Ensemble Kalman Filter and 4D-Var Intercomparison with the Japanese Operational Global Analysis and Prediction System

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

YOSHIAKI SATO AND TAKASHI KADOWAKI
Numerical Prediction Division, Japan Meteorological Agency, Tokyo, Japan

(Manuscript received 31 August 2009, in final form 2 February 2010)

ABSTRACT

The local ensemble transform Kalman filter (LETKF) is implemented and assessed with the experimental operational system at the Japanese Meteorological Agency (JMA). This paper describes the details of the LETKF system and verification of deterministic forecast skill. JMA has been operating a four-dimensional variational data assimilation (4D-Var) system for global numerical weather prediction since 2005. The main purpose of this study is to make a reasonable comparison between the LETKF and the operational 4D-Var. Several forecast–analysis cycle experiments are performed to find sensitivity to the parameters of the LETKF. The difference between additive and multiplicative error covariance inflation schemes is investigated. Moreover, an adaptive bias correction method for satellite radiance observations is proposed and implemented, so that the LETKF is equipped with functionality similar to the variational bias correction used in the operational 4D-Var. Finally, the LETKF is compared with the operational 4D-Var. Although forecast verification scores of the two systems relative to each system’s own analyses and to radiosonde observations show some disagreement, the overall conclusion indicates that the LETKF and 4D-Var have essentially comparable performance. The LETKF shows larger temperature bias in the lower troposphere mainly over the ocean, which is related to a well-known JMA model bias that plays an important role in the significant degradation of the forecast scores in the SH. The LETKF suffers less of a performance degradation than 4D-Var in the absence of satellite radiance assimilation. This suggests that better treatment of satellite radiances would be important in future developments toward operational use of the LETKF. Developing both LETKF and 4D-Var at JMA has shown significant benefits by the synergistic effect and is the recommended strategy for the moment.

1. Introduction

The ensemble Kalman filter (EnKF), first proposed by Evensen (1994), is now a feasible choice for use with operational numerical weather prediction (NWP). The Canadian Meteorological Centre (CMC) started to use an EnKF method with perturbed observations as an operational ensemble prediction system (EPS) in January 2005 (Houtekamer and Mitchell 1998, 2001, 2005; Houtekamer et al. 2005). In the summer of 2005, the Met Office started to use the ensemble transform Kalman filter (ETKF; Bishop et al. 2001), which does not use perturbed observations, as the operational EPS (Bowler et al. 2008); the system is the Met Office Global and Regional Ensemble Prediction System (MOGREPS). EnKF assimilates observations to generate ensemble initial conditions. It has two goals: 1) to obtain an accurate ensemble mean analysis and 2) to obtain ensemble perturbations well representing the analysis uncertainty. Canadian EPS uses EnKF for both aspects, whereas MOGREPS uses EnKF only to generate ensemble perturbations; the ensemble mean is provided by a separate data assimilation system. The current mainstream procedure in operational NWP is to use a variational data assimilation method to obtain an accurate analysis. Separately, bred vectors (BVs) or singular vectors (SVs) are computed to obtain initial ensemble perturbations. It is important to investigate if
EnKF provides as accurate an analysis as variational methods, as well as to investigate if EnKF initial perturbations provide better ensemble prediction than BV or SV. This study addresses only the former aspect by verifying deterministic forecast skills.

The EnKF methods that do not use perturbed observations are categorized as ensemble square root filters (EnSRF; Anderson 2001; Whitaker and Hamill 2002; Tippett et al. 2003). The Japanese Meteorological Agency (JMA) started the development of the local ensemble transform Kalman filter (LETKF; Hunt et al. 2007), a kind of EnSRF, in the summer of 2005. LETKF is an efficient upgrade of the local ensemble Kalman filter (LEKF; Ott et al. 2004). Szunyogh et al. (2005) applied the LEKF to the National Centers for Environmental Prediction (NCEP) Global Forecasting System (GFS) and found it performs well in a perfect model scenario. Whitaker et al. (2008) applied the LETKF to the NCEP GFS at a reduced resolution and found it outperforms the three-dimensional variational data assimilation (3D-Var) using real observations. Independently, Szunyogh et al. (2008) also applied the LETKF to the NCEP GFS and assimilated real observations successfully, with the exception of satellite radiances. J. Aravequia and I. Szunyogh (2009, personal communication) upgraded this system to include the assimilation of satellite radiances. As for Japanese developments, Miyoshi and Yamane (2007) reported the successful application of the LETKF to an atmospheric general circulation model (AGCM) for the Earth Simulator (AFES; Ohfuchi et al. 2004). Miyoshi and Aranami (2006) implemented the LETKF with JMA’s operational nonhydrostatic mesoscale model and assimilated precipitation observations in a perfect model scenario. Miyoshi and Sato (2007) applied the LETKF to JMA’s operational global model and reported successful assimilation of real observations including satellite radiances. More recently, Kondo and Tanaka (2009) reported successful application of the LETKF to the Nonhydrostatic Icosahedral Atmospheric Model (NICAM; Satoh et al. 2008). In this study several upgrades are made to the LETKF system of Miyoshi and Sato (2007), and the performance is compared with the currently operational four-dimensional variational data assimilation (4D-Var) system at JMA (JMA 2007).

4D-Var is the mainstream of current operational NWP, while the EnKF has emerged more recently as a viable alternative. Conceptual comparison between 4D-Var and the EnKF has been presented by Lorenc (2003) and Kalnay et al. (2007a). Gustafsson (2007) pointed out, and Kalnay et al. (2007b) agreed, that full-scale tests of the EnKF must be the responsibility of operational NWP centers in order to make more rigorous the discussion on “4D-Var or EnKF?” a recently controversial question about the future of NWP systems. At CMC, both 4D-Var and EnKF with perturbed observations are in operation. Buehner et al. (2010a,b) have made intercomparisons through reasonable modifications to the operational CMC systems, such as resolution changes, so that both systems are executed in essentially identical environments. The present study is another independent attempt to compare 4D-Var and EnKF at an operational NWP center, but without any modification to the operational 4D-Var system, and using the LETKF, which does not use perturbed observations. We design the LETKF system to be comparable to the operational 4D-Var. Section 2 describes the LETKF system, and experimental settings are described in section 3. Results are presented in section 4. The summary and discussion are provided in section 5.

2. System description

Based on research by Miyoshi and Yamane (2007) on the LETKF with the Earth Simulator global model called AFES, Miyoshi and Sato (2007) upgraded the FORTRAN90 codes of LEKF by Miyoshi (2005) to parallel LETKF and applied it to JMA’s global model. In this study, the following major upgrades are made to the system:

1) The method of error covariance localization is modified so that the error covariance is localized purely by geographical distances (Miyoshi et al. 2007b).
2) An additive covariance inflation method as in Whitaker et al. (2008) is applied, but using the Japanese 25-year reanalysis (JRA-25; Onogi et al. 2007) instead of the NCEP—National Center for Atmospheric Research (NCAR) reanalysis (Kalnay et al. 1996).
3) An efficient parallel algorithm is developed.
4) An adaptive bias correction method for satellite radiance observations is proposed and implemented.

In this section, the quasi-operational experimental system is described first, followed by detailed descriptions of the LETKF system including the error covariance localization and inflation, parallel algorithm, and adaptive bias correction.

a. Quasi-operational experimental system

Miyoshi and Sato (2007) developed a quasi-operational experimental system of the LETKF with the JMA global model, based on the operational 4D-Var system with incremental formulation (JMA 2007, as of March 2007). The forecast–analysis cycle is constructed as shown in Fig. 1. A 9-h forecast is performed to generate the first guess. Observing time is rounded to the nearest hour, so that all observations are assimilated as if they were made.
on the hour every hour. Observations on hours from 3- to 9-h forecast times, 7 slots total, are assimilated to compute and output the analysis valid at the 6-h forecast time. In 4D-Var, the number of observing slots is reduced to six in order to save computational time, but this does not imply a reduced number of observations; the details are described by JMA (2007).

Figure 2 compares data flows and job sequences of 4D-Var and LETKF. Decoded observation data used for this study are exactly identical to the ones used in operations. The observation dataset includes satellite radiances, surface reports, radiosondes, wind profilers, aircraft reports, atmospheric motion vectors (AMVs), the Quick Scatterometer (QuikSCAT) sea winds, dropsondes, and manually generated sea level pressure data by the Australian Bureau of Meteorology. If typhoons exist, bogus observation data are generated and assimilated around the typhoons. Satellite radiance data include the Advanced Microwave Sounding Unit-A (AMSU-A) channels of the National Oceanic and Atmospheric Administration (NOAA)-15/16 and Aqua satellites, AMSU-B channels of the NOAA-15/16/17 satellites, SSM/I channels of the Defense Meteorological Satellite Program (DMSP)-13/14/15 satellites, Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) channels of the TRMM satellite, and AMSR-E channels of the Aqua satellite. The details of the channels used for each satellite/sensor are described by Miyoshi and Sato (2007). The quality control (QC) system is essentially identical between 4D-Var and LETKF, except for the difference of the first-guess resolutions: T₄₃₁₉/₄₀ (semi-Lagrangian spectral truncation at 319 wavenumbers and 40 vertical layers, similarly hereafter) is used for the operational 4D-Var, whereas T₄₁₅₉/₄₀ is used for LETKF. Figure 3 shows 40 model levels of sigma-pressure hybrid vertical coordinates plotted in pressure coordinates. The resolution difference is favorable for 4D-Var since the higher resolution first guess would have less representativeness error and therefore better fit the observations. In spite of the resolution difference of the first guess, the tangent-linear and adjoint model (TLM/ADJ) resolution (i.e., the inner resolution) for 4D-Var is Eulerian T₁₀₆/₄₀, which has the same representation grid of 320 × 160 × 40 as semi-Lagrangian T₁₅₉/₄₀. Thus, there is no substantial difference in the resolution of the analysis increments between the LETKF and 4D-Var. Although the nonlinear outer model has a semi-Lagrangian dynamical core, the TLM/ADJ is based on an Eulerian dynamical core of an older version of the JMA global model. The physical processes were rebuilt from scratch for TLM/ADJ, thus these are essentially different from those used in the nonlinear outer model. Out of 70 minimizing iterations of 4D-Var, only a simplified boundary layer parameterization is included in the first 35 iterations. Then,
large-scale condensation, simplified convective parameterization, longwave radiation, and gravity wave drag are activated in the latter 35 iterations. There is no outer-loop update, thus the background field for the TLM/ADJ remains identical in the entire 70 iterations. The QC process corrects bias for radiosonde observations and satellite radiances, rejects erroneous observation data, thins observations appropriately for assimilation, and converts the file format. Although the LETKF first guess is composed of ensemble members, only the ensemble mean state is used for QC. In this way, the LETKF ensemble analysis system is seamlessly embedded into the JMA operational experimental system called the Numerical Analysis and Prediction Experiment system (NAPEX) first developed by K. Onogi.

b. Error covariance localization

Observation localization is efficiently applied within the LETKF (Hunt et al. 2007). Specifically, the inverse of the localization function is multiplied by the observation errors so that observations located farther from the analyzed grid point have less weight. For the horizontal localization, the localization function is given by the Gaussian function of the distance between the analyzed grid point and observation location. To avoid the infinitely long tails of the Gaussian function, the localization function is forced to be zero (i.e., no observations are assimilated beyond a certain distance). This critical distance is chosen to be identical to that of the fifth-order piecewise rational function of Gaspari and Cohn (1999), a commonly used localization function in EnKF studies (cf. Hamill et al. 2001). Although Miyoshi and Yamane (2007) measured distances in grid coordinates, geographical distances are used in this study as the computation became efficient (Miyoshi et al. 2007b). The globally constant localization length scale, defined by the distance at which the Gaussian function drops to $e^{-0.5}$, is the tuning parameter.

The vertical localization is performed similarly for nonradiance observations, where the vertical distance is measured in ln$p$ coordinates. Since radiance observations have a continuous observing sensitivity in the vertical, the Gaussian localization may not be appropriate. Therefore, following Miyoshi and Sato (2007), the vertical sensitivity function itself is used as the localization function for satellite radiances, after it has been normalized so that the largest value is equal to 1. Namely, the satellite radiance observations are assimilated essentially at all vertical levels using the localization weighting function given by the generally flow-dependent observing sensitivity function.

c. Error covariance inflation

Two kinds of covariance inflation methods are implemented: the multiplicative inflation and additive inflation methods. Miyoshi and Yamane (2007) as well as Miyoshi and Sato (2007) tested only the multiplicative inflation, with which the forecast or first-guess ensemble spread is inflated by a factor slightly greater than unity. In this study, the inflation parameters are defined separately in the NH ($20^\circ$–$90^\circ$N) and SH ($20^\circ$–$90^\circ$S) to account for the large difference of observing density. The parameters for each hemisphere are linearly interpolated in the tropics.
(20°S–20°N). Since the upper layers have smaller observing density, the inflation is linearly reduced at the top eight layers so that the top layer has no inflation.

Whitaker et al. (2008) reported better performance by using an additive inflation scheme; this technique is therefore implemented in this study. Whitaker et al. (2008) used the 6-h tendency of the NCEP–NCAR reanalysis (Kalnay et al. 1996) on randomly chosen dates in different years within a window of 30 days centered on the analyzed date. The tendency fields are multiplied by 0.33 and added to each member of the analysis ensemble, not to the first guess. The tendency fields have an appropriate power spectrum and balance for atmospheric perturbation fields, which are difficult to obtain from random numbers. In this study, the JRA-25 reanalysis (Onogi et al. 2007) is used instead of the NCEP–NCAR reanalysis. Parameters are chosen to be similar to Whitaker et al. (2008), as shown in section 3.

d. Parallel algorithm

The LETKF was designed to be efficient with a parallel architecture (Ott et al. 2004; Hunt et al. 2007). LETKF computations at each grid point are independent. Miyoshi and Yamane (2007) distributed computations at each grid point to each computational node according to geographical regions as shown schematically in Fig. 4a. This parallelization is subject to significant loss of efficiency when the observing density is irregular in the horizontal, which is the case in the real world. In the case of Fig. 4a, node 7 over Europe, for example, contains many more observations than node 1 over Antarctica and the Southern Ocean, so that the computational load imbalance is problematic.

Since LETKF computations at each grid point are independent, we do not need to parallelize by geographical areas. Figure 4b shows a revised node distribution, which resolves the load imbalance almost completely. This requires a major upgrade of the parallel FORTRAN90 code, which has been developed in this study.

e. Adaptive bias correction method for satellite radiances

Since satellite radiance observations are subject to bias, it is important to estimate and correct the observation bias prior to assimilation. The bias in satellite radiances is separated into two components: \( b = b_{\text{scan}} + b_{\text{air}} \), where \( b_{\text{scan}} \) is the bias according to scan positions; and \( b_{\text{air}} \) denotes the airmass bias, which is explained by predictors such as surface temperature, satellite zenith angle, and a constant. That is, the airmass bias is written by \( b_{\text{air}} = p^T \beta \), where \( p \) and \( \beta \) denote column vectors composed of the predictors and corresponding coefficients, respectively. The scan bias \( b_{\text{scan}} \) is estimated offline using many training samples of past observations. Similarly, the airmass bias coefficients \( \beta \) can be estimated offline, which gives constant \( \beta \) in time. Alternatively, Derber and Wu (1998) proposed the variational bias correction method (VarBC; cf. Dee 2004, 2005), which estimates \( \beta \) adaptively within variational data assimilation and allows temporal variations that are expected to account for the aging of satellite instruments. The temporal variance also helps to identify sudden instrument changes such as unexpected malfunction. Moreover, the adaptive estimation allows automatic adjustment without many training samples; this enables the use of new satellite observations as soon as possible when a new satellite is launched.

Fertig et al. (2009) proposed an adaptive bias correction method within the EnKF by augmenting the bias correction coefficients \( \beta \) to the ensemble state vectors. Specifically, \( \beta \) is analyzed simultaneously with the model state vector using an ensemble of \( \beta \). An alternative approach without an ensemble of \( \beta \) is proposed in this research, which is described in detail as follows.

VarBC solves the minimization problem of the cost function \( J \):

\[
J(x, \beta) = (x - x')^T B_{x'}^{-1} (x - x') + (\beta - \beta')^T B_{\beta'}^{-1} (\beta - \beta') \\
+ (y - Hx - p^T \beta)^T R_{p}^{-1} (y - Hx - p^T \beta),
\]

where \( x \) and \( y \) denote vectors of state and observation, respectively; and the superscript \( f \) indicates forecast or first guess. Here \( B \) and \( R \) indicate the background and observation error covariances, respectively; and the subscripts \( x \) and \( \beta \) denote the state vector and bias coefficients. The \( H \) is a generally nonlinear observation operator that maps a state vector to the equivalents of observations. The predictors \( p \) and coefficients \( \beta \) are defined separately for each satellite sensor–channel. The \( -p^T \beta \) term is applied only for corresponding satellite radiance observations. Differentiating Eq. (1) with respect to \( x \) and \( \beta \), we obtain the analytical solution of the minimizer \( (x, \beta) \):

\[
\delta x = (B_{x}^{-1} + H^T R_{p}^{-1} H)^{-1} H^T R_{p}^{-1} (d - p^T \delta \beta),
\]

\[
\delta \beta = (B_{\beta}^{-1} + p R_{p}^{-1} p^T)^{-1} p R_{p}^{-1} (d - H \delta x),
\]

where the perturbations \( \delta x = x - x' \) and \( \delta \beta = \beta - \beta' \), and the debiased innovation \( d = y - Hx - p^T \beta' \) are defined. Here \( H \) is the Jacobian or tangent-linear operator of generally nonlinear \( H \).

Alternatively, EnKF gives an analysis solution that ideally satisfies

\[
\delta x = B_{x} H^T (H B_{x} H^T + R)^{-1} d.
\]
which is equivalent to Eq. (2), except for the term $-p^T \delta \beta$. Then, Eq. (3) is solved explicitly using the ensemble-mean solution of Eq. (4) to obtain updated bias correction coefficients. To satisfy Eqs. (2) and (3) simultaneously, the process needs to be iterated. However, we solve it only once as a first-step approximation. Assuming the bias correction coefficients of each satellite sensor are independent, we solve Eq. (3) separately

![FIG. 4. Schematics showing parallel algorithms: (a) as in Miyoshi and Yamane (2007) and (b) as upgraded in this study. Each small box corresponds to a grid point, and the numbers in each box indicate the indices of computational nodes for the case of nine nodes.]

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Predictors</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMSU-A</td>
<td>Integrated weighted lapse rate (IWLR), surface temperature, cloud water content, 1/cos$\theta$ (satellite zenith angle), constant</td>
</tr>
<tr>
<td>AMSU-B</td>
<td>IWLR, surface temperature, 1/cos$\theta$, constant</td>
</tr>
<tr>
<td>SSM/I, TMI, AMSR-E</td>
<td>Precipitable water content, SST, square value of SST, sea surface wind speed, 1/cos$\theta$, constant</td>
</tr>
</tbody>
</table>
for each sensor. The computational cost to solve Eq. (3) is very low and negligible in practice since we use only several predictors. The same predictors, shown in Table 1, and $B_\beta$ as the operational 4D-Var (Sato 2007; JMA 2007) are chosen in this study. In the operational VarBC, Sato (2007) introduced diagonal components of $B_\beta$, denoted by $\sigma_{b\alpha}^2$, as follows:

$$
\sigma_{b\alpha}^2 = \begin{cases} 
\sigma_\beta^2/N_{\text{min}} & N < N_{\text{min}} \\
\sigma_\beta^2/N_{\text{min}}N\left[\log_{10}(N/N_{\text{min}}) + 1\right] & N \geq N_{\text{min}}
\end{cases}
$$

(cf. Sato 2007). Here, $\sigma_\beta^2$ and $N$ denote the observational error variance and the number of observations, respectively. The parameter $N_{\text{min}}$ was chosen to be 400. The nondiagonal components of $B_\beta$ were assumed to be zero.

In summary, the adaptive bias correction proposed in this study is composed of the following procedures:

1) Subtract satellite radiance biases using the bias correction coefficients estimated at the previous analysis.
2) Compute EnKF analysis as if there were no adaptive bias correction.
3) Solve Eq. (3) explicitly to update the bias correction coefficients.

In this study, we simply implement and apply this method to estimate the bias of satellite radiances adaptively. Although this is a new approach, detailed investigations on the method’s characteristics and comparisons with VarBC are beyond the scope of this study. Separately, T. Miyoshi and J. S. Whitaker (2009, unpublished manuscript) performed those studies with the NCEP GFS; these results will be reported in a separate paper.

### 3. Experimental settings

Using the experimental operational system described in section 2, LETKF forecast–analysis cycle experiments are performed for August 2004. The cycle begins on 20 July to allow for a 10-day spinup period. The ensemble initial conditions are chosen from the JMA operational analysis fields on randomly chosen dates in a different year in a similar season; the initial condition is an analog of the climatological mean.

Parameter values, including ensemble size as well as covariance localization and inflation parameters, are shown in Table 2. The ensemble size is fixed at 50 members for all experiments in this study. For comparison, experiments with multiplicative covariance inflation are performed. Here, the first-guess ensemble spread is inflated by 25% in the NH and 15% in the SH in consideration of the denser observing network in the NH. In addition, 10%, 20%, and 30% globally constant multiplicative inflation values are also tested. Although sophisticated multiplicative inflation methods such as adaptive inflation (Anderson 2007a,b; Li et al. 2009) and spatially dependent multiplicative inflation (Anderson 2009; Bonavita et al. 2008, 2010) would improve the performance of multiplicative inflation, these are not tested in this study. Moreover, an experiment without the adaptive bias correction is performed, so that the impact of this technique can be investigated. Initial values for adaptive bias correction coefficients are chosen to be the same as the 4D-Var experiment; that is, the values are estimated by 4D-Var. The list of LETKF experiments are shown in Table 3. Hereafter, the short names of experiments indicated in Table 3 are used without further explanation.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Short name</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additive inflation with adaptive bias correction</td>
<td>ABC</td>
<td>Parameters as in Table 2 with adaptive bias correction for satellite radiances</td>
</tr>
<tr>
<td>Additive inflation without adaptive bias correction</td>
<td>ADD</td>
<td>As with ABC, but without adaptive bias correction</td>
</tr>
<tr>
<td>Multiplicative inflation</td>
<td>MUL</td>
<td>As with ABC, but with multiplicative inflation instead of additive inflation, and with the forecast ensemble spread inflated by 25% in the NH and 15% in the SH</td>
</tr>
<tr>
<td>Global-constant multiplicative inflation</td>
<td>G10 (G20 or G30)</td>
<td>As with MUL, but with global constant 10% (20% or 30%) spread inflation</td>
</tr>
</tbody>
</table>
A 4D-Var experiment is performed using an experimental operational system, which is equivalent to the actual operational system as of March 2007. Differences between the 4D-Var and LETKF system configurations are summarized in Table 4. It is important to notice that the radiative transfer models are different, as the radiative transfer model (RTTOV-8) contains a significantly different surface emissivity model from RTTOV-7. Therefore the satellite radiances computed from the first guess for the channels sensitive to surface emissivity would be subject to bias between LETKF and 4D-Var. To avoid this problem, additional experiments using both LETKF and 4D-Var are performed without satellite radiances. For additional comparison, an experiment with 3D-Var is performed using the identical 4D-Var system but without involving the integration of the forecast model in the inner loop. “4DV” and “3DV” are used as short names for 4D-Var and 3D-Var experiments.

For verification purposes, deterministic forecast experiments are performed. The LETKF deterministic forecast is constructed from a single forecast (not an ensemble) initiated by the ensemble mean analysis. Thus, this study does not address LETKF’s probabilistic aspects. The deterministic forecasts are verified by computing anomaly correlations and by comparing with each system’s own analyses and radiosonde observations, all of which are common verification measures of NWP systems. These measures [i.e., anomaly correlations, root-mean-square errors (RMSEs), mean errors (i.e., bias)] are computed by averaging for 31 days from 1 to 31 August. The 31-day period is used for time-mean statistics in this study. The geographical regions, the NH, tropics (TR), and SH, are defined as 20°–90°N, 20°S–20°N, and 20°–90°S, respectively.

Table 4. Differences between 4D-Var and LETKF system configurations.

<table>
<thead>
<tr>
<th>Analysis resolution</th>
<th>4D-Var</th>
<th>LETKF</th>
</tr>
</thead>
<tbody>
<tr>
<td>9-day forecast resolution</td>
<td>T1319/L40 for first guess T106/L40 for tangent-linear and adjoint models</td>
<td>T1159/L40</td>
</tr>
<tr>
<td>Initialization</td>
<td>T1319/L40</td>
<td>T1159/L40</td>
</tr>
<tr>
<td>Radiative transfer model</td>
<td>Incremental nonlinear normal mode initialization (NNMI)</td>
<td>RTTOV-7</td>
</tr>
<tr>
<td>Radiative transfer model</td>
<td>Variational bias correction (VarBC)</td>
<td>Adaptive bias correction proposed in this study</td>
</tr>
<tr>
<td>Initial condition on the first date of cycle experiments</td>
<td>Operational analysis valid at the actual time</td>
<td>Operational analysis on randomly chosen dates in a different year in the same season</td>
</tr>
</tbody>
</table>

Fig. 5. Time series of 500-hPa geopotential height analysis RMS differences (m) between LETKF (ABC) and 4D-Var.
FIG. 6. Horizontal map of temperature analysis spread (K) at the 31st model level at 1200 UTC 29 Aug 2004 in the cases of the (a) MUL and (b) ADD experiments.
4. Results

First, we confirm that the LETKF system performs appropriately by comparing the analysis fields of an LETKF experiment with those of 4D-Var (Fig. 5). Throughout this paper, we use the short names of the LETKF experiments in Table 3. As in Miyoshi et al. (2007a), the LETKF fields rapidly approach the 4D-Var analysis fields in about

![Fig. 7. Horizontal maps of temperature analysis spread (K) at the (top) 7th, (middle) 15th, and (bottom) 31st model levels, close to the 850-, 500-, and 50-hPa levels, respectively, on 1200 UTC 1 Aug 2004 in the cases of the (left) MUL and (right) ADD experiments.](image-url)
5 days, and the system performs stably for the entire experimental period. The 500-hPa geopotential height RMS differences are about 7 m in the NH, 14 m in the SH, and 5 m in the tropics, which are similar orders of magnitude to typical analysis RMS errors. Sections 4a–4d present detailed results.

a. Multiplicative and additive inflation

The multiplicative inflation experiment (MUL) provided accurate analyses until it ended abnormally just after 1200 UTC 29 August, following 41 days of cycling. The time series of the vertical profile of horizontally averaged temperature ensemble spread shows a sudden increase around layers 31, 38, and 40, but the case with additive inflation (ADD) indicates no such strange behavior. If we look at the horizontal map of the temperature ensemble spread at the 31st level (≈50 hPa; Fig. 6), MUL shows unrealistic values of about 25 K over the North Pacific, but ADD does not show such abnormal values. Moreover, ADD indicates a much smaller land–sea pattern than MUL. The results suggest that multiplicative inflation tends to exaggerate the spatial pattern of observation density, although it may also be possible that additive inflation tends to underestimate the difference between data-dense and data-sparse areas.

Miyoshi and Yamane (2007) and Miyoshi et al. (2007a,b) used the globally constant value of 10% multiplicative spread inflation and reported stable performance. Therefore, it is expected that the multiplicative inflation method would provide a stable EnKF run if the parameter value is chosen properly. In fact, the multiplicative G10 and G20 experiments performed stably, whereas the G30 experiment resulted in an abnormal termination; the inflation of 25% in the NH in MUL was apparently too large. More sophisticated multiplicative inflation methods such as spatially adaptive multiplicative inflation (e.g., Anderson 2009; Bonavita et al. 2008, 2010) would be more appropriate for this highly irregular observing network both in the horizontal and in the vertical.

FIG. 8. Forecast anomaly correlations (%) of 500-hPa geopotential height for ADD (solid), G10 (dashed), and G20 (dotted) experiments in the (a) NH, (b) SH, and (c) tropics.
scales also agree well between ADD and MUL. These structural agreements in the ensemble spread suggest that both MUL and ADD experiments capture flow-dependent error structures.

To compare the overall performance of the forecast–analysis systems, forecast anomaly correlations of 500-hPa geopotential height, one of the most common measures to verify NWP systems, are computed and shown in Fig. 8. The G10 experiment shows the worst performance in the NH, thus a larger inflation parameter would be desired in a densely observed area. In the SH, the G20 experiment performs slightly better than the others up to a 7-day (i.e., 168 h) forecast, while it is the worst in the tropics.

If we look at 850-hPa temperature forecast bias relative to each system’s own analyses, additive inflation shows the largest negative forecast bias (Fig. 9). The negative bias in G20 is the smallest, which suggests that larger inflation contributes to reduce the forecast bias. However, the forecast bias reduction by larger inflation disappears if the bias is relative to radiosonde observations. We will see similar inconsistencies between various verification measures in subsequent results. The negative forecast bias at the 850-hPa level has been a well-known problem of the JMA global model for many years (Nakagawa 2003; Kazumori and Okamoto 2004; H. Kitagawa 2007, personal communication).

Overall, additive inflation performs as accurately as multiplicative inflation, although not enough evidence is found to conclude the relative advantage of additive inflation. Whitaker et al. (2008) found that additive inflation slightly outperforms multiplicative inflation, which is not clear in this study when the parameters are chosen appropriately. Since additive inflation performs as accurately as multiplicative inflation and provides more natural patterns of ensemble spread, additive inflation is selected for use in the following parts of this paper.

b. Adaptive bias correction for satellite radiances

In the ADD experiment, bias correction coefficients are fixed to those estimated by 4D-Var. In this subsection, the impact of the adaptive bias correction scheme is described. The time series of bias correction coefficients in the ABC experiment for NOAA-16 satellite’s AMSU-A channels (a) 4 and (b) 6. The vertical scale on the left indicates bias coefficients for predictors: integrated weighted lapse rate (IWLRA, solid), surface temperature (Tsurf, long-dashed), satellite zenith angle (ZANG, short-dashed), and constant (CONST, dotted). The scale on the right indicates the numbers of assimilated observations (number of OBS, dash–dot).
by the difference in the radiative transfer model (i.e., RTTOV-7 in 4D-Var and RTTOV-8 in LETKF). Thus, ABC would have better bias correction consistent with RTTOV-8.

The temperature analysis at the 850-hPa level becomes colder by about 0.5 K with the adaptive bias correction, which reduces the forecast bias of ABC to a similar level as that of G20 in Fig. 9. The only difference between ABC and ADD is the bias correction coefficients for satellite radiances, and the major cold bias pattern appears over the ocean where satellite radiances are assimilated. Although the major reason for the different bias coefficients is attributed to the radiative transfer model, the known cold bias of the JMA global model would also create a colder analysis, particularly because the adaptive bias correction scheme does not consider bias in the model explicitly.

The overall impact of applying the adaptive bias correction is positive in most forecast verification measures, mostly due to the adaptation to RTTOV-8, which is biased relative to RTTOV-7. For example, the 500-hPa geopotential height forecast anomaly correlations demonstrate the advantage of ABC. Although improvements in the NH are not so clear, the advantage of ABC is very evident in the SH and tropics. Although the 850-hPa forecast bias is similar between G20 and ABC, ABC is superior to G20 when considering anomaly correlations. Similar improvements are observed in other variables at other levels. However, if errors are relative to radiosonde observations, the forecast bias and RMS errors are not necessarily reduced. This inconsistency between verification measures will be shown more in detail by comparing ABC with 3D-Var and 4D-Var in section 4c.

\textit{c. Comparison with 3D-Var and 4D-Var}

The 500-hPa geopotential height anomaly correlations indicate that both 4D-Var and LETKF outperform 3D-Var (Fig. 11). The advantages of the advanced methods (i.e., 4D-Var and LETKF) relative to 3D-Var are greater in the SH and tropics where fewer observations exist, which agrees with previous studies (e.g., Thepaut 2006; Buizza et al. 2007). As for the comparison between LETKF and 4D-Var, LETKF demonstrates a slight advantage in the NH and tropics, although 4D-Var performs better in the SH.

The anomaly correlations of LETKF and 4D-Var for several variables are compared, and percent improvements of LETKF relative to 4D-Var are shown in Fig. 12. As with the results for 500-hPa geopotential height, LETKF generally outperforms 4D-Var in the NH and tropics, but not in the SH. LETKF has a considerable advantage with regard to wind forecasts, especially in the tropics. However, this is not necessarily true when
the RMS errors are relative to radiosonde observations, as we have seen in previous results.

The above results are generally true with regard to other levels and variables. The vertical profiles of temperature forecast errors indicate that LETKF slightly outperforms 4D-Var in the NH and tropics when errors are computed relative to each system’s own analyses (Fig. 13). However, when the errors are relative to radiosonde observations (Fig. 14), LETKF indicates larger errors than 3D-Var in the NH troposphere. In the SH and tropics, no significant difference is found between LETKF and 4D-Var, but 3D-Var shows significantly larger errors. LETKF indicates a larger negative temperature bias, especially around the 850-hPa level, in Fig. 13, which is not the case in Fig. 14. This is again an example of the results depending on the choice of verification measures.

We have noted several times the cold 850-hPa temperature forecast bias of the LETKF, demonstrated in Fig. 9. Figure 15 shows a similar plot, but for bias relative to each system’s own analyses and radiosonde observations in each geographical area. As expected from Fig. 9 and the well-known bias in JMA global model, we see generally negative forecast biases, except for the bias relative to radiosonde observations in the NH. LETKF indicates a larger negative bias relative to its own analyses, especially in the SH and tropics. The negative biases relative to their own analyses for 3/4D-Var are smaller in the SH and tropics than in the NH, but the opposite is true for LETKF. However, when bias is relative to radiosonde observations LETKF, 3D-Var, and 4D-Var all show similar results.

To further investigate the inconsistency of the bias relative to each system’s own analyses and radiosonde observations, horizontal patterns of biases are investigated (Fig. 16). The analysis bias between LETKF and 4D-Var shows that the LETKF has higher 850-hPa temperature by about 1.0 K mainly over the ocean where few radiosonde observations exist. This bias appears just after the beginning of the assimilation (i.e., 20 July), although the LETKF analysis is still spinning up. The 6-h forecast bias relative to each system’s own analyses indicates a large difference between LETKF and 4D-Var over the ocean. For example, the LETKF indicates negative values over the tropical Indian Ocean, whereas 4D-Var indicates large positive values, and similarly in the Pacific and Atlantic. However, 120-h forecast bias relative to 4D-Var analyses indicates very similar patterns for 4D-Var and LETKF for either ADD or ABC. In fact, the forecast bias of the LETKF relative to 4D-Var forecasts...
decreases in later forecast times (Fig. 17). Therefore, the 5-day forecast from initial conditions by either LETKF or 4D-Var would have a similar bias relative to nature. The results suggest that the forecast bias relative to each system’s own analyses may be simply the result of the biased difference between 4D-Var and LETKF analyses.

If we look at the horizontal pattern of 850-hPa temperature forecast improvements by LETKF relative to 4D-Var, LETKF is generally better over radiosonde...
locations (blue colors in Fig. 18). By contrast, most areas over the ocean indicate the disadvantage of LETKF (red colors in Fig. 18). Since the order of magnitude of the RMS difference is similar to the analysis bias, a major part of the red pattern could be explained by the analysis bias. In fact, this kind of land–sea pattern is not observed for other fields such as 500-hPa geopotential height.

Since the bias in the lower troposphere is strongly related to the radiative transfer models RTTOV-7 and -8, both 4D-Var and LETKF experiments are performed.
without assimilating satellite radiances to investigate the impact of radiance assimilation. Without satellite radiances, LETKF outperforms 4D-Var more clearly in almost all verification measures, although the forecast scores relative to radiosonde observations are neutral or slightly opposite. The anomaly correlation scores indicate improvements by the LETKF except for the sea level pressure in the NH (Fig. 19). One concludes from the results that presently satellite radiance assimilation has a relatively stronger positive impact on 4D-Var than on LETKF. Thus, it is important to improve the treatment of satellite radiances within the LETKF in future developments.

\section*{d. Computational time}

The computer used in this study is the JMA operational supercomputing system called the Numerical Analysis and Prediction System (NAPS). It contains a powerful supercomputing server, Hitachi SR11000, which is used in Japanese NWP operations. The wall clock time for computations is shown in Table 5. Although LETKF is slightly faster, the difference is below 10\%, which is not very significant. There must be room to accelerate the computing speed even more, and the parallelization efficiency would change depending upon the number of computing nodes. Since the ensemble model runs are
essentially perfectly parallel and the LETKF requires internode communication only twice, the LETKF system has significantly better parallel scalability than 4D-Var.

It is important to note that LETKF was at first much slower than 4D-Var, and major efforts, including the efficient parallel implementation described in section 2d, were made to achieve the current computing speed. As LETKF developers, the authors were highly motivated to improve the computational efficiency of the system when the 4D-Var system was faster. The opposite is also true with regard to the improvement of 4D-Var; now, the developers of the 4D-Var system try to accelerate their computations to be faster than LETKF. This synergistic effect to the development of both systems is important and discussed in section 5.

5. Summary and discussion

Four major upgrades were made to the LETKF system embedded in the JMA experimental operational system, with which several forecast–analysis cycle experiments have been performed as described in Table 3. The comparison between multiplicative and additive covariance inflation methods indicated that in general both schemes performed with comparable accuracy. However, significant structural differences were found in the ensemble perturbations; one of the most important differences was that the multiplicative inflation tended to exaggerate the spatial pattern of observing density in the ensemble spread, which led to unstable filtering when a large multiplicative inflation of 25% in the NH was used. Since such overestimation of ensemble spread in data-poor regions was not desired, additive inflation was chosen for the comparison with 4D-Var.

As the operational 4D-Var system is equipped with VarBC, an adaptive bias correction scheme to realize the VarBC functionality within the LETKF has been proposed and developed. The results indicated that the
Adaptive bias correction contributed to improve anomaly correlation forecast skill for most variables. Moreover, it reduced temperature forecast bias relative to each system’s own analyses, although 4D-Var showed an even smaller forecast bias. The negative temperature bias at the 850-hPa level is closely related to the radiative transfer model, as well as the cold bias in the JMA global model, which has been well known for many years.

Overall, the LETKF showed comparable results with 4D-Var. Most verification measures, except for forecast scores relative to radiosonde observations, indicated that the LETKF outperformed 4D-Var in the NH and tropics and that 4D-Var significantly outperformed the LETKF in the SH. From the sensitivity experiments without satellite radiance data, 4D-Var was improved more than LETKF through the assimilation of satellite radiances. Improving the treatment of satellite radiances would therefore be important in the future development of the LETKF. Throughout this study, we saw inconsistency between verifications relative to each system’s own analyses and radiosonde observations. Since the radiosonde locations are highly irregular, sometimes the verification could be misleading. It is important to look at many kinds of verification measures to have a better idea of the relative performance of systems.

An important difference between LETKF and 4D-Var systems is the version of the radiative transfer model. LETKF and 4D-Var used RTTOV-8 and -7, respectively, which are known to have significant bias in computed surface emissivity. This causes bias in AMSU-A channel 4, which is sensitive to surface emissivity as well as the lower troposphere. Although it is desired that both LETKF and 4D-Var
and 4D-Var use the same radiative transfer model, this study could not include it mainly because of technical difficulties.

The treatment of vertical coordinate assignment of observation data may not be optimal in the LETKF. The vertical coordinate system of the JMA global model is a sigma-pressure hybrid coordinate. Therefore, the vertical coordinate system depends on the surface pressure, which is changed by data assimilation. The vertical coordinate of each observation is determined at the beginning of the data assimilation to compute the innovations (i.e., observation minus first guess). The innovations are assimilated at the model coordinate assigned at the beginning, which may change after the analysis. This effect is considered in 4D-Var when the linearized observation operator is recomputed for increments in each iteration, but LETKF does not consider this at all.

The authors developed the LETKF as a possible future choice at JMA, while 4D-Var is currently used operationally. During the development, comparison with the operational 4D-Var played an essential role to identify any possible problems, and to improve the performance of the LETKF. As mentioned in section 4, the computing speed was a problem at first, but we succeeded in accelerating it so that it became faster than 4D-Var. In addition, we developed an adaptive bias correction for satellite radiances for the LETKF since 4D-Var had the advantage of VarBC. Thus, the competition between 4D-Var and LETKF made both systems stronger; this is an important synergetic effect. Although it would require more resources to keep developing both systems, the authors believe it pays off by the synergy. Until we reach an important synergetic effect. Although it would require more resources to keep developing both systems, the authors believe it pays off by the synergy. Until we reach an important synergetic effect.

The competition between 4D-Var and LETKF made both systems stronger; this is the advantage of VarBC. Thus, the competition between satellite radiances for the LETKF since 4D-Var had the advantage of VarBC. Thus, the competition between 4D-Var and LETKF made both systems stronger; this is an important synergetic effect. Although it would require more resources to keep developing both systems, the authors believe it pays off by the synergy. Until we reach an important synergetic effect.

The authors developed the LETKF as a possible future choice at JMA, while 4D-Var is currently used operationally. During the development, comparison with the operational 4D-Var played an essential role to identify any possible problems, and to improve the performance of the LETKF. As mentioned in section 4, the computing speed was a problem at first, but we succeeded in accelerating it so that it became faster than 4D-Var. In addition, we developed an adaptive bias correction for satellite radiances for the LETKF since 4D-Var had the advantage of VarBC. Thus, the competition between 4D-Var and LETKF made both systems stronger; this is an important synergetic effect. Although it would require more resources to keep developing both systems, the authors believe it pays off by the synergy. Until we reach an important synergetic effect.

The authors believe that developing a clear conclusion to choose either one or a possible combination of both, the authors believe it pays off by the synergy. Until we reach a clear conclusion to choose either one or a possible combination of both, the authors believe that developing both systems would be the right strategy.

Acknowledgments. This work was done while the first author was a scientist at the Numerical Prediction Division (NPD) at JMA. The source codes of the radiative transfer model RTTOV-7 and -8 were provided by EUMETSAT. Results of the 4D-Var experiment were provided by Dr. Kozo Okamoto of NPD/JMA. The first author is grateful to Masahiro Kazumori of NPD/JMA, Dr. Jeff Whitaker of NOAA/ESRL, Prof. Eugenia Kalnay, Steven Greybush, and other members of weather/chaos group at the University of Maryland for useful discussions. This work was partly supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan (21244074). “Study of advanced data assimilation and cloud resolving ensemble technique for prediction of local heavy rainfall”.

REFERENCES

Dee, D. P., 2004: Variational bias correction of radiance data in the ECMWF system. Proc. ECMWF Workshop on Assimilation of High Spectral Resolution Sounders in NWP, Reading, United Kingdom, ECMWF, 97–112.


Nakagawa, M., 2003: Recent development of a cumulus parameterization scheme for the operational global model at JMA. BMRC Research Rep. 93, 29–32.


