Geophysical Research Abstracts
Vol. 15, EGU2013-7029-1, 2013
EGU General Assembly 2013

Comparison of two ways for representation of the forecast probability density function in ensemble-based sequential data assimilation

Shinya Nakano
The Institute of Statistical Mathematics, Tachikawa, Japan (shiny@ism.ac.jp)

In the ensemble-based sequential data assimilation, the probability density function (PDF) at each time step is represented by ensemble members. These ensemble members are usually assumed to be Monte Carlo samples drawn from the PDF, and the probability density is associated with the concentration of the ensemble members. On the basis of the Monte Carlo approximation, the forecast ensemble, which is obtained by applying the dynamical model to each ensemble member, provides an approximation of the forecast PDF on the basis of the Chapman-Kolmogorov integral.

In practical cases, however, the ensemble size is limited by available computational resources, and it is typically much less than the system dimension. In such situations, the Monte Carlo approximation would not well work. When the ensemble size is less than the system dimension, the ensemble would form a simplex in a subspace. The simplex can not represent the third or higher-order moments of the PDF, but it can represent only the Gaussian features of the PDF. As noted by Wang et al. (2004), the forecast ensemble, which is obtained by applying the dynamical model to each member of the simplex ensemble, provides an approximation of the mean and covariance of the forecast PDF where the Taylor expansion of the dynamical model up to the second-order is considered except that the uncertainties which can not represented by the ensemble members are ignored. Since the third and higher-order nonlinearity is discarded, the forecast ensemble would provide some bias to the forecast.

Using a small nonlinear model, the Lorenz 63 model, we also performed the experiment of the state estimation with both the simplex representation and the Monte Carlo representation, which corresponds to the limited-sized ensemble case and the large-sized ensemble case, respectively. If we use the simplex representation, it is found that the estimates tend to have some bias which is likely to be caused by the nonlinearity of the system rather than the Monte Carlo sampling errors. This bias is observed even if the dynamical model has no model errors, and it tends to be larger when the uncertainty of the estimate is larger as predicted by the theory. The effectiveness of some techniques representing the model errors such as the random system noise and the covariance inflation are also discussed in terms of the bias correction.