Insights about the hybrid error covariance models

Francesco Sardelli\textsuperscript{1,2,3} and Craig Bishop\textsuperscript{1,2,3}
\textsuperscript{1}University of Melbourne, Faculty of Science, School of Earth Sciences, Australia
\textsuperscript{2}ARC Centre of Excellence for Climate Extremes, Australia
\textsuperscript{3}Bureau of Meteorology, Australia

Hybrid error covariance models construct the covariance matrix to be used in variational data assimilation methods through a linear combination $P_h = \alpha_c P_c + \alpha_e P_l$ of the climatological error covariance matrix $P_c$ and the localized ensemble covariance matrix $P_l = C \otimes P$ with scalar weights $\alpha_c$ and $\alpha_e$.

This work aims to provide a theoretical justification for current hybrid error covariance models and identify a critical issue in them in order to improve them in future research. In the framework of Bayes' theorem, a theory is developed by modelling the climatological distribution of true forecast error covariance matrix $P_f$ as an inverse matrix gamma distribution (prior distribution) and the distribution of the localized ensemble covariance matrix $P_l$ given a true forecast error covariance matrix $P_f$ as a Wishart or matrix gamma distribution (likelihood distribution). The following formulas for the expected values of the prior and likelihood distributions are assumed: $E[P_f] = P_c$ and $E[P_l P_f] = P_f$, respectively. The posterior distribution for the true forecast error covariance matrix $P_f$ given the localized covariance matrix $P_l$ is derived: it turns out to be an inverse matrix gamma distribution. Within this theory, a formula for the expected value $E[P_l P_f]$ of the true forecast error covariance matrix $P_f$ given the ensemble covariance matrix $P$ is derived: $E[P_l P_f] = \beta_c P_c + \beta_e P_l$ (where $\beta_c$ and $\beta_e$ are scalar weights). This provides a theoretical justification for hybrid error covariance models. Moreover, expressions (and thus an interpretation) for the scalar weights $\beta_c$ and $\beta_e$ in terms of the relative variances of the diagonal elements of the prior and likelihood distributions are obtained.

Hence, the consistency of current hybrid covariance models with the assumption $E[P_l P_f] = P_f$ is showed. This assumption is, in turn, inconsistent with $E[PP_f] = P_f$, which ensemble DA schemes are meant to satisfy, and it is falsifiable.

To illustrate the above theory, an experiment is run to simulate 3200 replicate Earth's all having the same true state trajectory, weather prediction system and observational network, but different realizations of the observations. Each replicate Earth is simulated through a 10-variable Lorenz '96 model with an ETKF data assimilation system. From the set of the true forecast errors of all replicate Earth's, the (otherwise hidden) true forecast error covariance matrix $P_f$ is computed at each time step and the (dis)similarity of its climatological distribution from the best-fit inverse matrix gamma distribution is considered. It is found that $P_f$ the inverse-matrix gamma distribution...
overestimates the probability of significant error correlations between widely separated model variables; (ii) it is the un-localized ETKF ensemble covariance matrix that equals the mean climatological covariance matrix, not the localized ensemble covariance matrix. These findings motivate research to discover more accurate approximations to the climatological distribution of the true forecast error covariance matrix and more accurate hybrid covariance models.