**Coupled data assimilation and ensemble initialisation with application to multi-year ENSO prediction**

## Terence J. O’Kane1, Paul A. Sandery1, Didier P. Monselesan1, , Pavel Sakov3, Matt. A. Chamberlain1, Richard J. Matear1, Mark Collier2, Lauren Stevens2

*1CSIRO, Ocean and Atmosphere, Hobart*

*2CSIRO, Ocean and Atmosphere, Aspendale*

*3Environment and Research Division, Bureau of Meteorology, Hobart*

*terence.okane@csiro.au*

We develop and compare variants of coupled data assimilation (DA) systems based on ensemble optimal interpolation (EnOI) and ensemble transform Kalman filter (ETKF) methods. The assimilation system is first tested on a small paradigm model of the coupled tropical-extratropical climate system, then implemented for a coupled general circulation model (GCM). Strongly coupled DA was employed to perform an observing system simulation experiment (OSSE) and to assess the impact of assimilating ocean observations (SST, SSH, SSS, Argo, XBT, CTD, moorings) on the atmospheric state analysis update via the cross-domain error covariances from the coupled-model background ensemble. We examine the relationship between ensemble spread, analysis increments and forecast skill in multi-year ENSO prediction experiments with a particular focus on the atmospheric response to tropical ocean perturbations. Initial forecast perturbations generated from bred vectors (BV) project onto disturbances at and below the thermocline with similar structures to ETKF perturbations. BV error growth leads ENSO SST phasing by 6 months whereupon the dominant mechanism communicating tropical ocean variability to the extra-tropical atmosphere is via tropical convection modulating the Hadley circulation. We find that bred vectors specific to tropical Pacific thermocline variability were effective choices for ensemble initialization and ENSO forecasting.

We first consider a paradigm model of tropical - extratropical interactions. In this small 9D model, three versions of the famous Lorenz 63 model are coupled to mimic the temporal behavior of an extratropical atmosphere weakly coupled to a tropical atmosphere which in turn is strongly coupled to a slow tropical ocean. We compare EnOI and ETKF data assimilation where only “ocean” observations are assimilated but cross-domain covariances are included between all state variables. These simulations point to the potential for a well constrained ocean state to also constrain the tropical atmosphere, in large part due to the strong coupling between ocean and tropical atmosphere. Here, and in the context of the paradigm model, well constrained requires that flow dependent information be captured in the ocean-tropical atmosphere cross covariance. While flow dependent information constrains the ocean - tropical atmosphere, despite being weakly coupled, cross covariance information between the ocean, tropical and the extra-tropical atmosphere attractors led to a suppression of the variance in the analyzed extra-tropics. In contrast, where static rather than flow dependent cross covariances are employed, the analyzed tropical atmospheric state, despite being strongly coupled to the ocean attractor, fails to track the truth, but the variance of the analyzed extra-tropical attractor was largely unchanged even with increments due to the cross covariance included. These simple model experiments suggest that one might best initialize ensemble climate forecasts by constraining the slow modes of the ocean with only a relatively weak large scale projection of ocean observations into the fast extratropical atmospheric circulation. It is on this basis that we then examined strongly coupled DA variants applied to the GCM, where the ocean is constrained, either with static or flow dependent cross covariances, and where the large scales of the atmosphere are modified based on suitably scaled ocean-atmosphere cross covariances.

Our focus is on seasonal and longer timescales, and in particular ENSO. Therefore, our premise underpinning the OSSE’s is that predictability primarily resides in the oceans and the fast atmosphere acts as a stochastic driver on the longer timescale ocean variability. We again considered two approaches to DA based on ETKF and EnOI, assimilating a wide range of ocean observations into a GCM. Outside of the tropics the ETKF system produced dramatically lower forecast bias and forecast mean absolute deviations (MAD) relative to the EnOI system however, these improvements were substantially reduced in the tropics. The reason for the low analysis error in the EnOI system in the tropics was found to be a result of seasonally dependent fixed ensemble spread at times producing larger observation impacts relative to the tropical ETKF where interannual variations in the background covariances can lead to periods of relatively reduced spread.

Initial forecast perturbations using bred vectors (BV's) representative of growing coupled tropical instabilities were found to modify tropical convection, particularly in the region of the maritime continent, which in turn generate a coherent modulation of the Hadley circulation. A direct renormalization of thermocline disturbances was found to be most effective in communicating information from the tropical ocean to the extra-tropical atmosphere on timescales of a couple of weeks to a month. Comparison of ensemble forecasts based on two types of bred vectors (masked and unmasked) centered about the EnOI analyzed state reveal a substantial reduction in uncertainties (forecast spread), when disturbances not directly associated with thermocline variability are eliminated. In particular, excluding SST disturbances led to a significant reduction in forecast errors in multi-year ENSO predictions and noticeably increased skill at lead-times out to two years. These results affirm the utility of using BV's explicitly constructed to project onto forecast errors entirely due to tropical subsurface ocean disturbances where the appropriate variance resides in application to ENSO prediction.

The OSSEs and methods discussed form a basis for coupled DA relevant to multi-year near term climate forecasts. The masked isosurface BV approach allows for the specific targeting of regions of large scale variability pertinent to dynamical processes that determine predictability on seasonal to interannual spatio-temporal scales. Beyond a season, strongly coupled data assimilation, where the slow ocean modes are explicitly constrained including projection onto the background atmospheric states (i.e. jets, cells etc) while leaving the fast atmospheric dynamics (synoptic scales) free, including targeted forecast perturbations, offers a pragmatic approach to determining the mechanisms and predictability of the key climate modes.

This work further highlights the complexity of data assimilation and forecast initialization in nonlinear multiscale systems. While we have demonstrated the advantages of flow dependent ocean data assimilation and the usefulness of ocean observations to constrain the large scales of the atmosphere, it is apparent that assimilation of atmospheric observations is further required to guarantee the correct extratropical variability. This is a focus of our ongoing work, as are methods to identify an appropriate theoretical basis necessary for identifying causal relationships between climate modes and for determining predictability on given spatio-temporal scales of interest and as a basis for developing a generalizable approach to multiscale forecast initialization.