Localization in ensemble data assimilation

P. Kirchgessner¹, L. Nerger² and A. Bunse-Gerstner²

¹Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany
²University of Bremen, Germany

Introduction

In data assimilation using ensemble Kalman filter methods, localization is an important technique to get good assimilation results. For the LETKF [1], the domain localization (DL) and observation localization (OL) are typically used. Depending on the localization method, one has to choose appropriate values for the localization parameters, such as the localization length, the inflation factor or the weight function. Although being frequently used, the properties of the localization techniques are not fully investigated. Thus, up to now an optimal choice for these parameters is a priori unknown and they are generally found by doing expensive numerical experiments.

The relationship between the localization length and the ensemble size in DL and OL is studied using twin experiments with the Lorenz-96 model [3]. It is found that for DL, the optimal localization length depends linearly on the local observation dimension. This also holds for the localization length at which the filter diverges. A similar behavior was observed for OL by considering an effective local observation dimension.

Experimental Setup

Filter Configuration

Assimilations were performed by using the LETKF [1] with DL and OL. In each step the whole state was observed. The ensemble was generated by choosing random states from a long model run. The domain decomposition was made by calculating a separate analysis for every single state component. Observations within the localization radius \( l \) were used for the assimilation each model grid point. The localization radius \( l \) was varied from 1 to 20 and the number of ensemble members from 5 to 30.

For OL, the observations were weighted by using the fifth order polynomial introduced by Gaspari and Cohn [3], for several localization radii.

Description of experiments

Twin experiments for various sets of parameters for OL have been performed. The observations, generated with a standard deviation \( \sigma_o = 1 \), have been assimilated for 5000 consecutive time steps. For statistical significance, all experiments were repeated 10 – 20 times. The experiments have been performed with PDAF [4].

Experimental results. The optimal localization radius is nearly linear dependent on the number of ensemble members. The region where the difference is less than 1% from the optimal configuration widens for increasing ensemble size. In the case where the localization radius is much smaller then the ensemble size, the optimal interval is very narrow and the localization radius has to be carefully chosen in order to get optimal results.

Sampling quality

Left: The improved analysis correlates with an improved estimate of the covariance matrix. This was observed by considering the difference \( \delta_i \) between an ideal covariance matrix and the estimate. If the localization radius is to small, the analysis is improved, but the covariance is not well estimated. For moderate localization radii the covariances are better estimated, therefore the analysis becomes better.

Conclusion

By considering the sum of the weights of the weighting function as an approximation to the observation dimension, it is possible to relate the results for both localization techniques. For both methods the curves show similar behavior. This explains the difference in observed behavior between the two methods.

References