Geophysical Research Abstracts
Vol. 14, EGU2012-9181, 2012
EGU General Assembly 2012

Applicability of the particle filter for high-dimensional problems using a massively parallel computer

S. Nakano and T. Higuchi
The Institute of Statistical Mathematics, Tachikawa, Japan (shiny@ism.ac.jp)

The particle filter (PF) is one of ensemble-based algorithms for data assimilation. The PF obtains an approximation of a posterior PDF of a state by resampling with replacement from a prior ensemble. The procedure of the PF does not assume linearity or Gaussianity. Thus, it can be applied to general nonlinear problems. However, in order to obtain appropriate results for high-dimensional problems, the PF requires an enormous number of ensemble members. Since the PF must calculate the time integral for each particle at each time step, the large ensemble size results in prohibitive computational cost.

There exists various methods for reducing the number of particle. In contrast, we employ a straightforward approach to overcome this problem; that is, we use a massively parallel computer to achieve sufficiently large ensemble size. Since the time integral in the PF can be readily be parallelized, we can notably improve the computational efficiency using a parallel computer. However, if we naively implement the PF on a distributed computing system, we encounter another difficulty; that is, many data transfers occur randomly between different nodes of the distributed computing system. Such data transfers can be reduced by dividing the ensemble into small subsets (groups). If we limit the resampling within each of the subsets, the data transfers can be done efficiently in parallel. If the ensemble are divided into small subsets, the risk of local sample impoverishment within each of the subsets is enhanced. However, if we change the grouping at each time step, the information held by a node can be propagated to all of the nodes after a finite number of time steps and the local sample impoverishment can be avoided.

In the present study, we compare between the above method based on the local resampling of each group and the naive implementation of the PF based on the global resampling of the whole ensemble. The global resampling enables us to achive a slightly better effective sample size and thus it provides slightly better estimates under a fixed ensemble size. However, in terms of the computational efficiency, the local resampling is much better than the global resampling. The characteristics of the effective sample size in the local resampling method are also discussed.