



Accounting for Model Errors in Ensemble Data Assimilation

HONG LI

Laboratory of Typhoon Forecast Technique, Shanghai Typhoon Institute of CMA, Shanghai, China

EUGENIA KALNAY AND TAKEMASA MIYOSHI

Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland

CHRISTOPHER M. DANFORTH

*Department of Mathematics and Statistics, Complex Systems Center, Vermont Advanced Computing Center,
University of Vermont, Burlington, Vermont*

(Manuscript received 26 August 2008, in final form 14 March 2009)

ABSTRACT

This study addresses the issue of model errors with the ensemble Kalman filter. Observations generated from the NCEP–NCAR reanalysis fields are assimilated into a low-resolution AGCM. Without an effort to account for model errors, the performance of the local ensemble transform Kalman filter (LETKF) is seriously degraded when compared with the perfect-model scenario. Several methods to account for model errors, including model bias and system noise, are investigated. The results suggest that the two pure bias removal methods considered [Dee and Da Silva (DdSM) and low dimensional (LDM)] are not able to beat the multiplicative or additive inflation schemes used to account for the effects of total model errors. In contrast, when the bias removal methods are augmented by additive noise representing random errors (DdSM+ and LDM+), they outperform the pure inflation schemes. Of these augmented methods, the LDM+, where the constant bias, diurnal bias, and state-dependent errors are estimated from a large sample of 6-h forecast errors, gives the best results. The advantage of the LDM+ over other methods is larger in data-sparse regions than in data-dense regions.

1. Introduction

After more than 10 years of research, variants of the ensemble Kalman filter (EnKF) proposed by Evensen (1994) are now becoming viable candidates for the next generation of data assimilation in operational NWP. The advance is primarily due to the fact that 1) they include a flow-dependent background error covariance; 2) they are easy to code and implement; and 3) they automatically generate an optimal ensemble of analysis states to initialize ensemble forecasts. Many studies to date have tested EnKF systems under the perfect-model assumption with simulated observations, and only within the last few years have there been tests of the EnKF assimilating real atmospheric observations (e.g.,

Houtekamer et al. 2005; Whitaker et al. 2008; Szunyogh et al. 2008; Torn and Hakim 2008; Meng and Zhang 2008). Notably, an EnKF has been operational since 2005 in the Canadian Meteorological Centre (Houtekamer and Mitchell 2005) and when using the same model and observations, the EnKF scores are about the same as the scores with a four-dimensional variational data assimilation (4DVAR) system (see online at http://4dvarenkf.cima.fcen.uba.ar/Download/Session_7/Intercomparison_4D-Var_EnKF_Buehner.pdf).

Numerical weather forecast errors grow as a result of errors in the initial conditions and model deficiencies. Model errors can result from approximate parameterizations of physical processes, the failure to represent sub-grid-scale events, and numerical discretization, etc. When preparing a data assimilation scheme for the inclusion of real observations, assuming that the model is perfect is overly optimistic (Dee 1995). In this case, the “true” forecast error covariance should include uncertainties

Corresponding author address: Hong Li, Shanghai Typhoon Institute, 166 Puxi Rd., Shanghai 200030, China.
E-mail: lih@mail.typhoon.gov.cn

from both inaccurate initial condition and model errors. As a result, if the background error covariance in the EnKF attributes errors only to the initial conditions, it will be smaller than the true forecast error covariance. To address this problem, approaches include inflating the background error covariance with multiplicative inflation (Anderson and Anderson 1999), additive inflation (Mitchell and Houtekamer 2000; Mitchell et al. 2002; Hamill and Whitaker 2005; Corazza et al. 2007; Whitaker et al. 2008) and the covariance relaxation method (Zhang et al. 2004). Though these techniques are easy to implement, they account for model errors in the second moment of the ensemble, not the mean. Other methods have been proposed to deal with the mean of model errors. Fujita et al. (2007) and Meng and Zhang (2007, 2008) reported that using different physical parameterization schemes resulted in both better background error covariance and mean estimates in their mesoscale EnKF experiments. Aksoy et al. (2006a,b) proposed a simultaneous state and parameter estimation method to estimate uncertain model parameters in an EnKF environment. Dee and da Silva (1998) proposed a method for the online estimation and correction of model bias. This bias correction method [i.e., the Dee and da Silva bias estimation method (DdSM)] has been successfully tested, for example, by Dee and Todling (2000), Carton et al. (2000), Chepurin et al. (2005) in three-dimensional variational data assimilation (3DVAR) and optimal interpolation (OI) data assimilation systems, and by Keppenne et al. (2005) in an EnKF system. Recently, Baek et al. (2006) developed another bias correction method similar to the DdSM except that it allows for “adjusting” the observations, rather than the model bias. It also accounts for the cross correlation of uncertainties in model state and bias ignored by the DdSM. They successfully tested this approach with the Lorenz-96 model. However, these two bias correction methods assume that model errors are constant in time, resulting in estimates of only the slowest varying component of model errors. In reality, model errors are likely to vary with time (e.g., errors in the diurnal cycle) or with the state of the atmosphere (e.g., biases are different during an El Niño episode).

Danforth et al. (2007) proposed an approach where the state-independent model error (bias) was estimated from the average of a large ensemble of 6-h forecast minus analysis fields (i.e., 6-h apparent forecast errors). Diurnal errors were estimated from the dominant empirical orthogonal functions (EOFs), and state-dependent errors were determined using the leading terms in a singular value decomposition (SVD). They found this low-dimensional method (LDM), where the state-dependent errors are expressed in terms of very few degrees of freedom (d.o.f.), to be very successful and computation-

ally efficient. This method has been successfully tested with the coupled Lorenz-96 model (Danforth and Kalnay 2008) and the simple but realistic global Simplified Parametrization Primitive Equation Dynamics (SPEEDY) model (Danforth et al. 2007). However, neither study addressed data assimilation; the initial conditions were taken to be a system trajectory and the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis, respectively. Here we expand the application of the LDM to more realistic situations where the forecast–analysis is cycled and, as a result, forecast error includes both model error and dynamical growing error due to the imperfect initial condition.

In this study, we investigate the performance of an ensemble Kalman filter in a perfect model and in the presence of significant model errors, and then compare approaches for dealing with model errors in ensemble data assimilation. The local ensemble transform Kalman filter (LETKF; Hunt et al. 2007) is used as a representative of other EnKF systems. A review of the LETKF is given in section 2. The different techniques for treating model errors are described in section 3. Section 4 reports the performance of the LETKF in a perfect-model scenario. In section 5, the LETKF performance in the presence of model errors due to assimilating observations generated from the NCEP–NCAR reanalysis fields is examined. Two inflation schemes (multiplicative and additive inflation) and two bias correction methods (DdSM and LDM) are applied to account for and/or correct model errors. Their results are compared and discussed in both a uniform and a rawinsonde-like observation network. Section 6 gives our summary and discussion.

2. LETKF data assimilation scheme

The LETKF is one of the ensemble square root filters (Tippett et al. 2003) in which the observations are assimilated to update only the ensemble mean by

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{K}[\mathbf{y}^o - H(\bar{\mathbf{x}}^f)], \quad (1)$$

where \mathbf{x}^a and \mathbf{x}^f are the ensemble mean of analysis and background (forecast) respectively; \mathbf{K} is the Kalman gain; \mathbf{y}^o is the observations; and $H()$ is the observation operator. Here we use the superindex f rather than b to denote the background (forecasts) in order to avoid confusion with the subindex b used to denote the bias in the next section.

The ensemble perturbations are updated by transforming the background perturbations through a transform matrix \mathbf{T} , as proposed by Bishop et al. (2001):

$$\mathbf{X}^a = \mathbf{X}^f \mathbf{T}, \quad (2)$$

where \mathbf{X}^a and \mathbf{X}^f are the analysis and background ensemble perturbations (matrices whose columns are the difference between the ensemble members and the ensemble mean), and the error covariances are given by $\mathbf{P}^{a,f} = \mathbf{X}^{a,f}(\mathbf{X}^{a,f})^T$.

In the LETKF, the Kalman gain and transform matrix are given by

$$\mathbf{K} = \mathbf{X}^f \tilde{\mathbf{P}}^a (\mathbf{H}\mathbf{X}^f)^T \mathbf{R}^{-1} \quad \text{and} \quad (3)$$

$$\mathbf{T} = [(\mathbf{K} - 1)\tilde{\mathbf{P}}^a]^{1/2}, \quad (4)$$

where $\tilde{\mathbf{P}}^a$, the analysis error covariance in ensemble space, is given by

$$\tilde{\mathbf{P}}^a = [(\mathbf{K} - 1)\mathbf{I} + (\mathbf{H}\mathbf{X}^f)^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{X}^f)]^{-1}, \quad (5)$$

with dimension $K \times K$, where the ensemble size K is usually much smaller than both the dimension of the model and the number of observations. As a result, the LETKF performs the analysis in the space spanned by the forecast ensemble members, which greatly reduces the computational cost. The LETKF is computed locally for each grid point by choosing the observations that will influence that grid point. More details about the LETKF are available in Hunt et al. (2007) and Szunyogh et al. (2008).

3. Methods to deal with model errors

a. Multiplicative inflation

Multiplicative inflation simply inflates the ensemble error covariance \mathbf{P}_e^f by a factor $1 + \Delta$ to approximate the true error covariance \mathbf{P}^f :

$$\mathbf{P}^f \leftarrow (1 + \Delta)\mathbf{P}_e^f, \quad (6)$$

where Δ is a tunable parameter. Equation (6) provides for an increase in the ensemble covariance \mathbf{P}_e^f to account for the model errors not included in the original \mathbf{P}_e^f . This method implicitly assumes that model errors have the same error structure as the internal errors so that their error covariance is proportional to the dynamically evolved error covariance \mathbf{P}_e^f .

b. Additive inflation

Additive inflation parameterizes model errors by adding random perturbations with a certain covariance structure to each ensemble member. Following Whitaker et al. (2008), in this study we randomly select samples from a subset of NCEP–NCAR reanalysis (NNR; Kalnay et al. 1996) 6-h tendency fields in January and February for the years 1982–86. Unlike random numbers, these

randomly selected tendency fields are geostrophically balanced. In each analysis cycle, we randomly select K tendency fields, remove their mean, scale these zero-mean fields (we tune the scaling factor), and add each of these scaled fields to one background ensemble member:

$$\mathbf{x}_k^f = \mathbf{x}_{e(k)}^f + r\mathbf{q}_k, \quad (7)$$

so that

$$\bar{\mathbf{q}}_k = 0 \quad (8)$$

and denote

$$\mathbf{q}_k \mathbf{q}_k^T = \mathbf{Q}. \quad (9)$$

Here k is the index for each ensemble member, $\mathbf{x}_{e(k)}^f$ denotes the k th ensemble forecast, \mathbf{q}_k is the field added to ensemble member k , and r is the globally uniform scaling factor, a tunable parameter. Our procedure is similar to that of Whitaker et al. (2008), ensuring that the added fields will only enlarge the background error covariance by \mathbf{Q} and will not change the ensemble mean.

c. DdSM

Dee and da Silva (1998) developed a two-stage bias estimation algorithm, in which the estimation procedures for the bias and the state are carried out successively. At the first step bias is estimated on every model grid point by

$$\mathbf{b}^a = \mathbf{b}^f - \mathbf{K}_b [\mathbf{y}^o - H(\mathbf{x}^f - \mathbf{b}^f)] \quad \text{and} \quad (10)$$

$$\mathbf{K}_b = \mathbf{P}_{bb}^f \mathbf{H}^T (\mathbf{H}\mathbf{P}_{bb}^f \mathbf{H}^T + \mathbf{H}\mathbf{P}_{xx}^f \mathbf{H}^T + \mathbf{R})^{-1}, \quad (11)$$

where the matrix \mathbf{P}_{xx}^f and \mathbf{P}_{bb}^f is the forecast error covariance for the state variables and for the bias, respectively.

In practice the bias forecast error covariance \mathbf{P}_{bb}^f is unknown, so that following Dee and da Silva (1998) we assume that

$$\mathbf{P}_{bb}^f = \alpha \mathbf{P}_{xx}^f. \quad (12)$$

Substituting (12) into (11), we have

$$\mathbf{K}_b = \alpha \mathbf{P}_{xx}^f \mathbf{H}^T [(1 + \alpha)\mathbf{H}\mathbf{P}_{xx}^f \mathbf{H}^T + \mathbf{R}]^{-1}, \quad (13)$$

where α is a tunable parameter.

In the second step, the analysis for the state variables is obtained using the standard analysis procedure with the unbiased forecast state $\mathbf{x}^f - \mathbf{b}^a$:

$$\mathbf{x}^a = (\mathbf{x}^f - \mathbf{b}^a) + \mathbf{K}_x [\mathbf{y}^o - H(\mathbf{x}^f - \mathbf{b}^a)] \quad \text{and} \quad (14)$$

$$\mathbf{K}_x = \mathbf{P}_{xx}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}_{xx}^f \mathbf{H}^T + \mathbf{R})^{-1}. \quad (15)$$

As for the bias forecast model, following Carton et al. (2000) we use damped persistence:

$$\mathbf{b}_i^f = \mu \mathbf{b}_{i-1}^a, \quad (16)$$

where i denotes time step, and $\mu < 1$ is a tunable parameter.

The cost of the DdSM is about twice that of no bias estimation, since the updated equations are solved twice, first for the bias estimation and then for the state variables. However, this doubling of the cost can be avoided if $\alpha \ll 1$ in which case Eq. (13) becomes

$$\mathbf{K}_b \approx \alpha \mathbf{P}_{xx}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}_{xx}^f \mathbf{H}^T + \mathbf{R})^{-1} = \alpha \mathbf{K}_x. \quad (17)$$

Reversing the order of the bias estimation step and that of the state analysis step, we obtain a simplified version of the DdSM (Radakovich et al. 2001):

$$\mathbf{x}^a = (\mathbf{x}^f - \mathbf{b}^f) + \mathbf{K}_x [\mathbf{y}^o - H(\mathbf{x}^f - \mathbf{b}^f)] \quad \text{and} \quad (18)$$

$$\mathbf{b}^a = \mathbf{b}^f - \alpha \mathbf{K}_x [\mathbf{y}^o - H(\mathbf{x}^f - \mathbf{b}^f)]. \quad (19)$$

In this approach the computation of Eq. (19) is almost cost free after the state analysis \mathbf{x}^a has been updated by Eq. (18), since $\mathbf{K}_x [\mathbf{y}^o - H(\mathbf{x}^f - \mathbf{b}^f)]$ is simply the analysis increment for the state variables.

Above we give the general equations for the DdSM and its simplified version. In the application of the DdSM to an EnKF system, no additional ensemble members are required for the bias since the bias forecast error covariance \mathbf{P}_{bb}^f is obtained directly from the state forecast error covariance \mathbf{P}_{xx}^f . Therefore, the term \mathbf{x}^f in Eqs. (10) and (19) is actually the ensemble mean forecast.

d. LDM

We assume the NNR as a reference field to approximate the true atmosphere and create a reference 6-h model forecast initialized from the NNR without data assimilation. The (apparent) 6-h forecast error \mathbf{x}^e is then defined as the difference between the reference forecast and the NNR valid at the same time.

The low-dimensional scheme (Danforth et al. 2007) assumes that model error \mathbf{x}_m^e has three components, namely the bias, periodic errors dominated by errors in the diurnal cycle, and model errors that are state-dependent:

$$\mathbf{x}_m^e(t) = \mathbf{b} + \sum_{l=1}^L \beta_l(t) \mathbf{e}_l + \sum_{n=1}^N \gamma_n(t) \mathbf{f}_n, \quad (20)$$

in which t denotes the time step. The forecast bias \mathbf{b} is obtained by averaging the forecast errors \mathbf{x}^e over a certain training time period $\mathbf{b} = \langle \mathbf{x}^e \rangle$, and the leading EOFs \mathbf{e}_l from the anomalous error field $\mathbf{x}^{e'} = \mathbf{x}^e - \langle \mathbf{x}^e \rangle$ are used to estimate diurnal or other periodic errors. The state-dependent model error component is given by the leading SVD modes \mathbf{f}_n of the covariance of the coupled model state anomalies $\mathbf{x}^{f'} = \mathbf{x}^f - \langle \mathbf{x}^f \rangle$ and corresponding error anomalies $\mathbf{x}^{e'}$. Here L and N are the number of retained leading modes of EOFs and SVDs, respectively. The spatial fields \mathbf{b} , \mathbf{e}_l , and \mathbf{f}_n are time independent and computed offline using the training period samples. We call this approach low-dimensional because the spatial shape of the model errors is preestimated separately and only the amplitudes $\beta_l(t)$ and $\gamma_n(t)$, which have a much lower dimension (L and N) than the full model dimension, are estimated online.

During the training period, the time series of β_l is calculated by projecting $\mathbf{x}^{e'}$ onto the EOFs \mathbf{e}_l . Since β_l is dominated by the bias in the diurnal cycle, the time-dependent $\beta_l(t)$ can be estimated by averaging the β_l over the diurnal cycle in the training period. For example, we can use the samples in the training period to calculate the average of β_l for 0000, 0600, 1200, and 1800 UTC separately and time interpolate them in the current time step. For the calculation of the state-dependent component $\sum_{n=1}^N \gamma_n(t) \mathbf{f}_n$ using SVD, the readers are referred to Danforth et al. (2007).

4. The LETKF performance in perfect-model experiments

The AGCM used in this study, SPEEDY (Molteni 2003), solves the primitive equations for prognostic variables of zonal wind (u), meridional wind (v), temperature (T), specific humidity (q), and surface pressure (p_s) at triangular truncation T30, corresponding to 96×48 grid points at each of the 7 sigma levels. First, we assume a perfect model where a nature run is generated by integrating the SPEEDY model from 0000 UTC 1 January 1987 to 1800 UTC 15 February 1987. The observations are simulated by adding normally distributed random noise to the nature run, and are available at each model grid point for p_s and every second model grid point for u, v, T, q , in both the zonal and meridional directions (i.e., 25% of the number of model grid points). The amplitudes of the observation errors are 1 m s^{-1} for u, v wind, 1 K for T , $10^{-4} \text{ kg kg}^{-1}$ for q , and 1 hPa for p_s . For each forecast–analysis cycle, the SPEEDY model is used to generate the 6-h forecast and the observations are assimilated by the LETKF with 30 ensemble members and a $\Delta = 0.05$ multiplicative

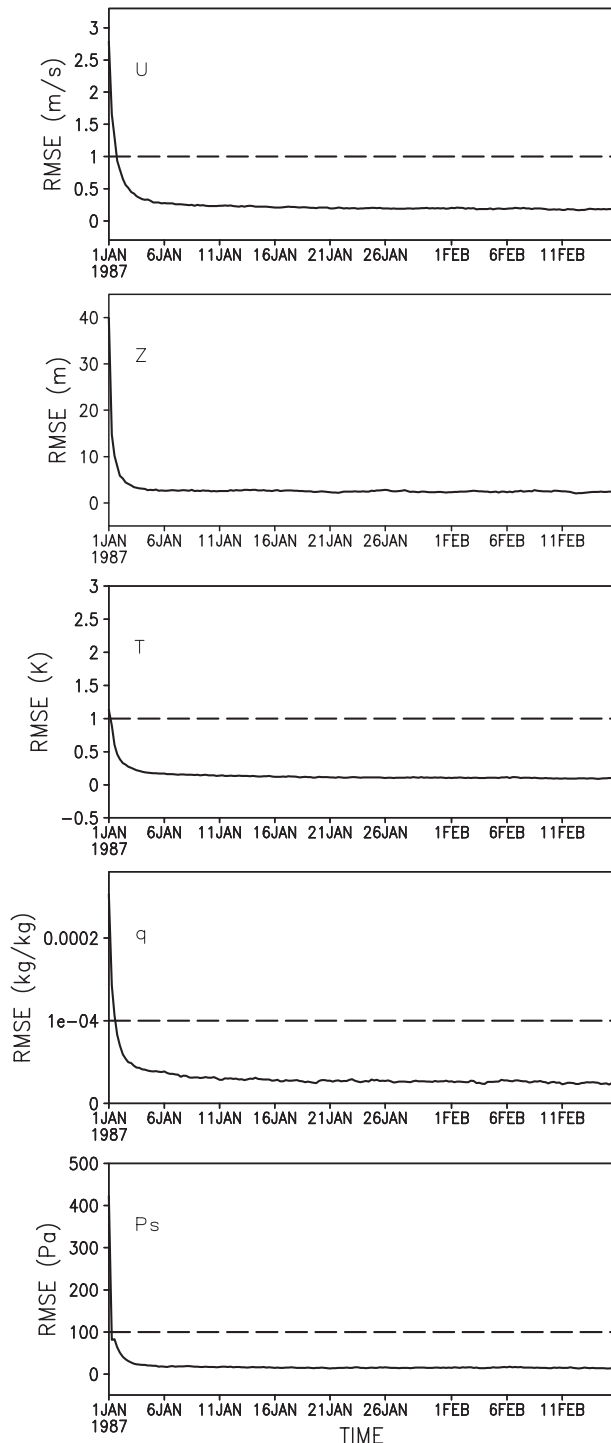


FIG. 1. Time series of global-averaged analysis RMSE (solid curve) for the period between 0000 UTC 1 Jan 1987 and 1800 UTC 15 Feb 1987 in a perfect-model experiment. The observational error standard deviations are shown as dashed lines wherever applicable. (from top to bottom) Zonal wind, geopotential height, temperature, specific humidity at 500 hPa, and surface pressure, respectively.

inflation parameter to deal with sample errors in the covariance. Figure 1 shows the time series of analysis root-mean-square error (RMSE) with respect to the nature run, averaged over the whole globe, for zonal wind (u), geopotential height (Z), temperature (T), specific humidity (q) at 500 hPa, and surface pressure (p_s). It is clear that after the initial spinup period, the analysis RMSE for all the variables is much smaller than the observational error standard deviations.

To investigate why the LETKF performs well, the background ensemble spread is compared with the background error. Carrying out perfect-model experiments, the true background error with respect to the nature run can be calculated. In order for the LETKF to perform well, the ensemble spread should be representative of the true background error. Figure 2 shows the 500-hPa height background ensemble mean error field (shaded) and the ensemble spread (contour) at 1200 UTC 3 February 1987, an arbitrarily chosen time. Although noisy, these two fields generally agree with each other in location. The background RMSE of the height field and the corresponding spread averaged over the whole globe are then compared in Fig. 3. The spread is slightly smaller at low levels and slightly larger in the upper levels compared to the background RMSE, but they are still quite close. Figures 2 and 3 together demonstrate that the spread among 30 ensemble members has captured the true background error well in both structure and magnitude.

5. Accounting for and correcting model errors in the LETKF

To assess the performance of the LETKF in the presence of model errors, we replace the nature run in the perfect-model experiments by the NNR fields. Since the NNR assimilated real observations, we assume that the NNR fields are an approximate estimate of the unknown true atmosphere (a quantitative validation of this assumption is beyond the scope of this research). Random noise with the same standard deviation used in the perfect-model experiments is added to simulate “NNR observations.” The density of observations remains the same as that in the perfect-model experiments.

a. Effects of model errors on the LETKF

Serving as a benchmark, the control run is carried out using the same configuration ($\Delta = 0.05$ and 30 ensemble members) as that in section 4 but replacing the observations generated from the nature run with the NNR observations. No additional method is applied to deal

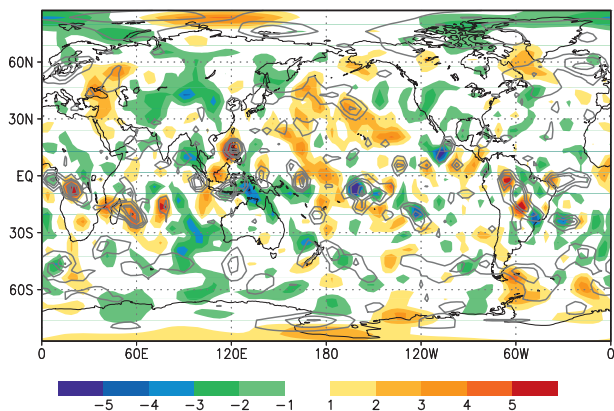


FIG. 2. The background (6-h forecast) error field (m, shaded) and the ensemble spread of 500-hPa height field (m, contour) at 1200 UTC 3 Feb 1987, an arbitrarily chosen time, in a perfect-model experiment.

with model errors. After an initial spinup period of 15 days, the analyses and forecasts are verified against the NNR.

The strong negative influence of the model errors on the performance of the LETKF is very clear in the control run (results not shown). In the presence of model errors and without accounting for them, the 500-hPa height analysis RMSE increases from 2.4 (in the perfect model) to 50 m because of the model errors and their accumulated effects. An increase by more than an order of magnitude is also observed for other variables and regions. However, it should be noted that with a more sophisticated and high-resolution numerical model, such as those currently used in operational centers, the negative impact of model errors would be much smaller.

The background ensemble spread of the height field is plotted in Fig. 4 to investigate why the LETKF performs poorly in the presence of large model errors. It is worth noting that the spread is very close to that obtained in the perfect-model experiment, while ideally it should be as large as the actual forecast error, as it happens for a perfect model. This result indicates that when using the same model, the forecast ensembles are blind to model errors, and therefore the ensemble spread underestimates the actual forecast error, leading to excessive confidence in the forecasts, less weight given to the observations, and, as a result, to the poor performance of the LETKF.

b. Accounting for and correcting model errors

The methods described in section 3 are then applied to the SPEEDY-LETKF system to account for model errors. As in the control run, experiments are implemented for the period 0000 UTC 1 January–1800 UTC

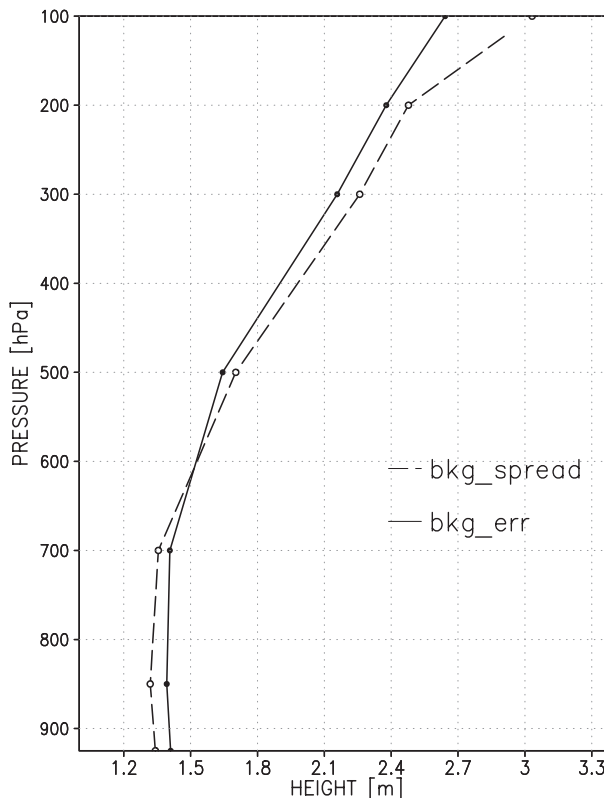


FIG. 3. Height background RMSE at all pressure levels (solid) and background ensemble spread (dashed), temporally averaged for one-month after the initial 15-day spinup period in a perfect-model experiment.

15 February 1987 and the verification statistics are computed for analyses and forecasts against the NNR fields after the initial spinup period of 15 days.

1) MULTIPLICATIVE INFLATION VERSUS ADDITIVE INFLATION

Figure 5 shows the analysis RMSE of 200-hPa u , 500-hPa Z , 850-hPa q , and 925-hPa T fields, from multiplicative inflation (with an optimal parameter of $\Delta = 1.5$), additive inflation (with an optimal amplitude of $r = 1.5$), and the control run. Both inflation schemes result in much better analyses than the control run for all the fields. As found in previous studies (e.g., Hamill and Whitaker 2005; Whitaker et al. 2008), additive inflation outperforms multiplicative inflation.

2) DEE AND DA SILVA METHOD COMBINED WITH ADDITIVE INFLATION (DdSM+)

The DdSM aims to estimate and correct model bias, but does not account for state-dependent and random errors. Since the performance of the additive inflation is better than that of the multiplicative inflation, we

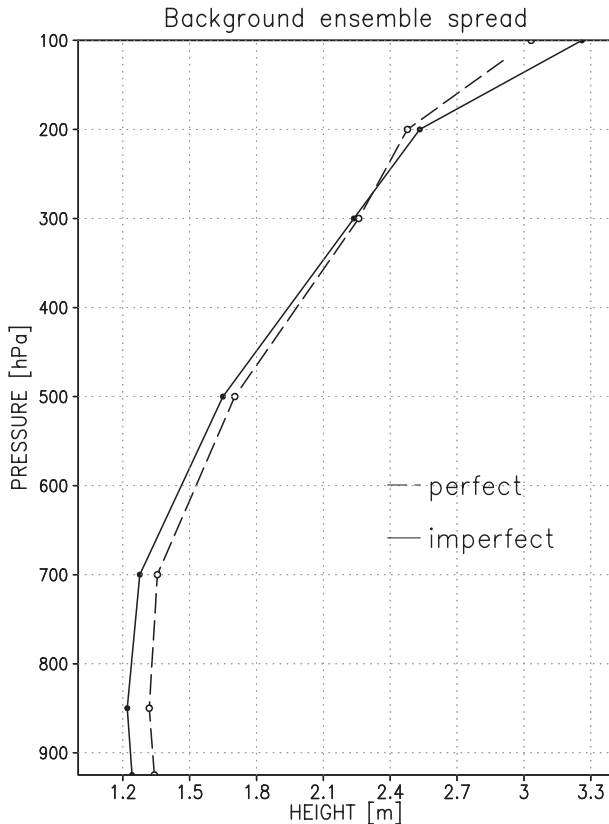


FIG. 4. Height background ensemble spread at all pressure levels temporally averaged over a month after the initial 15-day spinup period in the cases of perfect model, assimilating observations generated from the SPEEDY nature run (dashed), and imperfect model assimilating observations generated from the NNR field (solid).

combine the DdSM with additive inflation to account for system noise. The additive noise is obtained in the same manner as for the additive inflation scheme. We refer to the DdSM augmented with additive noise as DdSM+.

Recall that there are two variables (μ and α) to be tuned in the pure DdSM. If additive inflation is used to model the system noise, the amplitude (r) of additive noise is another parameter that also needs to be tuned. To simplify the task of tuning these three parameters, first we fix $\alpha = 0.5$ [following the recommendation of Dee and da Silva (1998)] and $\mu = 1.0$ (assuming a persistence model for bias prediction) and then tune the amplitude (r) of the additive noise. We start at 0000 UTC 1 January 1987 by assuming zero bias and run the SPEEDY-LETKF system. However, no matter how small r is, the filter diverges, especially for temperature fields in the lower levels, and the bigger r is, the faster the divergence. Using the same SPEEDY-LETKF system but with pure DdSM ($r = 0$), Miyoshi (2005) observed similar filter divergence for all choices of α when fixing

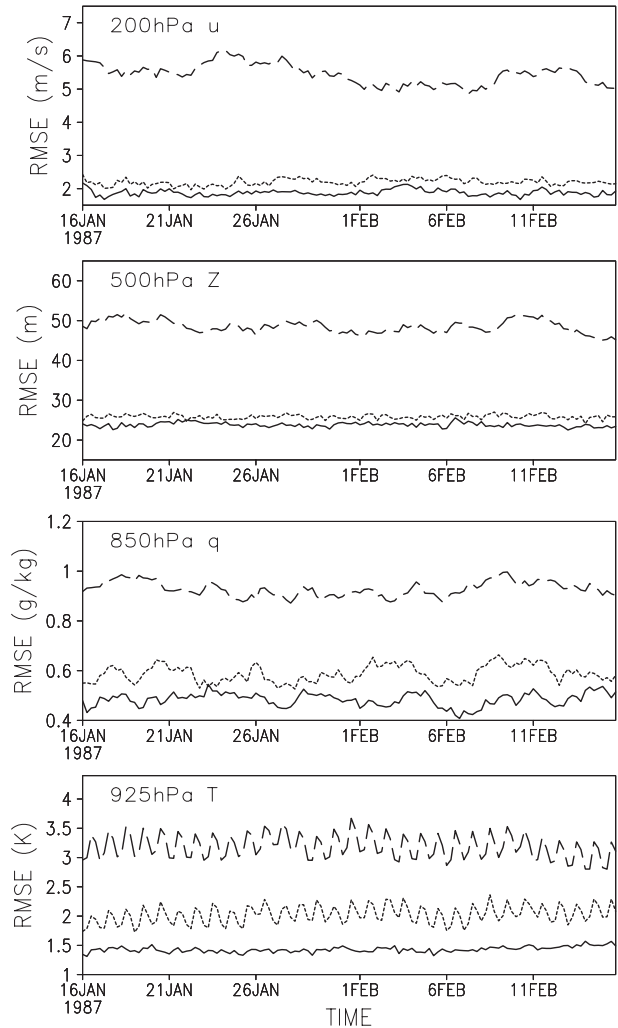


FIG. 5. Time series of the global-averaged analysis RMSE in the cases of the control run (dashed curve), $\Delta = 1.5$ multiplicative inflation (dotted curve), and $r = 1.5$ additive inflation (solid curve) assimilating the NNR observations. (from top to bottom) 200-hPa u , 500-hPa Z , 850-hPa q , and 925-hPa T fields, respectively.

$\mu = 1.0$. Thus, we let μ be less than 1 to reduce the impact of bias from the previous time step and find that $\mu = 0.9$ is successful for a wide range of choices for r and gives better results than $\mu = 0.8$. Fixing $\mu = 0.9$, the pairs of (α , r) are tuned. The results in terms of 500-hPa RMSE computed over the 1-month period after the initial 15-day spinup period are summarized in Table 1. It is clear that accounting for random system noise is essential in order for the LETKF to perform well. Without additive noise, the pure DdSM ($r = 0$) is not competitive with pure additive inflation ($\alpha = 0$) with an optimal amplitude of $r = 1.5$. Using a small additive noise ($r = 0.25$), the DdSM+ outperforms the pure additive inflation scheme, but the optimal choice of α is

TABLE 1. Analysis RMSE of 500-hPa height (m) using the DdSM+ with different choices of (α, r) with a fixed $\mu = 0.9$. When $r = 0$ (i.e., pure DdSM), a small parameter ($\Delta = 0.05$) of multiplicative inflation is applied in order to prevent filter divergence. For the other choices of r , no multiplicative inflation is used. For comparison, the pure additive inflation application is also shown ($\alpha = 0$ and $r = 1.5$).

| α | 0.0 | 0.25 | 0.50 | 0.75 | 1.00 |
|------------|------|------|------|------|------|
| $r = 0$ | | | 37.1 | 35.0 | 33.9 |
| $r = 0.25$ | | 22.5 | 22.0 | 18.9 | 19.2 |
| $r = 0.5$ | | 19.8 | 17.6 | 20.4 | 20.1 |
| $r = 0.6$ | | | 17.3 | | |
| $r = 1.5$ | 23.8 | | | | |

large ($\alpha = 0.75$). When r is increased to 0.5, the value of the optimal α reduces to 0.5. These results can be better understood by the expression of $\mathbf{P}_{bb}^f = \alpha \mathbf{P}_{xx}^f$ where α is an explicit parameter and r is an implicit factor (through affecting \mathbf{P}_{xx}^f) to determine the bias forecast error covariance \mathbf{P}_{bb}^f . When r is small, the system requires a large value of α to obtain an optimal \mathbf{P}_{bb}^f , while as r increases, the optimal value of α decreases because the forecast error covariance \mathbf{P}_{xx}^f for the state variables has already been increased. By increasing r from 0 to 0.5, a large improvement is found. Beyond $r = 0.5$, there is little further improvement. Therefore, we choose ($\mu = 0.9, \alpha = 0.5, r = 0.6$) as the optimal setting of the parameters for the DdSM+.

As for the simplified version of DdSM [Eqs. (18) and (19)], we did similar tuning experiments for parameters μ, α , and r , the optimal result (not shown) is worse than that from the optimal DdSM+.

3) LDM WITH ADDITIVE INFLATION (LDM+)

Since experiments are carried out for January and February in 1987, the 5-yr climatology of the same months for the years 1982–86 is chosen as the training period, following Danforth et al. (2007). In this training period, the 6-h SPEEDY forecasts initialized with the NNR fields are conducted; the samples of forecast error \mathbf{x}^e are then obtained by taking the differences between the SPEEDY 6-h forecasts and the NNR fields valid at the same time.

Three types of model errors in Eq. (20) with LDM are corrected but, as was the case with the pure DdSM, the pure LDM without accounting for system noise was unable to beat pure additive inflation (Fig. 6). To parameterize system noise, randomly selected NNR 6-h tendency fields are added to each background ensemble member and their amplitude is tuned as was done in the DdSM+ scheme. As seen in Fig. 6, the LDM, plus a small amount ($r = 0.4$) of additive noise (hereinafter LDM+)

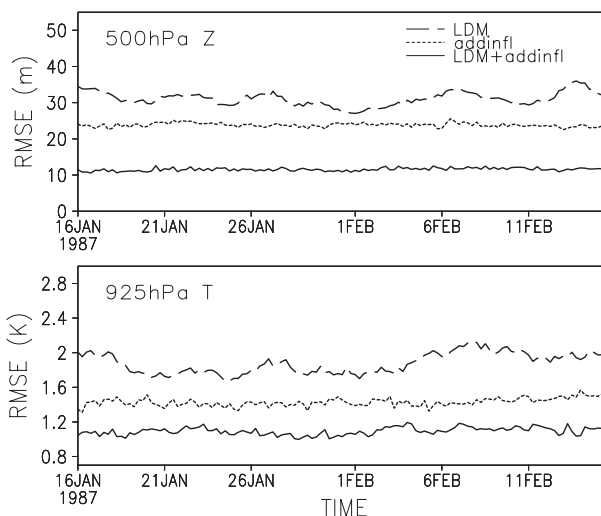


FIG. 6. Time series of the global-averaged analysis RMSE of the 500-hPa Z and 925-hPa T fields, in the cases of the LDM alone (dashed curve), $r = 1.5$ additive inflation (dotted curve), and the LDM together with additive inflation with an amplitude $r = 0.4$ (solid curve).

significantly outperforms the pure additive inflation scheme, suggesting the necessity to deal with both model bias and system noise in the presence of complicated model errors.

4) OVERALL COMPARISON

Finally all the methods with their optimal configurations are compared with each other. As before, the results are verified against the NNR fields.

(i) Analysis verification

Figure 7 shows the analysis RMSE comparison. Among all methods, the LDM+ gives the best results: it outperforms other methods in all the fields throughout all pressure levels, especially at lower levels. The DdSM+ generally outperforms both pure inflation schemes. Though the pure multiplicative inflation scheme produces the worst results of the four methods, it should be noted that all the methods have made major analysis improvements compared to the control run (gray dotted curve in Fig. 8). These results indicate that accounting for model errors is essential for the EnKF and correcting model biases is, in general, better than only accounting for their effects in the second moment of the ensemble, assuming we have a good method for estimating them.

(ii) 48-h forecast verification

So far we have focused on the comparisons in terms of the analysis accuracy. However, the goal of developing

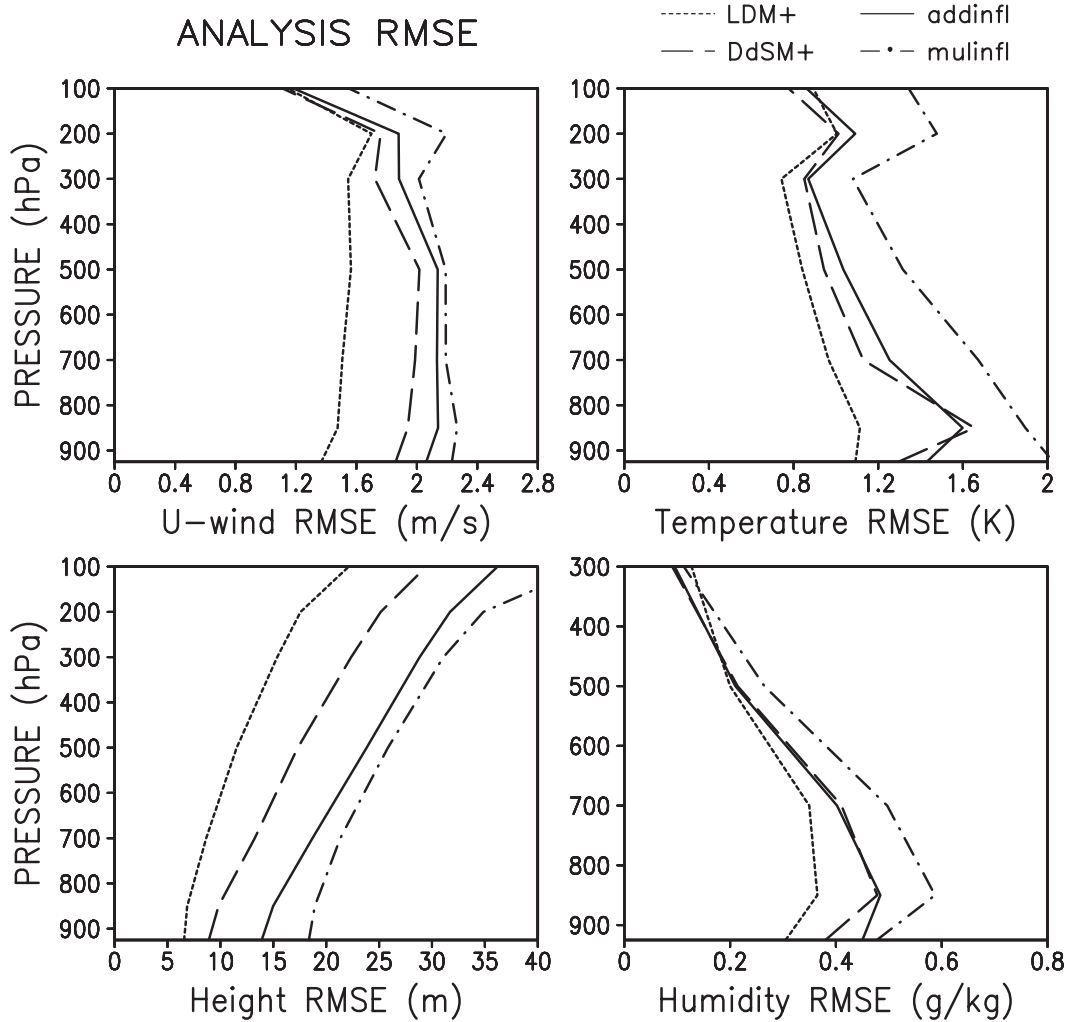


FIG. 7. Global-averaged analysis RMSE at all pressure levels in the cases of the LDM+ (dotted), the DdSM+ (dashed), additive inflation (solid), and multiplicative inflation (dotted-dashed). (top left) The u -wind field, (top right) temperature field, (bottom left) height field, and (bottom right) specific humidity field. The averages are taken for a month after the initial half-month spinup period.

more accurate analyses is to improve the short-term forecasts. Within an imperfect model, the short-term forecast errors come from both growing errors in the initial condition and model deficiencies. Although the model errors could be corrected within the forecast model, here we would like to see if the advantage of one method in the data assimilation process can be retained over a forecast period. Otherwise, there would be no benefit in improving the initial analysis on short-term forecasts. Figure 9 shows the global-averaged 48-h forecast RMSE at all pressure levels. The advantage of DdSM+ over additive inflation becomes less obvious for most fields, but remains significant for geopotential height fields at all levels. The large advantage of the LDM+ over the other two methods also decreases due to the con-

tamination of the model errors. Nevertheless, it is still quite obvious and significant, except for the zonal wind above 200 hPa and the humidity above 700 hPa.

Above we focused on the impact of initial analysis on the short-term forecast and did not attempt to correct the model errors during the forecast process. The LDM can also be used to estimate and correct the short-term model errors in the forecast phase, as shown by Danforth et al. (2007) with the SPEEDY model.

c. Sensitivity to observational network

Thus far, different methods have been compared using a globally uniform observation network (the upper-right panel in Fig. 10), as if they were based only on

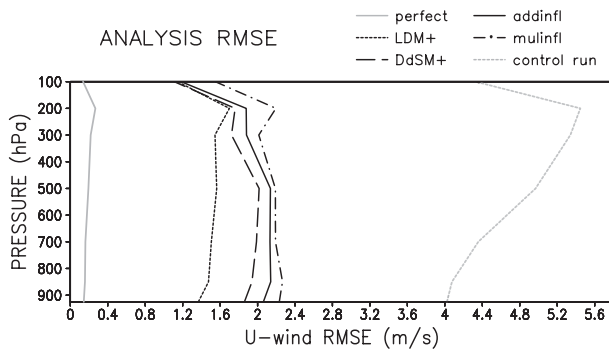


FIG. 8. As in top-left in Fig. 7, but also shows the result for the control run (gray dotted) and for the perfect-model experiment (gray solid). We note that in a more realistic operational model, the negative effect of model errors would not be as large as in this control run and a significant amount of inflation leads to better results even in the absence of bias correction (e.g., Szunyogh et al. 2008).

satellite observations, resulting in uniform error fields (not shown). However, in reality there are more rawinsonde observations over land and fewer over the ocean. To investigate the sensitivity of each method to the choice of observational system, we assimilate a rawinsonde-like network (the upper-left panel in Fig. 10) for u, v, T, q (p_s is still available everywhere) using pure additive inflation, the DdSM+, and the LDM+. We retune the parameters and choose the optimal setting for each method. Figure 10 (lower-left panel) shows the zonal and time-averaged latitudinal profiles of geopotential height analysis RMSE at 500 hPa for the three methods with a rawinsonde-like network. For comparison, the results from experiments described in section 5b with a uniform observation network are shown in the lower-right panel. With uniform observations, the performance of each method is less latitude dependent. We note that the sawtooth behavior of the error observed in the DdSM+ and pure additive inflation methods is due to the fact that observations are available every other latitude. With rawinsonde-like observations, though the DdSM+ is still better than pure additive inflation, both are far more sensitive to observation density than LDM+. In the Southern Hemisphere and the northern polar region where few upper-air observations are available, the DdSM+ and pure additive inflation behave significantly worse than that of LDM+. Figure 10 demonstrates that the DdSM+ and pure additive inflation schemes are more sensitive to the choice of observation network, and perform poorly in regions with sparse upper-air observations, while the LDM+ is more robust. This was confirmed in an additional intercomparison in which surface pressure was only observed at rawinsonde locations. Although there was a slight further degradation

in skill, the relative performance of the methods remained similar to that in Fig. 10 (not shown), indicating again that the advantage of the LDM+ over other two methods is larger in data-sparse regions than in data-dense regions. These results are not unexpected since the LDM+ bias correction is done in model space while the DdSM+ is performed in observation space, and additive inflation accounts for model errors through an inflated background error covariance that relies on the presence of observations to affect the final analysis.

6. Summary and discussion

In this study we addressed the issue of model errors with the ensemble Kalman filter. Though focusing on the LETKF, an efficient approach within the EnKF family, the results are applicable to other EnKF systems. First, we performed data assimilation experiments using the LETKF with the SPEEDY model under a perfect-model scenario. It was found that the LETKF works very well in this ideal case. Then we relaxed the perfect-model assumption and assimilated observations generated from the NCEP–NCAR reanalysis fields. Without any additional effort to handle model errors, the performance of the LETKF was seriously degraded. The background ensemble spread was similar to that in the perfect model and therefore much smaller than the true forecast error (that now includes model errors). The “blindness” of the LETKF to model error is likely due to the fact that each ensemble member is integrated with the same model. If forecasts from different systems are available, as in the superensemble of Krishnamurti et al. (2000), we would expect to be able to at least partially represent model errors.

We investigated two simple ways to represent the effect of model errors and two methods to estimate and remove model bias. Our results suggest that multiplicative inflation is worse than additive inflation. The pure bias removal methods (DdSM and LDM) correct model bias, but cannot remove system noise; as a result, they are unable to beat inflation schemes that account for the total model errors. Supplemented by additive noise to represent the random errors, bias removal methods generally outperform the pure inflation schemes. Of all these methods, the low-dimensional method with additive inflation (LDM+) where the time-averaged model bias, diurnal bias, and state-dependent errors are estimated from a large number of 6-h forecast errors, gives the most accurate analyses and 48-h forecasts. The advantage of the LDM+ over other methods is larger in data-sparse regions than in data-dense regions due to the fact that its bias correction is done in model space and is

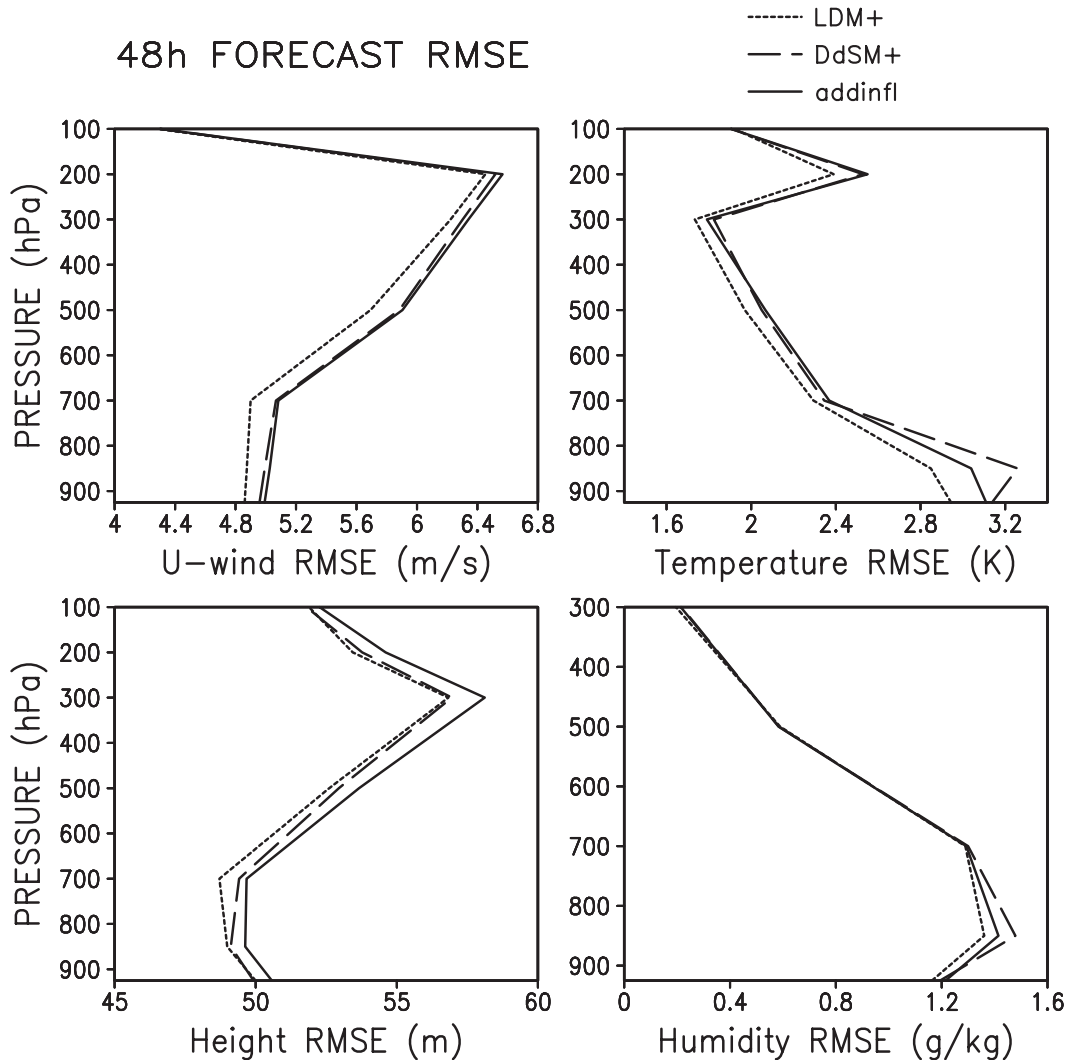


FIG. 9. 48-h forecast RMSE at all pressure levels in the cases of the LDM+ (dotted), the DdSM+ (dashed), and additive inflation (solid). (top left) The u -wind field, (top right) temperature field, (bottom left) height field, and (bottom right) specific humidity field. The averages are taken over all forecasts started every 6 h between 0000 UTC 1 Feb 1987 and 1800 UTC 15 Feb 1987.

thus less sensitive to the absence of abundant observations. We note that in training the LDM we have used a long reanalysis that has spread information from both satellite and rawinsonde observations throughout the atmosphere.

Although the DdSM combined with inflation (DdSM+) produces results not as good as the LDM+, it is generally superior to both pure inflation schemes. The disadvantages of DdSM+ are the doubling of the computational cost and exclusive reliance on observations. When the observations are sparser, the impact of the bias correction is limited. In the worst case, where the observations themselves are biased, it is not at all obvious that this algorithm can work correctly. Our results of LDM esti-

mation may be too optimistic, since in our applications we assume the NNR field is an approximation of the unknown truth and use it to generate the samples of model errors during the training period. In reality, the NNR field could be biased, and generating good samples of model errors is a challenge to the LDM+. It is not clear whether the model error samples generated from the NNR fields are good enough to represent the true model errors. In practice, we could use a more advanced reanalysis, like the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40; Uppala et al. 2005) or the Japanese 25-yr Re-Analysis (JRA-25; Onogi et al. 2007) to replace the NNR. We could also verify the analysis against a more

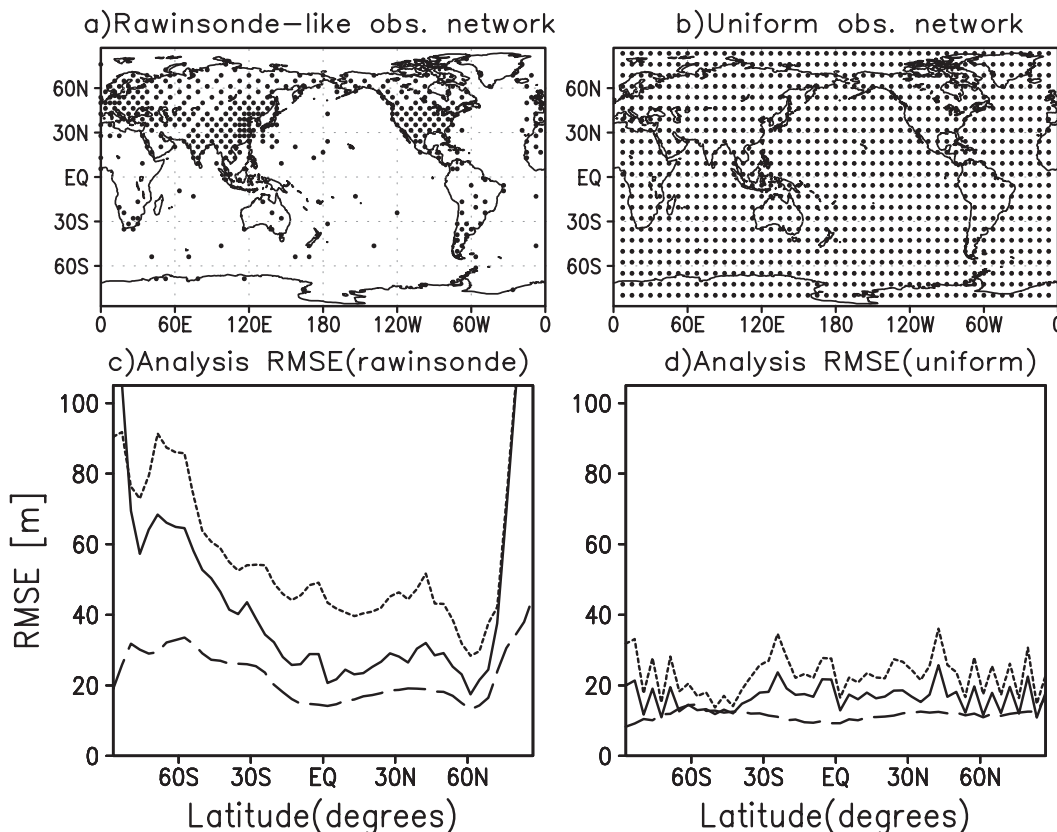


FIG. 10. (top) Upper-air observation locations in (left) rawinsonde-like and (right) uniform networks (p_s is available at each model grid point in both cases). (bottom) Zonal- and time-averaged latitudinal profiles of geopotential height analysis RMSE at 500 hPa for additive inflation (dotted), the DdSM+ (solid), and the LDM+ (dashed), in the case of (left) rawinsonde-like and (right) uniform observation networks.

advanced reanalysis, and thus simulate the fact that the truth is not available for training. Alternatively, the EnKF analysis increments could also be used as model error samples instead of using a reanalysis. For the training of error samples to derive model bias, this would imply a spinup of the bias correction since at first the analyses are also biased, so that the analysis increments cannot initially sample model error well. We intend to explore this idea with the SPEEDY model to see whether the final model error samples after convergence are good enough to represent the true model errors.

Acknowledgments. The authors are very grateful to the members of the Weather-Chaos group at the University of Maryland for helpful discussions, and to two anonymous reviewers for their constructive suggestions that helped us to improve the manuscript. This research was partially supported by 973 Program (2009CB421500), NASA Grant NNG04G29G, NOAA Grant NA04OAR4310103, and CMA Grant GYHY200806029.

REFERENCES

- Aksoy, A., F. Zhang, and J. W. Nielsen-Gammon, 2006a: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Mon. Wea. Rev.*, **134**, 2951–2970.
- , —, and —, 2006b: Ensemble-based simultaneous state and parameter estimation with MM5. *Geophys. Res. Lett.*, **33**, L12801, doi:10.1029/2006GL0261286.
- Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758.
- Baek, S.-J., B. R. Hunt, E. Kalnay, E. Ott, and I. Szunyogh, 2006: Local ensemble Kalman filtering in the presence of model bias. *Tellus*, **58A**, 293–306.
- 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.
- Carton, J. A., G. Chepurin, X. Cao, and B. Giese, 2000: A simple ocean data assimilation analysis of the global upper ocean 1950–95. Part I: Methodology. *J. Phys. Oceanogr.*, **30**, 294–309.
- Chepurin, G. A., J. A. Carton, and D. Dee, 2005: Forecast model bias correction in ocean data assimilation. *Mon. Wea. Rev.*, **133**, 1328–1342.

- Corazza, M., E. Kalnay, and S.-C. Yang, 2007: An implementation of the local ensemble Kalman filter in a quasi geostrophic model and comparison with 3D-Var. *Nonlinear Processes Geophys.*, **14**, 89–101.
- Danforth, C. M., and E. Kalnay, 2008: Using singular value decomposition to parameterize state-dependent model errors. *J. Atmos. Sci.*, **65**, 1467–1478.
- , —, and T. Miyoshi, 2007: Estimating and correcting global weather model error. *Mon. Wea. Rev.*, **135**, 281–299.
- Dee, D. P., 1995: Online estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145.
- , and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, **124**, 269–295.
- , and R. Todling, 2000: Data assimilation in the presence of forecast bias: The GEOS moisture analysis. *Mon. Wea. Rev.*, **128**, 3268–3282.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10 143–10 162.
- Fujita, T., D. J. Stensrud, and D. C. Dowell, 2007: Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Mon. Wea. Rev.*, **135**, 1846–1868.
- Hamill, T. M., and J. S. Whitaker, 2005: Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Mon. Wea. Rev.*, **133**, 3132–3147.
- Houtekamer, P. L., and H. L. Mitchell, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269–3289.
- , —, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620.
- 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.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Keppenne, C. L., M. M. Rienecker, N. P. Kurkowski, and D. A. Adamec, 2005: Ensemble Kalman filter assimilation of temperature and altimeter data with bias correction and application to seasonal prediction. *Nonlinear Processes Geophys.*, **12**, 491–503.
- Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4196–4216.
- Meng, Z., and F. Zhang, 2007: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments. *Mon. Wea. Rev.*, **135**, 1403–1423.
- , and —, 2008: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVAR in a real-data case study. *Mon. Wea. Rev.*, **136**, 522–540.
- Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416–433.
- , —, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 2791–2808.
- Miyoshi, T., 2005: Ensemble Kalman filter experiments with a primitive-equation global model. Ph.D. thesis, University of Maryland, College Park, 197 pp.
- Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments. *Climate Dyn.*, **20**, 175–191.
- Onogi, K., and Coauthors, 2007: The JRA-25 Reanalysis. *J. Meteor. Soc. Japan*, **85**, 369–432.
- Radakovich, J. D., P. R. Houser, A. M. da Silva, and M. G. Bosilovich, 2001: Results from global land-surface data assimilation methods. Preprints, *Fifth Symp. on Integrated Observing Systems*, Albuquerque, NM, Amer. Meteor. Soc., 132–134.
- Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. Yorke, 2008: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus*, **60A**, 113–130.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.
- Torn, R. D., and G. J. Hakim, 2008: Performance characteristics of a pseudo-operational ensemble Kalman filter. *Mon. Wea. Rev.*, **136**, 3947–3963.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.
- Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP global forecasting system. *Mon. Wea. Rev.*, **136**, 463–482.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253.