State Estimation with Huge Ensembles: Year 1 Progress Report

PI: Craig H. Bishop, Post-Doc: Justin Mclay, Masters Student: Kevin Shanley.

- A starting point for the project is a good non-linear global ensemble. To this end, Justin McClay, Carolyn Reynolds and I have developed and begun testing a NOGAPS Ensemble Transform (ET) ensemble. The ET ensemble produces initial perturbations that are (a) orthogonal under an analysis error variance norm, and (b) have initial amplitudes consistent with analysis error variance estimates produced by the NAVDAS data assimilation scheme.

- The software to augment and enlarge the ensemble using over 1000 historical perturbations has been developed. The first application of this software has been to represent model error in the 32 member ET ensemble. It had been found that after a week or two of cycling 32 ET perturbations, the ensemble became overdispersive in midlatitudes and underdispersive in the tropics. To address this problem, we developed a method of transforming 1000 or so historical perturbations into a "model error" ensemble that when drawn from and added to the cycling ET ensemble largely removes systematic latitudinal dependent biases in the ensemble variance. A figure illustrating this improvement is included in the attached powerpoint file.

- A new procedure for generating space-time deformation ensemble members was developed based on spherical-harmonics. This new code is an order of magnitude faster than the old code.

- A data assimilation scheme can only be as good as its error covariance models. For state estimation in huge ensembles we envisage combining historical perturbations with space-time deformation perturbations with NOGAPS ET perturbations. Consequently, it is vitally important to be able to measure the accuracy of the variance predictions made by the covariance models individually and collectively. However, in flow-dependent, heteroskedastic, chaotic systems the error covariance given the forecast is not directly observable. All one observes is a single realization from the distribution whose error covariance you are trying to predict. In our view, the most important scientific result from our THORPEX funded research this past year has been the development of a new method of assessing variance prediction accuracy. The new method views variance predictions as random draws from a Gamma distribution of variances whose mean variance is a linear function of the true variance. (Chi-square is a special case of Gamma). The viewpoint leads to a variance prediction accuracy regression problem with multiplicative non-Gaussian noise which has now been solved. We are hopeful that this (new?) solution will enable us to optimally tune the raw and hybrid error covariance models that are a cornerstone of the envisioned huge ensemble data assimilation framework. Amongst other things, the solution enables one to use the outcomes of many independent trials to estimate the distribution of true error variances given an ensemble sample variance. An example is shown on the powerpoint file.