Asia-Pac. J. Atmos. Sci., 54(s), 351-360, 2018
DOI:10.1007/s13143-018-0022-2

pISSN 1976-7633 / eISSN 1976-7951

Real Data Assimilation Using the Local Ensemble Transform Kalman Filter (LETKF) System for a Global Non-hydrostatic NWP model on the Cubed-sphere

Seoleun Shin[1], Jeon-Ho Kang[1], Hyoung-Wook Chun[1], Sihye Lee[1], Kwangjae Sung[1], Kyoungmi Cho[1], Youngsoon Jo[1], Jung-Eun Kim[1], In-Hyuk Kwon[1], Sujeong Lim[1], and Ji-Sun Kang[2]

[1]*Korea Institute of Atmospheric Prediction Systems (KIAPS), Seoul, Korea*
[2]*Korea Institute of Science and Technology Information (KISTI), Daejeon, Korea*

(Manuscript received 28 February 2018; accepted 19 June 2018)

Abstract: An ensemble data assimilation system using the 4-dimensional Local Ensemble Transform Kalman Filter is implemented to a global non-hydrostatic Numerical Weather Prediction model on the cubed-sphere. The ensemble data assimilation system is coupled to the Korea Institute of Atmospheric Prediction Systems Package for Observation Processing, for real observation data from diverse resources, including satellites. For computational efficiency in a parallel computing environment, we employ some advanced software engineering techniques in the handling of a large number of files. The ensemble data assimilation system is tested in a semi-operational mode, and its performance is verified using the Integrated Forecast System analysis from the European Centre for Medium-Range Weather Forecasts. It is found that the system can be stabilized effectively by additive inflation to account for sampling errors, especially when radiance satellite data are additionally used.

Key words: Ensemble data assimilation, local ensemble transform Kalman filter (LETKF), numerical weather prediction (NWP), atmospheric global model (AGM)

1. Introduction

The Ensemble Kalman Filter (EnKF) was first suggested by Evensen (1994) for atmospheric and oceanic data assimilation. Thereafter, EnKF was applied to an atmospheric system by Houtekamer and Mitchell (1998), and extensive investigations of the ensemble technique with operational interests have been made. For a numerical weather prediction (NWP) system, data assimilation provides optimal analysis of atmospheric states by making uses of observation data and numerical solution from a mathematical model under Bayesian paradigm. Recently, most operational centers consider combinations of the two major methods, such as a hybrid ensemble-variational data assimilation (e.g., Buehner et al., 2013; Kleist and Ide, 2015; Lorenc et al., 2015).

At the Korea Institute of Atmospheric Prediction Systems (KIAPS), we are developing a hybrid 4-dimensional ensemble variational system (Song et al., 2017) aiming for operational uses. For the applications in the hybrid system, we have developed a Local Ensemble Transform Kalman Filter (LETKF) system (Shin et al., 2016) using the 4D-LETKF algorithms (Hunt et al., 2007). The LETKF system had been implemented to the hydrostatic global weather forecast model, which is based on the spectral element method for discretization of governing equations on the cubed-sphere (Sadourny, 1972). The performance of the LETKF system has been evaluated using the Observing System Simulation Experiment (OSSE), and real data assimilation using the National Centers for Environmental Prediction (NCEP) PREPBUFR files, containing conventional observation data, such as sonde, aircraft, satwind, and surface pressure observations (Shin et al., 2016). The experience in those implementation works indicated that the applications of covariance inflation methods were important for real data assimilation, for which we need to handle underestimation of background uncertainties properly, to stabilize the ensemble data assimilation system. More details on the implementation can be found in Shin et al. (2016).

In practice of ensemble data assimilation, it is required to use inflation methods to deal with limitations such as the underestimation of background error variance and a localization to treat sampling errors. Multiplicative inflation (Anderson and Anderson, 1999) changes the magnitude of the background error covariance by multiplying a factor. Adaptive multiplicative inflation methods have been suggested (e.g. Anderson, 2009; Miyoshi, 2011) and they demand less manual tuning to avoid unnecessary covariance inflation where observation is sparse or absent. In this study, we use the adaptive multiplicative inflation algorithm developed by Miyoshi (2011) and use inflation factors updated at each analysis time and at every grid point. Multiplicative inflation methods tend to amplify (or lessen) the pre-existing perturbations. Additive inflation has been suggested to help capturing unrepresented errors associated with dynamical activity by adding perturbations which may emphasize baroclinically growing error structures (Whitaker et al., 2008). There are also relaxation methods such as relaxation to prior spread and relaxation to prior perturbation, and their characteristics are being investigated (e.g. Zhang et al., 2004; Whitaker and Hamill, 2012; Kotsuki et al., 2017). These methods inflate perturbations that are affected by observations only so that deflation does not occur (Kotsuki et al.,

Corresponding Author: Seoleun Shin, Korea Institute of Atmospheric Prediction Systems (KIAPS), Seoul 07071, Korea.
E-mail: seoleuns@gmail.com

2017). Although we have tested the adaptive multiplicative and additive inflation in this study, we will further implement the relaxation methods in our system.

The non-hydrostatic version of the Korean Integrated Model (hereafter, KIM) is available for tests (Hong et al., 2018). Thus, we modify the LETKF framework for the forecast model and couple the LETKF system to the KIAPS Package for Observation Processing (hereafter, KPOP, Kang et al., 2018) system for real data assimilation, instead of using the NCEP PREPBUFR data. For this coupling, we split the observation operator from the LETKF system, and instead use the observation operator implemented in the KPOP system (Kwon et al., 2015). Moreover, we optimize the infrastructure of the system especially the handling of several output files for semi-operational uses. In this article, we explain the main features of the LETKF framework coupled with KIM and KPOP and discuss its performance in real data assimilation. It is found that the use of additive inflation would be essential, particularly when the volume of observations increases because the multiplicative inflation might not be sufficient to use satellite observations whose positions are temporally changing. Also, the effect of scaling factor for additive inflation on the performance of the system is discussed.

In the next section, we briefly introduce the forecast models, KIM, KPOP, and LETKF, respectively. Here, we also explain the software engineering aspects of the LETKF system focusing on its efficient file handling and computation. Section 3 describes experimental design for the evaluation of the performance of ensemble data assimilation. In section 4, we discuss the results from the real data assimilation experiments. The final section summarizes the current work and presents some future plans.

2. System components

a. KIM

A non-hydrostatic global atmospheric model for NWP has been developed at the KIAPS, which uses the spectral element method, and is formulated on the cubed-sphere with an unstructured grid system (Choi and Hong, 2016). The horizontal spatial resolution used for this study is based on the 360 elements per face and 4 Gauss-Legendre-Lobatto (GLL) points per element, and thereby the average grid spacing is about 50 km. The model has 50 vertical levels, which is extended up to 0.3 hPa. The performance of the model is verified in experiments, such as heavy rainfall cases and tropical-cyclone events (Choi and Hong, 2016).

b. KPOP

(a) Main KPOP platform

For an optimal estimation of initial conditions, observation data go through a sophisticated observation quality control (QC) and bias correction (BC) before they are used in a data assimilation system. The current version of the KPOP can handle almost all observation types that are currently used in the Korea Meteorological Administration (KMA) operational NWP system. These observation data include the sonde, surface, aircraft, Global Positioning System-Radio Occultation (GPS-RO), Infrared Atmospheric Sounding Interferometer (IASI), Advanced Microwave Sounding Unit-A (AMSU-A), Cross-track Infrared Sounder (CrIS), Microwave Humidity Sounder (MHS), Advanced Technology Microwave Sounder (ATMS), and Atmospheric Motion Vectors (AMVs).

All of the raw observation data are obtained from the KMA in Binary Universal Form for the Representation of Meteorological data (BUFR) that are merged or re-constructed after ingested from the Global Telecommunication System of the World Meteorological Organization. In the KPOP system, observations near poles are treated in the same way as in the other parts of the globe in our current system. Bilinear interpolation is performed to estimate background quantities on the position of observations. For possible problems of analysis discontinuities, Yamazaki et al. (2017) suggested an improved way to handle observations close to the poles. This issue is beyond the scope of this paper and will be examined in the future.

To assimilate the radiance data, Radiative Transfer for the Television Infrared Observation Vertical Sounder model (RTTOV) version 10.2 is used as the observation operator for computing the radiance from the model profiles of temperature and moisture. Among the microwave observations, the AMSU-A data onboard the five polar-orbiting satellites, NOAA-15/18/19 and MetOp-A/B, are used to obtain the atmospheric temperature sounding. AMSU-A channels 5 to 10 provide global coverage over the oceans. The QCs for clear sky assimilation of AMSU-A are strictly applied to remove the pixels contaminated by cloud, precipitation, and sea-ice, using algorithms suggested by Grody et al. (1999, 2001). The KPOP system employed an adaptive BC scheme for radiance data, based on Harris and Kelly (2001). The bias of radiance observation is represented as the sum of scan bias terms at different zenith angles and air-mass bias terms, depending on the thermodynamic properties of the underlying atmosphere. Global multiple linear regression of the scan-corrected innovations against two model predictors of 850-300 and 200-50 hPa thicknesses is employed to correct the air-mass bias. These scan and air-mass bias correction coefficients are automatically calculated and updated at each assimilation cycle, i.e., every 6 h. Bias-corrected AMSU-A data are selected by a thinning process with the resolution of 1.5^{o} latitude-longitude grid. The values of observation error for AMSU-A are given as follows: 0.5 K for channel 5~8; 0.8 K for channel 9~10. As for ATMS, we apply the same conditions as AMSU-A, except that we use 5.0 K as the value of the observation error for water vapor.

Likewise, the MHS onboard the three polar-orbiting satellites, NOAA-18/19 and MetOp-A are used to obtain the atmospheric water vapor sounding. MHS channels 3 to 5 with land masking are used for data assimilation. Fundamentally, the

MHS and AMSU-A have a very similar QC process to each other. Cloud-contaminated pixels of MHS are removed with the empirical thresholds of the first-guess departure. After processing BC and thinning with 1.0° latitude-longitude grid, observation data from MHS is finally prepared to be passed to the data assimilation system. The thinning resolution is chosen individually depending on the type of satellite observation at most of operational NWP center and the choice is made in consideration of the characteristics of each observation type. It is found that the analysis accuracy has been improved when we select higher thinning resolution (1.0°) for MHS than the thinning resolution (1.5°) for AMSU-A in the tests using the 3D-VAR system at KIAPS. Since spatial correlation error is depending on the type of observation and MHS data is used for humidity, it might be reasonable to choose a higher thinning resolution for MHS than for AMSU-A. Nevertheless, this reasoning needs to be reexamined by using the latest version of the data assimilation system at KIAPS. Some other operational institution like UK-Met Office chooses thinning resolution differently also for equatorial, mid-latitude, and polar regions, respectively, but we apply a fixed value of thinning resolution everywhere. The observation error for MHS is set to be 5 K in this study.

Among the infrared observations, the IASI onboard MetOp-A/B and CrIS onboard Suomi NPP are used to provide the observed atmospheric temperature profile. In this study, we use 60 IASI channels, taken from the Met Office dataset of 138 channels (Hilton et al., 2009). Due to the difficulties in specifying surface emissivity over land and ice, the mid and upper atmospheric temperature sounding channels are selected, whose sensitivity peaks of weighting function are located around 50-700 hPa. In the case of CrIS, 20 channels with peaks at higher levels (50-200 hPa) were taken from the Met Office dataset of 134 channels in Smith et al. (2015). Although the IASI and CrIS channels differ from each other, the quality control processes are identical. IASI and CrIS observations at a particular spectrum in which they are free from the effects of clouds are identified, using the cloud detection scheme based on the background departure in McNally and Watts (2003). For the BCs we use the same scheme as that for the AMSU-A, but the cloud screening and BC are repeated five times because those two processes are heavily dependent on each other's quality. Data has been thinned at a scale of about 300 km so that the residual data are 2% of the full data set. We had tested a thinning distance like the microwave observations for the infrared observations, but the performance was rather worsened than when using a longer thinning distance. For this reason, the thinning distance about 300 km (3.0°) has been chosen for the infrared observation. However, these tests were carried out with the old versions of KIM and we plan to optimize the thinning distance and observation error for individual observations using the latest KIM. For the LETKF system at KIAPS, we currently use 60 channels for the upper to middle part of the troposphere and the observation errors are 1.0 K for ch 16~185 (upper air temperature), 0.7 K for ch 187~386 (middle air temperature). As for CrIS, we use 0.3~0.4 K for the upper channels used in our study.

Clear Sky Radiance (CSR) from geostationary satellites was also used. CSR data are available from Himawari, Meteosat, GOES, and Communication, Ocean and Meteorological Satellite (COMS), but currently we use only CSR data provided by COMS. The data is obtained from a water vapor sensitive channel in COMS. Since CSR is already cloud-screened data, we performed simple outlier removal quality control and bias correction without going through additional cloud screening processes. As CSR data are for water vapor, the observation error is set to be 4 K. We use diagonal part of observation error covariance only currently. However, it is a trend to decide observation errors by considering inter-channel correlation (e.g. Bormann et al., 2010; Weston et al., 2014; Bormann et al, 2016) and we may consider this treatment in future.

AMVs are derived by tracking the features of infrared, water vapor, and visible-channel imagery onboard sequential geostationary (MetOp-7/10, GOES-13/15, Himawari, and COMS) and polar-orbiting (NOAA-15/18/19, MetOp-A/B, and Terra/Aqua MODIS) satellites. However, the AMV fields from geostationary satellites are not derived poleward beyond 60° latitude, and those from polar-orbiting satellites are not available equatorward below 70° latitude. The vertical coverage of AMVs is up to approximately 100 hPa. Some AMVs of poor quality are eliminated, namely those with quality indicators (QIs) lower than the threshold of 0.85, and with the innovation check for each level. AMV errors originate mainly from two sources: errors in the height assignment and errors in the wind vector tracking (Salonen et al., 2015). The AMV errors were estimated simply based on innovation statistics in this study, although in some other operational institution the error was estimated in consideration of some other factors such as wind shear and retrieval algorithms. The AMV error in this study ranges 2.8~11.8 m s^{-1} depending on the reference height.

(b) KPOP-Mini for ensemble background states

There is a simple version of the observation operator for LETKF that we term it "KPOP-Mini." With this tool, we obtain background states interpolated at the position of observations. Our strategy is to apply the observation operator to background states (first-guess) of each ensemble member in such a way that all members use the same observations. To achieve this, the coupling of the KPOP system to the LETKF system consists of two parts: main KPOP and KPOP-Mini.

For the execution of the main KPOP we prepare 7 (± 3 h centered analysis time) ensemble mean background states. Regarding the ensemble mean states we proceed thinning, BC and QC, and identify positions of observations using the main KPOP. The information provided by the main KPOP also includes observation error, observation values, a detailed description of observations, among others. Then, the KPOP-Mini is performed for each ensemble member state. The KPOP-Mini also comprises forward models, RTTOV for satellite radiance data, and the Radio Occultation Processing Package

(ROPP) for GPS-RO data. By running the KPOP-Mini interpolated background states on the position of each observation are estimated for each member. Subsequently, the LETKF system can use a consistent number of observations for all members.

c. LETKF

(a) LETKF DA system

The LETKF system at the KIAPS (Shin et al., 2016) is based on the 4D-LETKF algorithm formulated by Hunt et al. (2007). The OSSE and real data assimilation, using conventional data, such as sonde, aircraft, satwind, and surface pressure has been done successfully with the LETKF system implemented to the hydrostatic version of the global NWP model, KIM. Recently, we have coupled the LETKF system to the non-hydrostatic version of KIM described briefly in section 2.1. Currently, we update a set of analysis variables, including zonal wind (U), meridional wind (V), temperature (T), and specific humidity (Q), while two of the prognostic variables of the forecast model are potential temperature (θ) instead of temperature (T), and mixing ratio (W) instead of Q. Therefore, we need to conduct a transform between those variables. Currently, we do not assimilate surface pressure because we need to do some corrections for surface observation. Since the differences in the vertical location of observation station and model surface are not ignorable in some places, we are considering an adjustment (correction) of the location of observation to the height of model surface given by the topography used for the forecast model. This correction is being implemented to the KPOP system, and the implementation and its impact will be further examined in future work.

The spatial localization weights are given by a Gaussian-like piecewise fifth-order rational function (Gaspari and Cohn, 1999; Miyoshi et al., 2007) so that the function drops to zero at about 1800 km. The vertical localization function for conventional data are defined by the Gaussian-like rational function, with the localization scale of $2\sqrt{10/3}\cdot\sigma_v$, where σ_v is chosen to be 0.1 natural-log-pressure for the standard deviation parameter in the function, currently.

Another main change added to the LETKF system described in Shin et al. (2016), is the inclusion of radiance data assimilation. For the use of radiance data, we use the RTTOV implemented in the KPOP to transform the state variables of the ensemble background into the observed variable of brightness temperature. We have implemented five options for the vertical localization of the column-integrated radiance information into the vertical levels of the model. Those options are (1) the direct use of weighting function defined by a gradient of transmittance of the measured radiance (Thépaut, 2003), (2) vertical localization identical to pointwise observation such as sonde (Houtekamer et al., 2005; Houtekamer and Mitchell, 2005), (3) to use a square of the weighting function normalized in such a way that its maximum value is equal to unity (Miyoshi and Sato, 2007), (4) the use of weighting function normalized in such a way that its vertical sum is equal to unity and use radiance data only if the value of the normalized weighting function at the level was greater than a tunable threshold (Fertig et al. 2007; Aravequia et al. 2011), and (5) a way similar to 4), but without the truncation of the normalized weighting function (Kang et al., 2012). Among them, we select the option 1) in this study and simply use the vertical weighting function (Thépaut, 2003) directly. Although this might lead to an overuse of information far apart from a model grid, we begin with this approach and plan to further improve the localization process for proper use of radiance satellite data.

An adaptive multiplicative inflation algorithm by Miyoshi (2011) is used to prevent the filter divergence, especially where observations are dense. We additionally implement additive inflation by using randomly sampled differences between 6- and 12-h forecasts for one valid time per ensemble perturbation to handle such problems as sampling and model errors (e.g., Whitaker et al., 2008). In section 4.1, we describe the implementation of the additive inflation in more detail and discuss its impact on the performance of the data assimilation system.

(b) Software engineering aspects of the LETKF system

We have developed a framework called "DaPy" for the LETKF system, which is implemented using the Python script language mixed with the Fortran programming language. All of the main components, namely KIM, KPOP, and LETKF, are integrated as the model operator in DaPy. The framework DaPy is parallelized using the Message Passing Interface (MPI) and the multiprocessing package of Python.

The ensemble data assimilation in our study needs to handle multiple states with enormous complexity. Consequently, high-performance computing systems, as well as the parallelization and optimization for the ensemble system are required. Figure 1 shows the workflow that consists of tasks for the real-time operation of LETKF including the 9-h model forecasts ("*letkf_fcst*"), execution of the KPOP ("*kpop*"), the KPOP-Mini forward operator ("*kpop_mini*"), data assimilation ("*letkf_da*"), and file transfers and check output files ("*file_tx*"). The scheduling of the tasks is operated and controlled by the software "Cylc" (available at https://cylc.github.io/cylc/), which is a Python-based workflow engine for cycling tasks. We continually update the forecast-analysis cycling system based on LETKF for the effective management of many data files and the optimization of processes suitable for high-performance computing resources.

We create many data files including 50-member background data files, 50 analysis data files, their mean and spread data files, an adaptive multiplicative inflation data file at each cycle and so on. After we finish the cycle, output files need to be transferred to a storage place. Among them, the file size of one ensemble analysis is about 700 Mbytes. Since we have 50 members, the total file size for analysis states is about 35 Gbytes. Hence, it would be time-intensive if those files are transferred using shell scripts executed in one thread (Fig. 2a).

To overcome this problem, we have developed a copy tool called "multicopy," to copy multiple files to a target directory using many cores (threads) in parallel as described in Fig. 2b. This tool is implemented using the multiprocessing module of Python (available at https://docs.python.org/2/library/multiprocessing.html).

Figure 3 shows that the time-consumption is much reduced by "multicopy" using many threads when copying 205 LETKF data files to another data server. When the number of threads increases, the time for the copy decreases until the number of threads reaches 25, and thereafter, the transfer time is nearly unchanged. We have used the Intel Haswell 2.6GHz (24 cores per node) supercomputer located at the KMA, equipped with Cray XC40 and multi-petabyte Cray Sonexion Lustre file system.

3. Experimental design

We use the following types of observations: sonde, aircraft, GPS-RO, AMV, AMSU-A, IASI, CrIS, COMS CSR, and ATMS. First, we test the system using only the adaptive multiplicative method (Miyoshi, 2011) for covariance inflation. With this experimental set-up, we have experienced a kind of "filter divergence." It means that error keeps growing in time, even if data assimilation is carried out every 6 h. By implementing additive inflation to the LETKF framework we attempt to manage this problem. For additive inflation in this study we obtain a state difference between two adjacent 6-h forecasts that are randomly chosen from the pool, which contains outputs from the model run for 3 months. The three-month model run starts from 1200 UTC 23 September to 1800 UTC 23 December 2016, and the initial condition for the simulation is obtained from the Unified Model (UM) analysis interpolated on the KIM grid points. Before those state differences are added to the analysis ensemble, they are multiplied by a scaling factor less than 1.0. Unless the inflation is scaled down, ensemble runs can become unstable and blow up. We begin with a small scaling factor of 0.1, and then examine the effect of the size of added perturbation by increasing the factor up to 0.4. On the performance of the LETKF system, we will discuss the results from those experiments with 1) adaptive multiplicative inflation only and 2) adaptive multiplicative inflation + additive inflation with scaling factors 0.1~0.4, respectively.

4. Evaluation

a. Experiments with/without additive inflation

To evaluate the performance of the system, we run a forecast-analysis cycle for the period between 0600 UTC 21 March and 1800 UTC 04 April 2017. Initially, we use the adaptive multiplicative inflation only. Figure 4 shows the inflation field at model level 15 at around 500 hPa at 0000 UTC on 04 April 2017. The value of inflation is typically high over

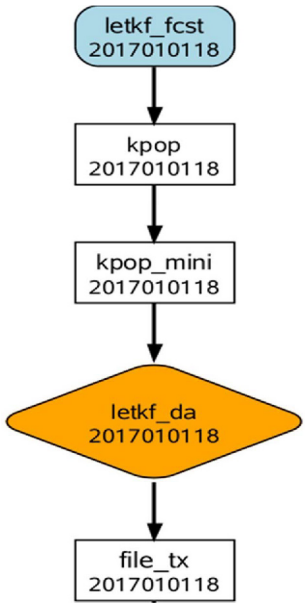

Fig. 1. Schematic diagram of the DaPy-based LETKF workflow.

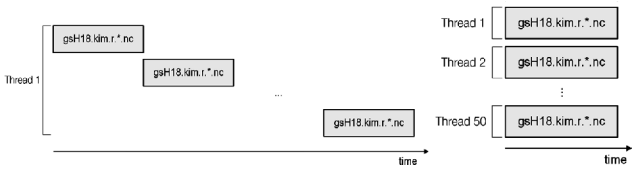

Fig. 2. Two ways of file transfer: (a) A sequential copy using a single thread, and (b) a multicopy method using 50 threads.

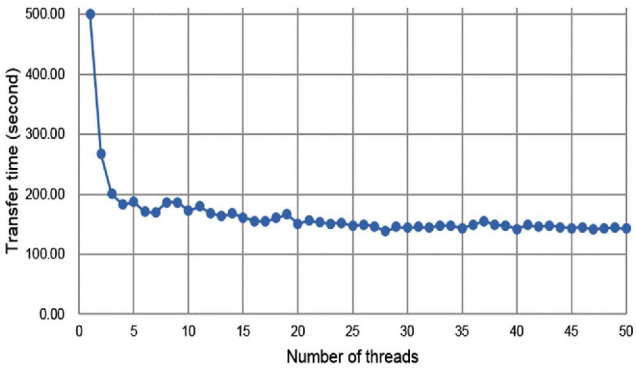

Fig. 3. Transfer time for the copy of multiple files using single and multiple threads.

lands where conventional data are dense at that time. For example, the inflation factor is greater than 1.05 over Northeastern America and East Asia. On the contrary, the inflation factor is below 0.95 over the ocean. It means that ensemble perturbations are deflated in the region. Kotsuki et al. (2017)

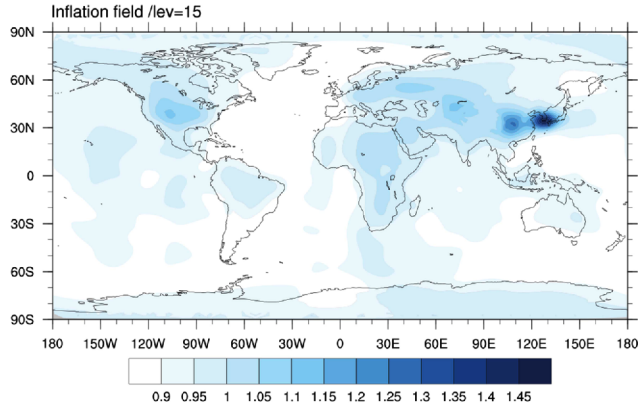

Fig. 4. An example of the distribution of multiplicative inflation factor: Inflation field at the fifteenth model level around 500 hPa at 0000 UTC on 04 April 2017.

showed deflation over oceans in the experiment using both conventional and AMSU-A data. They discussed that larger values of observation error of AMSU-A than the true value can be related to the underestimated inflation factors. The observation error of AMSU-A was also 0.5 K in the experiment shown in Fig. 10d of Kotsuki et al. (2017). Fair and direct comparison with our experiments might be difficult but we assume that the deflation in Fig. 4 can be related to the characteristics of adaptive multiplicative inflation sensitive to the observation errors of radiance satellite observations (Kotsuki et al., 2017). Also, the adaptive multiplicative inflation might be better estimated in areas where observation data, such as sonde and surface which can be provided continuously in time. The position of satellite data varies in time, and these data covering oceanic areas may not be fully used in updating the adaptive multiplicative inflation factor. This reasoning might need to be reexamined in a simplified setting where a single type of radiance satellite observation is used for data assimilation.

The low inflation value in those areas of deflation implies that the background covariance is possibly far underestimated. To manage this issue, we use the additive inflation method to perturb the subspace spanned by the original ensemble. As in Whitaker et al. (2008), we use a randomly sampled and the normalized difference between two adjacent forecasts (6- and 12-h forecasts, respectively) for one valid time, per ensemble perturbation. For comparison, we experiment two cases: one test only with adaptive multiplicative inflation, and the other test additionally with additive inflation using the scaling factor of 0.3. We evaluate the performance of the newly developed LETKF system using the Integrated Forecast System (IFS) analysis data from the European Centre for Medium-Range Weather Forecast (ECMWF). We regard the IFS analysis as a reference for a current atmospheric state and compute the Root-Mean-Square Difference (RMSD) of our analysis from the IFS analysis.

Figure 5 illustrates the RMSD time-series of the temperature

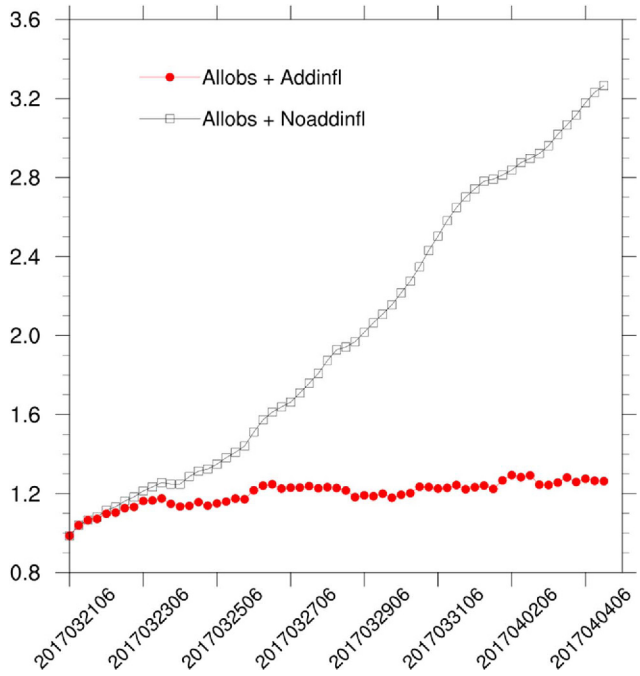

Fig. 5. Time-series of globally-averaged Root-Mean-Square Difference (RMSD) of temperature (T) analysis from the test using additive inflation with the factor 0.3 (red) and without additive inflation (black), during the forecast-analysis cycle 0600 UTC 21 March and 0600 UTC 04 April 2017.

analysis from the two tests. During the early stage of the cycle, the difference between the RMSDs from the two tests is not significant but growing continuously as the cycle progresses. The RMSD from the test using both additive and multiplicative inflation remains almost plateau in the later stage of the cycle but the value from the test using only multiplicative inflation keeps growing, which indicates a kind of "filter divergence." As the analyzed states drift away from the reference states, the number of radiance satellite observations used for the data assimilation decreases in time. It is because the BC of the radiance observation relies on the background model states. The decrease in the use of the radiance satellite data becomes more severe as the background states drift away from the true atmospheric states and observations. This result suggests that the use of additive inflation in addition to the adaptive multiplicative inflation can help the system deal with the underestimation of background uncertainties in real data assimilation using various types of observation including radiance satellite data.

b. Impact of the magnitude of additive inflation

We also examine the impact of the magnitude of the scaling factor on the performance of the LETKF system. We test the scaling factors 0.1, 0.3, and 0.4, and run the forecast-analysis cycle for the period between 0600 UTC 21 March and 1800 UTC 09 April 2017. Figure 6 shows the RMSD time-series of

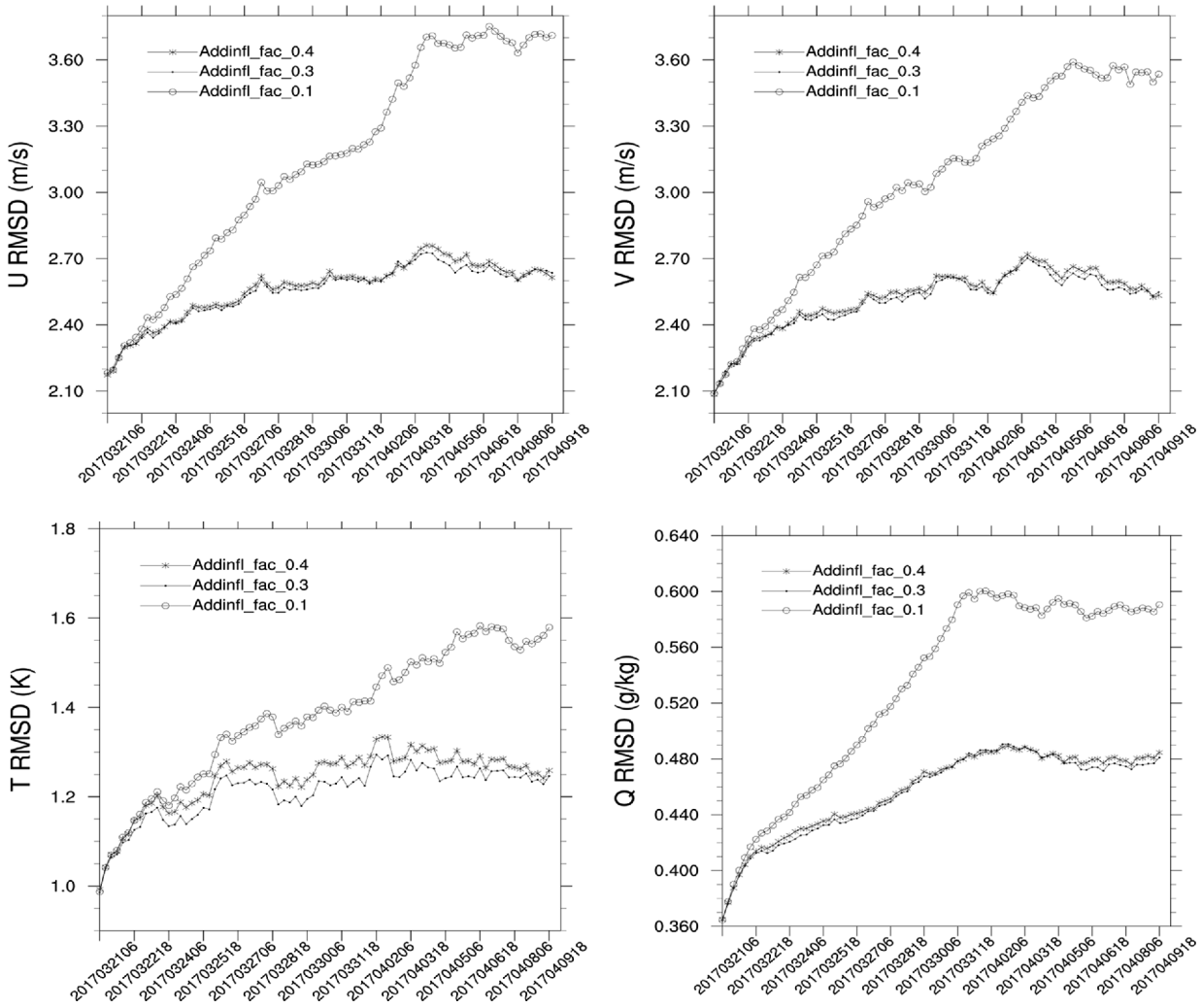

Fig. 6. Time-series of the Root-Mean-Square Difference (RMSD) of analysis variables, zonal wind (U), meridional wind (V), temperature (T), and specific humidity (Q) from the test using both multiplicative and additive inflation with the scaling factors 0.1, 0.3, and 0.4.

zonal wind (U), meridional wind (V), and temperature (T), and specific humidity (Q) from each test. The RMSD values of all variables continue growing steadily in the test with the scaling factor 0.1 but slower than in the test without additive inflation. This result indicates that the system can still experience the "filter divergence" or requires an extended spin-up if we add a relatively small amount of perturbation to the original ensemble. It means that underestimation of uncertainties has not been sufficiently handled by the addition of perturbation rescaled with the factor 0.1. A larger amount of perturbation seems to be required to handle better the uncertainties associated with the sampling errors.

We increase the scaling factor from 0.1 to 0.3 and re-evaluate the performance. The RMSD of the temperature analysis is already compared with that from the test without additive inflation in Fig. 5. The increase in the scaling factor markedly improves the performance of the DA system. The RMSD values fluctuate but they do not grow persistently in time. If we increase the factor from 0.3 to 0.4, the RMSD values of analysis variables become larger again (Fig. 6). Larger additive inflation perturbations might lead to an unnecessary accumulation of noises for analysis variables. We assume that an optimal factor can be around 0.3 for our current system but at the same time acknowledge that the factor is a kind of tuning parameter and more careful investigation would be beneficial for us to obtain a clearer answer. Nevertheless, this additive inflation method can be quite powerful, to stabilize an ensemble data assimilation system with system errors.

We examine the impact of additive inflation on the ensemble spread and on the gap between the spread and RMSD. Figure 7 demonstrates the time-series of the globally-averaged ensemble spread and RMSD of the zonal wind (U) analysis from the test using only multiplicative inflation, and from the test using both the multiplicative and additive inflation with the scaling

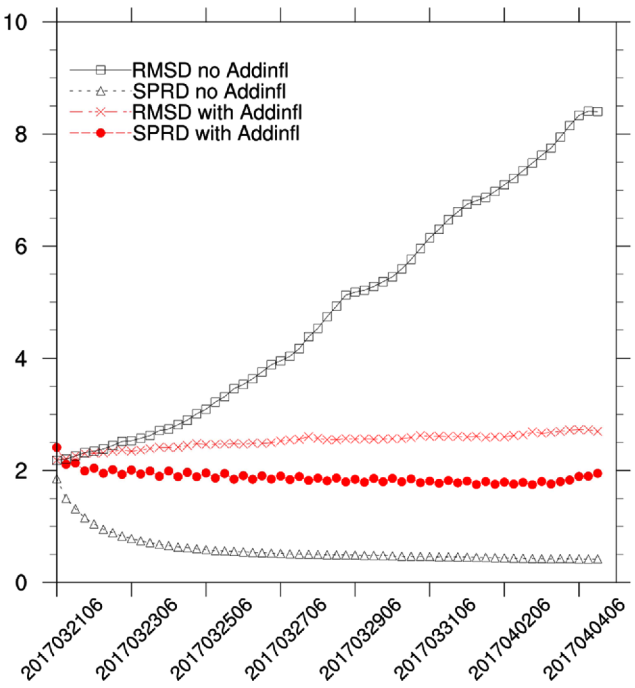

Fig. 7. Time-series of the Root-Mean-Square Difference (RMSD) and ensemble spread (SPRD) of zonal wind (U) analysis from the test using multiplicative inflation only, and the test using both multiplicative and additive inflation, with the scaling factor 0.3.

factor 0.3. If we use the adaptive multiplicative inflation only, the ensemble spread drops rapidly in time and remains about 0.5 m s^{-1} while RMSD keeps growing in time and approaches 8 m s^{-1} at the end of the cycle. Consequently, the gap between the spread and RMSD also keeps increasing and it becomes around 7.5 m s^{-1}. In contrast, the ensemble spread drops slightly, remaining about 2.0 m s^{-1} when we use the additive inflation. The RMSD also remains plateau with relatively low values around 2.5~3.0 m s^{-1} so that the gap between the ensemble spread and RMSD is maintained at about 0.5~1.0 m s^{-1}. This result supports that the extra use of additive inflation would help our system to manage the problem of underestimation of ensemble covariance.

5. Summary

We have developed the LETKF framework for the non-hydrostatic version of the forecast model KIM and coupled the system to the KPOP developed at KIAPS. The coupling with the KPOP facilitated the use of diverse types of observations including satellite radiance observation for the LETKF framework. Moreover, we optimized the infrastructure, such as handling multiple files for ensemble states, scripts for running forecast-analysis cycles, and scheduling for semi-operational runs. The infrastructure called DaPy has been implemented using a mix of the Python script and Fortran programming languages and parallelized using the MPI and the Python multiprocessing package.

We used the RTTOV implemented in the KPOP to transform the state variables of the ensemble background into the observed variable of brightness temperature and included several options for the vertical localization of the column-integrated radiance information into the vertical levels of the model. In this study, the following observation types were used: sonde, aircraft, GPS-RO, AMV, AMSU-A, IASI, CrIS, CSR, and ATMS.

For the evaluation of the system, we ran a forecast-analysis cycle for the period between 0600 UTC 21 March and 1800 UTC 04 April 2017 and compared the results with the IFS analysis as a reference. First, we tested the system only using the adaptive multiplicative method (Miyoshi, 2011) for covariance inflation. The error continued growing in time, and the analyzed state drifted away from the reference state. To handle the far underestimated ensemble covariance, we implemented additive inflation to the LETKF system, using randomly sampled adjacent forecast differences, which were added to the ensemble perturbations (Whitaker et al., 2008). The 6-h state differences were multiplied by a scaling factor before they are added to each ensemble perturbation. We began with a small scaling factor of 0.1 and then examined the effect of the size of added perturbations by increasing the factor up to 0.4. We ran cycling experiments for the period between 0600 UTC 21 March and 1800 UTC 09 April 2017.

The use of additive inflation effectively improved the performance of the system. More observations could be utilized for corrections and the accuracy of the analysis states was increased. The positive impact of additive inflation on the performance increased with the size of the scaling factor. However, the gain in the performance became smaller as the factor was further strengthened, and the system was even somewhat degraded if the factor was 0.4. We acknowledge that the scaling factor is a kind of tuning parameter and need to perform several tests to find an optimal value. Nevertheless, this additive inflation method can be a practical and powerful approach, to stabilize an ensemble data assimilation system. In future, we consider the implementation of an additive inflation method by which fast-growing errors can be captured more selectively (e.g. Houtekamer and Zhang, 2016; Yang et al., 2015).

Acknowledgement. We appreciate two anonymous reviewers for comments that helped us improve this manuscript greatly. This work has been carried out through the R&D project on the development of global numerical weather prediction systems of the Korea Institute of Atmospheric Prediction Systems (KIAPS) funded by the Korea Meteorological Administration (KMA).

Edited by: Seon Ki Park

References

Anderson, J. L., 2008: Spatially and temporally varying adaptive co-

variance inflation for ensemble filters. *Tellus*, **61**, 72-83.

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

Aravequia, J. A., I. Szunyogh, E. J. Fertig, E. Kalnay, D. Kuhl, and E. J. Kostelich, 2011: Evaluation of a strategy for the assimilation of satellite radiance observations with the Local Ensemble Transform Kalman Filter. *Mon. Wea. Rev.*, **139**, 1932-1951, doi:10.1175/2010MWR3515.1.

Bormann, N., A. Collard, and P. Bauer, 2010: Estimates of spatial and inter-channel observation error characteristics for current sounder radiances for NWP, part II: Application to AIRS and IASI. *Quart. J. Roy. Meteor. Soc.*, **136**, 1051-1063, doi:10.1002/qj.615.

_____, M. Bonavita, R. Dragani, R. Eresmaa, M. Matricardi, and A. McNally, 2016: Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quart. J. Roy. Meteor. Soc.*, **142**, 1767-1780, doi:10.1002/qj.2774.

Buehner, M., J. Morneau, and, C. Charette, 2013: Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlin. Processes Geophys.*, **20**, 669-682, doi:10.5194/npg-20-669-2013.

Choi. S.-J., and S.-Y. Hong, 2016: A global non-hydrostatic dynamical core using the spectral element method on a cubed-sphere grid. *Asia-Pac. J. Atmos. Sci.*, **52**, 291-307, doi:10.1007/s13143-016-0005-0.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143-10162.

Fertig, E. J., B. R. Hunt, E. Ott, and I. Szunyogh, 2007: Assimilating non-local observations with a local ensemble Kalman filter. *Tellus*, **59**, 719-730.

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

Grody, N., F. Weng, and R. Ferraro, 1999: Application of AMSU for obtaining water vapor, cloud liquid water, precipitation, snow cover and sea ice concentration. *Proc. the Tenth International ATOVS Study Conference*, Colorado, USA, BMRC, 230-240.

_____, J. Zhao, R. Ferraro, F. Weng, and R. Boers, 2001: Determination of precipitable water and cloud liquid water over oceans from the NOAA 15 advanced microwave sounding unit. *J. Geophys. Res.*, **106**, 2943-2953.

Harris, B. A., and G. Kelly, 2001: A satellite radiance-bias correction scheme for data assimilation. *Quart. J. Roy. Meteor. Soc.*, **127**, 1453-1468.

Hilton, F., N. C. Atkinson, S. J. English, and J. R. Eyre, 2009: Assimilation of IASI at the Met Ofce and assessment of its impact through observing system experiments. *Quart. J. Roy. Meteor. Soc.*, **135**, 495-505.

Hong, S.-Y., and Coauthors, 2018: The Korean Integrated Model (KIM) system for global weather forecasting (in press). *Asia-Pac. J. Atmos. Sci.*, **54**, doi:10.1007/s13143-018-0028-9.

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

_____, and _____, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269-3289.

_____, and F. Zhang, 2016: Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.,* **144**, 4489-4532, doi: 10.1175/MWR-D-15-0440.1.

_____, H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations. *Mon. Wea. Rev.*, **133**, 604-620.

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

Kang, J.-S., E. Kalnay, T. Miyoshi, J. Liu, and I. Fung, 2012: Estimation of surface carbon fluxes with an advanced data assimilation methodology. *J. Geophys. Res.*, **117**, D24101.

Kang, J.-H., H.-W. Chun, S. Lee, H.-J. Song, J.-H. Ha, I.-H. Kwon, H.-J. Han, H. Jeong, and H.-N. Kwon, 2018: Development of an observation processing package for data assimilation in KIAPS. *Asia-Pac. J. Atmos. Sci.*, **54**, doi:10.1007/s13143-018-0030-2.

Kleist, D. T. and K. Ide, 2015: An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Mon. Wea. Rev.*, **143**, 433-451, doi:10.1175/MWR-D-13-00351.1.

Kotsuki, S, Y. Ota, and T. Miyoshi, 2017: Adaptive covariance relaxation methods for ensemble data assimilation: experiments in the real atmosphere. *Quart. J. Roy. Meteor. Soc.*, **143**, 2001-2015, doi:10.1002/qj.3060.

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, doi:10.5194/amt-8-1259-2015.

Lorenc, A. C., N. E. Bowler, A. M. Clayton, S. R. Pring, and D. Fairbairn, 2015: Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. *Mon. Wea. Rev.*, **143**, 212-229, doi: 10.1175/MWR-D-14-00195.1.

McNally, A. P., and P. D. Watts, 2003: A cloud detection algorithm for high-spectral-resolution infrared sounders. *Quart. J. Roy. Meteor. Soc.*, **129**, 3411-3423.

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, doi:10.1175/2010MWR3570.1.

_____, and Y. Sato, 2007: Assimilating satellite radiances with a Local Ensemble Transform Kalman Filter (LETKF) applied to the JMA global model (GSM). *Sci. Online Lett. Atmos.*, **3**, 37-40.

_____, S. Yamane, and T. Enomoto, 2007: Localizing the error covariance by physical distances within a Local Ensemble Transform Kalman Filter (LETKF). *Sci. Online Lett. Atmos.*, **3**, 89-92.

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

Salonen, K., J. Cotton, N. Bormann, and M. Forsythe, 2015: Characterizing AMV height-assignment error by comparing best-fit pressure statistics from the Met Office and ECMWF data assimilation systems. *J. Appl. Meteor. Climatol.*, **54**, 225-242, doi:10.1175/JAMC-D-14-0025.1.

Shin, S., J.-S. Kang, and Y. Jo, 2016: The Local Ensemble Transform Kalman Filter (LETKF) with a Global NWP Model on the Cubed Sphere. *Pure Appl. Geophys.*, **173**, 2555-2570, doi:10.1007/s00024-016-1269-0.

Smith, A., N. Atkinson, W. Bell, and A. Doherty, 2015: An initial assessment of observations from the Suomi-NPP satellite: data from the Cross-track Infrared Sounder (CrIS). *Atmos. Sci. Lett.*, **16**, 260-266, doi:10.1002/asl2.551.

Song, H.-J., S. Shin, J.-H. Ha, and S. Lim, 2017: The advantages of hybrid 4DEnVar in the context of the forecast sensitivity to initial conditions. *J. Geophys. Res.*, **122**, 12226-12244, doi:10.1002/2017JD027598.

Thépaut, J.-N., 2003: Satellite data assimilation in numerical weather prediction: An overview. *Proc. the Annual Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean,* Reading, UK, ECMWF, 75-94.

Weston, P., W. Bell, and J. Eyre, 2014: Accounting for correlated error in the assimilation of high resolution sounder data. *Quart. J. Roy. Meteor. Soc.*, **140**, 2420-2429, doi:10.1002/qj.2306.

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, doi:10.1175/MWR-D-11-00276.1.

_____, _____, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.*, **136**, 463-481.

Yamazaki, A., T. Enomoto, T. Miyoshi, A. Kuwano-Yoshida, and N. Komori, 2017: Using Observations near the poles in the AFES-LETKF data assimilation system. *Sci. Online Lett. Atmos.*, **13**, 41-46, doi:10.2151/sola.2017-008.

Yang, S.-C., E. Kalnay, and T. Enomoto, 2015: Ensemble singular vectors and their use as additive inflation in ENKF. *Tellus A*, **67**, 26536, doi:10.3402/tellusa.v67.26536.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238-1253.