# Applying a Four-dimensional Local Ensemble Transform Kalman Filter (4D-LETKF) to the JMA Nonhydrostatic Model (NHM)

Takemasa Miyoshi and Kohei Aranami Numerical Prediction Division, Japan Meteorological Agency, Tokyo, Japan

#### Abstract

A four-dimensional local ensemble transform Kalman filter (4D-LETKF) is applied to the Japan Meteorological Agency (JMA)'s nonhydrostatic model (NHM) with explicit cloud microphysics to enable mesoscale ensemble prediction and data assimilation. Convective-scale data assimilation experiments in a perfect model scenario with 5-km grid spacing are performed, which indicates that the 4D-LETKF system works appropriately. Observations are taken every 10 minutes and every  $2 \times 2 \times 2$  grid points for horizontal winds, temperature, relative humidity, surface pressure, and precipitation rate. Although fixed lateral boundary conditions cause error reduction even without data assimilation, the advantages of 4D-LETKF are clear. When precipitation-rate observations are assimilated, some convective systems are better captured, although the impact is not always positive. Overall, 4D-LETKF shows encouraging results; it would be a tool adopted in future researches in convective-scale data assimilation and ensemble prediction.

#### 1. Introduction

Mesoscale ensemble prediction is a relatively new field of study in numerical weather prediction (NWP). Ensemble prediction tackles the problem of forecast uncertainties and predictability, and the key technique is to generate a limited number of initial ensemble perturbations that efficiently represent errors in a high-dimensional atmospheric system. The technique has been mainly explored with global models, where several methods such as the bred-vector and singular-vector methods are well-established (e.g., Kalnay 2003). Due to the facts that the general concept of ensemble prediction is applicable to any dynamical models and that uncertainties in mesoscale NWP have direct effects in short-term severe weather forecasts and warnings, there are substantial needs to perform ensemble prediction with regional convective-scale atmospheric models. However, the existence of lateral boundaries in regional models prevents applications of ensemble techniques in an early stage of development. In fact, it is only recently that some researches were brought up to apply wellestablished ensemble techniques to regional models.

The ensemble Kalman filter (EnKF) technique is a relatively new method of data assimilation and generation of ensemble perturbations. Similarly to other ensemble techniques, it has been mainly investigated with global models. In January 2005, Canadian Meteorological Centre (CMC) applied an EnKF method with perturbed observations to its operational global ensemble prediction system (EPS). This is a sign of establishment of the technique, although there are several approaches. They are divided into two categories: the perturbed observation method and the square root filter (SRF) method (e.g., Whitaker and Hamill 2002). Whitaker and

Corresponding author: Takemasa Miyoshi, Numerical Prediction Division, Japan Meteorological Agency, 1-3-4 Otemachi, Chiyoda-ku, Tokyo 100-8122, Japan. E-mail: miyoshi@naps. kishou.go.jp. ©2006, the Meteorological Society of Japan.

Hamill (2002) pointed out that the latter would be advantageous because it does not require perturbing observations. In the category of SRF, there are still several approaches (e.g., Tippett et al. 2004); no clear difference has been known among them (e.g., comparison between two SRF methods: a local ensemble transform Kalman filter (LETKF, Hunt 2005) and a serial ensemble SRF (serial EnSRF, Whitaker and Hamill 2002), Kalnay and Harlim, pers. comm.; Whitaker and Hamill, pers. comm.). After many experiences with global models, the EnKF technique is well-matured to be applied to regional models.

There have been several researches thus far applying EnKF to regional convective-scale models. Snyder and Zhang (2003) pioneered this field of study by applying an EnSRF to a mesoscale nonhydrostatic model by Sun and Crook (1997) with 2-km grid spacing to assimilate simulated Doppler radar observations taken every 5 minutes. The forecast/assimilation cycle lasts until 100 minutes to assimilate a single supercell convective storm. Zhang et al. (2004) expanded the study to a more realistic situation. Zhang et al. (2006) used the same EnSRF system as the above studies, but applied to a mesoscale model known as MM5 with a larger domain covering about half of the lower 48 states with 30-km grid spacing, where observations are assimilated every 3 hours and forecast/assimilation cycle lasts up to 36 hours.

In this study, we apply four-dimensional expansion of LETKF (4D-LETKF) to the JMA's operational version of NHM (Saito et al. 2006) with a smaller domain and investigate the feasibility and stability of the method. We perform experiments in the time-range of several hours with 5-km grid spacing to include several convective systems, the situation different from the studies by Synder and Zhang (2003) and Zhang et al. (2004, 2006). A goal of this study is to provide a working EnKF system with NHM, the tool adopted in future researches in convective-scale data assimilation and ensemble prediction. In order for 4D-LETKF to be considered as a research tool, it is important to ensure that the filter converges to provide accurate analyses. Therefore, this study is focused on the performance as a data assimilation method. Section 2 describes the experimental design, and the results are presented in section 3. Summary and discussions are provided in section 4.

## 2. Experimental design

Perfect-model twin experiments are performed in this study. The forecast model is the latest development version of the JMA's operational NHM (Saito et al. 2006), but with a smaller domain:  $65 \times 63$  horizontal grids (almost a square of about 300 km by 300 km) to cover the Tokyo metropolitan area, the entire domain shown in Fig. 1 and following figures with the horizontal map in this paper. The other model parameters are the same as the current operational settings (updated in March 2006), including 5-km horizontal grid spacing, 50-level vertical resolution from surface to 21800 meters high, and parameters for Kain and Fritsch's convective parameterization scheme. The prognostic variables are three-dimensional wind components (u, v, w), tempera-

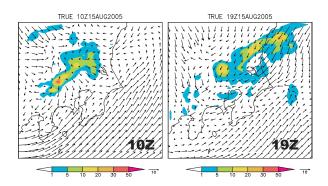


Fig. 1. Hourly accumulated precipitation (shades, mm/hr) and horizontal wind fields at the 3<sup>rd</sup> level near surface (arrows, m/s) of the true nature run at 10Z and 19Z on August 15, 2005.

ture (T), three-dimensional pressure (p), water-related quantities (vapor, cloud water, rain, cloud ice, snow, graupel), and ground temperature. The true nature run is generated by integrating the model initiated by the operational analysis on 06Z August 15, 2005, until 21Z August 15, 2005. The true run contains a few heavy rainfall events with the time-scale of about an hour, examples at 10Z and 19Z shown in Fig. 1. Observations are generated by converting the true state to an observing element and adding simulated observational noise. An observation is taken every  $2 \times 2 \times 2$  grid points in space and every 10 minutes in time after 07Z with an hour spin-up. Observing elements are horizontal wind components (u, v), temperature (T), relative humidity (RH), surface pressure (ps), and precipitation rate with observational error standard deviations of 1.0 m/s, 1.0 K, 10%, 1.0 hPa, and 1.0 mm/hr, respectively. The precipitation rate is defined as 10-minute accumulated precipitation amount (mm) multiplied by 6, so that it represents precipitation amount in an hour. A simple quality control process is applied to avoid assimilating observations with large departures from the first guess; we do not assimilate observations if the observational departures are more than 3 times as large as observa-tional error standard deviation. Three kinds of experiments are performed, one with all observing elements (w/rain), another with all observing elements except precipitation rate (w/o rain), and the other with only wind observations (only u, v), so that impacts of precipitation and wind observations are investigated. The initial ensemble members are chosen from operational analysis fields at 06Z on arbitrary dates in July 2004. In this way, the initial condition for data assimilation experiments is an analogue of climatological mean. The boundary conditions are assumed to be a part of the model, so that for ensemble runs they are all fixed to the ones generating the true run.

The MPI/OpenMP-parallelized FORTRAN90 codes of the 4D-LETKF system are originally developed by Miyoshi and Yamane (2006) for the AFES (AGCM for the Earth Simulator), which detailed system descriptions are given therein. It is modified for NHM; for example, water-related variables are added, pressure variable is expanded to three-dimensions, and lateral boundary conditions are changed not to be cyclic. Analyzed variables are the prognostic variables and precipitation rate, except ground temperature. Hourly cycle data assimilation is performed with six 10-minute slots for observation inputs, that is, to analyze at 08:00, for example, observations at 07:10, 07:20, 07:30, 07:40, 07:50, 08:00 are assimilated. Localization parameters are fixed to 11x11x9 local patch with 2.0-grid horizontal and vertical Gaussian-type localization. In addition, temporal localization (four-dimensional localization) is applied in the same way as the spatial localization, with

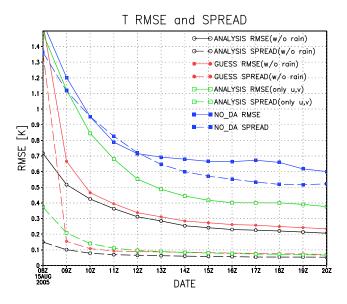


Fig. 2. Time series of temperature RMSE (K, solid lines) and ensemble spread (K, broken lines), averaged horizontally except boundary points and vertically from the 2<sup>nd</sup> to 40<sup>th</sup> levels. Black and red lines show analysis and first guess in the case assimilating all observing elements except precipitation rate (w/o rain). Green lines show the case assimilating only wind observations (only u, v). For comparison, RMSE and ensemble spread in the case without data assimilation are shown in blue lines. Solid and broken lines indicate RMSE and spread, respectively.

a time-scale of 20 minutes. The ensemble size is chosen to be 10, and the 10% multiplicative spread inflation (i.e., 21% covariance inflation) is employed.

#### 3. Results

To confirm that 4D-LETKF works appropriately, time series of root mean square errors (RMSE) for temperature are plotted in Fig. 2. The errors are defined as the difference between the true state and the most probable ensemble mean state. RMSE are decreased rapidly after the beginning of data assimilation. Ensemble spreads are much smaller than RMSE due to the dense observing network and 10% spread inflation. The first guess ensemble spread is greater than the analysis ensemble spread, which proves that analysis ensemble perturbations are actually growing in an hour. Because of the fixed boundary conditions for all ensemble runs, the error decreases even without data assimilation. Still, 4D-LETKF shows clear advantages; the w/o rain case shows about half the error size of the case without data assimilation. Even if observations except horizontal wind components (u, v) are not assimilated, temperature RMSE is clearly smaller than the case without data assimilation. Assimilating precipitation rate has a slight negative impact in temperature analysis RMSE, the difference less than 0.05 K (not shown). The above results shown for temperature are generally true for other variables including winds and water-related quantities.

Figure 3 shows analysis and true fields after the first analysis step at 08Z. At this first analysis, the case without data assimilation coincides with the first guess. Due to data assimilation, analyses become much closer to the true state, especially in wind fields. Assimilating precipitation rate reduces ensemble spreads, but the localized heavy precipitation, indicated by the brown arrow in the true state, is not well-assimilated. A possible reason is the quality control process since the

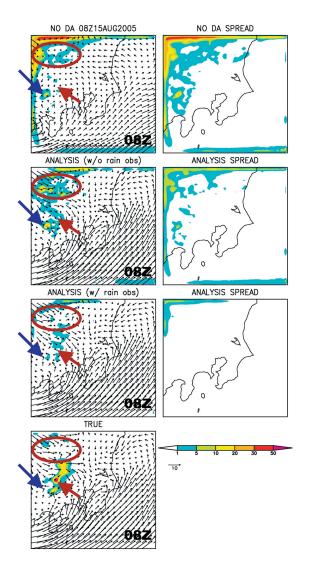


Fig. 3. Precipitation rate (shades, mm/hr) and horizontal wind fields at the  $3^{\rm rd}$  level near surface (arrows, m/s) on 08Z, August 15, 2005 (the first assimilation step). The left and right columns show analysis fields and ensemble spreads, respectively. The top row (NO DA) shows the case without data assimilation, the second shows the case assimilating u, v, T, RH, ps, and the third shows the case assimilating precipitation rate additionally, and the bottom shows the true state.

difference between the first guess and observation is so large that the heavy precipitation observations are not used. In this analysis step, assimilation of weak or zero precipitation is effective; if the w/o rain case shows precipitation in some places where the true state does not, the spurious precipitations are suppressed by observations of no or little precipitation. In fact, in the area shown by the brown circle, the w/o rain case shows a large precipitation area, but no such precipitation in the true state. Precipitation assimilation clearly suppresses the spurious precipitation. As another example, a convective precipitation noted by the blue arrow in the w/o rain case is analyzed very weak in the w/rain case. Note that this localized precipitation exists in the first guess (no data assimilation case) with a large ensemble spread. Without data assimilation, precipitation appears near boundaries because of the imbalance between initial and boundary conditions. Data assimilation makes the boundary precipitations smaller. Overall, precipitation patterns become closer to the true state due to

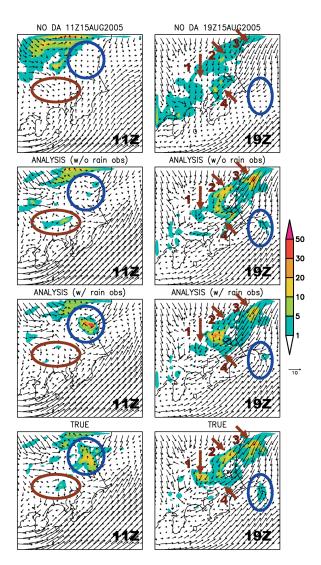


Fig. 4. Similar to the left column of Fig. 3, but for 11Z (left) and 19Z (right).

data assimilation even after the first assimilation step.

To verify analysis fields after forecast/analysis cycle processes, fields at 11Z and 19Z are shown in Fig. 4. Overall, analyzed wind fields show good correspondence to the true states as in Fig. 3. Precipitation pattern has different conclusions for each convective system. At 11Z, the system denoted by the blue circle is wellcaptured in the w/rain case. However in the w/o rain case, this system does not appear, although we see a little precipitation in the area. Without data assimilation, there is no signal of precipitation. Another system at 11Z is shown by the brown circle. This system caused heavy rainfall before 10Z as shown in Fig. 1 and is already at a decaying stage at 11Z in Fig. 4. In fact, we see a divergent flow near surface. This system is well-captured in the w/o rain case, but in the w/ rain case, the cell is already too weak and the position is a little north. This system is also absent in the case without data assimilation. Some precipitation comes from the northern boundary, but even the pattern of such precipitation systems is better captured by data assimilation.

Although 11Z is still in an adjusting stage of 4D-LETKF, 19Z is in the stabilized stage, as shown in Fig. 2. At this stage, Fig. 2 shows temperature RMSE as small

as 0.6 K even without data assimilation, thus it is expected that the boundary made a strong influence. In fact, we see the general pattern of the north-east to south-west line of the precipitation area in all panels at 19Z in Fig. 4. Still, we see advantages of 4D-LETKF. An example is the blue-circled convergence. Precipitation in the area is caused by the convergent flow, where the assimilation of winds plays an essential role. Even without temperature, relative humidity, and surface pressure observations, assimilating only horizontal winds are sufficient to capture this precipitation system (not shown). Another example is indicated by the brown arrows, the intrinsic structures of the broad precipitation area. There are several strong precipitation areas, indicated by numbers 1 to 4. The system 1 is captured by 4D-LETKF, but the intensity is a little weak. The system 2 is well-captured in the w/o rain case, but not in the w/rain case. The system 3 is well-captured in both 4D-LETKF cases, but the intensity is too strong in the w/o rain case. The system 4 is well-captured in the w/o rain case, but with precipitation assimilation, it is too strong and the position is a little different, around the middle between the systems 2 and 4.

## 4. Summary and discussions

In this study, 4D-LETKF has been applied and tested with NHM to provide a tool for future experiments in convective-scale data assimilation and ensemble prediction. It is confirmed, for the first time, that 4D-LETKF works appropriately with a nonhydrostatic regional model with explicit cloud microphysics in a convectivescale simulation. We see that the assimilation of precipitation observations helps to capture convective systems, although in some cases it has negative impacts. Overall, the results are encouraging, so that the 4D-LETKF system would be applied to future experiments.

A main limitation in this study is that the observing network is too dense, that is, it is unrealistic to have temperature and humidity profile observations every 10 km x 10 km. As a result, the ensemble spreads are too small after the first analysis step. We employed 10% multiplicative spread inflation to avoid the underestimation, but the spreads are still far smaller than the actual error size. Thus, we could reduce the number of observations, so that ensemble spreads become larger, until they are comparable to actual error size.

An open question is on the effect of precipitation assimilation and more generally, assimilation of waterrelated quantities. In mesoscale models with cloud microphysics, there are many water-related quantities, but the direct observations are limited. Currently, remote-sensing observations such as radar reflectivity are available. For example, precipitation-rate calibrated by rain-gauge network is operationally available in Japan. In this study, this kind of densely-distributed precipitation-rate observation is simulated and assimilated, but the impacts are not always positive. Moreover, negative impacts are generally observed in most threedimensional variables including winds and temperature. A possible reason is the fact that the error distribution of precipitation is far from Gaussian, especially near zero precipitation. In addition, the effects of crosscovariance among water-related quantities and other variables are not clear. Since assimilating precipitation rate showed non-negligible effects in precipitation patterns, it is important to find a right way to use the observations. Analyzing water-related quantities is a challenging field of study, so that further investigations are suggested.

Although this study is focused on EnKF as a data assimilation method, it is important to note that EnKF is a method of ensemble prediction. If EnKF converges, it is expected that the analysis members well represent actual error structures. In this way, EnKF may be an ideal method to generate ensemble perturbations. Further research is suggested to investigate how the 4D-LETKF system works in the aspect of ensemble prediction.

There are some other applications of EnKF. Since EnKF links numerical models and observations, it is useful in verifying forecast models. Moreover, EnKF can treat many kinds of observations by augmenting the observing elements to analyzed variables. The treatment of precipitation-rate observations in this study is an example. Furthermore, if model parameters are augmented in the state vector, they are adaptively esti-EnKF mated within assimilating observations (Appendix G in Evensen 2003; Annan 2005). These are further possible directions to improve mesoscale NWP using the 4D-LETKF system.

### Acknowledgments

The authors are grateful to Professor Eugenia Kalnay and Dr. Kazuo Saito for insightful discussions. The authors also thank the anonymous reviewers for useful comments to improve the manuscript. This work was partly supported by the Grant-in-Aid for Scientific Research.

#### References

- Annan, J. D., 2005: Parameter estimation using chaotic time series. *Tellus*, **57A**, 209–214. Evensen, G., 2003: The ensemble Kalman filter: theoretical for-
- mulation and practical implementation. Ocean Dyn., 53,
- Hunt, B. R., 2005: Efficient Data Assimilation for Spatiotemporal Chaos: a Local Ensemble Transform Kalman Filter. arXiv: physics/0511236v1, 25 pp.
- Kalnay, E., 2003: Atmospheric Modeling, Data Assimilation and
- Predictability. Cambridge, 341 pp.
  Miyoshi, T., and S. Yamane, 2006: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. submitted.
- Saito, K., T. Fujita, Y. Yamada, J. Ishida, Y. Kumagai, K. Aranami, S. Ohmori, R. Nagasawa, S. Kumagai, C. Muroi, T. Kato, H. Eito and Y. Yamazaki, 2006: The operational JMA nonhydrostatic mesoscale model. Mon. Wea. Rev., 134, 1266-1298.
- Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. Mon. Wea. Rev., 131, 1663-1677.
- Sun, J., and N. A. Crook, 1997: Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments. J. Atmos. Sci., 54, 1642-1661.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill and J. S. Whitaker, 2003: Ensemble square root filters. Mon. Wea. Rev., 131, 1485-1490.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. Mon. Wea. Rev., 130, 1913-1924.
- Zhang, F., C. Snyder and J. Sun, 2004: Impacts of initial estimate and observation availability with an ensemble
- Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253. Zhang, F., Z. Meng and A. Aksoy, 2006: Test of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Mon. Wea. Rev., in review.

Manuscript received 2 June 2006, accepted 23 August 2006 SOLA: http://www.jstage.jst.go.jp/browse/sola/