Effect of inaccurate specification of time-correlated model error in an Ensemble Smoother

Haonan Ren¹, Peter Jan Van Leeuwen¹,², and Javier Amezcua¹
¹University of Reading, Department of Meteorology, Reading, United Kingdom of Great Britain and Northern Ireland (h.ren@pgr.reading.ac.uk)
²Colorado State University, Department of Atmospheric Science, United States of America (p.j.vanleeuwen@reading.ac.uk)

Data assimilation has been often performed under the perfect model assumption known as the strong-constraint setting. There is an increasing number of researches accounting for the model errors, the weak-constrain setting, but often with different degrees of approximation or simplification without knowing their impact on the data assimilation results. We investigate what effect inaccurate model errors, in particular, the an inaccurate time correlation, can have on data assimilation results, with a Kalman Smoother and the Ensemble Kalman Smoother.

We choose a linear auto-regressive model for the experiment. We assume the true state of the system has the correct and fixed correlation time-scale $\omega_r$ in the model errors, and the prior or the background generated by the model contains the model error with the fixed, guessed time-scale $\omega_g$ which differs from the correct one and is also used in the data assimilation process. There are 10 variables in the system and we separate the simulation period into multiple time-windows. And we use a fairly large ensemble size (up to 200 ensemble members) to improve the accuracy of the data assimilation results. In order to evaluate the performance of the EnKS with auto-correlated model errors, we calculate the ratio of root-mean-square error over the spread of all ensemble members.

The results with a single observation at the end of the simulation time-window show that, using an underestimated correlation time-scale leads to overestimated spread of the ensemble, and with an overestimated time-scale, the results show underestimation in the ensemble spread. However, with very dense observation frequency, observing every time-step for instance, the results are completely opposite to the results with a single observation. In order to understand the results, we derive the expression for the true posterior state covariance and the posterior covariance using the incorrect decorrelation time-scale. We do this for a Kalman Smoother to avoid the sampling uncertainties. The results are richer than expected and highly dependent on the observation frequency. From the analytical solution of the analysis, we find that the RMSE is a function of both $\omega_r$ and $\omega_g$, and the spread or the variance only depends on $\omega_g$. We also find that the analyzed variance is not always a monotonically increasing function of $\omega_g$, and it also depends on the observation frequency. In general, the results show the effect of the correlated model error and the incorrect correlation time-scale on data assimilation result, which is also affected by the observation frequency.