter, which we call the Four-Dimensional Local Ensemble Transform Kalman Filter (4D-LETKF) for assimilating asynchronous observations. Through a change of coordinates, three-dimensional LETKF (Hunt et al., 2007) is equivalent to the Local Ensemble Kalman Filter (LEKF) of Ott et al. (2004), and locally its analysis is equivalent to the Ensemble Transform Kalman Filter (ETKF) of Bishop et al. (2001) and Wang et al. (2004). Unlike variational-based data assimilation schemes, ensemble Kalman filter (EnKF) schemes represent the forecast uncertainty with an ensemble of forecasts. Ensemble based data assimilation is a natural approach since numerical weather prediction centres, such as NCEP, Meteorological Service of Canada (MSC) and ECMWF, already employ ensemble forecasting operationally to assess the uncertainty in their forecasts. Using this information in the data assimilation procedure has the potential to provide better initial conditions, for both the main forecast and the ensemble forecast. The goal of an EnKF

is to generate at regular time intervals an *analysis* ensemble. This analysis ensemble of model states should reflect both an estimate of the true atmospheric state (through its mean) and the uncertainty of this estimate (through its spread). If successful, then applying the forecast model to the analysis ensemble from one time yields a background ensemble at the next time. In this case, the background ensemble represents a probabilistic estimate of the atmospheric state before new observations are assimilated. The analysis cycle is completed by adjusting the background ensemble to better fit the new observations. In particular, the analysis ensemble mean is a weighted average of the background ensemble mean and the observations, with the weights determined from the background and observation uncertainties. More precisely, the analysis ensemble mean is the model state that best fits the given background and observation probability distributions.

In a Kalman filter, these distributions are assumed to be Gaussian and the uncertainties are thus characterized by covariance matrices. The background and observation covariances determine, via Bayes' rule, the analysis covariance. In ensemble-based schemes, the background covariance is computed as the sample covariance of the background ensemble, and thus to be consistent one must choose an analysis ensemble whose sample covariance matches the analysis covariance determined by the Kalman filter. Early versions of EnKF (Evensen, 1994; Burgers et al., 1998; Houtekamer and Mitchell, 1998) do this

*Corresponding author.
e-mail: jharlim@cims.nyu.edu
Present address: Courant Institute of Mathematical Sciences, New York University, 251 Mercer st., New York, NY 10012, USA*
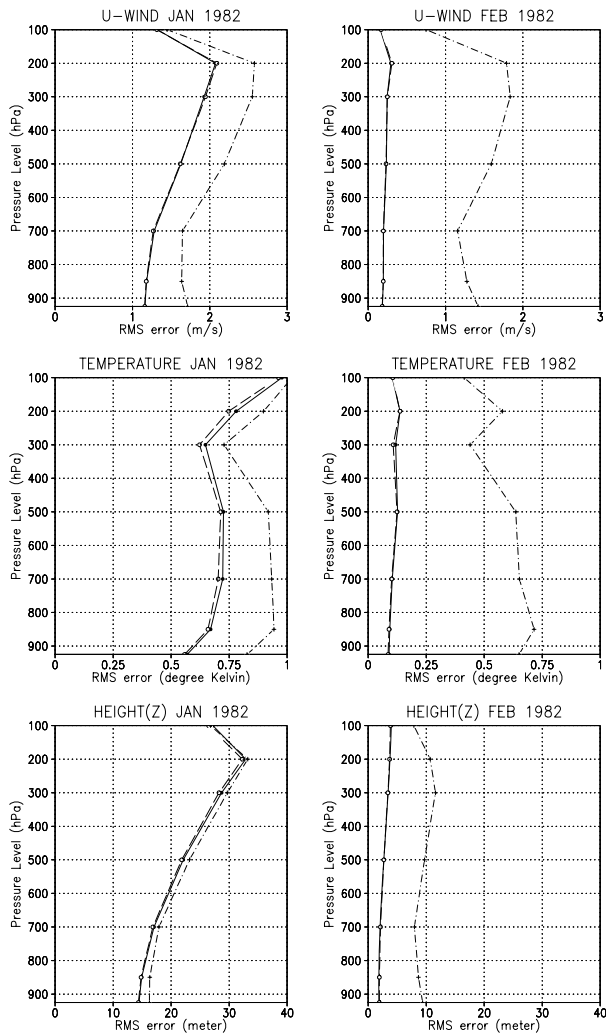
Fig. 1. Average analysis errors as functions of pressure (in hPa) for 12-h analysis cycle with 4D-LETKF (solid), FGAT-LETKF (dashes) and 3D-LETKF (dash–dotted line). Errors are averaged over period of January 1982 (on all left subfigures) and over period of February 1982 (right subfigures). Variables U-wind (first row) are in metres per second, temperature (second row) are in Kelvin and height (third row) are in metres.

stochastically, in the sense that they perturb the observations randomly and independently when generating each ensemble member. Another approach is to add the columns of a square root of the analysis error covariance matrix to the analysis ensemble mean. This approach is deterministic, provided that a specific choice of the matrix square root is specified, though many different choices for the square root are possible (Tippett et al., 2003). Examples of a deterministic EnKF include the Ensemble Adjustment Kalman Filter (EAKF) of Anderson (2001), ETKF of Bishop et al. (2001), Ensemble Square Root Filter of Whitaker and Hamill (2002) and LEKF of Ott et al. (2004).

In an operational setting, the analyses are typically generated every 6 h, though many observations are available more frequently. Limited computational time is allowed for each analysis (less than 10 min at NCEP). Given such constraints, an efficient algorithm to assimilate asynchronous observations becomes important. One approach is the 4D-EnKF of Hunt et al. (2004). This four-dimensional extension of EnKF finds the analysis ensemble mean by fitting the linear combinations of the trajectories of the background ensemble to the asynchronous observations. This scheme may be thought of as an approximation to 4D-VAR (Fertig et al., 2007), the four-dimensional data assimilation technique used operationally by ECMWF (see for example, Le Dimet and Talagrand, 1986; Courtier et al., 1994; Rabier et al., 1998, 2000). The main advantages of 4D-EnKF over 4D-VAR are that, as with other ensemble Kalman filters, it does not require computing the linear adjoint model for the (non-linear) forecast model, and it propagates background covariance information from one analysis cycle to the next.

Fertig et al. (2007) show that 4D-LETKF is comparable with 4D-VAR in a perfect model simulations with Lorenz-96 model (Lorenz, 1996). Here, we describe 4D-LETKF in more detail and present results of simulations with the Simplified-Parametrized primitive Equation Dynamics (SPEEDY) model (Molteni, 2003). In Section 2, we derive 4D-LETKF from the variational point of view, and describe how we apply the analysis locally and incorporate variance inflation. In Appendix A, we show that when the observation operator is linear, the analysis of 4D-LETKF is equivalent to that of 4D-EnKF (Hunt et al., 2004) and, in case of synchronous observations, LEKF (Ott et al., 2004). For an ensemble of size k, LEKF and 4D-EnKF perform the analysis in a $(k-1)$-dimensional space E, using an orthonormal basis consisting of eigenvectors of the background covariance matrix. Here, we show that performing the analysis in a k-dimensional space S with the background ensemble perturbations as the 'basis', each analysis becomes computationally more efficient. This choice of coordinates is equivalent to that used in the analysis of ETKF (Bishop et al., 2001). While our choice of matrix square root is different than in Bishop et al. (2001), it agrees with the square root used by Wang et al. (2004). A step-by-step pseudo-algorithm for implementing 4D-LETKF is given in Appendix B.

Simpler approaches for dealing with asynchronous observations are to treat the observations as if they occur at the analysis time (Gustafsson et al., 2001; Lindskog et al., 2001) or use First Guess at the Appropriate Time (FGAT) of Huang et al. (2002). FGAT computes the innovations, the differences between the observations and the model state, at the asynchronous times and treats these innovations as if they occur at the analysis time. Both approaches were implemented with 3D-VAR (Courtier et al., 1998) in a High Resolution Limited Area Model (HIRLAM). In Section 3, we adapt these approaches to the LETKF formulation and compare both schemes with 4D-LETKF in perfect model simulations with the SPEEDY model (Molteni, 2003). Our
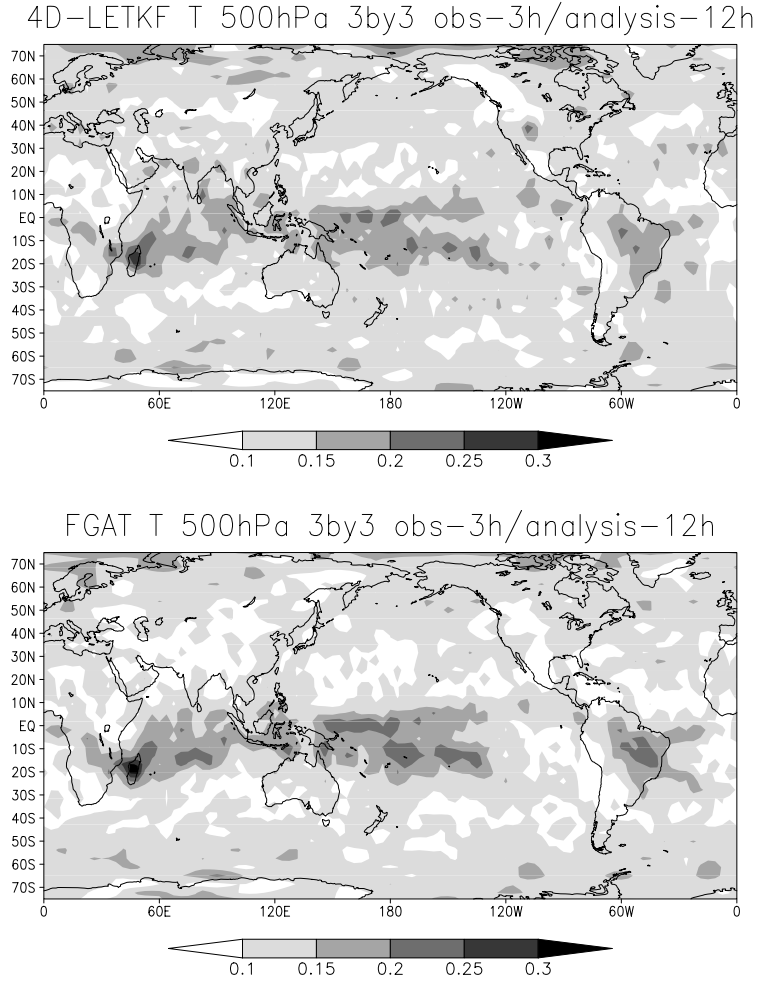
Fig. 2. Average analysis errors of temperature (in Kelvin) at 500 hPa for 12-h analysis cycle with 4D-LETKF (top) and FGAT-LETKF (bottom) during February 1982. In each analysis, the local region size is 3×3 grid points and the observations frequency is 3 h.

numerical results show that 4D-LETKF produces better analyses than these two simpler approaches when the analysis time interval is sufficiently long, while the results are similar to FGAT approach over shorter analysis intervals. We conclude this paper with a short summary in Section 4.

2. Formulations

The goal of data assimilation is to estimate the true state $\mathbf{x}_n^t$ of a system, such as the atmosphere, at current time t_n given noisy observations

$$\mathbf{y}_l^o = H_l(\mathbf{x}_l^t) + \epsilon_l^o, \tag{1}$$

where H_l and ϵ_l^o are the observation operator and observation error, respectively, at times $\{t_l: l = 1, 2, \ldots, n\}$, where $t_l < t_{l+1}$. Typically, the true state $\mathbf{x}_n^t$ and its underlying dynamics are unknown. We assume that the evolution of $\mathbf{x}_n^t$ is modelled by a chaotic dynamical system:

$$\mathbf{x}_{n+1}^t = \mathcal{M}(\mathbf{x}_n^t), \tag{2}$$

where $\mathcal{M}$ denotes a non-linear operator. In the derivations below, we assume that $\mathcal{M}$ is invertible. Indeed this is always the case if $\mathcal{M}$ is the evolution operator for a system of ordinary differential equations. However, $\mathcal{M}^{-1}$ may be difficult to compute in practice because integrating the model backward in time can be highly unstable, for example, in atmospheric models where the underlying system of partial differential equations is irreversible. We emphasize that we use $\mathcal{M}^{-1}$ only from a theoretical point of view below, and that the filter we develop in Section 2.1 does not require $\mathcal{M}^{-1}$ to be computed.

From a probabilistic point of view, we want not just a specific estimate of $\mathbf{x}_n^t$, but a probability distribution $p(\mathbf{x}_n)$ representing the likelihood that a particular model state $\mathbf{x}_n$ is equal to $\mathbf{x}_n^t$. Assuming that observations prior to times t_1 have yielded a background distribution $p(\mathbf{x}_n)$, the goal of data assimilation is to find the *analysis probability distribution* $p(\mathbf{x}_n \mid \mathbf{y}_1^o, \mathbf{y}_2^o, \ldots, \mathbf{y}_n^o)$ given observations $\{\mathbf{y}_l^o : l = 1, \ldots, n\}$ in addition to the background information. The analysis state $\mathbf{x}_n^a$ is typically chosen to be the mode of this distribution, that is, the most likely state. If the distribution of observation errors is known, then applying Bayes
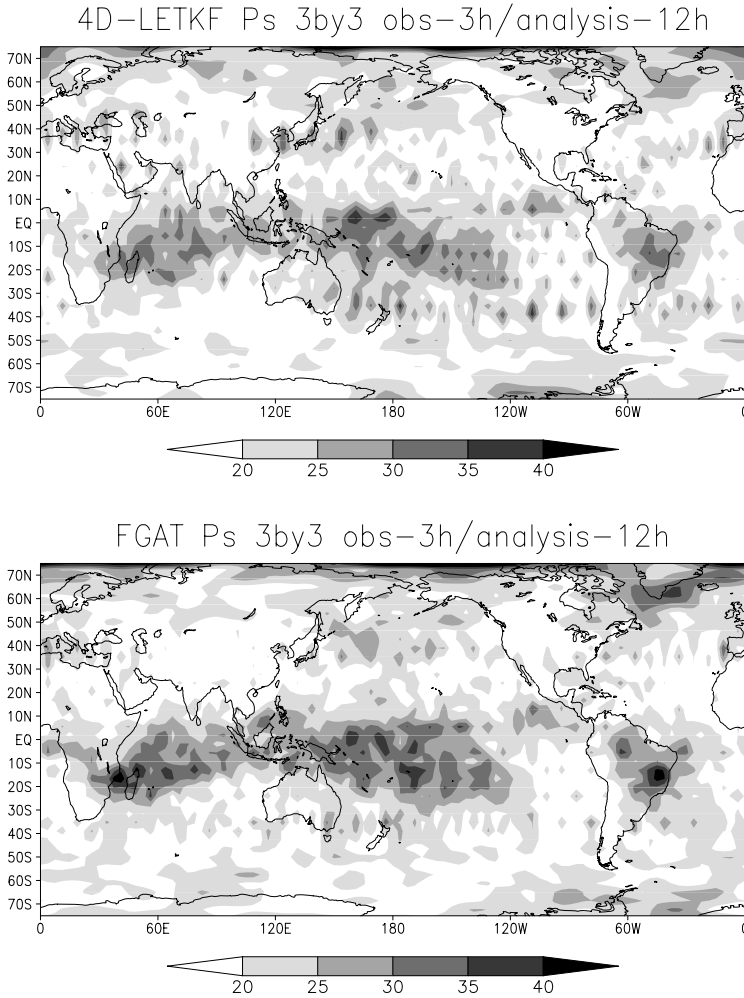
4D–LETKF Ps 3by3 obs–3h/analysis–12h

FGAT Ps 3by3 obs–3h/analysis–12h

Fig. 3. Average analysis errors of surface pressure (in Pa) for 12-h analysis cycle with 4D-LETKF (top) and FGAT-LETKF (bottom) during February 1982. In each analysis, the local region size is 3×3 grid points and the observations frequency is 3 h.

rule gives

$$p\big(\mathbf{x}_n \big| \mathbf{y}_1^o, \mathbf{y}_2^o, \dots, \mathbf{y}_n^o\big) \propto p(\mathbf{x}_n) p\big(\mathbf{y}_1^o, \mathbf{y}_2^o, \dots, \mathbf{y}_n^o \big| \mathbf{x}_n\big). \qquad (3)$$

Hereafter, we denote $\mathbf{y}^o = [(\mathbf{y}_1^o)^T, (\mathbf{y}_2^o)^T, \dots, (\mathbf{y}_n^o)^T]^T$, $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_n^T]^T$ and define operator $H(\mathbf{x}) = [H_1(\mathbf{x}_1)^T, H_2(\mathbf{x}_2)^T, \dots, H_n(\mathbf{x}_n)^T]^T$. Kalman filters generally assume Gaussian background and observation error distributions: $p(\mathbf{x}_n) \sim \mathcal{N}(\mathbf{x}_n^b, \mathbf{B})$ and $p(\mathbf{y}^o|\mathbf{x}_n) \sim \mathcal{N}(H(\mathbf{x}), \mathbf{R})$, respectively. Here, components of $\mathbf{x}$ depend on $\mathbf{x}_n$ through the relation $\mathbf{x}_{n-l} = \mathcal{M}^{-l}\mathbf{x}_n$ and $\mathbf{x}_n^b$ is the background state at time t_n, obtained by feeding the prior analysis state $\mathbf{x}_0^a$ at time t_0 into (2) and iterating. $\mathbf{B}$ is the background error covariance matrix and $\mathbf{R}$ is the observation error covariance matrix. We assume that the observation errors at different times $\{t_l : l = 1, \dots, n\}$ are uncorrelated, so the matrix $\mathbf{R}$ is block diagonal where each block is an $s_l \times s_l$ covariance matrix $E[\epsilon_l^o(\epsilon_l^o)^T] = \mathbf{R}_l$. Here, s_l denotes number of observations at time t_l. With these assumptions, maximizing (3) is equivalent

to minimizing the cost function:

$$\begin{aligned} J(\mathbf{x}) &= J^b(\mathbf{x}) + J^o(\mathbf{x}) \\ &= \frac{1}{2}\left(\mathbf{x}_n - \mathbf{x}_n^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}_n - \mathbf{x}_n^b\right) \\ &\quad + \frac{1}{2}[\mathbf{y}^o - H(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y}^o - H(\mathbf{x})], \\ &= \frac{1}{2}\left(\mathbf{x}_n - \mathbf{x}_n^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}_n - \mathbf{x}_n^b\right) \\ &\quad + \frac{1}{2}\sum_{l=1}^{n}\left[\mathbf{y}_l^o - H_l(\mathbf{x}_l)\right]^T \mathbf{R}_l^{-1}\left[\mathbf{y}_l^o - H_l(\mathbf{x}_l)\right]. \end{aligned} \qquad (4)$$

with constraints defined by the dependence of $\mathbf{x}_l$ at times $\{t_l : l = 1, \dots, n-1\}$ on $\mathbf{x}_n$ via the inverse of $\mathcal{M}$.

The cost function (4) is the same as for 4D-VAR, except that in 4D-VAR the background term and the vector $\mathbf{x}$ are expressed in term of model state $\mathbf{x}_1$ at the beginning of the analysis time interval. (This distinction will become irrelevant for the approximate cost function we minimize in the next section.) Operational implementations of 4D-VAR (e.g. at ECMWF) generally use the
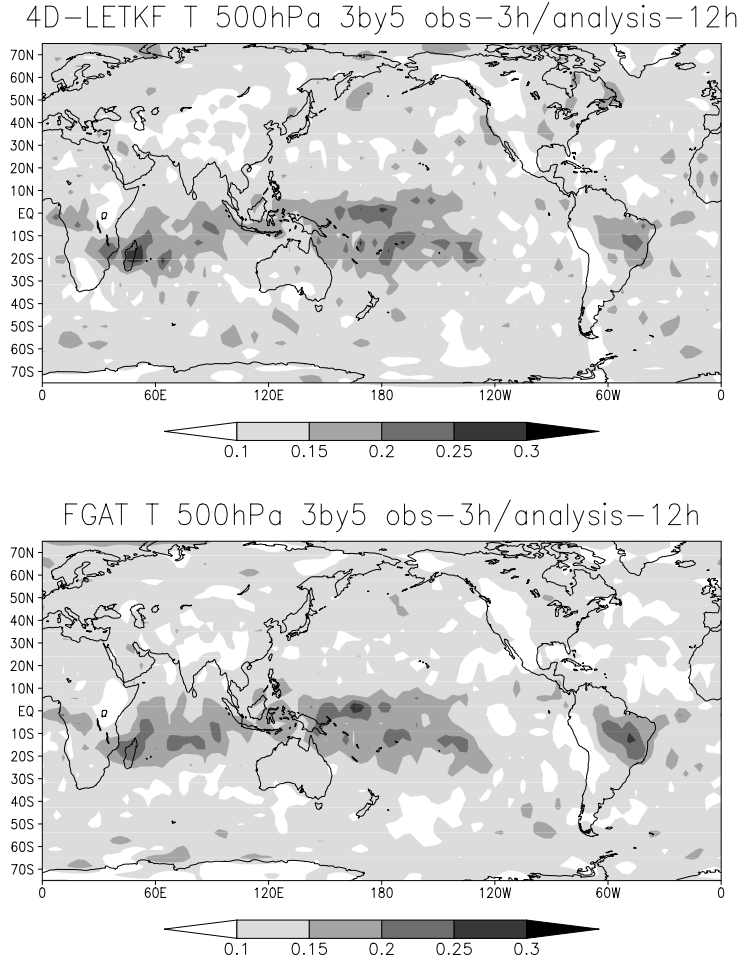
4D-LETKF T 500hPa 3by5 obs-3h/analysis-12h

FGAT T 500hPa 3by5 obs-3h/analysis-12h

Fig. 4. Average analysis errors of temperature (in Kelvin) at 500 hPa for 12-h analysis cycle with 4D-LETKF (top) and FGAT-LETKF (bottom) during February 1982. In each analysis, the local region size is 3 × 5 grid points and the observations frequency is 3 h.

same **B** for each analysis. However, the background uncertainty can vary considerably from time to time, so it is desirable to allow **B** to vary from one analysis to the next.

Kalman filters take the covariance from the analysis probability function $p(\mathbf{x}_n \mid \mathbf{y}_1^o, \ldots, \mathbf{y}_n^o)$ and use it to determine the background covariance for the next analysis cycle. In an ensemble Kalman filter, an ensemble of initial conditions distributed according to the analysis probability distribution is propagated by the model to produce an ensemble of background states whose sample mean and covariance determine the background probability distribution for the next analysis. In the next section, we use this approach to derive our four-dimensional ensemble Kalman filter. Like other Kalman filters, we will make a linear approximation that makes the cost function quadratic; allowing us to minimize the cost function exactly. In this four-dimensional setting, our approximation also turns the constrained minimization problem described in this section into an unconstrained problem.

2.1. Filter derivation through variational approach

Ensemble Kalman Filters estimate the true state $\mathbf{x}_n^t$ with an ensemble whose mean represents an estimate of $\mathbf{x}_n^t$ and covariance

represents the uncertainty in the estimate. Starting with a background ensemble $\{\mathbf{x}_n^{b(i)}, i = 1, \ldots, k\}$, the analysis step assimilates the observations to produce an analysis ensemble $\{\mathbf{x}_n^{a(i)}, i = 1, \ldots, k\}$ (which is used to provide the background for the next analysis cycle, as described above). In the cost function (4), we replace the background state $\mathbf{x}_n^b$ with the sample mean $\bar{\mathbf{x}}_n^b$ of the background ensemble, and the background error covariance matrix **B** with the sample covariance matrix

$$\mathbf{P}_n^b = (k-1)^{-1}\mathbf{X}_n^b\left(\mathbf{X}_n^b\right)^T, \tag{5}$$

where k is the number of ensemble members and

$$\mathbf{X}_n^b = \left[\mathbf{x}_n^{b(1)} - \bar{\mathbf{x}}_n^b, \mathbf{x}_n^{b(2)} - \bar{\mathbf{x}}_n^b, \ldots, \mathbf{x}_n^{b(k)} - \bar{\mathbf{x}}_n^b\right] \tag{6}$$

is the $m \times k$ matrix of background ensemble perturbations. Note that this approximation is problematic for $k < m$ since $\mathbf{P}_n^b$ is not a full rank matrix, and hence is not invertible. However, $(\mathbf{P}_n^b)^{-1}$ is well defined on the 'ensemble subspace' spanned by the columns of $\mathbf{X}_n^b$. Thus, an ensemble Kalman filter minimizes the cost function J over all model states $\mathbf{x}_n$ such that $\mathbf{x}_n - \bar{\mathbf{x}}_n^b$ is in this ensemble subspace, so that J is well defined.

4D−LETKF Ps 3by5 obs−3h/analysis−12h

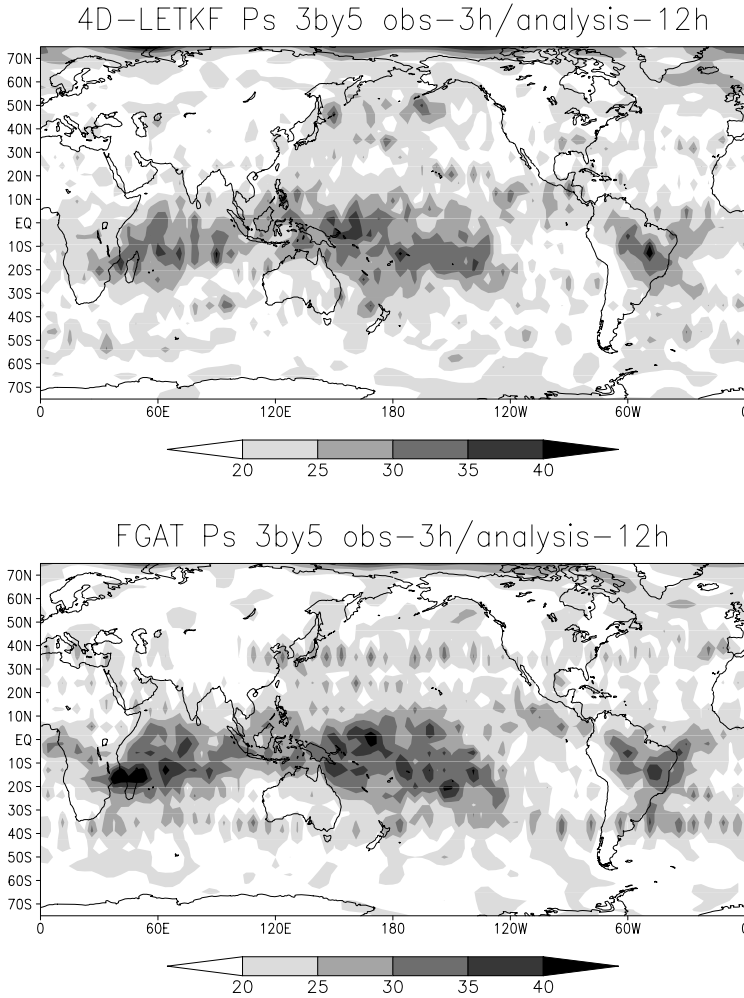## FGAT Ps 3by5 obs−3h/analysis−12h

Fig. 5. Average analysis errors of surface pressure (in Pa) for 12-h analysis cycle with 4D-LETKF (top) and FGAT-LETKF (bottom) during February 1982. In each analysis, the local region size is 3×5 grid points and the observations frequency is 3 h.

As we discussed earlier, computing the cost function J as a function of the model state $\mathbf{x}_n$ at time t_n would require using $\mathcal{M}^{-1}$, which may be impractical. However, by storing the background ensemble $\{\mathbf{x}_l^{b(i)}, i = 1, \ldots, k\}$ at the observation times t_l as the model is integrated from the previous analysis time, we can approximate $\mathcal{M}^{-1}$ as follows. To each $\mathbf{x}_n$ for which J is well defined, we will associate a corresponding model state $\mathbf{x}_l$ in the space spanned by the background ensemble at time t_l, using the same linear combination of the ensemble members as for $\mathbf{x}_n$. The background distribution of $\mathbf{x}_l$ will have mean equal to $\bar{\mathbf{x}}_l^b$, the sample mean of the ensemble at time t_l, and covariance $\mathbf{P}_l^b = (k-1)^{-1} \mathbf{X}_l^b (\mathbf{X}_l^b)^T$, where

$$\mathbf{X}_l^b = \left[\mathbf{x}_l^{b(1)} - \bar{\mathbf{x}}_l^b, \mathbf{x}_l^{b(2)} - \bar{\mathbf{x}}_l^b, \ldots, \mathbf{x}_l^{b(k)} - \bar{\mathbf{x}}_l^b \right]. \quad (7)$$

In this sense, at each time t_l we will use the Gaussian background distribution associated with the ensemble at that time.

To make this approach more precise, we employ a precondition (or a coordinate change) by expressing the deviation of a state $\mathbf{x}_l$ from the background mean state $\bar{\mathbf{x}}_l^b$ as a linear com-

bination of the background ensemble of perturbations, that is,

$$\mathbf{x}_l = \bar{\mathbf{x}}_l^b + \mathbf{X}_l^b \mathbf{w}, \quad l = 1, 2, \ldots, n, \quad (8)$$

where the weight $\mathbf{w} \in \mathbb{R}^k$ is to be determined. Here $\mathbf{w}$ does not depend on l, so the sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is a linear combination of the ensemble trajectories. The resulting trajectory represents an approximate model trajectory, rather than an exact trajectory of (2) as described earlier. By making this approximation, we can consider the unconstrained problem of minimizing J as a function of $\mathbf{w}$.

Next, we approximate the observation vector corresponding to the model state $\mathbf{x}_l$ at time t_l by:

$$H_l(\mathbf{x}_l) = H_l\left(\bar{\mathbf{x}}_l^b + \mathbf{X}_l^b \mathbf{w}\right) \approx H_l\left(\bar{\mathbf{x}}_l^b\right) + \mathbf{Y}_l^b \mathbf{w}. \quad (9)$$

Here the ith column vector of the $s_l \times k$ matrix $\mathbf{Y}_l^b$ is the deviation of $H_l(\mathbf{x}_l^{b(i)})$ from its ensemble average. That is,

$$\mathbf{Y}_l^b = \left[H_l\left(\mathbf{x}_l^{b(1)}\right) - \bar{\mathbf{y}}_l^b, H_l\left(\mathbf{x}_l^{b(2)}\right) - \bar{\mathbf{y}}_l^b, \ldots, H_l\left(\mathbf{x}_l^{b(k)}\right) - \bar{\mathbf{y}}_l^b \right], \quad (10)$$
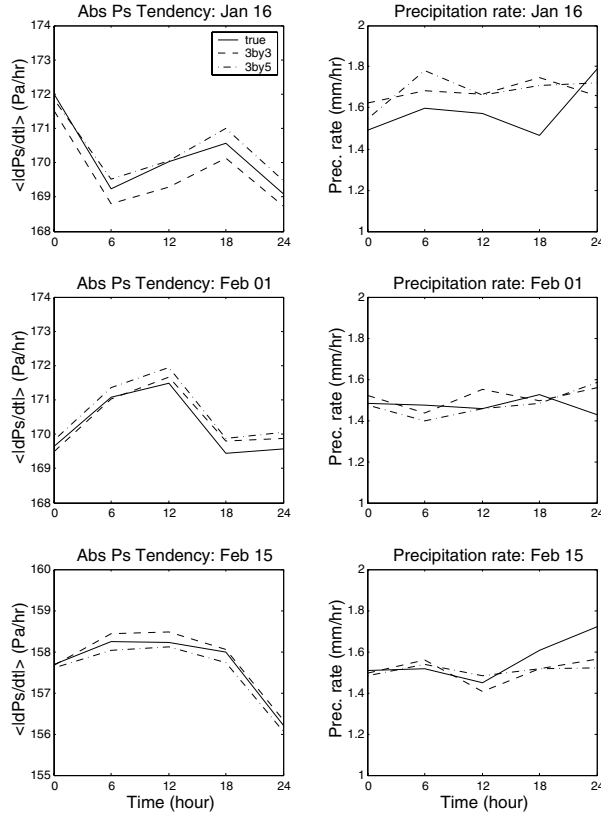
Fig. 6. The left figures show the average absolute value of surface pressure tendency (in Pa h^{-1}) as a function of time for forecasts initiated at time 0 on three arbitrarily chosen dates (January 16, February 1 and February 15) in our assimilation period. The right figures show the average of the precipitation rate (in mm h^{-1}) as a function of time. In each figure, the solid line represents the corresponding quantity for the true trajectory, the dashes for the analysis with a 3×3 local region, and the dash–dotted line for the analysis with a 3×5 local region.

with $\bar{\mathbf{y}}_l^b = \frac{1}{k} \sum_{i=1}^{k} H_l(\mathbf{x}_l^{b(i)})$.[1] Replacing $\mathbf{x}_l^b$ with $\bar{\mathbf{x}}_l^b$ and $\mathbf{B}$ with $\mathbf{P}_n^b$, and using (5), (8) and (9), reduce the cost function (4) to:

$$J(\mathbf{w}) = \frac{1}{2}(k-1)\mathbf{w}^T\mathbf{w} + \frac{1}{2}\sum_{l=1}^{n}\left[\mathbf{y}_l^o - H_l(\bar{\mathbf{x}}_l^b) - \mathbf{Y}_l^b\mathbf{w}\right]^T$$
$$\times \mathbf{R}_l^{-1}\left[\mathbf{y}_l^o - H_l(\bar{\mathbf{x}}_l^b) - \mathbf{Y}_l^b\mathbf{w}\right]. \qquad (11)$$

That is, we reduce the $(n \times m)$-dimensional constrained minimization problem to a k-dimensional unconstrained minimization problem. Note that in the $\mathbf{w}$ coordinate system, the background error covariance matrix becomes the identity and hence we do not have to invert it. This choice of coordinates is analogous to a preconditioning step that often done in variational methods, whereby the cost function is expressed in terms of a vector whose background covariance is the identity matrix. Lorenc

---

[1]If H$_l$ is linear, then $\bar{\mathbf{y}}_l^b = H_l(\bar{\mathbf{x}}_l^b)$. For non-linear H$_l$, these quantities are different.

(2003) suggests using cost function similar to (11), but without the linear approximation (9), to incorporate an ensemble-based background covariance into 3D-VAR or 4D-VAR, and Zupanski (2005) uses the same cost function in an ensemble filter. A similar cost function has also been used in hybrid ensemble/3D-VAR methods (Buehner, 2005; Wang et al., 2007).

The minimum of (11) is obtained by setting

$$\nabla J(\mathbf{w}) = (k-1)\mathbf{w} - \sum_{l=1}^{n}\left(\mathbf{Y}_l^b\right)^T\mathbf{R}_l^{-1}\left[\mathbf{y}_l^o - H_l(\bar{\mathbf{x}}_l^b) - \mathbf{Y}_l^b\mathbf{w}\right] = \mathbf{0}.$$

The solution of this equation is the analysis weight vector

$$\bar{\mathbf{w}}^a = \tilde{\mathbf{P}}^a\left\{\sum_{l=1}^{n}\left(\mathbf{Y}_l^b\right)^T\mathbf{R}_l^{-1}\left[\mathbf{y}_l^o - H_l(\bar{\mathbf{x}}_l^b)\right]\right\}, \qquad (12)$$

where

$$\tilde{\mathbf{P}}^a = \left[(k-1)\mathbf{I} + \sum_{l=1}^{n}\left(\mathbf{Y}_l^b\right)^T\mathbf{R}_l^{-1}\mathbf{Y}_l^b\right]^{-1}. \qquad (13)$$

The analysis error covariance matrix $\tilde{\mathbf{P}}^a$ in the ensemble space is the inverse of the Hessian of the cost function (11) (Fisher and Courtier, 1995; Zupanski, 2005). The analysis state is obtained by substituting (12) into (8):

$$\bar{\mathbf{x}}_n^a = \bar{\mathbf{x}}_n^b + \mathbf{X}_n^b\bar{\mathbf{w}}^a$$
$$= \bar{\mathbf{x}}_n^b + \mathbf{X}_n^b\tilde{\mathbf{P}}^a\left\{\sum_{l=1}^{n}\left(\mathbf{Y}_l^b\right)^T\mathbf{R}_l^{-1}\left[\mathbf{y}_l^o - H_l(\bar{\mathbf{x}}_l^b)\right]\right\}. \qquad (14)$$

In the case that H_l is linear, eqs (13) and (14) are equivalent to the standard Kalman filter equations, which minimize J in closed form.

To complete the analysis, we generate an analysis ensemble of model states whose mean is $\bar{\mathbf{x}}^a$ and whose error covariance matrix in the model space is $\mathbf{P}^a = \mathbf{X}^b\tilde{\mathbf{P}}^a(\mathbf{X}^b)^T$. To satisfy these constraints, we update the ensemble using

$$\mathbf{x}_n^{a(i)} = \bar{\mathbf{x}}_n^a + \mathbf{X}_n^b\mathbf{W}^{a(i)} = \bar{\mathbf{x}}_n^b + \mathbf{X}_n^b\left(\bar{\mathbf{w}}^a + \mathbf{W}^{a(i)}\right), \qquad (15)$$

where $\mathbf{W}^{a(i)}$ is the ith column of matrix $\mathbf{W}^a = [(k-1)\tilde{\mathbf{P}}^a]^{\frac{1}{2}}$,[2] where $M^{\frac{1}{2}}$ denotes the symmetric square root of a symmetric matrix M (We compute $M^{\frac{1}{2}}$ using its eigenvalues and eigenvectors; since $M = U\Sigma U^T$ where U is an (orthogonal) matrix of eigenvectors of M and Σ is a diagonal matrix of its eigenvalues, $M^{\frac{1}{2}} = U\Sigma^{\frac{1}{2}}U^T$, where $\Sigma^{\frac{1}{2}}$ is the diagonal matrix of the square roots of the eigenvalues).

As we will show in Appendix A, this four-dimensional filter is equivalent to 4D-EnKF of Hunt et al. (2004). The main difference is in the choice of coordinate (8). Next let us describe the localization and a way to do variance inflation.

---

[2]In order that the mean of the analysis ensemble be $\bar{\mathbf{x}}_n^a$, we need that $\mathbf{X}_n^a\mathbf{1} = \mathbf{X}_n^b\mathbf{W}^a\mathbf{1}$ to be zero, where $\mathbf{1} = (1, 1, \ldots, 1)^T$. Since $\mathbf{X}_l^b\mathbf{1} = \mathbf{0}$, it suffices that $\mathbf{1}$ be an eigenvector of $\mathbf{W}^a$. This is also true for the symmetric square root $\mathbf{W}^a$.

4D−LETKF T 500hPa 3by3 obs−6h/analysis−12h
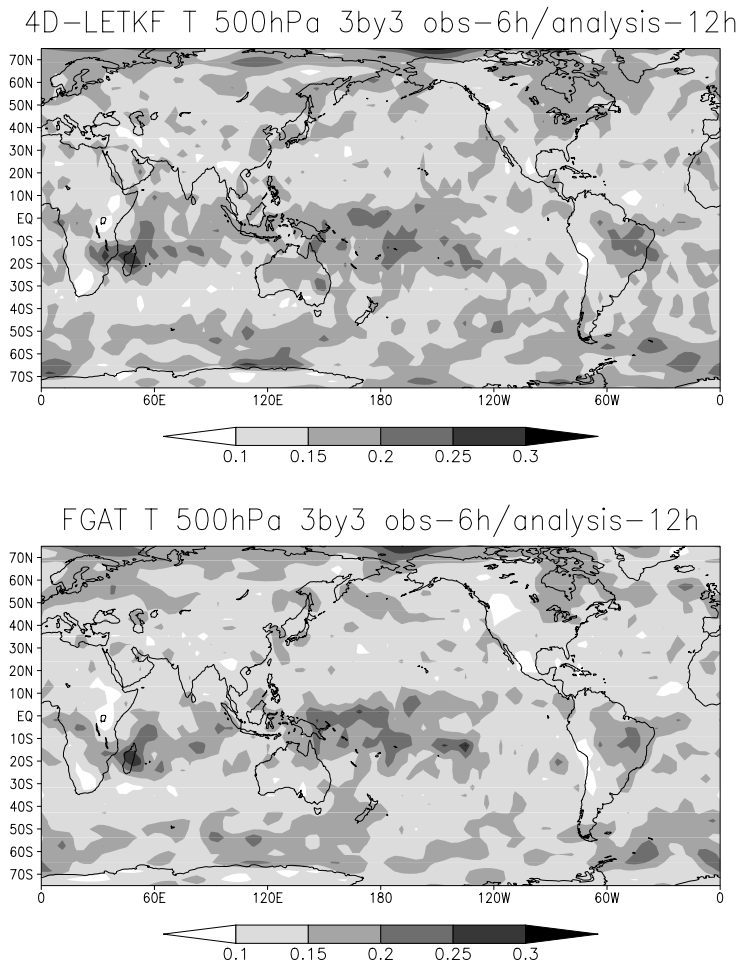

FGAT T 500hPa 3by3 obs−6h/analysis−12h

Fig. 7. Average analysis errors of temperature (in Kelvin) at 500 hPa for 12-h analysis cycle with 4D-LETKF (top) and FGAT-LETKF (bottom) during February 1982. In each analysis, the local region size is 3×3 grid points and the observations frequency is 6 h.

2.2. *Localization*

To perform ensemble data assimilation for a global atmospheric model with an ensemble of moderate size, some form of localization is necessary. As pointed out in Anderson (2001), Hamill et al. (2001) and Houtekamer and Mitchell (2001), the localization suppresses spurious long-range correlations produced by a limited ensemble size. On the other hand, it also improves the efficiency of the scheme because each local analysis involves much less data than a global analysis. Our localization is similar to that of Keppenne (2000) and Ott et al. (2004), in that the analysis is done separately and, if desired, in parallel for different local regions that cover the globe. In our formulation, the localization is relatively simple; for each grid point of the model, we choose a local subset of the global observations and apply the equations of Section 2.1 using only the local observations. To be more precise, see step (ii) of Appendix B.

In order to use the analysis ensemble members as initial conditions for the forecast model, it is essential that the results of the analysis be similar at nearby grid points. This can be ensured by choosing similar sets of observations for neighbouring grid points. As long as the observation sets overlap heavily, the analy-

ses will be similar. In practice, we find that such heavy overlap is not always necessary. However, our choice of the matrix square root is important in achieving consistency of nearby local analysis; the symmetric square root ensures that $\mathbf{W}^a$ depends continuously on $\tilde{\mathbf{P}}^a$. Indeed, with localization, other choices of the matrix square root can cause our filter to diverge (Harlim, 2006).

2.3. *Variance inflation*

In order to compensate for the tendency of a small ensemble to underestimate uncertainty, it may be desirable to artificially inflate the background error covariance matrix $\mathbf{P}^b$ before each analysis. (Or, one could instead inflate the analysis error covariance matrix $\mathbf{P}^a$ after each analysis.) A common approach is multiplicative variance inflation (Anderson and Anderson, 1999): multiply the background ensemble perturbations $\mathbf{X}_n^b$ by a constant factor $\sqrt{1+r}$ with $r > 0$, which effectively multiplies $\mathbf{P}^b$ by $(1 + r)$. A similar result can be achieved more efficiently by leaving $\mathbf{X}_n^b$ alone and replacing (13) by

$$\tilde{\mathbf{P}}^a = \left[(k-1)\mathbf{I}/(1+r) + \sum_{l=1}^{n} \left(\mathbf{Y}_l^b\right)^T \mathbf{R}_l^{-1} \mathbf{Y}_l^b \right]^{-1}. \qquad (16)$$
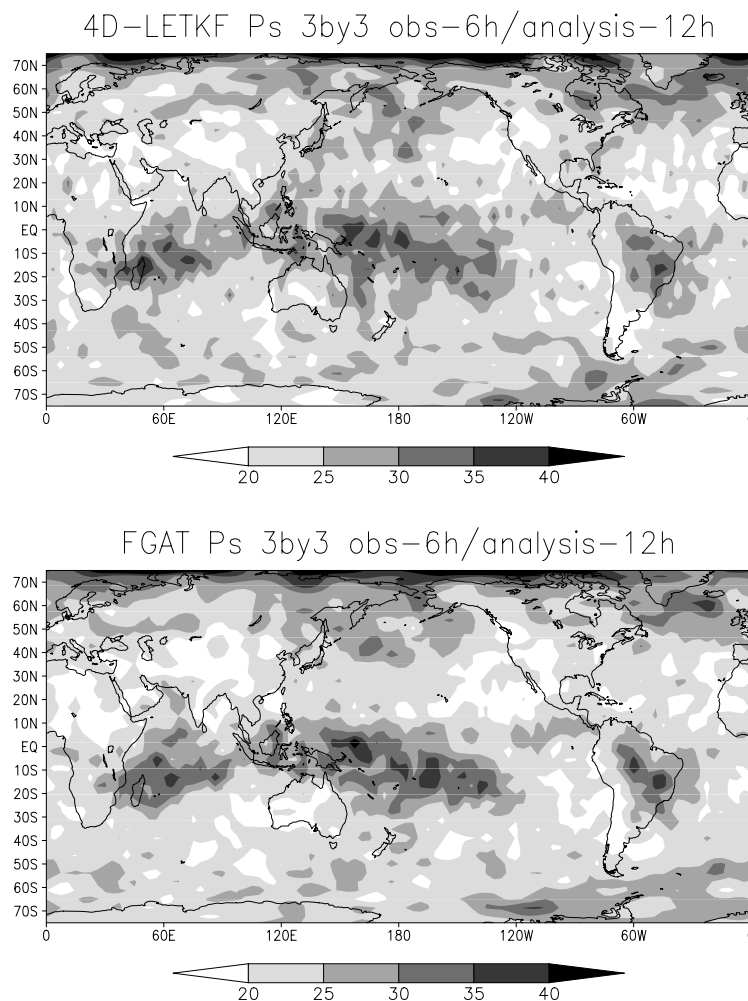
Fig. 8. Average analysis errors of surface pressure (in Pa) for 12-h analysis cycle with 4D-LETKF (top) and FGAT-LETKF (bottom) during February 1982. In each analysis, the local region size is 3×3 grid points and the observations frequency is 6 h.

Here $(k - 1)$ **I** represents the inverse of the background covariance matrix in the **w** coordinate system, and in (16) we have multiplied this covariance matrix by $(1 + r)$.

2.4. Relationship with other filters

Now, suppose that there are no observations at time $\{t_l : l = 1, 2, \ldots, n - 1\}$ and that all observations at time t_n are used (i.e. no localization is done). Equation (14) is reduced to the analysis mean update in a standard three-dimensional ensemble Kalman filter. The analysis ensemble update (15) is different from the ETKF of Bishop et al. (2001) because the choice of matrix square root is different. Our choice of the symmetric square root is the same as the Spherical-Simplex Ensemble Transform Kalman Filter (Wang et al., 2004). That is, our matrix $\mathbf{W}^a$ is equivalent to Wang et al. (2004) transform matrix $\mathbf{T} \in^T$ (eqs 5 and C1).

In the three-dimensional setting, the analysis used in the 4D-EnKF (Hunt et al., 2004) is the same as that in LEKF (Ott et al., 2004). We show in Appendix A that LETKF is identical to LEKF except in the coordinate representations of each analysis; LETKF

is more efficient since it requires no computations of the eigenvalues of the background error covariance matrix $\mathbf{P}^b$. In numerical experiments with the Lorenz-96 model (Lorenz, 1996), we are able to reproduce the results of Ott et al. (2004) with significant computational savings (Harlim, 2006).

3. Numerical experiments with SPEEDY model

3.1. SPEEDY model

In this paper, we apply our four-dimensional filter to a primitive-equation Global Circulation Model (GCM). This spectral model (nicknamed SPEEDY, for Simplified Parametrizations primitive-Equation Dynamics; see Molteni, 2003, for details) has 7 vertical levels (with sigma level 0.950, 0.835, 0.685, 0.510, 0.340, 0.200 and 0.080) and a horizontal resolution corresponding to a triangular spectral truncation at total wave number 30 (this yields 96×48 grid points in a standard Gaussian grid). There are five basic prognostic variables: vorticity, divergence, absolute temperature, specific humidity and logarithm of
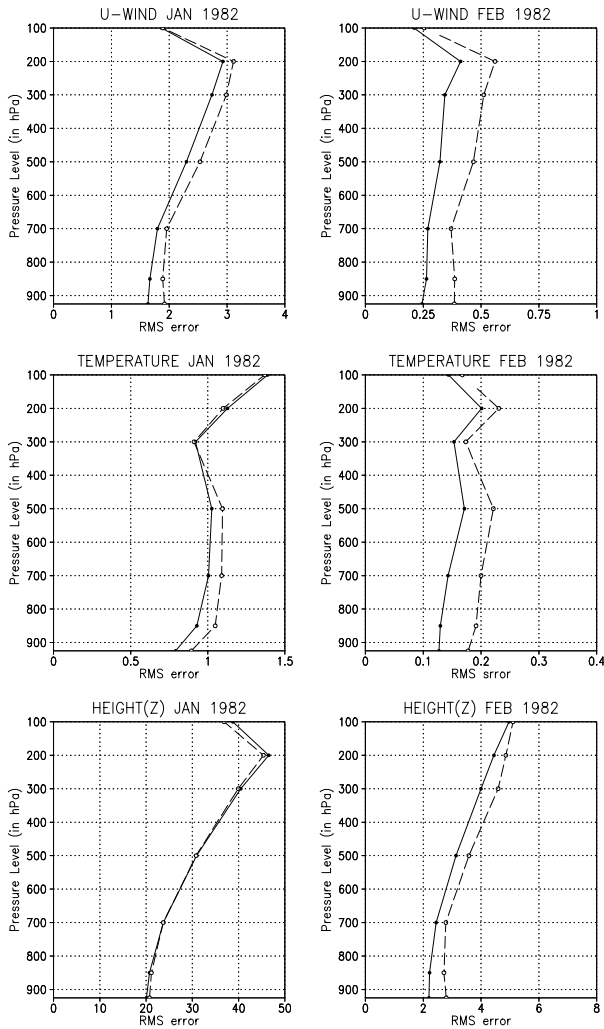
Fig. 9. Average analysis errors as functions of pressure (in hPa) for 24-h analysis cycle with 4D-LETKF (solid) and FGAT-LETKF (dashes). Errors are averaged over period of January 1982 (on all left subfigures) and over period of February 1982 (right subfigures). Variables U-wind (first row) are in metres per second, temperature (second row) are in Kelvin and height (third row) are in metres.

surface pressure. In addition to these variables, the model includes some diagnostic variables (such as saturation specific humidity, relative humidity, dry and moist static energy and saturation moist static energy) whose dynamics follow some simple models of physical processes (such as convection, large-scale condensation, clouds, short-wave and long-wave radiations and diffusion). With these simplified parametrizations, the model is designed to be (at least) an order of magnitude faster (in CPU time) than an operational GCM with similar horizontal resolutions.

The model dissipation and external forcing are determined by winter-time climatological fields sea-surface temperature, surface temperature and moisture in the top soil layer, snow

depth, bare-surface albedo, fraction of the sea-ice, land-ice and land-surface covered by vegetation. The model forcing are updated daily with no diurnal variations. The prognostic variables are post-processed into zonal and meridional wind components (U- and V-wind), absolute temperature (T), specific humidity (Q), geopotential heights (Z) on pressure levels (925, 850, 700, 500, 300, 200 and 100 hPa) and surface pressure (Ps).

Theoretically, the forecasts produced by SPEEDY model are less realistic than a more sophisticated model with higher resolution such as the NCEP GFS. However, this model serves our purpose in this paper since it is computationally inexpensive and it describes the atmospheric variability in the Northern Hemisphere during winter-time reasonably well.

3.2. Experimental design

In this paper, we compare 4D-LETKF to two approaches for handling asynchronous observations (implemented to LETKF). As we mentioned ealier, both approaches were formerly implemented in a 3D-VAR scheme.

The first approach (Gustafsson et al., 2001; Lindskog et al., 2001) is to treat all observations as if they occur at analysis time. In our variational formulation, this approach corresponds to replacing $\mathbf{x}_l^{b(i)}$ with $\mathbf{x}_n^{b(i)}$ in (10) and $\bar{\mathbf{x}}_l^b$ by $\bar{\mathbf{x}}_n^b$ in (11). In all of our experiments, we observe the same variables at each time t_l, so $H_l = H_n$ for all l. In this case, we replace $\mathbf{Y}_l^b$ by $\mathbf{Y}_n^b$ and the cost function (11) becomes

$$J(\mathbf{w}) = \frac{1}{2}(k-1)\mathbf{w}^T\mathbf{w} + \frac{1}{2}\sum_{l=1}^n \left[\mathbf{y}_l^o - H_l(\bar{\mathbf{x}}_n^b) - \mathbf{Y}_n^b\mathbf{w}\right]^T$$
$$\times \mathbf{R}_l^{-1}\left[\mathbf{y}_l^o - H_l(\bar{\mathbf{x}}_n^b) - \mathbf{Y}_n^b\mathbf{w}\right]. \tag{17}$$

Note that in the second term of (17), all model dependent variables are at time t_n. Here after, we refer to this approach as 3D-LETKF where asynchronous observations are treated as observations at analysis time.
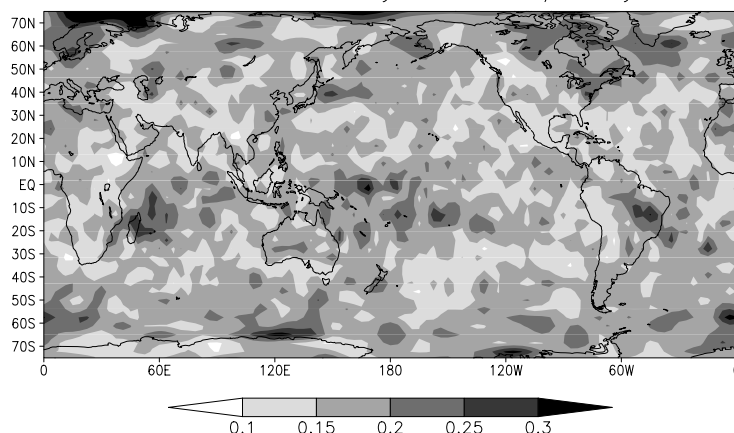
The second approach is to treat each observation innovation (the difference between the observation and the background model state at observation time) as if it occurs at the analysis time. In our variational formulation, this corresponds to replacing $\mathbf{x}_l^{b(i)}$ by $\mathbf{x}_n^{b(i)}$ in (10) but keeping $\bar{\mathbf{x}}_l^b$ in (11). In our scenario where $H_l = H_n$ for all l, the cost function then becomes

$$J(\mathbf{w}) = \frac{1}{2}(k-1)\mathbf{w}^T\mathbf{w} + \frac{1}{2}\sum_{l=1}^n \left[\mathbf{y}_l^o - H_l(\bar{\mathbf{x}}_l^b) - \mathbf{Y}_n^b\mathbf{w}\right]^T$$
$$\times \mathbf{R}_l^{-1}\left[\mathbf{y}_l^o - H_l(\bar{\mathbf{x}}_l^b) - \mathbf{Y}_n^b\mathbf{w}\right]. \tag{18}$$

Huang et al. (2002) call this way of handling observations as the First Guess at the Appropriate Time (FGAT). Hereafter, we refer it as FGAT-LETKF.

For each of the cost function (17) and (18), we perform the analysis using equations analogous to eqs (14)–(16) that we use for cost function (11). Specifically, for FGAT-LETKF we change $\mathbf{Y}_l^b$ to $\mathbf{Y}_n^b$ whenever it appears in these equations. For 3D-LETKF,
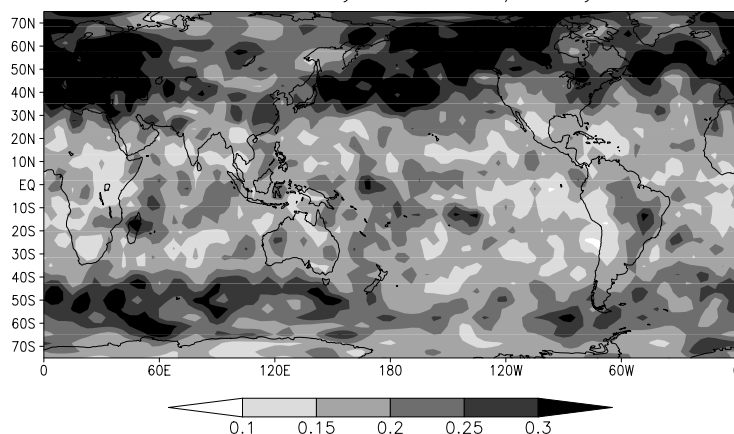
Fig. 10. Average analysis errors of temperature (in Kelvin) at 500 hPa for 24-h analysis cycle with 4D-LETKF (top) and FGAT-LETKF (bottom) during February 1982. In each analysis, the local region size is 3 × 3 grid points and the observations frequency is 6 h.

we additionally change $\bar{\mathbf{x}}_l^b$ to $\bar{\mathbf{x}}_n^b$ in (14). Thus, 3D-LETKF uses the background ensemble only at the analysis time, while both FGAT-LETKF and 4D-LETKF use the background ensemble states at each observation time, though FGAT-LETKF uses only the background mean at times $t_l < t_n$.

We compare these three schemes with a 12-h analysis cycle and a 24-h analysis cycle. A true trajectory is created by running the SPEEDY model for 2 months starting from NCEP reanalysis January 01 1982. Then, we generate simulated observations by adding a normally distributed noise with zero mean to the true states every 3 h for a 12-h analysis cycle and every 6 h for a 24-h analysis cycle. Here the observations errors for each variables are: 1 m s^{-1} for (both zonal, U-wind and meridional, V-wind) wind speed, 1 K for temperature, 0.0001 kg kg^{-1} for specific humidity, and 100 Pa for surface pressure, which are small compared to the model natural variability (standard deviation from its temporal mean without data assimilation). For pressure level 500 hPa, the natural variabilities are 6.78 m s^{-1} for U-wind, 6.84 m s^{-1} for V-wind, 2.92 K for temperature, 0.0005 kg kg^{-1} for specific humidity, 69.41 m for the geopotential height and 695.05 Pa for the surface pressure. At each (sigma) level, we

observe about 22% of the grid points (1008 locations) with observations uniformly distributed between 75°N and 75°S.

In each data assimilation experiment, we use an ensemble of size 20. The initial ensemble consists of states from a long integration of the SPEEDY model at 20 randomly chosen times. For each analysis, we use observations from a two-dimensional local region of size 3 × 3 grid points; that is, we use all observations from the same vertical level (recall that there are seven vertical levels in the model) and up to one grid spacing away in both latitude and longitude. For the analysis at each grid point, the number of observation locations varies among 1, 2 and 4 in each local region. For the 12-h analysis cycle, we also add results assimilated with a two-dimensional local region of size 3 × 5 grid points (3 grid points in latitude and 5 in longitude), so that in each local analysis the number of observation locations varies among 2, 3, 4 and 6. Other local region sizes we tried yield similar or worse results in all the cases we show.

Note that for every pair of horizontally neighbouring grid points, their local regions share a set of observations that constitute all of the observations for one region but only half or 2/3 of the observations for the other region. Furthermore, the local
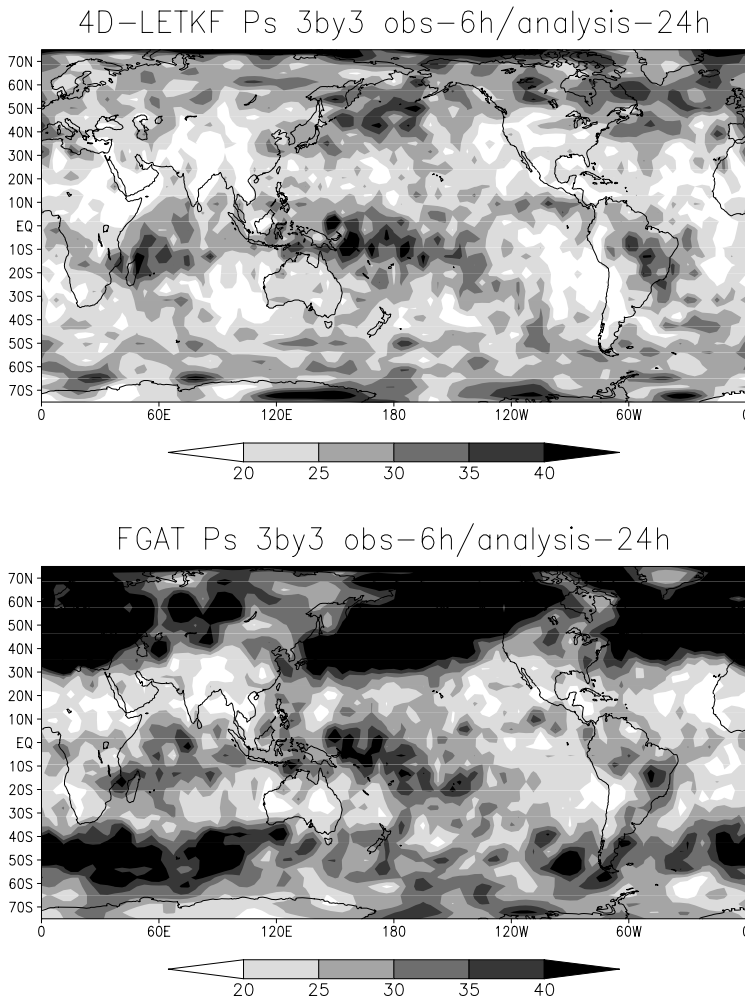
Fig. 11. Average analysis errors of surface pressure (in Pa) for 24-h analysis cycle with 4D-LETKF (top) and FGAT-LETKF (bottom) during February 1982. In each analysis, the local region size is 3×3 grid points and the observations frequency is 6 h.

regions do not overlap vertically. While this makes it possible to have incompatible analysis at neighbouring grid points, below we do not find the analysis fields to be imbalanced.

In the 12-h analysis cycle, we show simulations with variance inflation coefficient $r = 20\%$ in the case of 3×3 local region after comparing simulations with $r = 5, 15, \ldots, 30\%$. In the case of 3×5 local region simulations, we use variance inflation coefficient $r = 30\%$ after comparing simulations with $r = 20, 25, 30, \ldots, 40\%$. For the 24-h analysis cycle, we compared simulations with variance inflation coefficient $r = 30, 35, \ldots, 60\%$ and show results with $r = 40\%$. In each case, the chosen value of r yielded the best results for 4D-LETKF and FGAT-LETKF, and the results did not improve substantially with other values of r for 3D-LETKF.

3.3. Results

In the results below, we measure the quality of each analysis by calculating the root mean square (rms) difference between the true state and the analysis ensemble mean at analysis time

t_n. Here after, we refer to this quantity as the analysis error. The temporal average analysis error is calculated in the rms sense.

In Fig. 1, we show average analysis errors with a 12-h analysis cycle as functions of pressure (in hPa). We see that on average (during January and February 1982), the analysis errors of both 4D-LETKF (solid) and FGAT-LETKF (dashes) are very similar and significantly outperform 3D-LETKF (dash–dotted) on all shown variables (U-wind, temperature and geopotential height). In Fig. 2, we show for 4D-LETKF and FGAT-LETKF the geographical dependence of the average analysis temperature error for February 1982. While the FGAT-LETKF errors are smaller in some areas, the 4D-LETKF errors are notably better in the tropical regions where the errors are the largest (greater than 0.15). In Fig. 3, where the average analysis errors for surface pressure (in Pa) are shown, we see a similar pattern. That is, the errors of FGAT-LETKF are generally larger than those of 4D-LETKF in regions where both errors are large (greater than 25), mainly in the tropics but also near southern Greenland.
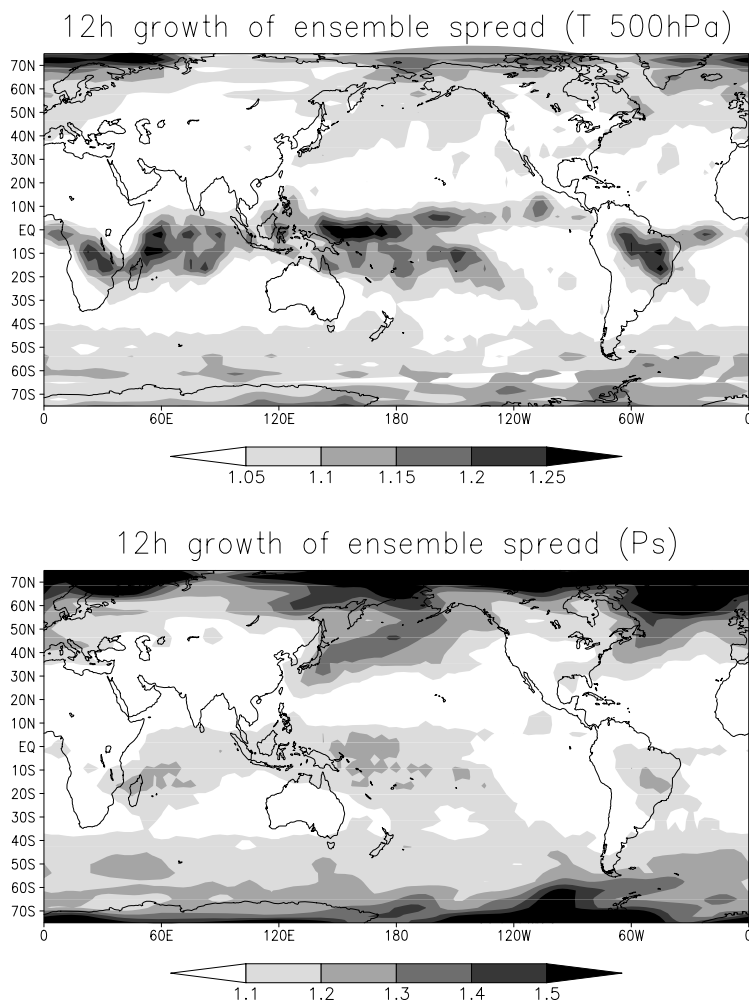
12h growth of ensemble spread (T 500hPa)

12h growth of ensemble spread (Ps)

Fig. 12. The average 12-h growth of the ensemble spread (as defined in the text) during February 1982. The top figure shows the growth for temperature at pressure level 500 hPa. The bottom figure shows the growth for surface pressure.
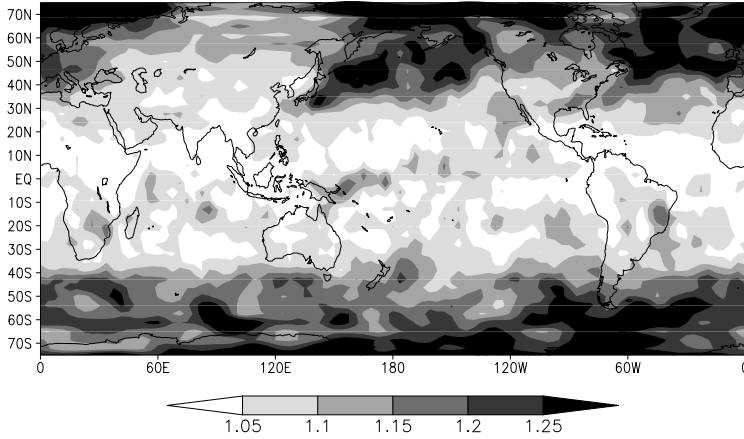
In Figs. 4 and 5, we show the analysis errors of assimilations with local region of size 3×5. Note that the experiments with larger local region require more variance inflation (as noted earlier, $r = 30\%$ for this choice, and 20% for those assimilated with local region of size 3×3) because the ratio of the ensemble size and the number of model variables in each local analysis decreases as a function of local region size. Secondly, the larger the window of local region suggests more observations are included in each local analysis. Thus, the larger the uncertainties are reduced which implies that each ensemble members are more alike the other members. In fact, we observed a worse errors when the window size is larger.

Note also that there are significant small-scale fluctuations in the surface pressure analysis error in Figs. 3 and 5. This is an artefact of our observation network, which provides observations only at every other model grid point, so that as noted above the number of observations assimilated varies significantly from one grid point to the next. The analysis ensemble fields do not seem to be imbalanced, however. To check the balance, in Fig. 6, we

show the average absolute value of the surface pressure tendency and the precipitation rate, both plotted as functions of forecast time. Here, both quantities are averaged over the globe excluding the polar regions above $75°$ latitude. In Fig. 6, we choose three different times (January 16, February 1 and February 15) from our assimilation period and compute a 24-h forecasts of both quantities with initial conditions from the true state (solid), from the analysis with local region of size 3×3 (dashes) and from the analysis with local region of size 3×5 (dash–dotted line). In these results we see no sign of imbalance at the analysis time for either local region size.

Below, we will show results of a 24-h analysis interval with local region of size 3×3. First, we include results (Figs. 7 and 8) that show the effect of using less frequent observations (every 6 h rather than 3 h) in a 12-h analysis cycle. For these figures, we use variance inflation $r = 20\%$. Comparing with Figs. 2 and 3, we see that using fewer observations increases the analysis errors significantly in the mid-latitudes but not in the tropics. We use the same observations (every 6 h) for the 24-h analysis interval

24h growth of ensemble spread (T 500hPa)

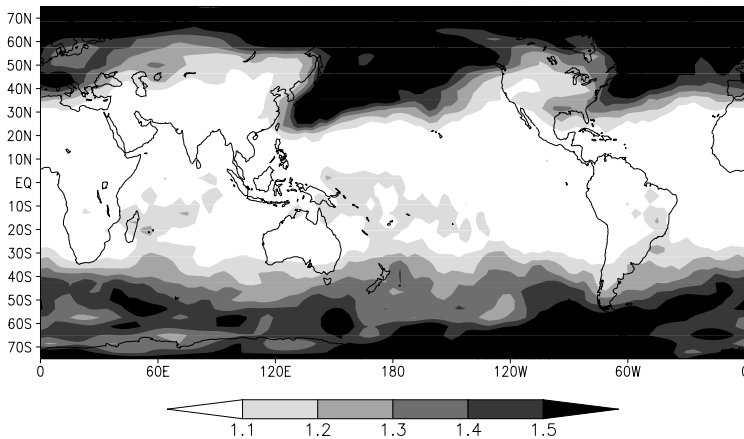## 24h growth of ensemble spread (Ps)

Fig. 13. The average 24-h growth of the ensemble spread (as defined in the text) during February 1982. The top figure shows the growth for temperature at pressure level 500 hPa. The bottom figure shows the growth for surface pressure.

results below. In Fig. 9, we see an identical representation of results as in Fig. 1, but all simulations are run with 24-h analysis cycle. We note that 4D-LETKF (solid), on average, yields a better analysis compared to FGAT-LETKF (dashes). For the 24-h analysis cycle, 3D-LETKF fails to bring the analysis errors to be at least comparable to observation errors even after cycling through 2 months of assimilations (results are not shown). Figures 10 and 11 show the average analysis errors of temperature at 500 hPa and surface pressure with 24-h analysis cycle during February 1982, respectively. These numerical results clearly suggest advantages of 4D-LETKF over FGAT-LETKF when the analysis time is longer.

To understand further the circumstances that cause 4D-LETKF and FGAT-LETKF to differ, consider (for the sake of exposition) a scenario where observations are available at a single time $t_l < t_n$. Then eq. (14) for the 4D-LETKF analysis mean becomes (substituting from eq. 16)

$$\bar{\mathbf{x}}_n^a = \bar{\mathbf{x}}_n^b + \mathbf{X}_n^b \big[(k-1)\mathbf{I}/(1+r) + \big(\mathbf{Y}_l^b\big)^T \mathbf{R}_l^{-1} \mathbf{Y}_l^b \big]^{-1} \big(\mathbf{Y}_l^b\big)^T \mathbf{R}_l^{-1} \big[\mathbf{y}_l^o - H_l(\bar{\mathbf{x}}_l^b)\big].$$

If in addition H$_l$ is linear, H$_l(\mathbf{x}) = \mathbf{H}_l \mathbf{x}$, then $\mathbf{Y}_l^b = \mathbf{H}_l \mathbf{X}_l^b$, and using the matrix identity

$$[c\mathbf{I} + \mathbf{Y}^T \mathbf{R}^{-1} \mathbf{Y}]^{-1} \mathbf{Y}^T \mathbf{R}^{-1} = \mathbf{Y}^T [c\mathbf{R} + \mathbf{Y}\mathbf{Y}^T]^{-1}$$

we can write

$$\bar{\mathbf{x}}_n^a = \bar{\mathbf{x}}_n^b + \mathbf{X}_n^b \big(\mathbf{X}_l^b\big)^T \mathbf{H}_l^T \big[(k-1)\mathbf{R}_l/(1+r)$$
$$+ \mathbf{H}_l \mathbf{X}_l^b \big(\mathbf{X}_l^b\big)^T \mathbf{H}_l^T \big]^{-1} \big(\mathbf{y}_l^o - \mathbf{H}_l \bar{\mathbf{x}}_l^b\big).$$

In FGAT-LETKF we replace $\mathbf{X}_l^b$ by $\mathbf{X}_n^b$, which affects the analysis in two ways. Changing $\mathbf{X}_n^b(\mathbf{X}_l^b)^T$ to $\mathbf{X}_n^b(\mathbf{X}_n^b)^T$ overestimates the cross-covariance between the background ensemble at the observation time t_l and at the analysis time t_n. Changing $\mathbf{X}_l^b(\mathbf{X}_l^b)^T$ to $\mathbf{X}_n^b(\mathbf{X}_n^b)^T$ ignores the change in the spread of the background ensemble between times t_l and t_n. Below we find in fact that the amount of change in ensemble spread over the analysis time window is a reasonable predictor of the size of the difference between the 4D-LETKF and FGAT-LETKF analysis error.

Figures 12 and 13 show the geographical distribution of respectively the 12- and 24-h growth of the ensemble spread. We compute this growth factor as the ratio of the average amplitude

of the ensemble perturbations $\mathbf{X}_n^b$ at analysis time t_n and the average amplitude of $\mathbf{X}_0^b$ at the start t_0 of each analysis window, where each (rms) average is taken over February 1982. This averaged ratio simply reflects the growth in forecast uncertainties. For temperature field, the largest growth occurs mostly in the tropics after 12 h (Fig. 12). After 24 h, however, the largest growth is in the extratropics (see Fig. 13). In both figures, we see the largest growth for the surface pressure in similar regions as for temperature, but the amount of growth is much higher in the extratropics. These higher growth factors, especially for the temperature, are geographically correlated with those regions with larger analysis errors produced by FGAT-LETKF (compare Fig. 12 with Figs. 2 and 3 for the 12-h analysis cycle or Fig. 13 with Figs. 10 and 11 for 24-h analysis cycle). This result suggests that the regions where the FGAT-LETKF analysis is worst compared to 4D-LETKF are those regions with the most rapid growth in the ensemble spread.

4. Summary

In this paper, we describe a four-dimensional ensemble Kalman filter. In our derivations, we show how our four-dimensional filter differs from related approaches (Bishop et al., 2001; Hunt et al., 2004; Ott et al., 2004; Wang et al., 2004). In addition to the mathematical formulation, we also prescribe a pseudo-algorithm for practical implementation.

In Section 3, we showed some results, simulated on a relatively low resolution global weather model (SPEEDY model). In our simulations, we compared three different approaches for handling asynchronous observations with LETKF. We conclude that 4D-LETKF is in general a better approach than 3D-LETKF and FGAT-LETKF, though with a short enough analysis time interval (12 h in this case), FGAT-LETKF and 4D-LETKF yield comparable results. In this scenario, we also found that our results are not sensitive to the size of local regions used; however, the larger the local regions is used, the more variance inflation is needed. Even with a relatively small local region, we do not find any problem with the balance of the analysis fields produced by 4D-LETKF.

As we increase the analysis interval to 24 h, we see the advantage of 4D-LETKF over FGAT-LETKF. We conclude that this is due to the fact that FGAT-LETKF ignores changes both in size of the background covariance over the analysis time interval and in the correlation between the background ensembles at the analysis time and the observation times, while 4D-LETKF accounts for these changes. In particular, we found that 4D-LETKF outperformed FGAT-LETKF significantly in the regions where the ensemble spread grows the fastest.

In other scenarios, we expect one would see similar results; 3D-LETKF as a very crude approximation to 4D-LETKF, while FGAT-LETKF should provide a much better approximation. How long the analysis time must be to see a sig-

nificant difference between 4D-LETKF and FGAT-LETKF will depend on the model and the time distribution of the observations. Within the LETKF framework, both methods have similar computational complexity, though FGAT-LETKF requires applying the observation operator only to the ensemble mean at the intermediate observation times whereas 4D-LETKF requires applying it to each of the ensemble members.

5. Acknowledgments

6. Appendix A: Variational derivation of 4D-EnKF

In this section, we show that 4D-LETKF and 4D-EnkF (Hunt et al., 2004) are equivalent for linear observation operators H_l, which is assumed to be the case in the latter paper. Consider $H_l(\mathbf{x}) = \mathbf{H}_l\mathbf{x}$ where $\mathbf{H}_l$ is an $s_l \times m$ matrix. In this case, 4D-LETKF is a more efficient way to compute the same analysis as 4D-EnKF. To see this, let us derive 4D-EnKF like in Section 2.1. The fundamental difference between these two schemes is that in 4D-EnKF, (8) is replaced by:

$$\mathbf{x}_n = \bar{\mathbf{x}}_n^b + \mathbf{U}\mathbf{v}, \tag{A1}$$

where $\mathbf{U}$ is an $m \times (k-1)$ matrix whose columns are the eigenvectors of $\mathbf{P}^b$ corresponding to nonzero eigenvalues. In matrix form

$$\mathbf{P}^b = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T \tag{A2}$$

where the diagonal component of the diagonal matrix $\boldsymbol{\Sigma}$ is the eigenvalue of $\mathbf{P}^b$ corresponding to eigenvector $\mathbf{u}_i$ (the ith column of $\mathbf{U}$). Note that unlike $\mathbf{w} \in \mathbb{R}^k$ in (8), $\mathbf{v} \in \mathbb{R}^{k-1}$.

The model state in the observation space, $\mathbf{H}_n\mathbf{x}_n$, at time t_n is:

$$\mathbf{H}_n\mathbf{x}_n = \mathbf{H}_n\big(\bar{\mathbf{x}}_n^b + \mathbf{U}\mathbf{v}\big) = \mathbf{H}_n\bar{\mathbf{x}}_n^b + \hat{\mathbf{H}}_n^b\mathbf{v}, \tag{A3}$$

where $\hat{\mathbf{H}}_n^b = \mathbf{H}_n\mathbf{U}$ is an $s_n \times (k-1)$ matrix. At time t_l, $l = 1, \ldots, n-1$, we need to represent model state $\mathbf{x}_l$ as a function of state $\mathbf{x}_n$ so that the asynchronous observation operator $\mathbf{H}_l\mathbf{x}_l$ can be approximated in the coordinate system defined by (A1).

Recall that in each analysis, we readjust the background ensemble state $\{\mathbf{x}_n^{b(i)}, i = 1, \ldots, k\}$ with eqs (13)–(15) to produce analysis ensemble state $\{\mathbf{x}_n^{a(i)}, i = 1, \ldots, k\}$, for which each analysis ensemble member lies in the space spanned by the

background ensemble state. Thus, we can write a prospective model state $\mathbf{x}_n$ at time t_n as

$$\mathbf{x}_n = \mathbf{X}_n \hat{\mathbf{w}}, \tag{A4}$$

where column vectors of $\mathbf{X}_n$ are the background ensemble states at time t_n and $\hat{\mathbf{w}}$ is a $k \times 1$ weight vector. Note that unlike $\mathbf{w}$ of Section 2.1, which was multiplied by the ensemble perturbations from their mean, $\hat{\mathbf{w}}$ is multiplied by ensemble states themselves. At the observation time t_l, we associate to $\mathbf{x}_n$ the model state

$$\mathbf{x}_l = \mathbf{X}_l \hat{\mathbf{w}}, \tag{A5}$$

where column vectors of $\mathbf{X}_l$ denote the ensemble background states at time t_l.

The orthogonal projection onto the space spanned by the column vectors of $\mathbf{X}_n$ is given by the matrix $\mathbf{X}_n (\mathbf{X}_n{}^T \mathbf{X}_n)^{-1} (\mathbf{X}_n)^T$. A vector $\mathbf{x}_n$ in this space is equal to its projection, so (A4) is satisfied with

$$\hat{\mathbf{w}} = \left(\mathbf{X}_n{}^T \mathbf{X}_n \right)^{-1} (\mathbf{X}_n)^T \mathbf{x}_n. \tag{A6}$$

Therefore, observations $\mathbf{H}_l \, \mathbf{x}_l$ at time t_l can be described as a function of $\mathbf{x}_n$ via the matrix $\mathbf{H}'_l = \mathbf{H}_l \, \mathbf{X}_l (\mathbf{X}_n^T \, \mathbf{X}_n)^{-1} \, \mathbf{X}_n^T$:

$$\begin{aligned} \mathbf{H}'_l \mathbf{x}_n &= \mathbf{H}_l \mathbf{X}_l \left(\mathbf{X}_n{}^T \mathbf{X}_n \right)^{-1} \mathbf{X}_n^T \mathbf{x}_n \\ &= \mathbf{H}_l \mathbf{X}_l \hat{\mathbf{w}} \\ &= \mathbf{H}_l \mathbf{x}_l. \end{aligned} \tag{A7}$$

Then following (A3), we have

$$\mathbf{H}_l \mathbf{x}_l = \mathbf{H}'_l \mathbf{x}_n = \mathbf{H}'_l \left(\bar{\mathbf{x}}_n^b + \mathbf{U}\mathbf{v} \right) = \mathbf{H}'_l \bar{\mathbf{x}}_n^b + \hat{\mathbf{H}}_l^b \mathbf{v}, \tag{A8}$$

with $\hat{\mathbf{H}}_l^b = \mathbf{H}'_l \mathbf{U}$.

Substituting (A1), (A2) and (A8) into cost function (4) yields

$$\begin{aligned} J(\mathbf{v}) = \frac{1}{2} \mathbf{v}^T \Sigma^{-1} \mathbf{v} + \frac{1}{2} \sum_{l=1}^n \left(\mathbf{y}_l^o - \mathbf{H}'_l \bar{\mathbf{x}}_n^b - \hat{\mathbf{H}}_l^b \mathbf{v} \right)^T \\ \times \mathbf{R}_l^{-1} \left(\mathbf{y}_l^o - \mathbf{H}'_l \bar{\mathbf{x}}_n^b - \hat{\mathbf{H}}_l^b \mathbf{v} \right). \end{aligned} \tag{A9}$$

Here, term $\mathbf{H}'_l \bar{\mathbf{x}}_n^b$ is calculated with (A7); note that $\mathbf{H}'_n = \mathbf{H}_n$ from (A7). The minimum of (A9) is given by

$$\mathbf{v}^a = \hat{\mathbf{P}}^a \sum_{l=1}^n \left(\hat{\mathbf{H}}_l^b \right)^T \mathbf{R}_l^{-1} \left(\mathbf{y}_l^o - \mathbf{H}'_l \bar{\mathbf{x}}_n^b \right) \tag{A10}$$

with

$$\hat{\mathbf{P}}^a = \left[\Sigma^{-1} + \sum_{l=1}^n \left(\hat{\mathbf{H}}_l^b \right)^T \mathbf{R}_l^{-1} \hat{\mathbf{H}}_l^b \right]^{-1}.$$

The analysis is completed by evaluating

$$\mathbf{x}_n^{a(i)} = \bar{\mathbf{x}}_n^b + \mathbf{U} \left(\mathbf{v}^a + \mathbf{V}^{a(i)} \right), \tag{A11}$$

where $\mathbf{V}^{a(i)}$ is the ith column vector of matrix $\mathbf{V}^a$ that satisfies $\mathbf{V}^a (\mathbf{V}^a)^T = \hat{\mathbf{P}}^a$. However, as pointed out by Ott et al. (2004), ensemble generation via the symmetric square root of $\hat{\mathbf{P}}^a$ may create a discontinuity between two adjacent analyses when local analysis is applied. To eliminate this issue, Ott et al. (2004) and

Hunt et al. (2004) generate ensemble perturbations in a similar manner to (15), and replace (A11) by

$$\begin{aligned} \mathbf{x}_n^{a(i)} &= \left(\bar{\mathbf{x}}_n^b + \mathbf{U}\mathbf{v}^a \right) + \mathbf{X}_n^b \hat{\mathbf{W}}^{a(i)} \\ &= \bar{\mathbf{x}}_n^a + \mathbf{X}_n^b \hat{\mathbf{W}}^{a(i)} \end{aligned} \tag{A12}$$

where $\hat{\mathbf{W}}^{a(i)}$ is the ith column of

$$\hat{\mathbf{W}}^a = \left[\mathbf{I} + (k-1)^{-1} \left(\mathbf{X}_n^b \right)^T \mathbf{U} \Sigma^{-1} (\hat{\mathbf{P}}^a - \Sigma) \Sigma^{-1} \mathbf{U}^T \mathbf{X}_n^b \right]^{1/2}.$$

One can show that $\hat{\mathbf{W}}^a$ is equal to $\mathbf{W}^a$ used in (15) in the case that the observation operators $\mathbf{H}_l$ are linear. In fact, one can also show that for both methods, the analysis error covariance $\mathbf{P}^a$ in the model state can be recovered by

$$\mathbf{X}_n^b \tilde{\mathbf{P}}^a \left(\mathbf{X}_n^b \right)^T = \mathbf{P}^a = \mathbf{U}\hat{\mathbf{P}}^a \mathbf{U}^T, \tag{A13}$$

with the first equality for 4D-LETKF and the second equality for 4D-EnKF.

Note that from the projection onto space spanned by the column vectors of $\mathbf{X}_n$ (see eqs A5 and A6) and the definitions of $\mathbf{X}_n^b$ in eq. (6) and $\mathbf{X}_l^b$ in eq. (7), we have identity

$$\mathbf{X}_l^b = \mathbf{X}_l \left(\mathbf{X}_n{}^T \mathbf{X}_n \right)^{-1} (\mathbf{X}_n)^T \mathbf{X}_n^b. \tag{A14}$$

Using (12), (A10), (A13) and (A14), we deduce

$$\begin{aligned} \mathbf{X}_n^b \mathbf{w}^a &= \mathbf{X}_n^b \tilde{\mathbf{P}}^a \sum_{l=1}^n \left(\mathbf{X}_l^b \right)^T \mathbf{H}_l \mathbf{R}_l^{-1} \left(\mathbf{y}_l^o - \mathbf{H}_l \bar{\mathbf{x}}_l^b \right) \\ &= \left[\mathbf{X}_n^b \tilde{\mathbf{P}}^a \left(\mathbf{X}_n^b \right)^T \right] \mathbf{X}_n \left(\mathbf{X}_n{}^T \mathbf{X}_n \right)^{-1} \\ &\quad \times \sum_{l=1}^n \mathbf{X}_n^T \mathbf{H}_l^T \mathbf{R}_l^{-1} \left(\mathbf{y}_l^o - \mathbf{H}'_l \bar{\mathbf{x}}_n^b \right) \\ &= \mathbf{P}^a \sum_{l=1}^n \left(\mathbf{H}'_l \right)^T \mathbf{R}_l^{-1} \left(\mathbf{y}_l^o - \mathbf{H}'_l \bar{\mathbf{x}}_n^b \right) \\ &= \mathbf{U}\hat{\mathbf{P}}^a \mathbf{U}^T \sum_{l=1}^n \left(\mathbf{H}'_l \right)^T \mathbf{R}_l^{-1} \left(\mathbf{y}_l^o - \mathbf{H}'_l \bar{\mathbf{x}}_n^b \right) \\ &= \mathbf{U}\mathbf{v}^a. \end{aligned} \tag{A15}$$

That is, the analysis ensemble mean $\bar{\mathbf{x}}_n^a$ is equivalent in both schemes, 4D-EnKF and 4D-LETKF. Moreover, the ensemble perturbations are exactly similar since $\hat{\mathbf{W}}^{a(i)} = \mathbf{W}^{a(i)}$. Though mathematically they are equivalent, 4D-LETKF is computationally more efficient since the analysis needs no calculations of the eigenvectors $\mathbf{u}_i$. When $n = 1$, this derivation reflects the similarities between LETKF and LEKF (Ott et al., 2004).

7. Appendix B: Pseudo-algorithm of 4D-LETKF

We now give a step-by-step description of how to implement the analysis described in the Section 2.1 together with the localization and variance inflation. Hereafter, we refer to the state at time t_l with a subscript l, where $l = 1, \ldots, n$. The inputs to the steps below are the m-dimensional vectors $\{\mathbf{x}_l^{b(i)} : i = 1, 2, \ldots, k\}$, a non-linear operator $\mathbf{H}_l$ from m variables to s_l variables, an s_l-dimensional vector $\mathbf{y}_l^o$ and an $s_l \times s_l$ matrix $\mathbf{R}_l$.

(i) Form $\{\mathbf{x}_l^{b(i)}\}$ into an $m \times k$ matrix $\mathbf{X}_l$, average the columns of $\mathbf{X}_l$ to get an m-dimensional vector $\bar{\mathbf{x}}_l^b$ and subtract this vector from each column of $\mathbf{X}_l$ to get $\mathbf{X}_l^b$. Apply H_l to each column of $\mathbf{X}_l$ to form an $s_l \times k$ matrix $\mathbf{Y}_l$, average columns of $\mathbf{Y}_l$ to get the s_l-dimensional vector $\bar{\mathbf{y}}_l^b$, and subtract this vector from each column of $\mathbf{Y}_l$ to get $\mathbf{Y}_l^b$. Apply H_l to $\bar{\mathbf{x}}_l^b$ and to get an s_l-dimensional vector $H_l(\bar{\mathbf{x}}_l^b)$.

(ii) From this point, perform the analysis locally at each model grid point j. For each j, choose the observations to be used in the local analysis and denote them $\mathbf{y}_l^o(j)$ (In this paper, we chose the observations from within a two-dimensional local region of size 3×3 grid points centered at grid point j). Denote the components of $\mathbf{Y}_l^b$, $H_l(\bar{\mathbf{x}}_l^b)$ and $\mathbf{R}_l$ corresponding to the local observations for grid point j by $\mathbf{Y}_l^b(j)$, $H_l(\bar{\mathbf{x}}_l^b)(j)$ and $\mathbf{R}_l(j)$, respectively.

(iii) Form the $k \times k$ matrix

$$(k-1)\mathbf{I}/(1+r) + \sum_{l=1}^{n} \left[\mathbf{Y}_l^b(j)\right]^T \mathbf{R}_l(j)\mathbf{Y}_l^b(j), \tag{B1}$$

where r is the desired amount of multiplicative variance inflation. Take the inverse of (B1) to get $\tilde{\mathbf{P}}^a(j)$ and calculate

$$\mathbf{w}^a(j) = \tilde{\mathbf{P}}^a(j) \left\{ \sum_{l=1}^{n} \left[\mathbf{Y}_l^b(j)\right]^T \mathbf{R}_l(j)^{-1}\left[\mathbf{y}_l^o(j) - H_l\left(\bar{\mathbf{x}}_l^b\right)(j)\right]\right\}. \tag{B2}$$

In this case, we ignore the correlation between observation errors at different times t_l. Generally, if one knows this correlation, then one can form matrix $\mathbf{R}(j)$ with diagonal blocks $\mathbf{R}_l(j)$. The correlation coefficients are subsequently assigned in the appropriate non-block diagonal components. Here, the second term in (B1) is replaced with $[\mathbf{Y}^b(j)]^T \mathbf{R}(j)\mathbf{Y}^b(j)$ where $\mathbf{Y}^b(j)$ be the matrix formed by stacking the matrices $\mathbf{Y}_l^b(j)$ vertically. Equation (B2) is replaced with

$$\mathbf{w}^a(j) = \tilde{\mathbf{P}}^a(j)[\mathbf{Y}^b(j)]^T \mathbf{R}(j)^{-1}[\mathbf{y}^o(j) - H(\bar{\mathbf{x}}^b)(j)], \tag{B3}$$

where $\mathbf{y}^o(j)$ and $H(\bar{\mathbf{x}}^b)(j)$ are obtained by (vertically) concatenating vectors $\mathbf{y}_l^o(j)$ and $H_l(\bar{\mathbf{x}}_l^b)(j)$, respectively. In this paper, since we assume zero correlation between observation errors at different time, we use (B1) and (B2).

(iv) Compute the $k \times k$ matrix $\mathbf{W}^a(j) = [(k-1)\tilde{\mathbf{P}}^a(j)]^{1/2}$. Add $\mathbf{w}^a(j)$ to each column of $\mathbf{W}^a(j)$, forming the $k \times k$ matrix $\mathbf{W}(j)$. The analysis ensemble state $\{\mathbf{x}_n^{a(i)}(j)\}$ at grid point j is obtained by adding the background mean state $\bar{\mathbf{x}}_n^b(j)$ at grid point j to the ith column of $\mathbf{X}_n^b(j)\,\mathbf{W}(j)$, where $\mathbf{X}_n^b(j)$ consists of the rows of $\mathbf{X}_n^b$ with components corresponding to variables at grid point j.

References

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

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

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

Buehner, M. 2005. Ensemble-derived stationary and flow-dependent background error covariances: evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Met. Soc.* **131**, 1013–1043.

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

Courtier, P., Thépaut, J.-N. and Hollingsworth, A. 1994. A strategy for operational implementation of 4D-VAR, using an incremental approach. *Quart. J. Roy. Met. Soc.* **120**, 1367–1387.

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D. and co-authors. 1998. The ECMWF implementation of three-dimensional variational assimilation. *Quart. J. Roy. Met. Soc.* **124**, 1783–1807.

Le Dimet, F.-X. and Talagrand, O. 1986. Variational algorithm for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* **38A**, 97–110.

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

Fertig, E. J., Harlim, J. and Hunt, B. R. 2007. A comparative study of 4D-VAR and a four-dimensional ensemble Kalman filter: perfect model simulations with Lorenz-96. *Tellus* **59A**, 96–100.

Fisher, M. and Courtier, P. 1995. Estimating the covariance matrix of analysis and forecast error in variational data assimilation. *ECMWF Tech. Memo.* **220**, 5pp.

Gustafsson, N., Berre, L., Hörnquist, S., Huang, X.-Y., Lindskog, M. and co-authors. 2001. Three-dimensional variational data assimilation for a limited area model. Part I: general formulation and the background error constraint. *Tellus* **53A**, 425–446.

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

Harlim, J., 2006. Errors in the initial conditions for numerical weather prediction: a study of error growth patterns and error reduction with ensemble filtering, *PhD dissertation*, University of Maryland.

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

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

Huang, X.-Y., Morgensen, K. S., and Yang, X. 2002. First-guess at the appropriate time: the HIRLAM implementation and experiments. In *Proceedings, HIRLAM workshop on variational data assimilation and remote sensing*, Finland, 28–43.

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

Hunt, B. R., Kostelich, E. J. and Szunyogh, I. 2007. Efficient data assimilation for a spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D*, **230**, 112–126.

Keppenne, C. 2000. Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.* **128**, 1971–1981.

Lindskog, M., Gustafsson, N., Navascuès, B., Mogensen, K. S., Huang, X.-Y. and co-authors. 2001. Three-dimensional variational data assimilation for a limited area model. Part II: observation handling and data assimilation experiments. *Tellus* **53A**, 447–468.

Lorenc, A. C. 2003. The potential of the ensemble Kalman filter for NWP - a comparison with 4D-VAR. *Quart. J. Roy. Met. Soc.* **129**, 3183–3203.

Lorenz, E .N., 1996. Predictability—a problem partly solved. In: *Proceedings on Predictability,* held at ECMWF on 4–8 September 1995, 1–18.

Molteni, F. 2003. Atmospheric simulations using a GCM with simplified physical parametrizations, I: model climatology and variability in multi-decadal experiments. *Clim. Dyn.* **20**, 175–191.

Ott, E., Hunt, B. R., 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.

Rabier, F., Thépaut, J.-N. and Courtier, P. 1998. Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quart. J. Roy. Met. Soc.* **124**, 1–39.

Rabier, F., Järvinen, H., Mahfouf, J.-F. and Simmons, A. 2000. The ECMWF operational implementation of four-dimensional variational assimilation: experimental results with simplified physics. *Quart. J. Roy. Met. Soc.* **126**, 1143–1170.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. 2003. Ensemble square-roots filters. *Mon. Wea. Rev.* **131**, 1485–1490.

Wang, X., Bishop, C. H. and Julier, S. J. 2004. Which Is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble?. *Mon. Wea. Rev.* **132**, 1590–1605.

Wang, X., Snyder, C. and Hamill, T. M. 2007. On the theoretical equivalence of differently proposed ensemble-3DVAR hybrid analysis schemes. *Mon. Wea. Rev.* **135**, 222–227.

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

Zupanski, M. 2005. Maximum likelihood ensemble filter: theoretical aspects. *Mon. Wea. Rev.* **133**, 1710–1726.