The Local Ensemble Transform Kalman Filter (LETKF) with a Global NWP Model on the Cubed Sphere

SEOLEUN SHIN,[1] JI-SUN KANG,[1,2] and YOUNGSOON JO[1]

Abstract—We develop an ensemble data assimilation system using the four-dimensional local ensemble transform kalman filter (LEKTF) for a global hydrostatic numerical weather prediction (NWP) model formulated on the cubed sphere. Forecast-analysis cycles run stably and thus provide newly updated initial states for the model to produce ensemble forecasts every 6 h. Performance of LETKF implemented to the global NWP model is verified using the ECMWF reanalysis data and conventional observations. Global mean values of bias and root mean square difference are significantly reduced by the data assimilation. Besides, statistics of forecast and analysis converge well as the forecast-analysis cycles are repeated. These results suggest that the combined system of LETKF and the global NWP formulated on the cubed sphere shows a promising performance for operational uses.

Key words: Ensemble data assimilation, Local ensemble transform Kalman filter (LEKTF), Numerical weather prediction (NWP), Atmospheric global model (AGM).

Abbreviations

LETKF	Local ensemble transform Kalman filter
NWP	Numerical weather prediction
AGM	Atmospheric global model
EnKF	Ensemble Kalman filter
KIAPS	Korea institute of atmospheric prediction systems
KIM	KIAPS integrated model
CAM	Community atmospheric model
GPS-RO	Global positioning system radio occultation
OSSE	Observing system simulation experiment
ESMF	Earth system modelling framework
HOMME	High order method modeling environment
IO	Input–output
GLL	Gauss-Legendre-Lobatto
CESM	Community earth system modeling
CAM-SE	Spectral element version of community atmospheric model
PDF	Probability density function
RMSE	Root mean square error
IFS	Integrated forecast system
RMSD	Root mean square difference
WRF	Weather research and forecast
GFS	Global forecast system
KMA	Korea meteorological administration

[1] Data Assimilation Team, Development Division, Korea Institute of Atmospheric Prediction Systems (KIAPS), Seoul 156-849, Korea. E-mail: seoleuns@gmail.com

[2] *Present Address*: Korea Institute of Science and Technology Information (KISTI), Daejeon, South Korea.

1. Introduction

Data assimilation is a key element in a numerical weather prediction system in that it provides an improved initial state for the next forecast by obtaining an optimal analysis state from statistical treatments of available observations and current forecasts. There are two major approaches in data assimilation: variational and ensemble methods. Since the ensemble Kalman filter (EnKF) technique suggested by EVENSEN (1994) was applied to an atmospheric system by HOUTEKAMER and MITCHELL (1998), investigations of the ensemble technique with operational interests have been made extensively. Thereby computationally efficient algorithms draw attention as high-performance computing resources for parallel implementation becomes an issue. The local ensemble transform Kalman filter (LETKF) considers observations, only within a specified local area surrounding each model grid point (HUNT *et al.* 2007). This nature of the LETKF algorithm can lead

to an ensemble data assimilation technique that scales and handles memory spaces efficiently (MIYOSHI and YAMANE 2007). The LETKF technique has been evaluated in applications such as for global atmospheric system, and shown good performance compared to other existing techniques in data assimilation systems (e.g., BUEHNER et al. 2010; MIYOSHI et al. 2010).

For the applications in weather forecasting it is advantageous to take account for the timing of the observations since changes in weather can be significant within a usual analysis time interval of 6 h (HUNT et al. 2007). A four-dimensional-LETKF (thereafter 4D-LETKF) system considers the time information of observation in such way that time-evolving background errors are counted for finding analysis that best fits to the observations. The 4D-LETKF algorithm has been reported to improve the performance of forecast and reduces analysis errors (e.g., MIYOSHI and ARANAMI 2006; HARLIM and HUNT 2007). MIYOSHI (2011) compared the LETKF and the operational 4D-Var system implemented to the global model at the Japanese Meteorological Agency (JMA), and their results suggested that the LETKF has comparable performance to the 4D-Var. At the Korea Institute of Atmospheric Prediction Systems (thereafter KIAPS), we have applied the 4D-LETKF algorithm derived in HUNT et al. (2007) for the development of data assimilation system coupled to a new global NWP model. The forecast model that is being developed at the KIAPS (KIAPS Integrated Model: KIM) is using the spectral element method for discretization of governing equations and formulated on the cubed sphere (SADOURNY 1972) so that a singularity problem at poles can be avoided. This leads to an unstructured grid system for the model (KIM) and we need to additionally develop a tool for the observation operator in the 4D-LETKF framework. We name the LETKF system KIAPS-LETKF for a spectral-element global NWP model on the cubed sphere. More details on the model grid system and the methodology to take account for forecast fields on the unstructured grid will be presented in the following sections.

Before our own forecast model KIM had been developed, we alternatively used the NCAR Community Atmospheric Model (CAM) for a test of KIAPS-LETKF, which has a dynamical core formulated using the spectral element method on the cubed sphere (CAM-SE, TAYLOR et al. 1997). From the KIAPS-LETKF implemented to the CAM, we obtained encouraging results assimilating conventional and Global Positioning System-Radio Occultation (GPS-RO) bending angle data (KANG et al. 2014; KWON et al. 2015). As the KIM becomes available for tests, we modified the KIAPS-LETKF, and coupled it to the hydrostatic version of the KIM. In this article we explain main features of the KIAPS-LETKF system implemented to the KIM and discuss about its performance in an Observing System Simulation Experiment (OSSE) and real data assimilation.

Our discussion focuses on the implementation of the LETKF to the KIM with unstructured grid system and the use of the adaptive multiplicative inflation method, which is especially useful when we implement the LETKF algorithm to a newly developed model. In this situation we could avoid a manual tuning for inflation, which can demand a lot of effort and time by using the adaptive multiplicative inflation method. Fixed inflation method cannot give different inflation factors in space and time, which cannot reflect background uncertainty effectively (e.g., LI et al. 2009; MIYOSHI 2011; WHITAKER and HAMILL 2012). This study may provide useful information to people who are testing the adaptive method in a LETKF framework.

In the next section we introduce the forecast model KIM with the unstructured grid system. In Sect. 3 the development of the KIAPS-LETKF system is described and in Sect. 4 the performances of data assimilation in the OSSE and real data assimilation experiment are discussed. In the final section we summarize the current work and present some future plans. We also provide a list of abbreviations in the Abbreviation group.

2. Forecast Model

An NWP system tends to require a large number of modules with different functions to produce weather forecasts. Recently many research and operational institutions have interest in making their models more flexible, high-performing, and robust. To build such an NWP system, it can be inevitable to

use the concept of "modelling framework", which can provide an infrastructure that connects modules seamlessly (HILL *et al.* 2004). There are a few well-known software framework in the community of NWP and one of them is the Earth System Modelling Framework (ESMF). The ESMF is applied for the construction of coupled modelling systems to use in the area of climate modelling and NWP. It is based on the idea that the complex applications can split into a "component", which is replaceable and modifiable easily (HILL *et al.* 2004). A modelling framework is designed at the KIAPS, which is based on the similar idea as the ESMF, but diminishes technical barriers for general users and developers of the preliminary version of the KIM.

For the hydrostatic version of the KIM, the same set of governing equations and their discretization methodology is used as the spectral element hydrostatic dynamical core of High Order Method Modeling Environment (HOMME, DENNIS *et al.* 2012). It is formulated in an unstructured quadrilateral grid system on the cubed sphere. It has an excellent scalability and can be free of polar singularity of latitude-longitude grid system (DENNIS *et al.* 2012).

Some major differences between the stand-alone HOMME and the hydrostatic version of the KIM are the modeling framework discussed earlier, a newly implemented infrastructure including an input–output (IO) system, a 50-level vertical coordinate extended up to about 60 km, and a physical parametrization package coupled to the dynamical core, etc. (SHIN *et al.* 2014). The horizontal spatial resolution used for this study is based on the 30 elements per face and 4 Gauss–Legendre-Lobatto (GLL) points per element, and thereby the average grid spacing at the equator is $1°$ and a minimum grid spacing is $0.83°$ (EVANS *et al.* 2013). Topographical information is obtained from the smoothed topography from the Community Earth System Modeling (CESM).

3. KIAPS-LETKF

The KIAPS-LETKF is based on the 4D-LETKF introduced by HUNT *et al.* (2007) and its technical implementation methodology is adopted from MIYOSHI and YAMANE (2007). Computational codes for infrastructure and main computations are obtained from http://www.code.google.com/p/miyoshi/ and we modified them for our purposes. A main feature of the KIAPS-LETKF is that a computational module containing observation operator has been established for a global atmospheric model with an unstructured grid system in the physical space such as the CAM-SE and the KIM. Choices of observations and local subset drawn from the global state for the local analysis are determined by the newly implemented modules. Since the observation operator is implemented in an independent computation module outside of the LETKF system, it is flexible in using any kind of nonlinear operator. Also an adaptive multiplicative inflation by (MIYOSHI 2011) is used to avoid underestimation of background error covariance where observation is densely distributed.

At first the KIAPS-LETKF was tested using the spectral element version of Community Atmospheric Model (CAM-SE) before the KIM became available, since it has the same horizontal grid systems as the KIM adopts (KANG *et al.* 2014). The CAM-SE that was used for the test has 30 vertical layers up to about 2.25 hPa (~ 40 km) and its horizontal grid spacing is about 250 km. Although the focus of the CAM-SE might be originally climate studies rather than NWP, it has an equivalent complexity for the test of a global data assimilation system for weather forecasts. Observing System Simulation Experiment (OSSE) and real data assimilation using conventional data such as sonde and surface pressure has been done successfully with the model.

As an early version of KIM with the hydrostatic governing equations was released, the KIAPS-LETKF was implemented to the KIM and its performance has been evaluated. For the KIM introduced in the previous section, input/output procedures and grid system information are updated for the implementation of the data assimilation codes based on the algorithms described in the following sections.

3.1. Local Ensemble Transform Kalman Filter

In this section, we introduce the main idea of Local Ensemble Transform Kalman Filter (LETKF)

briefly. More details on the LETKF algorithm and its implementation can be found in HUNT *et al.* (2007). Suppose that $\mathbf{x}$ is a state vector of dynamic variables at model grids. Ensemble analyses at the previous analysis step are used as initial conditions to generate background ensemble states $\mathbf{x}^{b(k)}$ at time t, $k = \{1, 2, \ldots K\}$ where K is the number of ensemble members. We denote $\mathbf{X}^b$ as the matrix whose columns contain a departure of each ensemble forecast $\mathbf{x}^{b(k)}$ from the ensemble mean $\bar{\mathbf{x}}^b$: the k-th column of $\mathbf{X}^b$ is $\mathbf{x}^{b(k)} - \bar{\mathbf{x}}^b$. Then, the observation operator h is applied to the ensemble forecast $\mathbf{x}^{b(k)}$ to transform the background states from the model grid space to the observation space, $\mathbf{y}^{b(k)} = h(\mathbf{x}^{b(k)})$. Let $\mathbf{Y}^b = \mathbf{y}^{b(k)} - \bar{\mathbf{y}}^b$ be the background perturbations in the observation space. Then, the background information is ready to be compared with observations in the same space. To update analysis states at every grid point, the LETKF assimilates only observations within a certain distance from each grid point. Here we use the subscript (l) to denote a quantity defined on such a local region centered at an analysis grid point. The analysis mean $\bar{\mathbf{x}}^a_{(l)}$, is given by

$$\bar{\mathbf{x}}^a_{(l)} = \bar{\mathbf{x}}^b_{(l)} + \mathbf{X}^b_{(l)} \bar{w}_{(l)}, \qquad (1)$$

where $\bar{\mathbf{w}}_{(l)}$ is the mean weighting vector calculated by

$$\bar{\mathbf{w}}_{(l)} = \tilde{\mathbf{P}}^a_{(l)} (\mathbf{Y}^b_{(l)})^T \mathbf{R}^{-1}_{(l)} (\mathbf{y}^o_{(l)} - \mathbf{y}^b_{(l)}). \qquad (2)$$

Here, $\tilde{\mathbf{P}}^a_{(l)} = [(\mathbf{Y}^b_{(l)})^T \mathbf{R}^{-1}_{(l)} (\mathbf{Y}^b_{(l)}) + (K-1)\mathbf{I}/\rho]^{-1}$ is the analysis error covariance in the ensemble space, $\mathbf{R}$ is the observation error covariance matrix, $\mathbf{y}^o$ is the observation vector, and ρ is the multiplicative inflation factor. Within a local region, space localization is carried out by multiplying the inverse observation error covariance matrix with a factor that decays from one to zero as the distance of the observations from the analysis grid point increases. The spatial localization weights are given by a Gaussian-like piecewise fifth order rational function (GASPARI and COHN 1999; MIYOSHI *et al.* 2007) with the localization scale of $2\sqrt{10/3} \cdot \sigma_h$, where we choose $\sigma_h = 500$ km for the horizontal localization so that the function drops to zero at about 1800 km. Likewise, the vertical localization function for conventional data is defined by the Gaussian-like

rational function, with the localization scale of $2\sqrt{10/3} \cdot \sigma_v$, where $\sigma_v = 0.2$ in the unit of the logarithm pressure.

Then ensemble perturbations of the analysis are determined by

$$\mathbf{X}^a_{(l)} = \mathbf{X}^b_{(l)} [(K-1) \tilde{\mathbf{P}}^a_{(l)}]^{\frac{1}{2}}. \qquad (3)$$

This provides an estimation of analysis uncertainty and the global analysis ensemble $\mathbf{x}^{a(k)}$ is obtained by gathering the values for $\bar{\mathbf{x}}^a_{(l)}$ and $\mathbf{X}^a_{(l)}$ at all the analysis grid points.

We adopt the 4D-LETKF formulation introduced by HUNT *et al.* (2007) and a time index needs to be added to denote time-dependent terms in above equations. See HUNT *et al.* (2007) for more detailed derivation of equations in 4D formulation. Besides, we use the adaptive multiplicative inflation suggested by MIYOSHI (2011) for covariance inflation. In Sect. 3.3 we briefly describe the implementation of the adaptive multiplicative inflation and parameter choices for a spin-up of inflation factor.

3.2. Modification of LETKF for an Unstructured Grid Model

The KIM is formulated with fully unstructured quadrilateral meshes based on the cubed sphere grid which is distributed irregularly when projected on the longitude and latitude grids. For example, two adjacent unstructured grid-points that seem to be located at the same latitude do not actually have the same latitude when projected on the global meshes of latitude and longitude. A tool to support such grid system has not been yet implemented in the LETKF framework (MIYOSHI and YAMANE 2007; MIYOSHI 2011). Thus, it is required to develop a new algorithm in the observation operator h for a spatial interpolation of quantities on such unstructured grids, and data search algorithm to collect information of $\mathbf{Y}^b_{(l)}$ and $\mathbf{y}^o_{(l)}$ for every observation point.

Original LETKF technique (e.g., MIYOSHI and KUNII 2012) defines a relative position of each observation data (r_i, r_j, r_k) with respect to the model grid of (i, j, k) for zonal, meridional, and vertical directions. Then background ensemble perturbations $\mathbf{y}^{b(k)}$ are computed by the bilinear interpolation using

the closest eight points surrounding (r_i, r_j, r_k). For this computation in the regular latitude-longitude grid systems, it is simple to find points surrounding (r_i, r_j) in the horizontal direction. However, more careful examination is required to search such surrounding points in an unstructured grid system since an extrapolation can occur if one simply chooses the closest four points from the position of observation data (Fig. 1). Therefore, a search algorithm is required to look for four points adjacent to each observation position in such way that those points are not only closest to the observation position but also surround the observation. We apply Jordan Curve Theorem for a search algorithm to sample grid points enclosing the position of observation nearby in the KIAPS-LETKF system. We confirmed that the modified spatial interpolation of the observation operator worked well in previous studies (e.g., KANG et al. 2014; KWON et al. 2015).

It is also possible that one can map the unstructured grid to a regular latitude and longitude grid system and perform data assimilation, and then remap back to the original model grid. Remapping between two different grid structures introduces errors because it is also an approximation. Fortunately, our approa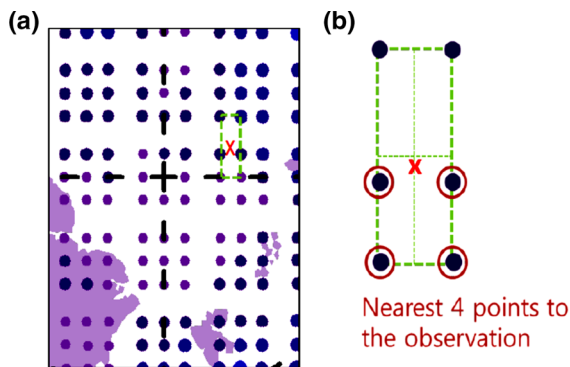ch does not require such interpolation; LETKF algorithm respects the model's own grid as it is, but just transforming model background at their own grid into the observation space before comparing background states and the observations. Therefore, there is no need to introduce errors due to the remapping between two grid structures during the data assimilation. Indeed, it can be one of advantageous characteristics in our data assimilation system because computational cost for remapping between two different grid systems will be increasing rapidly as the model resolution increases.

We recently found that TERASAKI et al. (2015) compared two versions of Non-hydrostatic Icosahedral Atmospheric Model (NICAM)-LETKF. One is with the remapping process between the icosahedral and longitude-latitude grids and the other is without the grid conversion in similar way as we are doing with the KIAPS-LETKF. They showed that the second version of the NICAM-LETKF without the grid conversion accelerates computation by 40 % and improves accuracy by about 10 %, compared to the version with the grid conversion. They assumed that the remapping may cause additional error through repeated interpolations, and add computational costs. Their results agree with our intuitive understanding that the remapping between two different grids will add interpolation errors and computational overhead. Also NERGER and HILLER (2013) used the background fields directly from the Finite Element Ocean Model (FEOM) with unstructured triangular grids in their ensemble data assimilation system, even if one might search neighboring grid points for a local observation domain more easily with a regular latitude-longitude grid system.

Figure 1
a Distribution of grid points (*violet dots*) of CAM-SE model over the North Pole. Suppose that observation is located at the mark of *red* x, then **b** it causes an extrapolation when using the nearest four points (with *pink circles*) to the observation for the bilinear interpolation. That is, model values at those four points would be used for computing $y^{b(k)}$ at the location of the observation x. Therefore, in order to avoid the extrapolation at this step, we have applied Jordan Curve Theorem to check whether the four closest model points surround the observation, and if not then search other points to satisfy the condition

3.3. Adaptive Multiplicative Inflation

The degree of freedom is $O(10^6)$, with our current model resolution but we can use a much fewer number of ensemble members in practice. A sampling error and underestimation of background error covariance are hardly avoidable in an Ensemble Kalman Filter (EnKF) system for the description of geophysical flows. Generally an "inflation" technique is used to treat the problem of underestimation of error variance and a localization method to deal with the sampling error. In this study we use the

adaptive covariance inflation suggested by MIYOSHI (2011) in the KIAPS-LETKF system. This technique is the implementation of adaptive inflation approach introduced by LI et al. (2009) within the LETKF in such way that the inflation parameters are updated with the ensemble transform matrix at each grid point (See more details in MIYOSHI 2011). This adaptive multiplicative inflation needs less effort for a tuning and it is independent of variable (LI et al. 2009). A prior Probability Density Function (PDF) of the inflation parameter is assumed to be a Gaussian in this approach and the PDF is $\Pr(\alpha_i^b) = N\left(\overline{\alpha_i^b}, v_i^b\right)$, where $\overline{\alpha_i^b}$ is the mean and v_i^b is the variance and their values are tunable and prescribed initially (MIYOSHI 2011). Then a posterior PDF of inflation parameter updated by using the Gaussian approach is given by

$$\Pr(\alpha_i^a) = N\left(\overline{\alpha_i^o}, v_i^o\right)\Pr(\alpha_i^b)/norm, \qquad (4)$$

In Eq. (5) in Miyoshi (2011), and the "norm" denotes the posterior PDF, $\Pr(y_i|y_{i-1}, y_{i-2}, \ldots, y_0)$ for ith observation y_i in discrete time. Here the posterior inflation parameter is denoted by α_i^a, i = 1, 2,…, p, and the updated inflation parameter from the newest observations p is denoted by $\overline{\alpha_i^o}$ and its variance by v_i^o. Avoiding sampling error of those estimation of inflation factors, LI et al. (2009) and MIYOSHI (2011) introduced temporal smoothing of the parameter using the prior variance v_i^b. The variance is a tuning parameter and the strength of temporal smoothing grows (weakens) if one sets it large (small). In this study we initially choose the prior variance of the inflation parameter $v_i^b = 0.01^2$ for the Gaussian approximation to the Bayesian estimates of covariance inflation $\Pr(\alpha_i^b)$ in real data assimilation. This value is small but realistic for the variance of the prior estimate of inflation in practice (MIYOSHI 2011). We observed that the forecast-analysis run is unstable if we use a larger parameter than $v_i^b = 0.01^2$ with an earlier version of the KIM model for the assimilation test using real observation data. The value of variance is an indicator of the strength of temporal smoothing in the spin-up of the inflation parameter and a larger value leads to more temporal fluctuations (MIYOSHI 2011). Since the cycle runs stably with the current version of the KIM model even when we increase the variance, we additionally test with the variance $v_i^b = 0.04^2$ and compare the performance with the test using $v_i^b = 0.01^2$ for data assimilation of real observation. More detailed explanation with respect to the parameter will be given along with corresponding results in Sect. 4.2.

4. Evaluation

4.1. Observing System Simulation Experiment (OSSE)

We first evaluate the performance of the KIAPS-LETKF implemented to the global NWP model KIM under the OSSE where we can easily find sources of errors. Since a true state is given in an OSSE, it is useful to evaluate a newly developed data assimilation system prior to carrying out real data assimilation. We assume a single model run using the KIM as a true state (nature) and generate simulated observations by projecting the true state into an observational space through a spatial interpolation and variable transformation. Certain observational errors of realistic scale are added to the simulated observations. We attempt to maintain the simulated observations close to real data by drawing temporal and spatial positions of NCEP PrepBufr data containing conventional observations such as sonde and surface pressure observations. In this study we generate a true state by integrating the forecast model KIM from 00 UTC 25 July 2011 for 15 days. Typically observational data can be obtained at 00 and 12 UTC more than 06 and 18 UTC temporally. Also more observations are distributed over the land than the ocean, and over the Northern Hemisphere than the Southern Hemisphere.

We use 30 members of ensemble so that the initial ensemble members are obtained by choosing 30 model states simulated by the forecast model KIM. Consequently, the initial error of the ensemble is supposed to be quite large. The purpose of this OSSE is to examine if analysis and forecast can converge to the true state in time when simulated observations are assimilated by the KIAPS-LETKF even though the initial ensemble states are far from the true state.

Figure 2 shows the difference of analysis and background from the nature when the forecast-analysis cycle is performed once, at 06 UTC on 25 July 2011. The upper panel shows the zonal winds at
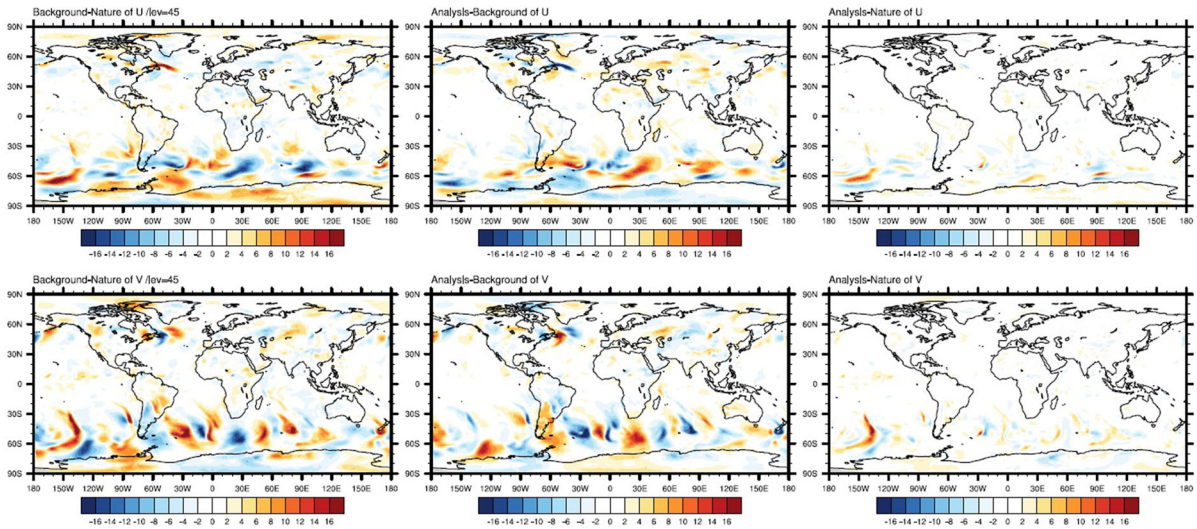
Figure 2

Background errors (*left column*), analysis increments (*middle*), and the analysis errors (*right column*) of U (*top*) and V (*bottom*). Here, analysis increments indicate analys minus background, and background/analysis errors are computed by true states subtracted from the background/analysis states

the 45th model level from the top (about 925 hPa) and lower panel shows the meridional winds at the same vertical level as the upper one. Since we assume the nature as the true state at a given time and space, we define the difference from the nature as an error. It is shown that the large background error is effectively compensated by the analysis increment as a result of data assimilation and eventually the analysis error becomes small in the regions of large background errors. While sonde observations are concentrated over lands in the Northern Hemisphere, surface pressure observations are distributed evenly in the whole globe, even in the Southern Hemispheric Ocean. Even if there are few sonde observations over the Southern Hemispheric Ocean, analysis increments induced by surface pressure data significantly reduce background errors of wind variables in addition to the surface pressure due to the multivariate background error covariance in the ensemble data assimilation technique. The analysis with reduced errors is then used as an initial condition for the next forecast-analysis cycle.

Figure 3 shows the time series of globally averaged value of Root Mean Square Error (RMSE) computed in the observational space. The background RMSE drops rapidly in the early stage of the forecast-
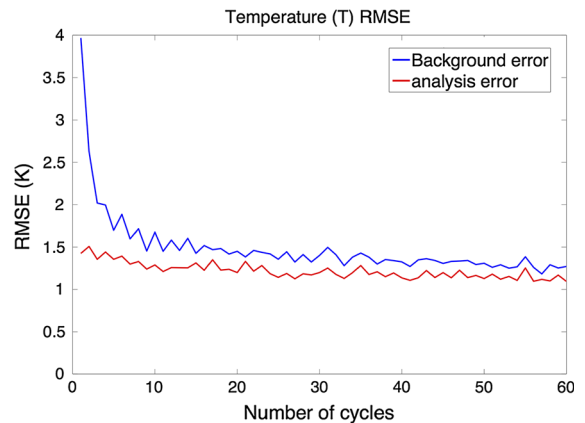

Figure 3

Time series of root mean square error (RMSE) for temperature during the first 15-day forecast-analysis cycles. The background RMSE is denoted in *blue*, and the analysis RMSE in *red lines*

analysis cycle and approaches to the level of the RMSE in analysis with time. Depending on the number of available observations and background uncertainties, the magnitude of the background RMSE fluctuates in a small scale, but converges to the level of the analysis RMSE. This behaviour indicates that the forecast-analysis cycle runs stably and ensemble forecast does not drift away from the true state.
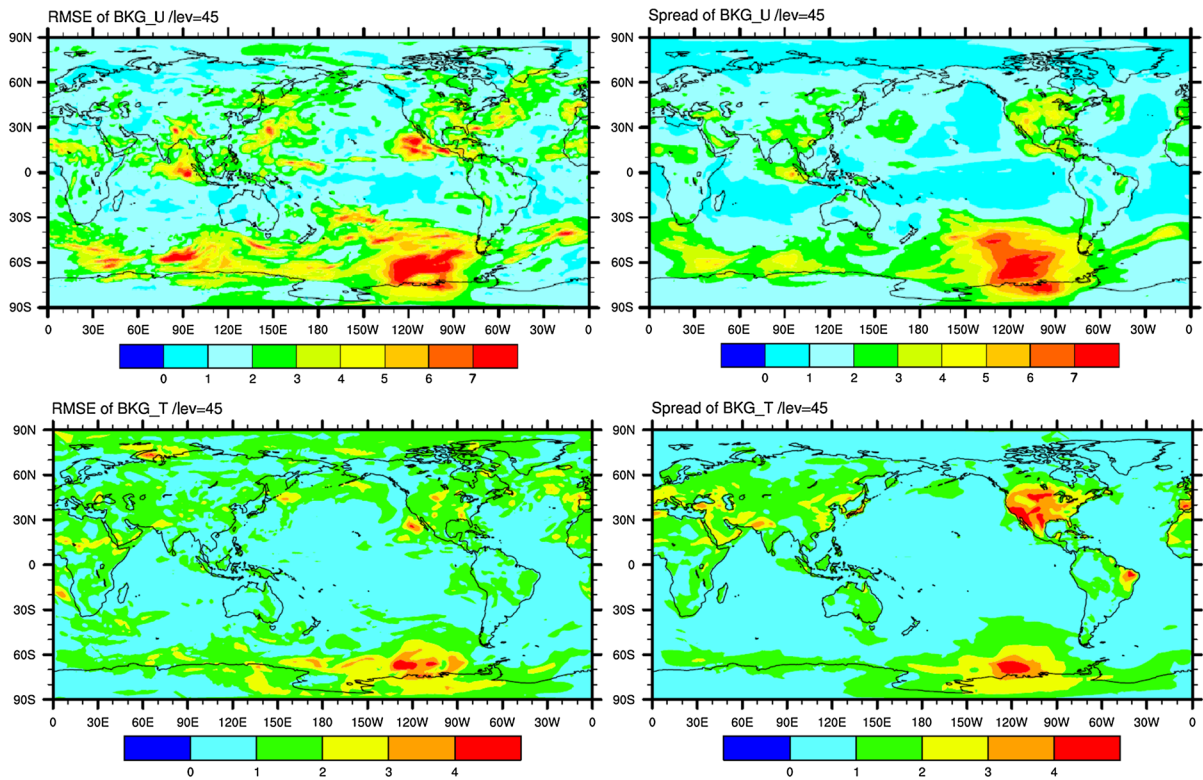
Figure 4

RMSE with respect to the truth (*left*) and ensemble spread (*right*) for zonal wind (*top*) and temperature (*bottom*) at 925 hPa averaged over the last 5 days during the 15-day forecast-analysis cycles

It is important for a stable forecast-analysis cycle run to estimate uncertainties of background reasonably and to represent background error covariance correctly. Thus, we further examine whether the scale of the ensemble spread is comparable to that of the background RMSE error. The ensemble spread is defined as a standard deviation of ensemble members with respect to the ensemble mean.

Figure 4 shows the background RMSE and the ensemble spread averaged over the last 5 days of the 15-day forecast-analysis cycle runs. Here we show the ensemble spread multiplied by the adaptive multiplicative inflation to reflect the effective ensemble spread that the data assimilation algorithm is actually identifying. The upper panel shows the RMSE and the ensemble spread of zonal wind at the level of 925 hPa, and the lower panel shows those of temperature at the same level. Mostly the ensemble spread is large in the area of large background error and the magnitude of spread is nearly equivalent to

that of errors in general. An outstanding feature in the pattern of the ensemble spread is the large spread in temperature over the northern America. That is, the analysis system tends to overestimate the background errors, while trying to avoid an underestimation with the inflation. This can result from that the adaptive multiplicative inflation is independent of variable and can be enhanced in the area of dense observation, where ensemble spreads decrease while the differences between the background and observation remains significantly larger than the spread of any analyzed variables (MIYOSHI 2011). We examine the magnitude difference between the background error and ensemble spread of the other variables, and found that the background error of specific humidity is significantly larger than its spread over the northern America (not shown here). This may lead to an enhanced inflation in that region for all variables used for analysis. In that way, the adaptive multiplicative inflation effectively hinders the underestimation of

the ensemble spread in that area of dense observation. Therefore, this result indicates that the estimation of background uncertainties is reasonable in this OSSE testing the KIAPS-LETKF implemented to the KIM, and the system properly avoids filter divergence.

4.2. Assimilation of Real Data

After we examine the performance of the KIAPS-LETKF implemented to the KIM in an ideal situation, real data assimilation has been carried out using the sonde and surface pressure observations. The real data assimilation experiment is performed for one-month period starting from 1 February 2014. We use NCEP PrepBufr synoptic and surface pressure observation data. As in the OSSE, we use 30 members of ensemble. The cluster that we use at the KIAPS has the Central Processing Unit (CPU) from INTEL Xeon 2.9 GHz RHEL 6.3. The computational time of LETKF is on average 14 min, and of 9-h model forecast is 45 min for 30 members when we use 20 computing nodes (16 processors per node).

For the evaluation of our analysis, we use the observation data that have been used for data assimilation and an independent data of ECMWF ERA-Interim reanalysis (DEE *et al.* 2011), respectively. The ECMWF reanalysis is produced by the Integrated Forecast system (IFS) at the ECMWF, which has been verified in long history and known as a qualified analysis through a data assimilation of diverse types of observations. Hence it is reasonable to assume that the ECMWF analysis can provide states of atmosphere close to reality. It might be desirable to use independent observation data for the evaluation, but we think that the evaluation using the NCEP PrepBufr can be complemented by that using the ECMWF reanalysis for a relevant interpretation.

Besides, comparison of short-range forecast with the observations is still useful to see the performance of data assimilation while comparison between analysis and the verification using observations could be a sanity check of the data assimilation system. For a quantitative monitoring of the error with reference to the ECMWF reanalysis, a remapping of data from the cubed sphere grid system onto the latitude-longitude grid system is required. We use a tool developed at the KIAPS for a conservative data
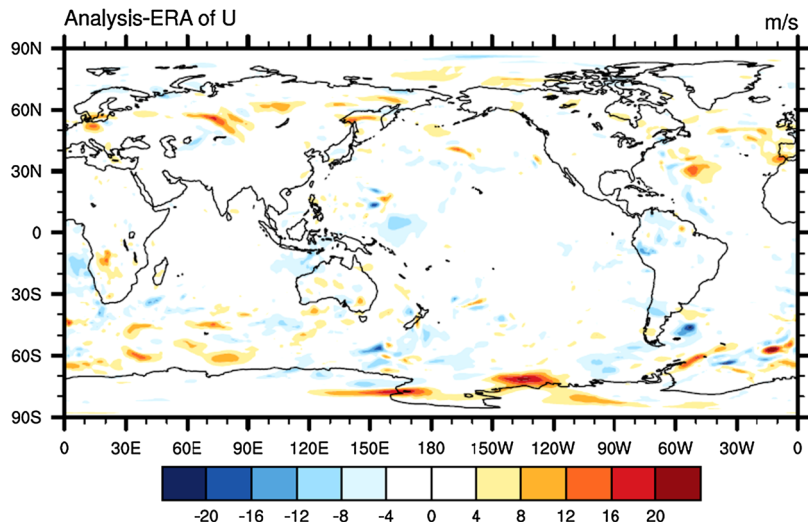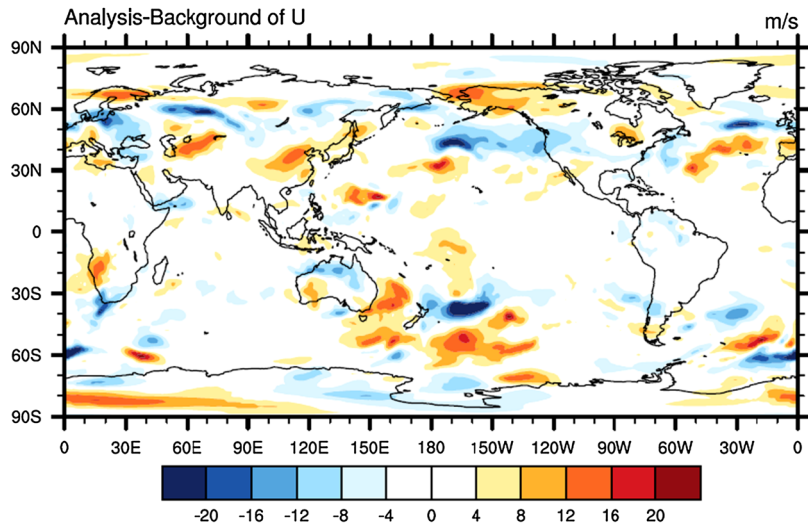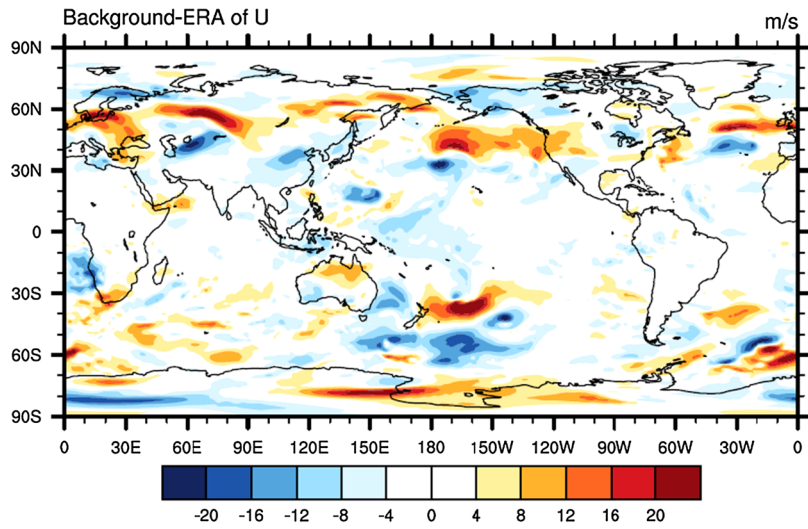
Figure 5 ▶

Differences between KIM background and ERA interim reanalysis (*upper panel*), between the background and KIAPS-LETKF analysis (*middle*), and between the KIAPS-LETKF analysis and the ERA reanalysis (*low panel*) for 925 hPa-zonal winds at 06 UTC on 01 February 2014

remapping on the sphere between two any grid systems (KIM *et al.* 2014). Also we interpolate the data from KIM vertical levels to the pressure levels defined as in the ECMWF reanalysis data. The number of vertical levels of pressure coordinate is 37 from 0.1 hPa to 1000 hPa. The initial ensemble is composed of model states that are obtained from model simulation in 12-h interval in order to have a sufficient spread at the initial time.

Figure 5 shows the background and analysis errors of zonal winds at 925 hPa after one forecast-analysis cycle at 06 UTC 01 February. Here we regard the ECMWF reanalysis as the true state for evaluation, and look at the Root Mean Square Difference (RMSD) between an ensemble mean and the reanalysis. Although the background error of initial ensemble is large (top of Fig. 5), the analysis increment compensates it significantly after the first analysis cycle (middle of Fig. 5). Therefore, analysis resulted from the KIAPS-LETKF gets closer to the ECMWF reanalysis after one cycle of data assimilation using conventional data only. If one considers that ECMWF assimilates various kinds of observations in addition to the conventional data, this shows promising performance of the KIAPS-LETKF data assimilation system as the first experiment under an operational setting. The decrease in the magnitude of RMSD is especially effective in the Northern Hemisphere where more dense observations are available.

We also examine changes in the vertical profile of background and analysis Root Mean Square Difference (RMSD) and bias in comparison to the sonde data during the forecast-analysis cycle (Fig. 6). The analysis of KIAPS-LETKF shows much smaller bias and RMSD than the background, and that difference is especially large in the middle troposphere after the first forecast-analysis cycle, at 06 UTC 01 February 2014. As the data assimilation cycles are repeated, the profile of bias and RMSD with respect to the observations becomes stabilized, and thus the gap of the profiles between background and analysis gets
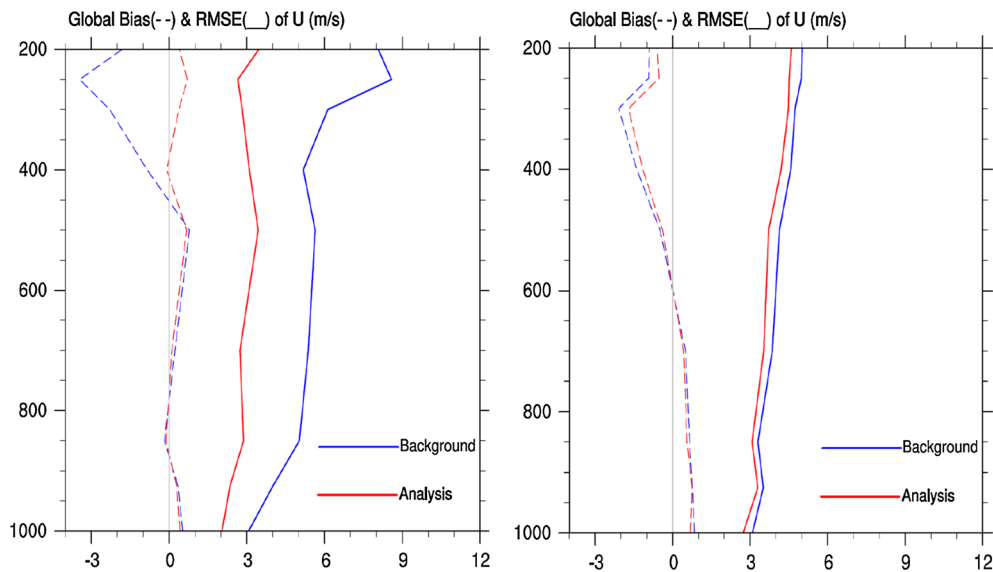
Background-ERA of U   m/s

Analysis-Background of U   m/s

Analysis-ERA of U   m/s

Figure 6
Vertical profiles of global mean bias (*dashed lines*) and RMSD (*solid lines*) with respect to the NCEP prepbufr data, for the background (*blue*) and the analysis (*red*) of the zonal wind at 06 UTC on 01 February 2014 (*left*) and at 06 UTC on 28 February 2014 (*right*)

small at 06 UTC on 28 February when the forecast-analysis cycle proceeds 4 weeks. This illustrates that KIAPS-LETKF data assimilation well reflects the observations overall.

Figure 7 shows the vertical profiles of background and analysis differences from the ECMWF reanalysis. The background and analysis are remapped onto the grid system defined in the ECMWF reanalysis and then compute the global mean values of the bias and RMSD at each pressure level of ECMWF data. Here we show the RMSD of zonal wind at 06 UTC 01 and at 06 UTC 28 February, respectively (Fig. 7). As shown in the comparison with the NCEP PrepBufr data (Fig. 6), the analysis RMSD of zonal wind with reference to the ECMWF data becomes much smaller than the background RMSD in the whole troposphere after the first cycle of forecast-analysis. After 4 weeks of forecast-analysis cycle, the profile of analysis from our assimilation and the background RMSDs become similar to each other, which means that the analysis increments are not as large as during the early stage of the forecast-analysis cycle.

The decrease of analysis increments with time may indicate the convergence of background and analysis, but also it can imply that the background uncertainties may be underestimated due to a decrease in the ensemble spread. However, in real data assimilation, background error covariance could not fully reflect forecast uncertainties and thereby ensemble spread is significantly limited compared to errors. Indeed, this is why we have incorporated the adaptive multiplicative inflation introduced by MIYOSHI (2011) to better represent background uncertainties when the difference between the background and observation is large. MIYOSHI and KUNII (2012) also used the adaptive multiplicative inflation in real data assimilation using the LETKF implemented to the Weather Research and Forecast (WRF) model (SKAMAROCK *et al.* 2005). The adaptive inflation is originally designed to balance between the departure of background from observation and ensemble spread. However, they found that the magnitude of ensemble spread became much smaller than the RMSD compared to the NCEP analysis as the forecast-analysis cycle runs repeat although the adaptive inflation method was applied. They assumed that the ensemble spread tended to be small and uncertainties of background states were underestimated when adaptive multiplicative inflation was not spun-up sufficiently. Therefore, we also tune the parameter of variance v_i^b of Eq. (4) for the estimation of adaptive inflation factor.
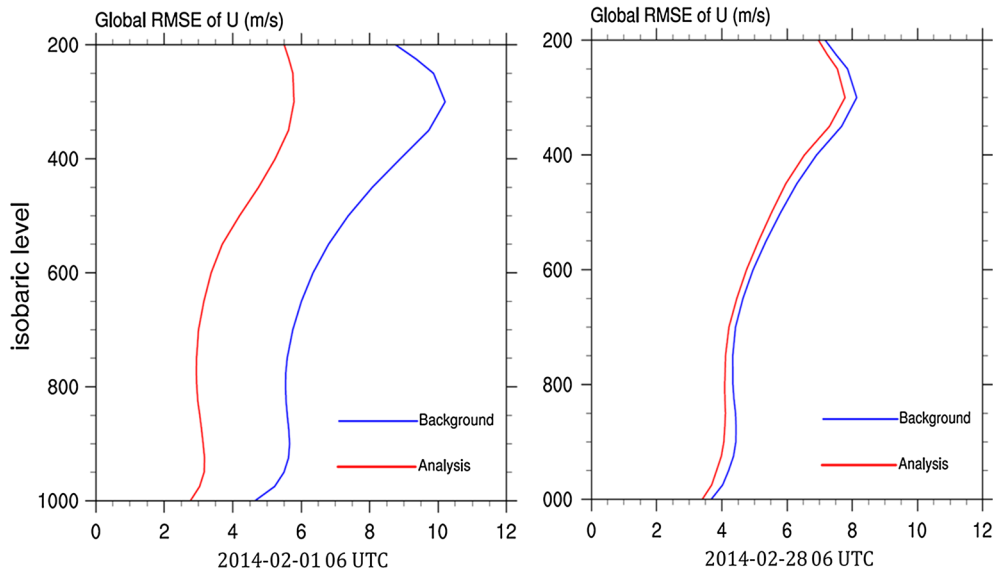
Figure 7
Vertical profiles of global mean RMSDs with respect to the ERA Interim data, for the zonal wind of background (*blue*) and analysis (*red*) at 06 UTC on 01 February 2014 (*left*) and at 06 UTC on 28 February 2014 (*right*)

The variance of inflation parameter represents the time-smoothing strength, and thus spin-up proceeds faster, but temporal fluctuation increases if the variance increases. As discussed in Sect. 3.3, we increase the variance from $v_i^b = 0.01^2$ to $v_i^b = 0.04^2$, to accelerate the spin-up and repeat the experiment using real observations. Figure 8 shows the time series of globally-averaged RMSD of the two KIAPS-LETKF analyses with different v_i^b in comparison to ECMWF reanalysis data. We examine the RMSD of zonal wind at 850 hPa. The black (red) line shows the result from the experiment using $v_i^b = 0.01^2$ ($v_i^b = 0.04^2$). At the beginning, the performance difference is negligible, but the performance becomes better in the analysis with the larger variance after about 25 cycles, although temporal fluctuations in RMSD are increased in that case. Figure 9 shows the vertical profiles of both bias and RMSD from the NCEP PrepBufr data averaged over the period 06 UTC 15 ∼ 18 UTC 27 February. The RMSDs of temperature and zonal wind are evidently smaller in the whole troposphere when we use $v_i^b = 0.04^2$. A similar pattern in the vertical profiles is found in the comparison with the ECMWF reanalysis (Fig. 10). We found that the results from the KIAPS-LETKF analysis are significantly
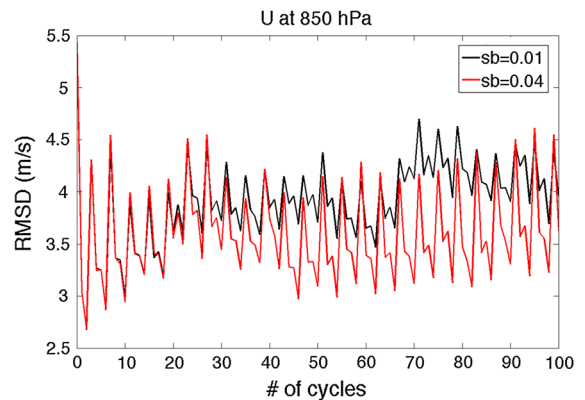

Figure 8
Time series of the root mean square difference (RMSD) of the zonal wind (U) analysis at 850 hPa with respect to the ECMWF reanalysis between 06 UTC 01 and 00 UTC 26 February 2014. The *black solid line* shows the result from the test using $v_i^b = 0.01^2$ which is denoted by sb = 0.01 in the legend, and *red line* with *diamond markers* shows the case using $v_i^b = 0.04^2$

improved when the spin-up of inflation parameter proceeds faster. We also examine if the increase of the variance also affects the scale of ensemble spread. Figure 11 shows the time series of horizontal mean ensemble spread of the zonal wind at 850 hPa in the experiments using the two different variances. The magnitude of ensemble spread is initially about half
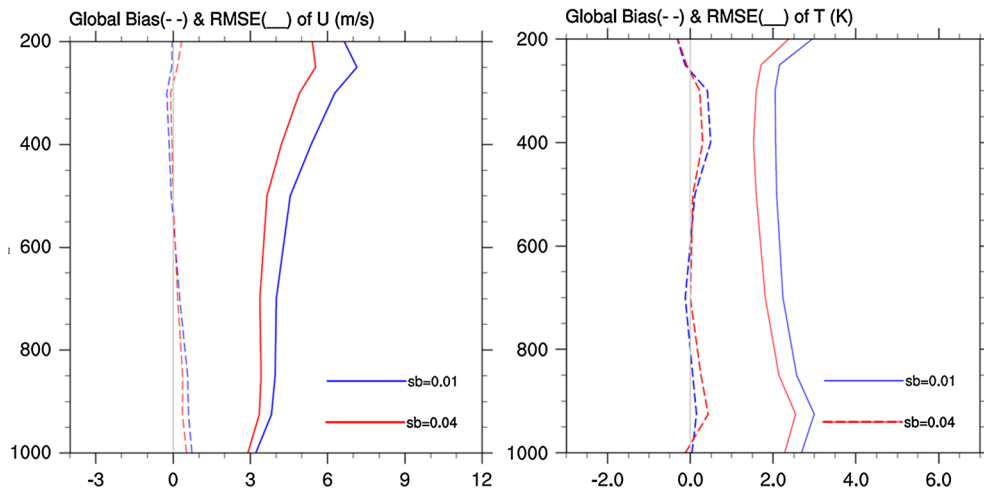
Figure 9

Vertical profiles of global mean bias (*dashed lines*) and RMSD (*solid lines*) of zonal wind (*left*) and temperature (*right*) analysis with reference to the sonde data from NCEP PrepBufr, for the case using the prior inflation variance $v_i^b = 0.01^2$ denoted by sb = 0.01 in blue, and for the case using $v_i^b = 0.04^2$ denoted by sb = 0.04 in red line. These are time-averaged values for the period between 06 UTC on 15 February and at 18 UTC on 27 February 2014
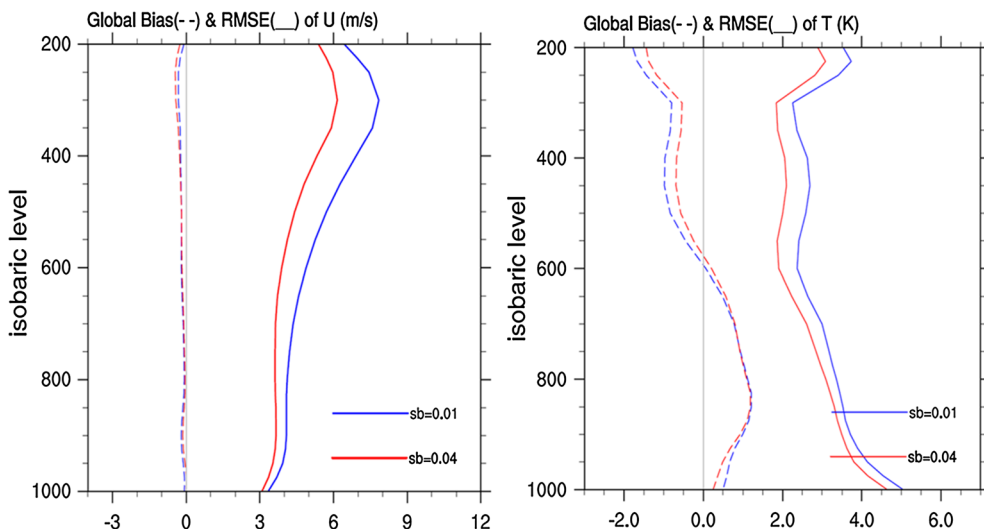

Figure 10
The same as Fig. 9, except with reference to the ECMWF reanalysis data

of the RMSD from the ECMWF analysis. However, the ensemble spread drops rapidly during the early stage of the cycle in both experiments. Once the spread drops, the spread remains nearly unchanged for long period when the variance of the adaptive inflation $v_i^b = 0.01^2$. It is the concerning case that may cause filter divergence at the end. Meanwhile, the spread becomes gradually grow again when the

raised variance for the inflation parameter is used. It may help the LETKF algorithm better estimate the uncertainties of backgrounds and take more observations to be reflected in the analysis and to avoid a filter divergence. This result indicates that we need to optimize a relevant growth rate for the multiplicative inflation as the performance of the system is affected by the choice of the parameter.

Finally we examine the geopotential height field in the analyses produced 6-hourly from the forecast-analysis cycle runs for a month, and compare with those from the ECMWF reanalysis. We compute the differences in the 500 hPa-geopotential height from the ECMWF reanalysis for the period 06 UTC 13 February $\sim$00 UTC 28 February, and then average the differences. Also we repeat that comparison for the forecasts produced by the integration of the KIM for one-month with the initial condition obtained from the Global Forecast System (GFS) reanalysis (Environmental Modeling Center 2003). The GFS reanalysis at the initial time is quite close to the ECMWF reanalysis (not shown here). Figure 12 shows that analysis from the forecast-analysis cycle run is much closer to the ECMWF reanalysis than the

forecast in most areas on the globe. This result indicates that the KIAPS-LETKF shows a promising performance in forecast-analysis cycle runs for weather forecast, given forecast model and observation data.

5. Summary

We develop the KIAPS-LETKF system with the KIM, a newly developed global NWP model at the KIAPS. The major tool added to the preexisting LETKF technique is the new interpolation algorithm for the observation operator h in using the forecast fields with unstructured grid system on the cubed sphere. Also we construct a computing routine for the forecast-analysis cycle, in harmony with the KIM modeling framework.

The KIAPS-LETKF system is evaluated using the OSSE and data assimilation using NCEP PrepBufr containing conventional observation data such as sonde and surface pressure observations. The forecast-analysis cycle proceeds fine in the OSSE and the analysis errors of prognostic variables are much lower than the background errors just after one cycle. The background errors of all variables decrease as the cycle repeats, and the magnitude of errors approaches to the level of analysis errors. This indicates that the ensemble data assimilation system shows a reasonable performance and motivates us to perform real data assimilation for further verification of the system.
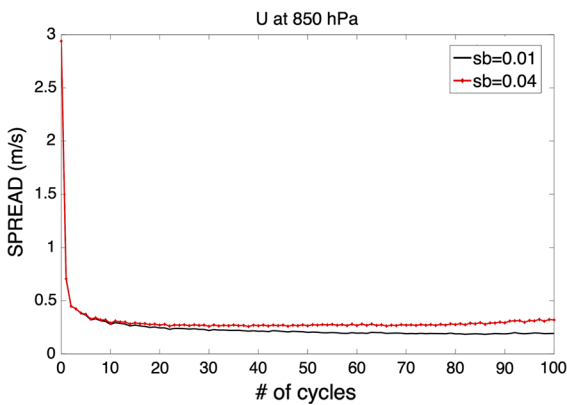

Figure 11
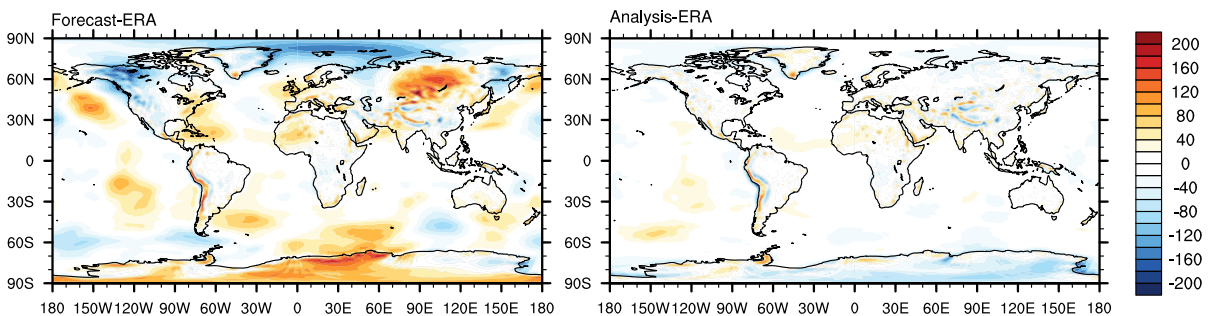The same as Fig. 8, but ensemble spread in each test case

Figure 12
Time-averaged geopotential height field difference (in meter) at 500 hPa between the single forecast by KIM and the ECMWF reanalysis (*left*), and between the analysis from the KIAPS-LETKF assimilation run and the ECMWF reanalysis (*right*) for the period between 06 UTC on 13 and 00 UTC on 28 February 2014

For the quantitative evaluation of the KIAPS-LETKF performance in real data assimilation experiment, NCEP PrepBufr data and ECMWF reanalysis are used and both bias and RMSD are computed. Results are consistently encouraging: (1) there is significant error reduction in the early stage and (2) background and analysis converge in time, and the forecast-analysis cycle runs stably. However, the difference between the magnitude of ensemble spread and RMSD is much larger than that estimated in the OSSE. This may imply that uncertainties of system can be underestimated in real data assimilation experiments. In the OSSE only initial errors of ensemble exist as the forecast model is assumed to be perfect. However, in real data assimilation background error covariance may not fully reflect forecast uncertainties and thereby ensemble spread is significantly limited compared to errors when the multiplicative inflation is not sufficiently spun-up. We increase the variance of inflation parameter to accelerate the spin-up of the multiplicative inflation, and this leads to a better performance of the KIAPS-LETKF. The value of RMSD from the ECMWF data is reduced and the ensemble spread grows up again gradually after it drops at the initial forecast-analysis cycle. We may need to take further consideration of using an optimal value of the prior variance of inflation parameter which can affect the performance of the KIAPS-LETKF system. In addition, we started investigating the use of an additive inflation (YANG et al. 2015) as a complement to the multiplicative inflation to handle such problems as sampling and model errors. Moreover, we intend to assimilate additional types of observations such as microwave radiance data and GPS-RO.

Acknowledgments

REFERENCES

BUEHNER, M., P. L. HOUTEKAMER, C. CHARETTE, H. L. MITCHELL, B. HE, 2010: *Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP, Part II: One-month experiments with real observations.* Mon. Wea. Rev., *138*, 1567-1586.

DENNIS, J. M, J. EDWARDS, K. J. EVANS, O. GUBA, P. H. LAURITZEN, A. A. MIRIN, A. ST-CYR, M. A. TAYLOR, and P. H. WORLEY, 2012: CAM-SE: *A scalable spectral element dynamical core for the Community Atmosphere Model.* Int. J. High Perf. Comput. Appl., *26*, 74–89.

DEE, D. P., UPPALA S. M., A. J. SIMMONS, P. BERRISFORD, P. POLI, S. KOBAYASHI, U. ANDRAE, M. A. BALMASEDA, G. BALSAMO, P. BAUER, P. BECHTOLD, A. C. M. BELJAARS, L. van de BERG, J. BIDLOT, N. BORMANN, C. DELSOL, R. DRAGANI, M. FUENTES, A. J. GEER, L. HAIMBERGER, S. B. HEALY, H. HERSBACH, E. V. HÓLM, L. ISAKSEN, P. KÅLLBERG, M. KÖHLER, M. MATRICARDI, A. P. MCNALLY, B. M. MONGE-SANZ, J.-J. MORCRETTE, B.-K. PARK, C. PEUBEY, P. de ROSNAY, C. TAVOLATO, J.-N. THÉPAUT, and F. VITART, 2011: *The ERA-interim reanalysis: Configuration and performance of the data assimilation system.* Quart. J. R. Meteorol. Soc., *137*, 553–597.

Environmental Modeling Center, 2003: The GFS Atmospheric Model. NCEP Office Note 442, Global Climate and Weather Modeling Branch, EMC, Camp Springs, Maryland.

EVANS, K. J., P. H. LAURITZEN, S. K. MISHRA, R. B. NEALE, M. A. TAYLOR, and J. J. TRIBBIA, 2013: AMIP *Simulation with the CAM4 spectral element dynamical core.* J. Climate., *26*, 689–709.

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, 10143–10162.

GASPARI, G., and S. E. COHN, 1999: *Construction of correlation functions in two and three dimensions.* Quart. J. Roy, Meteor. Soc., *125*, 723–757.

HARLIM, J., and B. R. HUNT, 2007: *Four-dimensional local ensemble transform Kalman filter: numerical experiments with a global circulation model.* Tellus, *59A*, 731–748.

HILL, C., C. DELUGA, V. BALAJI, M. SUAREZ, and A. da SILVA, 2004: *The architecture of the earth system modeling framework.* Computing in Science and Engineering Society, *6*, 18–28.

HUNT, B. R., E. J. KOSTELICH, and I. SZUNYOGH 2007: *Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter.* Physica D., *230*, 112–126.

HOUTEKAMER, P. L., and H. L. MITCHELL, 1998: *Data Assimilation Using an Ensemble Kalman Filter Technique*. Mon. Wea. Rev., *126*, 796–811.

KANG, J.-S, B. -J. JUNG, H. -W. JUN, J. -H. KIM, S. SHIN, Y. JO, H. KWON, 2014: Progress and Plans of KIAPS-LETKF: 4D-LETKF, *Radiance Data Assimilation, and Implementation to KIAPS-FM*. Fall meeting of K Korean Meteorological Society, 103–105

KIM, K.-H., D. JOO, Y. LEE, and S. SHIN, 2014: Geometrically conservative remapping between spherical grids using Voronoi diagram (V-GECoRe), *Spring meeting of Korean Meteorological Society*, 27–28.

KWON, H., J.-S. KANG, Y. JO, and J. H. KANG, 2015: *Implementation of a GPS-RO data processing system for the KIAPS-LETKF data assimilation system*. Atmos. Meas. Tech, *8*, 1259–1273.

LI, H., E. KALNAY, AND T. MIYOSHI, 2009: *Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter*. Quart. J. Roy, Meteor. Soc., *135*, 523–533.

MIYOSHI, T., 2011: *The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter*. Mon. Wea. Rev., *139*, 1519–1535.

MIYOSHI, T., and K. ARANAMI, 2006: *Applying a Four-dimensional Local Ensemble Transform Kalman Filter (4D-LETKF) to the JMA Nonhydrostatic Model (NHM)*. SOLA, *2*, 128–131.

MIYOSHI, T., and M. KUNII, 2012: *The Local Ensemble Transform Kalman Filter with the Weather Research and Forecasting Model: Experiments with Real Observations*. Pure Appl. Geophys., *169*, 321–333.

MIYOSHI, T., Y. SATO, and T. KANDOWAKI, 2010: *Ensemble Kalman Filter and 4D-Var Intercomparison with Japanese Operational Global Analysis and Prediction System*. Mon. Wea. Rev., *138*, 2846–2866.

MIYOSHI, T., and S. YAMANE, 2007: *Local ensemble transform Kalman filtering with an agcm at a T159/L48 resolution*. Mon. Wea. Rev., *135*, 3841–3861

MIYOSHI, T., S. YAMANE, and T. ENOMOTO, 2007: *Localizing the error covariance by physical distances within a Local Ensemble Transform Kalman Filter (LETKF)*. SOLA, *3*, 89–92.

NERGER, L., HILLER, W. 2013: *Software for Ensemble-based Data Assimilation Systems - Implementation Strategies and Scalability*. Computers and Geosciences, *55*, 110–118.

SADOURNY, R., 1972: *Conservative finite-difference approximations of the primitive equations on quasi-uniform spherical grids*. Mon. Wea. Rev., *100*, 136–144.

SHIN, S., J. KIM, S.-Y. JUN, D. JOO, K.-H. KIM, Y. LEE, and I.-S. SONG, 2014: Progress in the development of the 3D-KIAPSGM framework, *Fall meeting of Korean Meteorological Society*, 25–26.

SKAMAROCK, W. C., J. B. KLEMP, J. DUDHIA, D. O. GILL, D. M. BARKER, W. WANG, and J. G. POWERS, 2005: A description of the Advanced Research WRF version 2. NCAR Tech. Note TN-468 + STR

TAYLOR, M. A., J. TRIBBIA, and M. ISKANDARANI, 1997: *The spectral element method for the shallow water equations on the sphere*. J. Comput. Phys., *130*, 92-108.

TERASAKI K., M. SAWADA, and T. MIYOSHI, 2015: *Local Ensemble Transform Kalman Filter experiments with the Nonhydrostatic Icosahedral Atmospheric Model NICAM*. SOLA, *11*, 23-26.

WHITAKER, J. S., and T.M. HAMILL, 2012: *Evaluating methods to account for system errors in ensemble data assimilation*. Mon. Wea. Rev., *140*, 3078-3089.

YANG, S.-C., E. KALNAY, and T. ENOMOTO, 2015: *Ensemble singular vecotrs and their use as additive inflation in ENKF*. Tellus A, *67*, 26536, http://dx.doi.org/10.3402/tellusa.v67.26536