

# Applying the Local Ensemble Transform Kalman Filter to the Nonhydrostatic Icosahedral Atmospheric Model (NICAM)

Keiichi Kondo<sup>1</sup> and H. L. Tanaka<sup>2</sup>

<sup>1</sup>*Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan*

<sup>2</sup>*Center for Computational Sciences, University of Tsukuba, Tsukuba, Japan*

## Abstract

In this study, we apply the local ensemble transform Kalman filter (LETKF) to the Nonhydrostatic Icosahedral Atmospheric Model (NICAM) to develop the NICAM-LETKF. In addition, an algorithm to adaptively estimate the inflation parameter and the observational errors is introduced to the LETKF. The feasibility and stability of the NICAM-LETKF are investigated under the perfect model scenario.

According to the results, we confirm that the converged analysis errors of the NICAM-LETKF are smaller than the observational errors, and the magnitude and distribution of the root mean square errors (RMSEs) are comparable to those of the ensemble spreads. In our experiments, we find that the inflation parameter is optimally tuned and the observational errors are close to the true value.

It is concluded that the NICAM-LETKF works appropriately and stably under the perfect model scenario even if the inflation parameter and the observational errors are adaptively estimated within the LETKF.

## 1. Introduction

Data assimilation is one of the most important techniques in numerical weather predictions. It is important for the data assimilation to draw information from the observations as much as possible. Evensen (1994) suggested an ensemble Kalman filter (EnKF), which approximates the covariance matrix of the Kalman filter (KF; Kalman 1960) by using the ensemble predictions. The EnKF can consider the flow-dependent covariance matrix, so the EnKF can draw information which the observations have. There are a number of studies about the EnKF with the Lorenz-96 system (Lorenz 1996), the regional models and other models based on the primitive equations, e.g., the SPEEDY model (Molteni 2003). Zhang et al. (2006) investigated the EnKF with a nonhydrostatic regional model and gave details on the dependence of EnKF performance on error growth rate and scales. Hunt et al. (2007) developed the LETKF, which has an important advantage in assimilating the observations in each local patch. Due to this advantage, the sampling errors are filtered, and the LETKF has a higher performance for the implementation in parallel computers. In Miyoshi et al. (2007), they removed local patches of LETKF and applied the LETKF to AFES (AGCM (atmospheric general circulation model) for the Earth Simulator; Ohfuchi et al. 2004) with a T159L48 resolution, and they investigated the stability of the LETKF without the local patches. In the ensemble based KF, it is necessary to tune the covariance or spread inflation parameter, which often costs a lot. In Miyoshi (2005) the inflation parameter is estimated adaptively by means of the scalar KF algorithm in order to avoid the complicated tuning. The method, however, did not work properly in the experiments with the real observations because the observational errors are not perfectly known, and the errors also influence the accuracy of an analysis. Kalnay et al. (2007) and Li et al. (2009) then reported the algorithms to adaptively estimate not only the inflation parameter but also the observational

errors at a time within the LETKF.

On the other hand, a new type of ultra-high resolution atmospheric general circulation model is developed by the Center for Climate System Research, University of Tokyo and Frontier Research Center for Global Change/Japan Agency for Marine-Earth Science and Technology. The new model is designed to perform cloud-resolving simulations by directly calculating deep convection and meso-scale circulation, which plays a key role not only in the tropical circulations but also in the entire general circulation of the atmosphere. As the model adopts the non-hydrostatic equations and icosahedral grid structure, it is called the Nonhydrostatic Icosahedral Atmospheric Model (NICAM; Satoh et al. 2008). However, the assimilation system for the NICAM has not been developed and the optimum initial condition for the NICAM does not yet exist.

Therefore, in this study we apply the non-local patch version of the LETKF by Miyoshi et al. (2007) to the NICAM (referred to as NICAM-LETKF), and investigate the feasibility and stability of the NICAM-LETKF with the adaptive estimation of the inflation parameter and observational errors under the perfect model scenario. This is the first test of the LETKF with a global non-hydrostatic model, although the model should still behave hydrostatically with the horizontal resolution of 224 km. In Section 2, we describe our experimental design and original program code. The algorithms of the adaptive estimation for the inflation parameter and observational errors are presented. In Section 3, we show our numerical results of the assimilation experiments under the perfect model scenario. Finally, the conclusion is summarized with discussion in Section 4.

## 2. Experimental settings

In this study, the data assimilation experiments with the NICAM-LETKF are implemented under the perfect model scenario. The forecast model used is the NICAM mentioned in Section 1. Its horizontal resolution is 224 km (Glevel-5) and the number of vertical layers is 40. The prognostic variables are pressure, temperature, horizontal wind components, vertical wind and mixing ratio of water vapor, cloud water and rain water. In the NICAM, the horizontal wind is decomposed to three elements. So, the number of the prognostic variables is 9. The model physics used in this study are Louis's surface layer, Mellor and Yamada Level 2, Arakawa and Schubert's cumulus parameterization and third-order Runge-Kutta method for time integration.

The control run for the truth is generated with an initial data by the JMA/GSM operational analysis on 12Z 30 December 2006, and integrated until 12Z 22 January 2007. The hypothetical observations are generated by adding prescribed observational error to the truth. The error standard deviations are 1.0 hPa (pressure), 1.0 K (temperature), 1.0 m s<sup>-1</sup> (horizontal wind) and 0.5 g kg<sup>-1</sup> (mixing ratio of water vapor). The observations cover 10% grid points of the entire horizontal 2-dimensional grid space and cover about 3.3% grid points of the entire 3-dimensional grid space uniformly.

In this study, a Gaussian-like fifth order polynomial function (Gaspari and Cohn 1999) is adopted for the horizontal and vertical localization, by multiplying the function to the diagonal elements of the inverse of the observational covariance matrix. The localization scale is defined by the one standard deviation. The horizontal localization scale is

Corresponding author: Hiroshi L. Tanaka, Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan. E-mail: tanaka@sakura.cc.tsukuba.ac.jp. ©2009, the Meteorological Society of Japan.

500 km, and the Gaussian-like function drops to zero at about 1800 km from the analysis point. The vertical localization scale in 30°N–90°N and 30°S–90°S is 4.0 grid points, and that in 20°N–20°S is 3.0 grid points. In the other region the scale changes smoothly by a linear interpolation. The Gaussian-like function drops to zero at about 10 grid points to 15 grid points in the vertical. The NICAM-LETKF assimilation cycle is every 6 hours, and the period of the experimental assimilation is from 12Z 1 January 2007 to 12Z 22 January 2007. The ensemble size is fixed to 40, and the initial ensemble members are computed from the JMA operational analysis integrated for 2 days by the NICAM. The dates of the JMA operational analysis are chosen at random to avoid the similar values.

In this study, four experiments are performed, and in all experiments the multiple spread inflations are employed. The details of each experiment are summarized in Table 1. In the first experiment (Ex. 1), the spread inflation parameter is fixed temporally. In detail, the spread inflation is 1% in 30°N–90°N and 30°S–90°S, and is 3% in 20°N–20°S. In the other region the spread inflation changes smoothly by a linear interpolation. In the second experiment (Ex. 2), the spread inflation parameter is estimated adaptively, following the method shown in Miyoshi (2005). In Miyoshi (2005), an estimate of the covariance inflation parameter  $\Delta_o$  can be obtained from

$$\Delta_o = \frac{\mathbf{d}_{o-f}^T \mathbf{d}_{o-f}}{\text{trace}(\mathbf{HP}^T \mathbf{H}^T) + \text{trace}(\mathbf{R})} - 1, \quad (1)$$

where  $\mathbf{d}_{o-f}$  is the difference between the observation and forecast ensemble mean, and  $\mathbf{P}$ ,  $\mathbf{R}$  and  $\mathbf{H}$  denote the forecast error covariance, the observational error covariance and the linear observational operator, respectively. The superscript  $\top$ , the subscript  $o$  and  $f$  denote the matrix transpose, observation and forecast. The spread inflation parameter is obtained from the covariance inflation parameter. In the third and fourth experiments (Ex. 3 and Ex. 4), not only the inflation parameter but also the observational errors are estimated adaptively. The method of estimating the observational error is shown in Desroziers et al. (2005), and the method for the covariance inflation parameter and the observational error is shown in Li et al. (2009). In Desroziers et al. (2005), an estimate of the observational error variance  $\sigma_o^2$  can be obtained from

$$\sigma_o^2 = \text{trace}(\mathbf{d}_{o-a} \mathbf{d}_{o-a}^T) / p, \quad (2)$$

where  $\mathbf{d}_{o-a}$  is the difference between the observation and analysis ensemble mean, and  $p$  and subscript  $a$  denote the number of observations and analysis, respectively. In this study, the estimation method is implemented for each observation variable. Li et al. (2009) estimated the inflation parameter and the observational error variance adaptively at each analysis time step. However, if the number of observations is not enough, a large sampling error is introduced. Therefore, in Miyoshi (2005) and Li et al. (2009), they assumed that  $\Delta_o$  and  $\sigma_o^2$  are the same as the observation, respectively, to avoid this problem. They used a simple scalar KF approach, which usually uses the postprocess model output. By using this KF approach, the past information is accumulated, and the inflation parameter and observational error variance gradually converge to the optimum values while still allowing for time variations. The KF estimation is often ruined by an unrealistically large sampling error. So, to avoid this problem we impose reasonably wide upper and lower limits in the observed inflation  $\Delta_o$ , e.g., 0.0 to 0.2, before applying the KF approach. In Ex. 3 and Ex. 4, the initial specifications of the observational errors are 3.0 and 0.1 times the true value, respectively, for each variable. At each analysis time, we evaluate the analysis error using the RMSE between the true state and the analysis ensemble mean and compute the ensemble spread.

### 3. Results

Figure 1 shows the time series of the analysis RMSEs and ensemble spreads of 500 hPa geopotential height and 850 hPa temperature for Ex. 2. In the early period the

Table 1. Configuration of each experiment.

	inf. param.	obs. error (init.)
Ex. 1	Constant (1–3%)	True (–)
Ex. 2	Adaptive	True (–)
Ex. 3	Adaptive	Adaptive (3.0 times)
Ex. 4	Adaptive	Adaptive (0.1 times)

RMSEs decrease with time. The RMSEs are larger than the ensemble spreads. About three days after the data assimilation has been started, the RMSEs become comparable to the ensemble spreads. The RMSE and ensemble spread in the Tropics are larger than the other regions. This result shows that the uncertainty in the Tropics is large in spite of the perfect model experiments because the cumulus convection is active. In the temperature field, the RMSE and ensemble spread in the Southern Hemisphere are larger than that in the Northern Hemisphere. Such a tendency dominates in the lower troposphere, and there is no clear difference between the Southern Hemisphere and the Northern Hemisphere in the middle troposphere and above. This result implies that the cause is in the land-ocean distribution because the RMSE and ensemble spread are larger over the ocean, particularly in the temperature field. Moreover, for the other elements, e.g., SLP (Sea Level Pressure), wind components and water vapor, the RMSEs are apparently smaller than the observational errors. For example, the analysis RMSEs of SLP in the Northern Hemisphere, Southern Hemisphere, and Tropics are about 0.2 hPa, 0.2 hPa and 0.3 hPa, respectively.

To see the horizontal distribution of the analysis errors, Fig. 2 illustrates the temporally averaged analysis RMSE and ensemble spread of 500 hPa geopotential height for Ex. 2, in which the inflation parameter is estimated and the observational error is perfectly known. The shaded areas show the analysis RMSE or ensemble spread, and the contours show 500 hPa geopotential height. As shown in both fields in Fig. 2, the analysis error distribution is comparable to that of the ensemble spread very much. Particularly the peaks of the RMSE correspond to those of the ensemble spread. The pattern represents the area which indicates large uncertainty. For example, in the Tropics the RMSE is very large along the ITCZ (The Intertropical Convergence Zone) because of the active cumulus convection. Therefore, the ensemble spread becomes large by the chaotic nature which originates by the uncertainty. The result is consistent with that in Fig. 1. On the other hand, over the east part of the North Pacific Ocean the westerly jet is meandering, and there is a ridge along the West Coast. In the upstream of the ridge, the RMSE is large, where an extratropical cyclone is developing in the surface. Moreover, in the area with large RMSE, the ensemble spread is also large as the RMSE. Therefore, it is confirmed that the NICAM-LETKF captures the characteristics of the regional analysis errors.

Figure 3 shows the time series of the analysis RMSEs of the 500 hPa zonal wind in Exs. 1, 2, 3, and 4. In the early stage the error level of Ex. 2 is the smallest, because only the spread inflation parameter is estimated adaptively, and the observational errors are perfectly known. Ten days after the beginning of the assimilation, however, a clear separation of the analysis RMSEs among the experiments is not seen. In the Ex. 4 the long adjusting period is necessary for the LETKF to converge. About ten days are needed so that the Ex. 4 is adjusted by the NICAM-LETKF, but eventually the RMSE in the Ex. 4 becomes comparable to other experiments. In the other variables, similar results are obtained (not shown).

Figure 4 shows the adaptively estimated observational errors of pressure in the lowest layer of the NICAM in Exs. 3 and 4. The experiments start from the wrong observational error with 3.0 (Ex. 3) and 0.1 (Ex. 4) times the true value. In Ex. 3 the estimated observational error gradually decreases by assimilating data iteratively, and then the error converges to the true value. In Ex. 4 the estimated error becomes larger than the initial value in the early

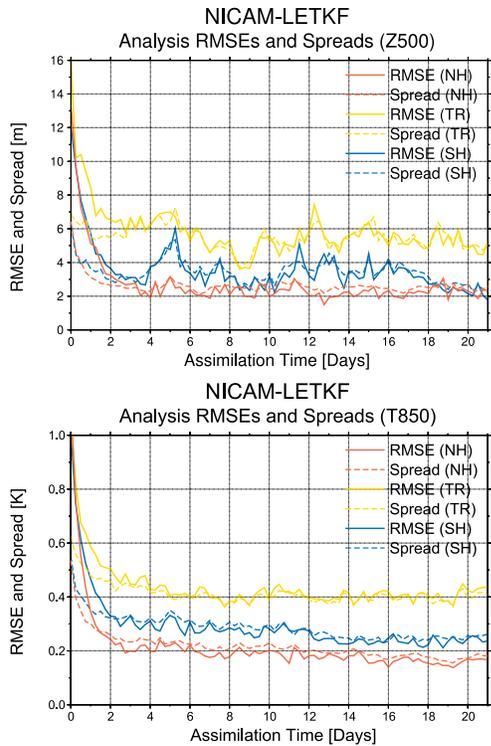


Fig. 1. Time series of the analysis RMSEs and ensemble spreads of 500 hPa geopotential height (m) (top panel) and 850 hPa temperature (K) (bottom panel) for the Ex. 2. Initial time is 12Z 1 Jan 2007. The red, blue, and yellow lines are for the Northern Hemisphere (NH; 20°N–90°N), the Southern Hemisphere (SH; 20°S–90°S), and Tropics (TR; 20°N–20°S), respectively.

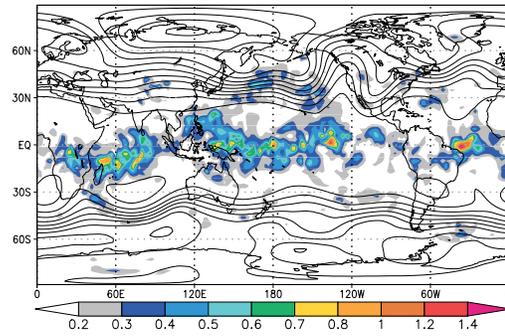
period. Then, the error converges to the true value such as in the case of the Ex. 3. About 18 days after the beginning of the assimilation, both of the estimated observational errors already become very close to the true value though the observational errors are not known. The estimated observational errors are slightly larger than the truth even though the initial observational error is very small in the case of the Ex. 4. Hence, it is confirmed that the algorithm which estimates the observational error adaptively works very well. In the other variables, similar results are obtained (not shown). As a result, we can get a good analysis by estimating the observational errors adaptively though the errors are not well known.

According to Figs. 3 and 4, it seems that the NICAM-LETKF has small sensitivity to the observational errors because in the early assimilation period the RMSE of the Ex. 3 is comparable to the RMSE of the Ex. 2 although the estimated observational errors are overestimated. This result is discussed in Section. 4.

#### 4. Conclusion and discussion

In this study we developed the LETKF and applied it to a nonhydrostatic and realistic global atmospheric model called NICAM. In addition, an algorithm which estimates not only the inflation parameter but also the observational errors adaptively was introduced to the NICAM-LETKF. We conducted three kinds of experiments to investigate the feasibility and stability of the NICAM-LETKF under the perfect model scenario: 1) the inflation parameter is fixed temporally (Ex. 1), 2) the inflation parameter is adaptively estimated, and the observational errors levels are perfectly specified (Ex. 2), and 3) the inflation parameter and the observational errors are adaptively estimated at a time (Exs. 3 and 4). It is confirmed that the LETKF works appropriately for the nonhydrostatic global model, although the model with the horizontal resolution of 224 km behaves

U500 ( $\text{m s}^{-1}$ ) Analysis RMSE (Average)  
Ensemble size=40



U500 ( $\text{m s}^{-1}$ ) Analysis Spread (Average)  
Ensemble size=40

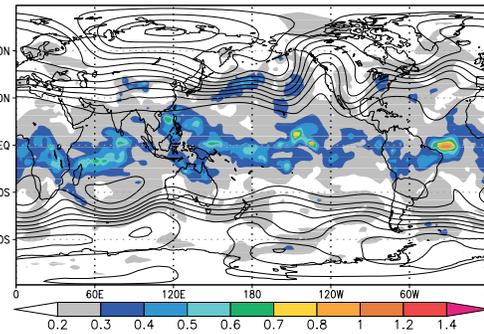


Fig. 2. Spatial distributions of the analysis RMSEs (top panel) and ensemble spreads (bottom panel) of 500 hPa zonal wind and 500 hPa geopotential height (contour) for the Ex. 2, temporally averaged for 1 day from 00Z 21 January 2007 to 18Z 21 January 2007.

hydrostatically.

First, we demonstrated that the NICAM-LETKF works stably without diverging, and the analysis errors become smaller than the observational errors in all variables. In addition, the magnitude and distribution of the analysis RMSEs are temporally and spatially comparable to those of the analysis ensemble spreads. The RMSEs are large in the area with large uncertainty such as the ITCZ or the developing extratropical cyclone. In such areas, each ensemble member spreads rapidly due to the strong chaotic nature, resulting in the large ensemble spread. According to the above result, the inflation parameter is tuned optimally in the each grid point so that the analysis errors become the smallest by adaptively estimating the parameter. These results indicate that the NICAM-LETKF combined with the adaptive estimation of the inflation parameter works appropriately and stably, and the NICAM-LETKF can capture the true analysis errors.

Second, in the case where the observational errors are perfectly known, and the inflation parameter is adaptively estimated (Ex. 2), the NICAM-LETKF converges fastest. On the other hand, in the case where the inflation parameter and observational errors are both estimated adaptively and the initial observational errors are 0.1 times the true value (Ex. 4), the NICAM-LETKF converges slowest. Since the observational errors are generally not perfectly known, the results are quite reasonable. In the early assimilation period, the accuracy of the analysis mainly depends on the inflation parameter and observational errors because the number of assimilated observations is insufficient. It may be important to note that the observational errors are overestimated by the present formulation, regardless of the magnitude of the initial observational errors. The observational error variance is obtained from Eq. (2). In the Eq. (2), it is assumed that there is no correlation between the analysis error  $\delta\mathbf{x}^a$  and the forecast error  $\delta\mathbf{x}^f$ . Actually, however, it is considered that there is a little correlation

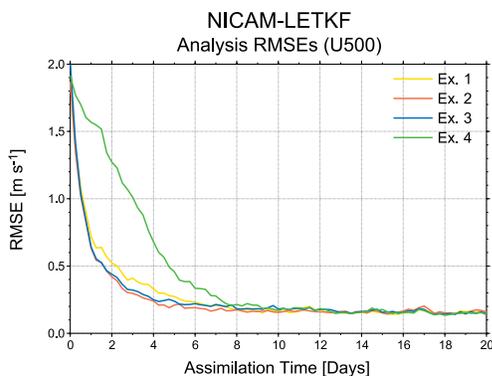


Fig. 3. Time series of the analysis RMSEs of 500 hPa zonal wind ( $\text{m s}^{-1}$ ) for the Exs. 1, 2, 3, and 4. The yellow, red, blue, and green lines are for the Exs. 1, 2, 3, and 4, respectively. Initial time is 12Z 1 Jan 2007.

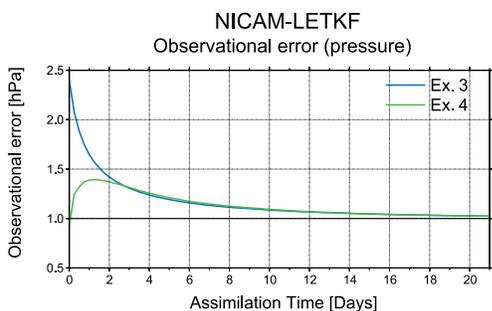


Fig. 4. Time series of adaptively estimated observational error of the pressure of the lowest layer of the NICAM. The blue and green lines are for the Exs. 3 and 4, respectively. The true observational error is 1.0 hPa. Initial time is 12Z 1 Jan 2007.

between the analysis error and the forecast error because the analysis fields and forecast field is similar in the area where the number of observations is not enough or the ensemble spread is small. So, it is considered that the estimated observational errors are overestimated. However, this result should be investigated in detail as a future work. In the latter half of the analysis period, there is no clear difference among the analysis RMSEs of the 4 experiments. It is demonstrated that the observational errors sufficiently converge to the true values, and the inflation parameter is tuned optimally because the NICAM-LETKF assimilates sufficient number of observations.

We discuss about the sensitivity of the NICAM-LETKF to the observational error settings. The accuracy of the Exs. 2 and 3 is the same level in early period though the observational errors are overestimated in the Ex. 3. Therefore, it seems that the NICAM-LETKF has small sensitivity to them. However, the adaptive estimation algorithm might have some advantages in the case such as assimilating real observations. The inflation parameter can be estimated appropriately only when the observational errors are adequately known. In addition, when the observation instrument is renewed and its observational error is changed, the LETKF can estimate its error by the adaptive estimation algorithm. Moreover, Li et al. (2009) mentioned the possibility of the estimation of the error cross-correlations in such as satellite data.

In the assimilation with the real observations, the method which estimates the inflation parameter does not work properly because the observational errors are not perfectly known. Li et al. (2009) reported the algorithm to adaptively estimate both the inflation parameter and the observational errors using the simple Lorenz-96 system and the SPEEDY model, which are based on the primitive equations. In this study, we demonstrated that the algo-

rithm works appropriately even for the realistic nonhydrostatic global model of NICAM. However, it is not clarified whether the NICAM-LETKF and the adaptive estimation algorithm works stably in the experiments with the real observations. It is suggested that this approach will suffer more in the real observations because of the model errors and the different sources of observations of similar variables. In Zhang et al. (2006), they tested the variance relaxation algorithm and reported that it is very effective and prevents filter divergence, preserving the structure of the local ensemble perturbations due to the error growth. It is considered that these algorithms are useful to draw information from the observations as much as possible even for the experiments with the real observations. These should be the subject of the future work.

## Acknowledgments

We are grateful to the member of our Meteorology and Climatology group at the University of Tsukuba for valuable comments. We appreciate Drs. T. Miyoshi and M. Matsueda for worthy discussion, and also thank Prof. Masaki Satoh of CCSR and Hirofumi Tomita of FRCGC/JAMSTEC for implementing the NICAM. This study is supported by the T2K-Tsukuba system.

## References

- Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 338–3396.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99C5**, 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.
- Hunt, B. R., E. Kostelich, and I. Szynogoh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126.
- Kalman, R., 1960: A new approach to linear filtering and predicted problems. *J. Basic Eng.*, **82**, 35–45.
- Kalnay, E., and co-authors, 2007: <http://www.atmos.umd.edu/~ekalnay/LiKalnayMiyoshiObserr-Pub.pdp>.
- 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.
- Lorenz, E. N., 1996: Predictability: A problem partly solved. *In Proc. Predictability, ECMWF, 4–8 September 1995*.
- Miyoshi, T., 2005: Ensemble Kalman filter experiments with a primitive-equation global model. Ph.D. dissertation, University of Maryland, 197 pp.
- 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.
- Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments. *Clim. Dyn.*, **20**, 175–191.
- Ohfuchi, W., H. Nakamura, M. K. Yoshioka, T. Enomoto, K. Takaya, X. Peng, S. Yamane, T. Nishimura, Y. Kurihara, and K. Ninomiya, 2004: 10-km mesh meso-scale resolving simulations of the global atmosphere on the Earth Simulator: Preliminary outcomes of AFES (AGCM for the Earth Simulator). *J. Earth Simulator*, **1**, 8–34.
- Satoh, M., T. Matsuno, H. Tomita, H. Miura, T. Nasuno, and S. Iga, 2008: Nonhydrostatic Icosahedral Atmospheric Model (NICAM) for global cloud resolving simulations. *Journal of Computational Physics*, the special issue on Predicting Weather, Climate and Extreme events. **227**, 3486–3514, doi:10.1016/j.jcp.2007.02.006.
- Zhang, F., Z. Meng, and A. Aksoy, 2006: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part I: Perfect model experiments. *Mon. Wea. Rev.*, **134**, 722–736.