Empirical determination of the covariance of forecast errors: an empirical justification and reformulation of Hybrid covariance models

Diego Saul Carrio Carrio[1], Craig Bishop[2], and Shunji Kotsuki[3]

[1]The University of Melbourne, Melbourne, Australia (diego.carriocarrio@unimelb.edu.au)

[2]The University of Melbourne, Melbourne, Australia (craig.bishop@unimelb.edu.au)

[3]Center for Environmental Remote Sensing (CEReS), Chiba University (shunji.kotsuki@chiba-u.jp)

The replacement of climatological background error covariance models with Hybrid error covariance models that linearly combine a localized ensemble covariance matrix and a climatological error covariance matrix has led to significant forecast improvements at several forecasting centres. To deepen understanding of why the Hybrid's superficially ad-hoc mix of ensemble based covariances and climatological covariances yielded such significant improvements, we derive the linear state estimation equations that minimize analysis error variance given an imperfect ensemble covariance. For high dimensional models, the computational cost of the very large sample sizes required to empirically estimate the terms in these equations is prohibitive. However, a reasonable and computationally feasible approximation to these equations can be obtained from empirical estimates of the true error covariance between two model variables given an imperfect ensemble covariance between the same two variables. Here, using a Data Assimilation (DA) system featuring a simplified Global Circulation Model (SPEEDY), pseudo-observations of known error variance and an ensemble data assimilation scheme (LETKF), we quantitatively demonstrate that the traditional Hybrid used by many operational centres is a much better approximation to the true covariance given the ensemble covariance than either the static climatological covariance or the localized ensemble covariance. These quantitative findings help explain why operational centres have found such large forecast improvements when switching from a static error covariance model to a Hybrid forecast error covariance model. Another fascinating finding of our empirical study is that the form of current Hybrid error covariance models is fundamentally incorrect in that the weight given to the static covariance matrix is independent of the separation distance of model variables. Our results show that this weight should be an increasing function of variable separation distance. It is found that for ensemble covariances significantly different to zero, the true error covariance of spatially separated variables is an approximately linear function of the corresponding ensemble covariance, However, for small ensemble sizes and ensemble covariances near zero, the true covariance is an increasing function of the magnitude of the ensemble covariance and reaches a local minimum at the precise point where the ensemble covariance is equal to zero. It is hypothesized that this behaviour is a consequence of small ensemble size and, specifically,

associated spurious fluctuations of the ensemble covariances and variances. Consistent with this hypothesis, this local minimum is almost eliminated by quadrupling the ensemble size.