EGU21-13764
https://doi.org/10.5194/egusphere-egu21-13764
EGU General Assembly 2021

Insights about the hybrid error covariance models

Francesco Sardelli[1,2,3] and Craig Bishop[1,2,3]
[1]University of Melbourne, Faculty of Science, School of Earth Sciences, Australia
[2]ARC Centre of Excellence for Climate Extremes, Australia
[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 $\mathbf{P}^{\mathbf{h}}= \alpha_c\mathbf{P}^{\mathbf{c}} + \alpha_e\mathbf{P}^{\mathbf{l}}$ of the climatological error covariance matrix $\mathbf{P}^c$ and the localized ensemble covariance matrix $\mathbf{P}^{\mathbf{l}} = \mathbf{C}\circ\mathbf{P}$ with scalar weights α_c and α_e.

This work aims to provide a theoretical justication 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 $\mathbf{P}^f$ as an inverse matrix gamma distribution (prior distribution) and the distribution of the localized ensemble covariance matrix $\mathbf{P}^{\mathbf{l}}$ given a true forecast error covariance matrix $\mathbf{P}^f$ as a Wishart or matrix gamma distribution (likelihood distribution). The following formulas for the expected values of the prior and likelihood distributions are assumed: $E\,[\mathbf{P}^f\,] = \mathbf{P}^c$ and $E\,[\mathbf{P}^{\mathbf{l}}\ \mathbf{P}^f\,]= \mathbf{P}^f$, respectively. The posterior distribution for the true forecast error covariance matrix $\mathbf{P}^f$ given the localised covariance matrix $\mathbf{P}^{\mathbf{l}}$ is derived: it turns out to be an inverse matrix gamma distribution. Within this theory, a formula for the expected value $E\,[\mathbf{P}^f\mathbf{P}\,]$ of the true forecast error covariance matrix $\mathbf{P}^f$ given the ensemble covariance matrix $\mathbf{P}$ is derived: $E\,[\mathbf{P}^f\mathbf{P}]= \beta_c\,\mathbf{P}^{\mathbf{c}}+ \beta_e\mathbf{P}^{\mathbf{l}}$ (where β_c and β_e are scalar weights). This provides a theoretical justication for hybrid error covariance models. Moreover, expressions (and thus an interpretation) for the scalar weights β_c and β_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 $E\,[\mathbf{P}^{\mathbf{l}}\ \mathbf{P}^f\,]= \mathbf{P}^f$ is showed. This assumption is, in turn, inconsistent with $E\,[\mathbf{P}\mathbf{P}^f]= \mathbf{P}^f$, which ensemble DA schemes are meant to satisfy, and it is falsiable.

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 $\mathbf{P}^f$ 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 **(i)** the inverse-matrix gamma distribution

overestimates the probability of signicant 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.