Correlation-Cutoff Method for Covariance Localization in Strongly Coupled Data Assimilation

Takuma Yoshida[1] and Eugenia Kalnay[1]

[1]University of Maryland College Park

November 26, 2022

Abstract

Due to its inherent ability to estimate the background error covariances, an ensemble Kalman filter (EnKF) is thought to be a practical approach to the strongly coupled data assimilation problems, where an entire coupled model state is estimated as if it was a single integrated system. However, increased complexity and the multiple time scale of the coupled system aggravate the rank-deficiency and spurious correlation problems caused by limited ensemble size available for the analysis. To alleviate these problems, a distance-independent localization method to systematically select the observations to be assimilated into each model variable has been developed and successfully tested with a nine-variable coupled model with slow and fast modes. This method, called correlation-cutoff method, utilizes the mean squared ensemble error correlation between each observable and model variable to identify where the cross-update should be used, and we cut off the assimilation of observations when the squared error correlation becomes small. To implement the method on a more realistic model, we thoroughly investigate inter-fluid background covariances in an atmosphere-ocean coupled general circulation model where the spatiotemporal scales of coupled dynamics significantly vary by latitudes and driving processes.

Correlation-Cutoff Method for Covariance Localization in Strongly Coupled Data Assimilation (A23I-2971)

Takuma Yoshida (tyoshida@terpmail.umd.edu) and Eugenia Kalnay

Department of Atmospheric and Oceanic Science, University of Maryland, College Park

1. Background

By using coupled models for the background state estimate, *coupled data assimilation* tries to provide more consistent and accurate state estimate of a coupled system (e.g., atmosphere-ocean system). We can broadly classify coupled DA methodologies into *weakly* coupled DA (coupled background, uncoupled update) and *strongly* coupled DA (coupled background and update) (Penny et al., 2017). Since strongly coupled DA tries to solve a larger problem, strongly coupled ensemble Kalman filters often suffer from the rank-deficiency problem and need to be localized appropriately.

Weakly coupled DA

$$\mathbf{B}_{\text{weak}} = \begin{bmatrix} \mathbf{B}_{\text{AA}} & \mathbf{O} \\ \mathbf{O} & \mathbf{B}_{\text{OO}} \end{bmatrix}$$

Consider background error cross-covariance →

Strongly coupled DA

$$\mathbf{B}_{\text{strong}} = \begin{bmatrix} \mathbf{B}_{\text{AA}} & \mathbf{B}_{\text{AO}} \\ \mathbf{B}_{\text{AO}}^{\text{T}} & \mathbf{B}_{\text{OO}} \end{bmatrix}$$

EnKF w/o localization

$$\mathbf{B} \propto \mathbf{X}'^{b}\mathbf{X}'^{b\text{T}}$$

← Ignore distant background error covariance

Localized EnKF

$$\mathbf{B} \propto \rho \circ \left[\mathbf{X}'^{b}\mathbf{X}'^{b\text{T}}\right]$$

2. Theory

With sequential assimilation of observations, the relative reduction of the analysis error variance by assimilation of each observation can be written as a product of two quantities (Yoshida and Kalnay, 2018):

a) relative accuracy of the observation to the background

b) squared background error correlation between the analyzed and observed variables

$$\frac{\sigma_{bi}^2 - \sigma_{ai}^2}{\sigma_{bi}^2} = \frac{\sigma_{y_b}^2}{\sigma_{y_b}^2 + \sigma_{y_o}^2}\,\text{corr}^2(\delta x_{bi}, \delta y_b)$$

Therefore, we should mutually assimilate observations (i.e., strongly coupled DA) only between pairs of variables that have well-correlated background errors.

3. "Correlation-cutoff method" procedure

(i) Run an EnKF analysis cycle in any form ("offline cycle")

(ii) Estimate the strength of background error correlation between each pair of observation and analysis variables from the mean squared background ensemble correlation of (i)

(iii) Conduct strongly coupled analysis, coupling only the pairs of variables with large mean squared background ensemble correlation

References

Jacob, R. L., 1997: Low Frequency Variability in a Simulated Atmosphere Ocean System. 159 pp.

Peña, M., and E. Kalnay, 2004: Separating fast and slow modes in coupled chaotic systems. *Nonlinear Process. Geophys.*, **11**, 319–327, doi:10.5194/npg-11-319-2004.

Penny, S. G., and Coauthors, 2017: Coupled Data Assimilation for Integrated Earth System Analysis and Prediction: Goals, Challenges and Recommendations. *World Meteorological Organization*, **50**.

Yoshida, T., and E. Kalnay, 2018: Correlation-Cutoff Method for Covariance Localization in Strongly Coupled Data Assimilation. *Mon. Wea. Rev.*, **146**, 2881–2889, doi:10.1175/MWR-D-17-0365.1.

4. Toy model experiments

The 9-variable model of Peña and Kalnay (2004) mutually couples three Lorenz63 model with different timescales, representing an "Ocean", a "Tropical atmosphere", and an "Extratropical atmosphere". The model offers a realistic and handy testing environment for coupled DA methodologies.

With the correlation-cutoff method, we test partially coupled analysis, where only tropical atmosphere and ocean observations are mutually assimilated (Figure 1). We achieved better analysis accuracy than both weakly coupled and fully coupled analysis, especially with smaller ensembles (Figure 2).

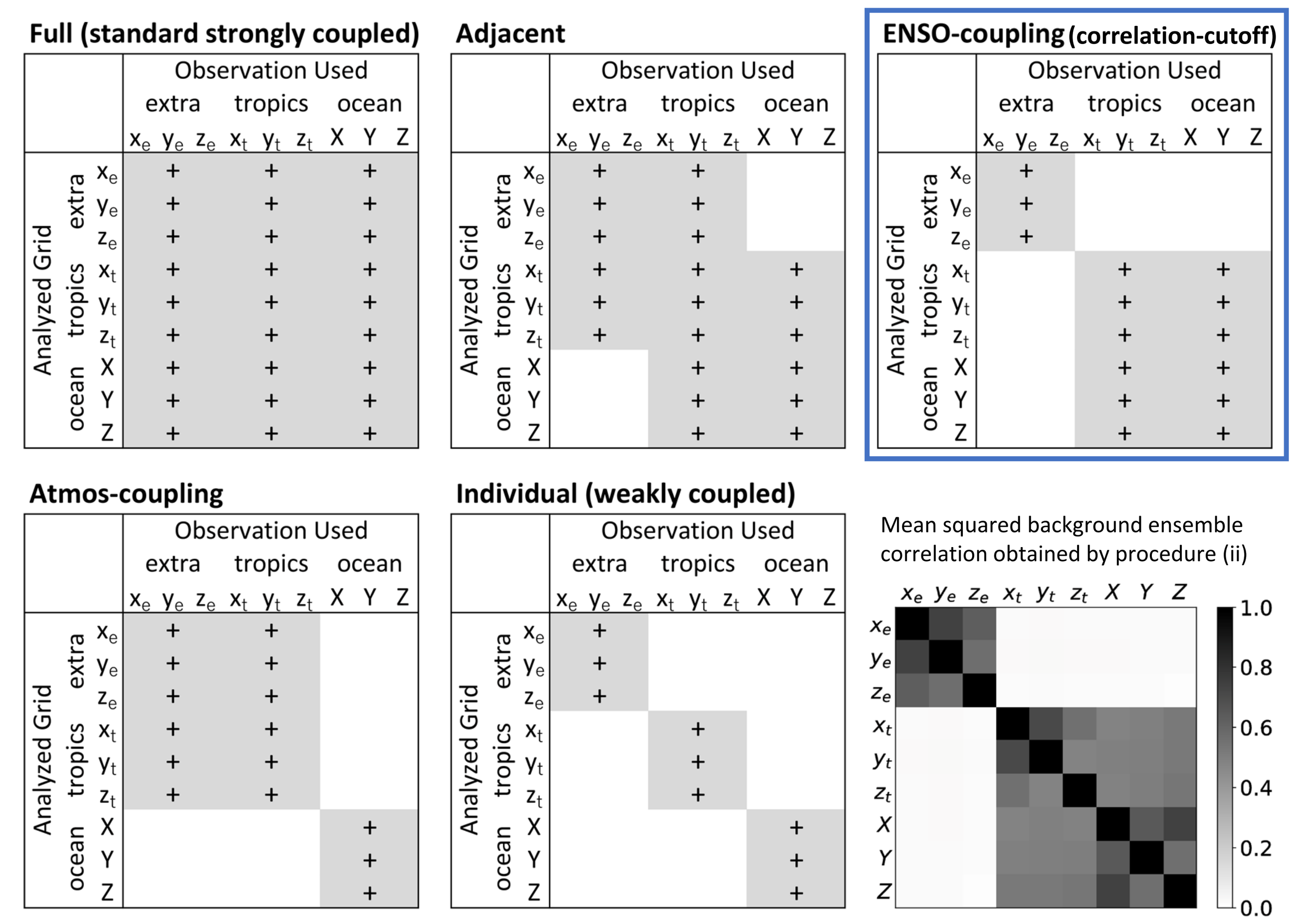

Figure 1: Covariance localization patterns tested for the toy model. Shading indicates allowed background error covariances between components. The lower right insertion shows the mean squared background ensemble correlation obtained by procedure (ii), with a 10-member, fully coupled LETKF, which indicates that the *ENSO-coupling* localization is optimal.

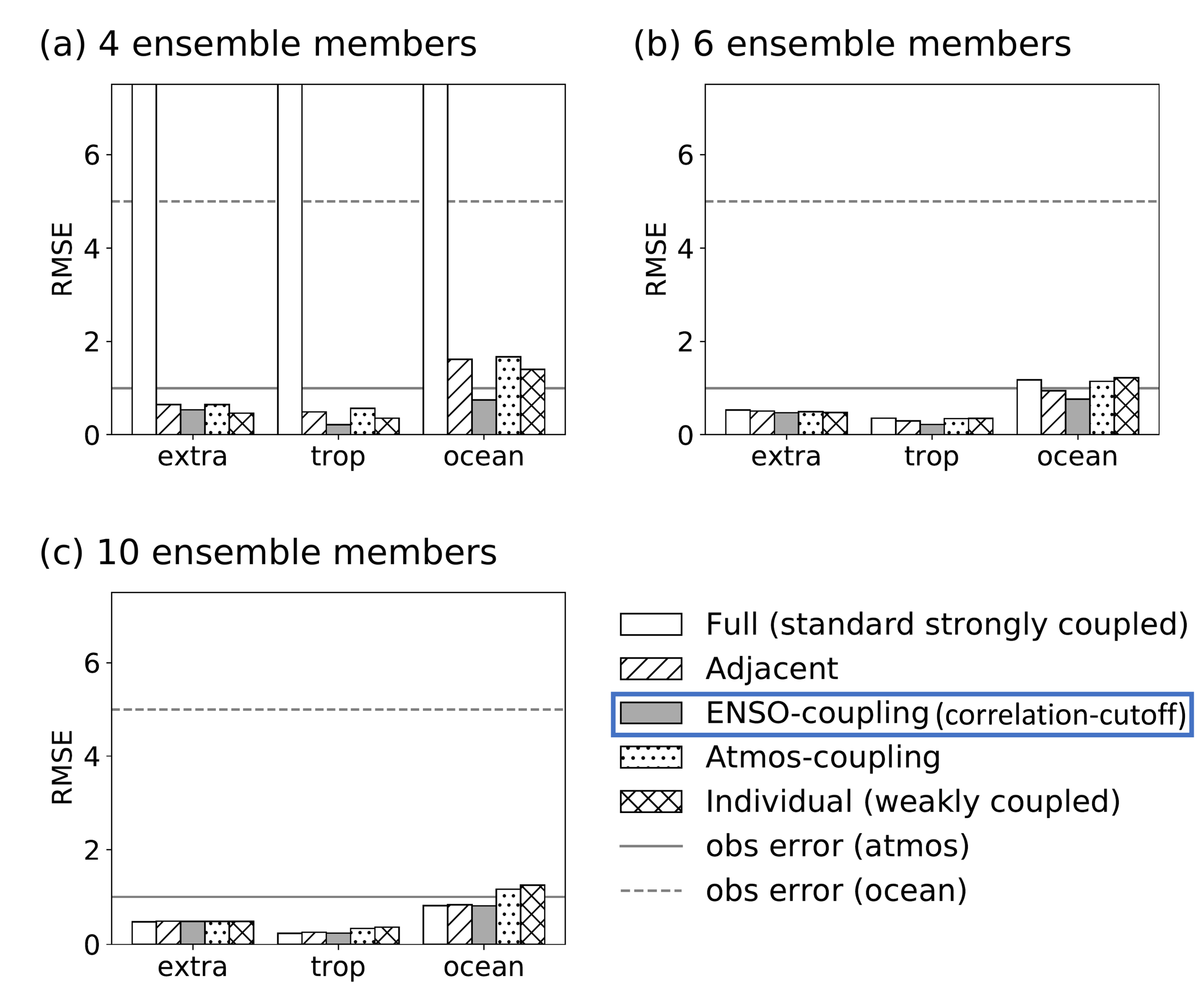

Figure 2: Temporal mean analysis root-mean-square error (RMSE) for each toy model analysis experiment. The shading indicates the covariance localization pattern used. Each panel is the result of experiments with (a) 4, (b) 6, and (c) 10 members. Note that the filter diverged in the 4-member Full experiment.

5. Global model experiments

We employ a global atmosphere-ocean coupled model (FOAM; Jacob 1997) and observation system simulation experiments (OSSEs) for the investigation of coupled background error correlations. The background ensemble correlations of a 64-member, weakly coupled LETKF cycle is examined as a proxy to the unknown background error correlations of the coupled system.

Figure 3 shows examples of temporary aggregated background error correlation (represented by an ensemble perturbations) between the atmosphere and the ocean. Panel (a) shows that a hypothetical surface air observation can constrain the subsurface temperature. Panel (b) shows that a hypothetical surface wind observation can constrain the subsurface current below, which are negatively correlated (explained by linearized Ekman layer dynamics). Such background ensemble correlations between atmosphere and ocean surface variables are globally found.

We use a small neural network to generalize the complex ensemble correlation structure as a function of distance, observation type, analysis variable type, etc. The network succeeded to reproduce the background ensemble correlations of the global system, including the coupled ones between the atmosphere and the ocean (Figure 4). Since the neural networks are compact and fast, we will be able to use them as an optimal localization function for strongly coupled ensemble Kalman filters (correlation-cutoff method).

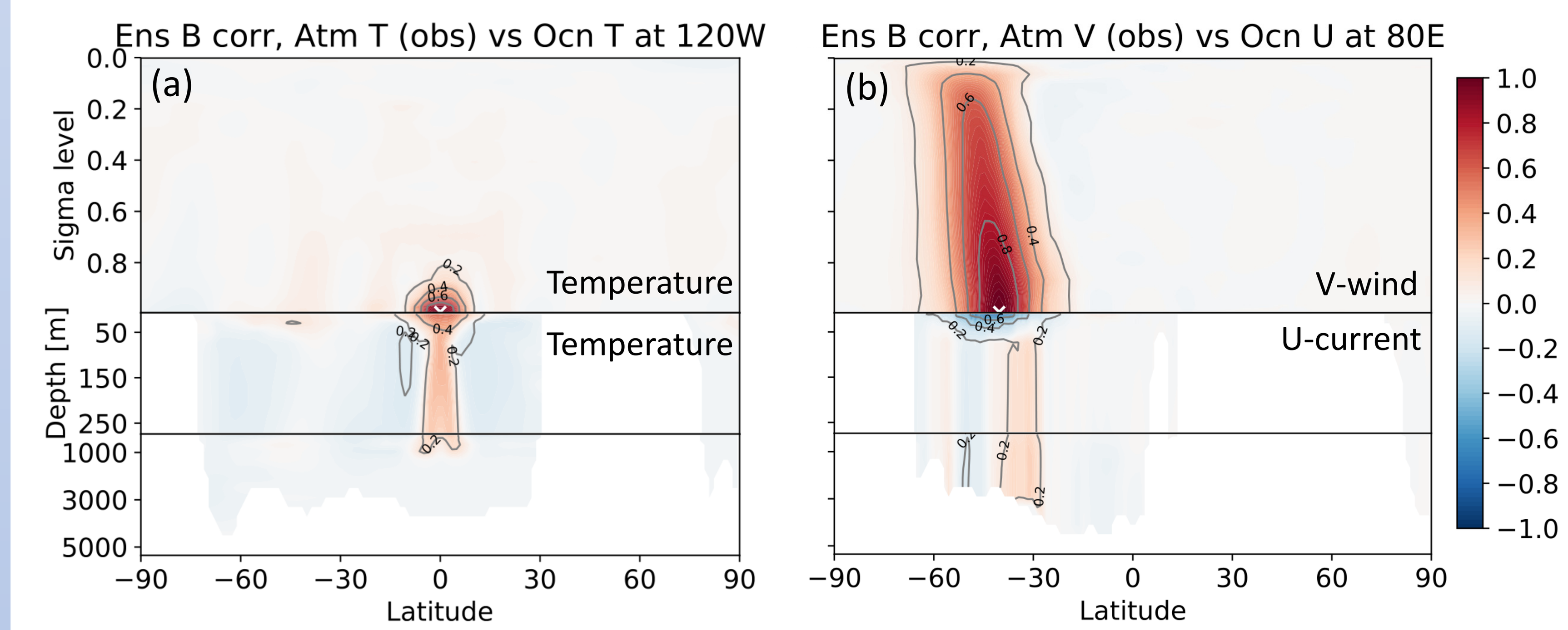

Figure 3: Temporarily aggregated background ensemble correlation of (a) a hypothetical surface air temperature observation at the equator with the ocean temperature field (and the air temperature field itself) (b) a hypothetical surface meridional wind observation to the zonal current field (and the meridional wind field itself). Shading and contours show temporal mean and temporal root-mean-squared respectively. The latter includes flow-dependent portion.

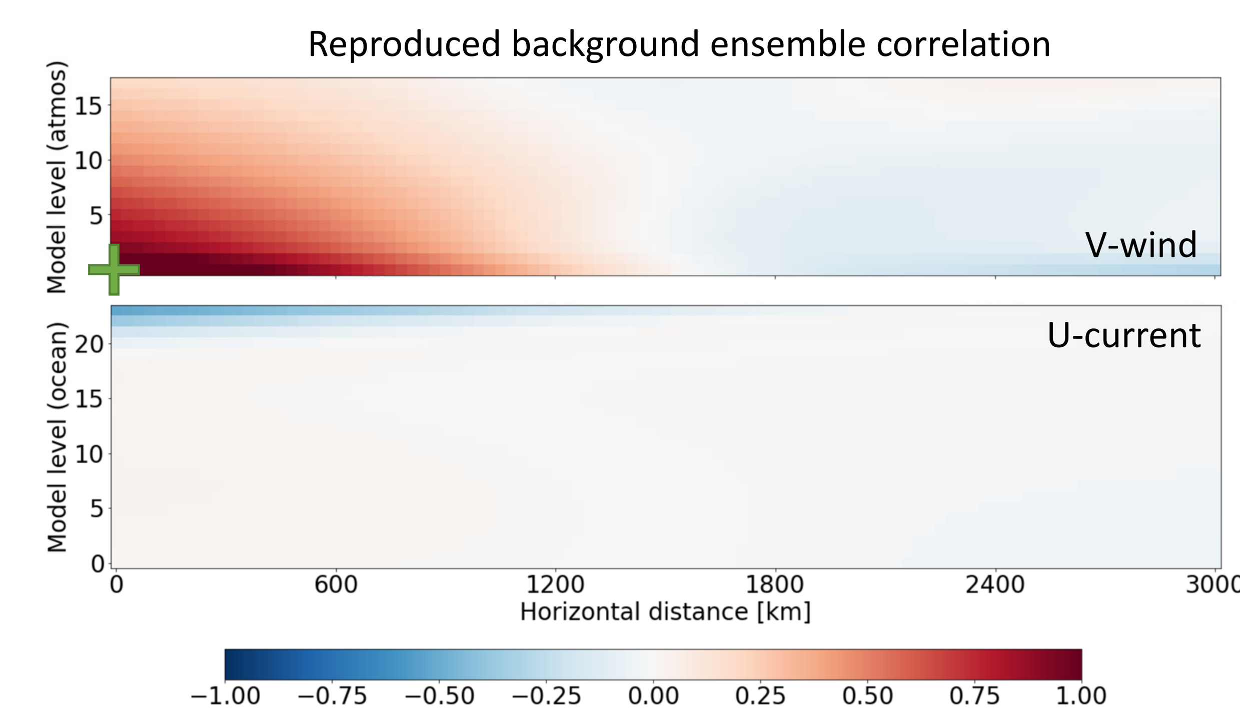

Figure 4: Background ensemble correlation to a hypothetical observation of surface meridional wind at 45S (green plus sign) reproduced by the neural network. The upper half shows the background ensemble correlation field of meridional wind (atmosphere) as a function of analysis model level and horizontal distance. The lower half shows the background ensemble correlation field of zonal current (ocean), as a function of analysis model level and horizontal distance. Compare with Figure 3b.