Data assimilation in a two-scale model with Kalman filters

Tadashi Tsuyuki
Meteorological Research Institute/JMA, E-mail: ttuyuki@mri-jma.go.jp

1. Introduction

Data assimilation in multi-scale models, such as a coupled ocean-atmosphere model and a global cloud-resolving atmospheric model, has a potential to improve large-scale analyses by assimilating small-scale observations through scale interactions. A problem is that multi-scale models are strongly nonlinear while most of the data assimilation methods used in geophysics assumes that probability density functions are nearly Gaussian.

Ballabrera-Poy et al. (2009) investigated the accuracy of analyses when both large- and small-scale variables are simultaneously assimilated with an ensemble Kalman filter by using a Lorenz-96 two-scale model (Lorenz, 1996). They found that assimilation of large-scale variables with a few small-scale variables significantly degraded the filter performance due to spurious correlations from sampled ensemble covariances.

The present study further investigated multi-scale data assimilation with Kalman filters. There is almost no difficulty in applying Kalman filters to numerical models which contain very different time scales, compared to 4-dimensional variational assimilation.

2. Method

Data assimilation experiments were conducted on a perfect model assumption with the Lorenz-96 two-scale model:

$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b}\sum_{j=1}^{J} Y_{jk} \qquad (k = 1, \cdots, K),$$

$$\frac{\mathrm{d}Y_{jk}}{\mathrm{d}t} = -cbY_{j+1\,k}\left(Y_{j+2\,k} - Y_{j-1\,k}\right) - cY_{jk} + \frac{hc}{b}X_k \qquad (j = 1, \cdots, J).$$

The parameter values used were $K = 36,\ J = 10,\ F = 10,\ h = 1,\ b = c = 10$ as in Lorenz (1996). The temporal and spatial scales and amplitude of Y were about a tenth of those of X in these parameters. The time integrations were conducted with a time step $\Delta t = 0.005$.

The assimilation methods used were a local ensemble transform Kalman filter (LETKF; Hunt et al., 2007) and an extended Kalman filter (EKF; Thornton and Bierman, 1980). EKF does not have sampling error, but has linearization error. The observational data of X were available at a time interval of $10\Delta t$ at all coarse grid points with observation error standard deviation of unity. The impact of the temporal and spatial densities and error standard deviation of observations of Y on the analyses of X were examined. The assimilation period was $40{,}000\Delta t$.

3. Results

Figure 1 shows the average analysis errors of X for the latter half period of assimilation experiments with LETKF with 10 ensemble members. The values of multiplicative inflation factor and the size of local patches were optimized so that the average analysis errors of X were minimized under a constraint that the local patch size of Y is a tenth of that of X. The figure demonstrates that if the spatial or temporal densities or accuracy of observational data of small-scale variables is not enough, a straightforward application of an ensemble Kalman filter ("with C.I.") tends to degrade the analysis of large-scale variables. In those cases better analyses are obtained by not assimilating the small-scale observations ("Obs of X only"). This result is consistent with that of Ballabrera-Poy et al. (2009), but it is not due to sampling error, because similar results were obtained from LETKF experiments with larger ensemble members such as 200 and 500. In those experiments, adaptive covariance inflation (Li et al., 2009) was used and local patches were not applied. Assimilation experiments with EKF with the adaptive covariance inflation also gave a qualitatively similar result. Those results suggest that the degradation of large-scale analysis is due to the Gaussian assumption of Kalman filters.

The figure also shows the results from assimilation experiments in which the cross analysis increments between large- and small-scale variables were neglected at the analysis step of LETKF ("w/o C.I."). Scale interactions in data assimilation were allowed only in the forecast step. If the spatial and temporal densities and accuracy of observational data of small-scale variables were enough, the large-scale analyses were degraded compared to the previous experiments. If that was not the case, however, the neglect of cross analysis increments lead to better large-scale analyses than the analyses obtained from assimilating large-scale observational data only.

The neglect of cross analysis increments may be easily applied to a coupled ocean-atmosphere model, since most of the state variables are separated into oceanic and atmospheric variables. On the other hand, a hierarchical approach may be necessary for a global cloud-resolving atmospheric model, since it is not easy to extract cumulus-scale variables from the state variables. In the hierarchical approach, large-scale observational data are assimilated first to obtain large-scale analysis. Then small-scale observational data are assimilated with the large-scale analysis given, and the multi-scale model is integrated in the forecast step. It was found from LETKF experiments with the Lorenz-96 two-scale model that this approach gave analyses of similar quality to those obtained by neglecting the cross analysis increments.

References

Ballabrera-Poy, J., E. Kalnay and S. C. Yang, 2009: Data assimilation in a system with two scales – Combining two initialization techniques, *Tellus*, **61A**, 539-549.

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.

Li, H., E. Kalnay and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter, *Q. J. R. Meteorol. Soc.*, **135**, 523-533.

Lorenz, E. D., 1996: Predictability – A problem partly solved, *Proceedings of the ECMWF Seminar on Predictability (4-9 September 1995, Reading, UK),* ECMWF, 1-18.

Thornton, C. L. and G. J. Bierman, 1980: UDUT Covariance Factorization for Kalman Filtering, *Control and Dynamic Systems (C.T. Leondes, Ed.)*, **16**, 177-248.

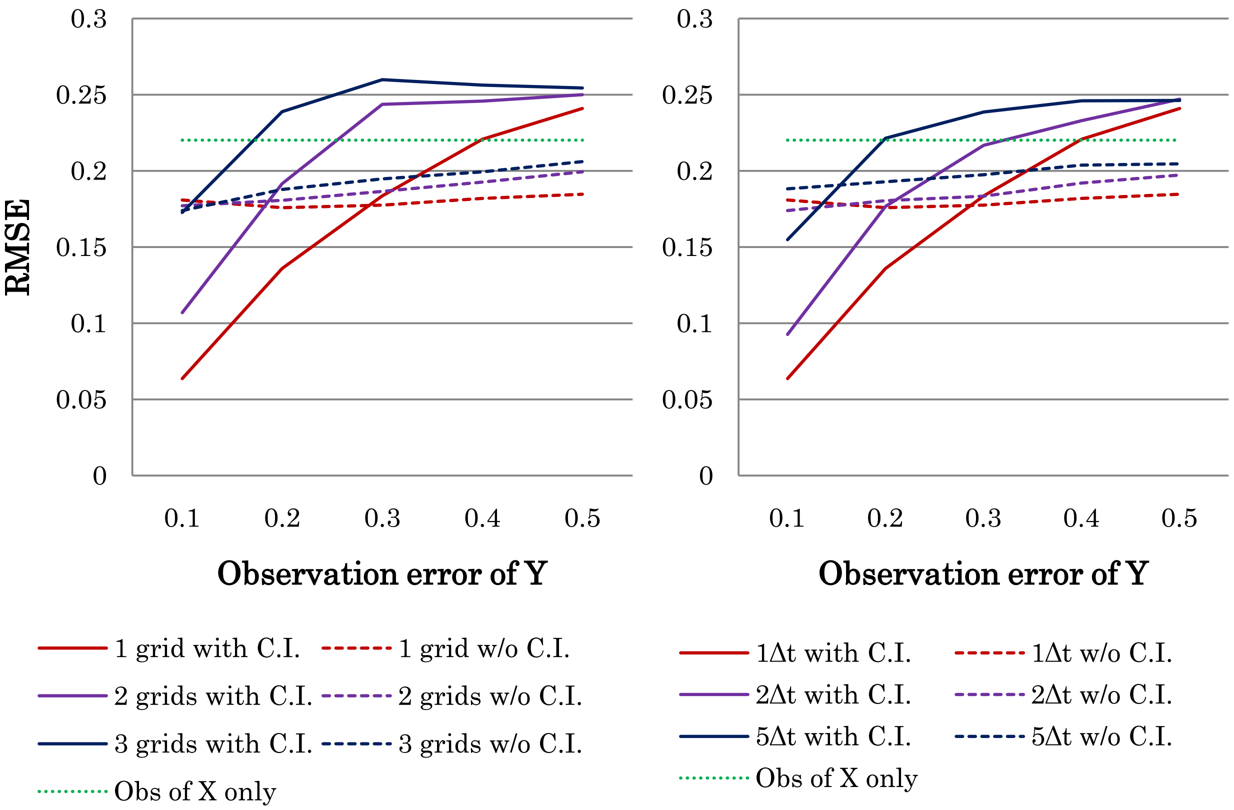

Figure 1. RMSE of analyses of X averaged over 20,000 time steps and all grid points from assimilation experiments with LETKF with 10 ensemble members. The left panel shows the impact of grid interval and error standard deviation of observations of Y when they are available at every time step. The right panel shows the impact of time interval and error standard deviation of observations of Y when they are available at every grid point. “C.I.” in the figure legends stands for the cross analysis increment between the variables of X and Y.