ABSTRACT

Title of Dissertation: EXPLORING THE OBSERVATION IMPACTS AND ENHANCING THE PREDICTABILITY FOR ENSEMBLE-BASED COUPLED DATA ASSIMILATION

Chu-Chun Chang, Doctor of Philosophy, 2023

Dissertation directed by: Professor Eugenia Kalnay
Dept. of Atmospheric and Oceanic Science

This research aims to explore the observation impacts in coupled data assimilation (CDA) and improve the predictability of coupled systems by advanced DA approaches. Three topics are discussed in this dissertation:

(1) An enhanced application of the correlation cutoff method (Yoshida and Kalnay, 2018) as a spatial localization is introduced. We investigated the feasibility and characteristics of the traditional distance-dependent (Gaspari and Cohn, 1999) and the correlation-dependent localizations preliminary on the Lorenz (1996) model with the local ensemble transform Kalman filter (LETKF). We further discussed the potential of integrative localization strategies and the application of the correlation cutoff method on Mars DA.

(2) We found that the surface sea temperature (SST) relaxation operationally used in the Climate Forecast System version 2 (CFSv2) is not effective in reducing existing SST biases. To

address this issue, we replaced the SST relaxation with the weakly coupled data assimilation (WCDA) of satellite-retrieved SST products. A series of experiments with real observations were conducted on the CFSv2-LETKF (Sluka et al., 2018) to investigate the impacts of SST WCDA on the CFSv2 analysis and the forecasts.

(3) The Ensemble Forecast Sensitivity to Observations (EFSO, Kalnay et al., 2012) is a powerful tool to identify the beneficial or detrimental impact of *every* observation and has been widely used in atmospheric ensemble-based DA. However, EFSO has not yet been applied to any ocean or coupled DA due to the lack of a proper error norm for oceanic variables. This study first introduces a novel density-based error norm that simultaneously includes sea temperature and salinity forecast errors, by which EFSO becomes available to ocean DA for the first time. We implemented the oceanic EFSO on the CFSv2-LETKF for quantifying the individual impact of ocean observations and explored the great potential of EFSO to be extended as a data selection criterion to improve the CFSv2 forecasts.

EXPLORING THE OBSERVATION IMPACTS AND ENHANCING THE PREDICTABILITY FOR ENSEMBLE-BASED COUPLED DATA ASSIMILATION

by

Chu-Chun Chang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Professor Eugenia Kalnay, Chair/Advisor
Professor James A. Carton
Professor Brian R. Hunt,
Professor Jonathan Poterjoy
Professor Steven J. Greybush
Dr. Tse-Chun Chen
Dr. Safa Mote

Acknowledgments

My journey toward a Ph.D. was challenging and full of difficulties, and I would not be able to complete it without the support and help of many people, so here I wish to express my gratitude to:

My advisor Prof. Eugenia Kalnay, for her mentoring, inspiring ideas, and support. She always encourages me and guides me when I get lost. I'm very lucky to be her student.

My co-advisors Prof. Steven Greybush and Dr. Tse-Chun Chen, for guiding me on the Mars and EFSO projects. I really enjoy working with them and have learned a lot from them.

Thanks to my dissertation committee members, Prof. James Carton, Prof. Jonathan Poterjoy, Prof. Brian Hunt, and Dr. Safa Mote. They gave me many great ideas and helpful suggestions to improve this dissertation work.

Thanks to my former advisors in Taiwan, Prof. Shu-Chih Yang and Dr. Mao-Chang Liang. They were my first teachers who taught me how to do research and inspired me to become a scientist.

Many thanks to my colleagues in the UMD Chaos group, Dr. Cheng Da, Dr. Takuma Yoshida, Dr. Luyu Sun, Dr. Travis Sluka, Dr. Kriti Bhargava, and Dr. Eviatar Bach, for all the exciting and helpful discussions on data assimilation.

Thanks to all my friends in Taiwan and the U.S. for their emotional support and accompany.

Finally, the deepest gratitude goes to my family, especially my mother and younger brother, for their endless support and patience.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AMSR	Advanced Microwave Scanning Radiometer
AVHRR	Advanced Very High Resolution Radiometer
CDA	Coupled data assimilation
CFSv2	Climate Forecast System version 2
DA	Data assimilation
ECMWF	European Centre for Medium-Range Weather Forecasts
EFSO	Ensemble Forecast Sensitivity to Observation
EnKF	Ensemble Kalman Filter
GFDL	Geophysical Fluid Dynamics Laboratory
GFS	Global Forecast System
LETKF	Local Ensemble Transform Kalman Filter
MOM	Modular ocean model
NASA	National Aeronautics and Space Administration
NCEP	National Centers for Environmental Prediction
OISST	optimum interpolation sea surface temperature
OSSE	observing system simulation experiment
RMSE	root mean square error
RTPS	relaxation to prior spread
SCDA	Srongly coupled data assimilation
SST	Sea surface temperature
SSS	Sea surface salinity
WCDA	Weakly coupled data assimilation

Chapter 1: Introduction

1.1 Data assimilation (DA) on coupled model initialization

Coupled atmosphere-ocean models have become an important tool for predicting phenomena with a wide range of time scales, such as the El Niño Southern Oscillation (ENSO) and the Madden Julian Oscillation (Woolnough et al., 2007; Zhang, et al., 2020; Moore & Kleeman, 1996), monsoons (Wang, 2005), and tropical cyclones (Bender et al., 1993). Coupled models tend to integrate various Earth system components, including the atmosphere, ocean, land, and sea ice, and have demonstrated remarkable improvements in the synoptic weather, seasonal-to-sub-seasonal (S2S), and interannual climate mesoscale forecasting (Penny & Hamill, 2017). Nowadays, many countries have developed their own fully coupled models for operational numerical weather/climate predictions, such as the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) and the National Centers for Environmental Prediction (NCEP) Climate Forecast System (CFS).

Due to the imperfections in the model dynamics and the initial condition used for forecasting, data assimilation (DA) becomes essential for providing a more precise initial condition for model initialization and enhancing the state estimation (e.g., analysis) by optimally combining observations and model states. In general, there are two ways to initialize the coupled model: uncoupled DA and coupled DA (CDA). Traditionally, in a convenient way, uncoupled DA is designed to generate atmosphere

and ocean state estimations separately for the coupled model, as shown in Figure 1.1 (a). The updated atmosphere and ocean analyses would then be used to initialize the coupled model. In other words, the observation information from one domain (e.g., atmosphere) would not be included in either the forecast step or the analysis step of another domain (e.g., ocean). Although the uncoupled DA is low-cost and easy to implement, it may induce a strong imbalance in the coupled model states due to the mismatch between the separate analyses, causing "initial shocks" in the initial condition and consequently degrading the analysis and forecast skills (Rosati et al.,1997; Zhang, 2011; Mulholland et al., 2015).

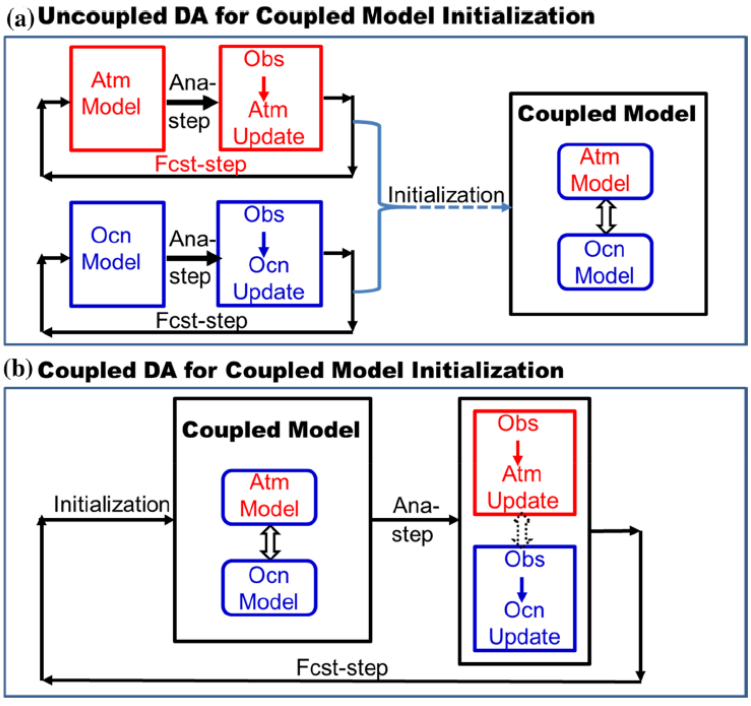

Figure 1.1 Schematic of (a) uncoupled DA and (b) coupled DA for coupled model initialization. (Zhang et al., 2020).

The coupled DA (CDA), unlike the uncoupled DA, is directly conducted with the coupled model, as shown in Figure 1.1 (b). For CDA, the air-sea coupling dynamics and the observation information from another domain would be involved in both the forecast and analysis steps. Therefore, CDA is shown to be a better solution for coupled model initialization and could significantly improve climate predictions as well as benefit a wide range of numerical predictions (Penny and Hamill, 2017).

CDA can generally be divided into weakly coupled DA (WCDA) and strongly coupled DA (SCDA). The main difference between these two types is the involvement of cross-domain corrections during DA. For WCDA, the atmosphere and ocean states are updated by their respective observations. Namely, there's no cross-domain error correlation between the atmospheric and oceanic variables. However, the increments made in one domain (e.g., atmosphere) would influence another domain (e.g., ocean) through the flux exchanges during coupled forecasting. That said, the signals of the air-sea coupling dynamics could be delivered in the background error covariance of WCDA, which is the major difference between WCDA and uncoupled DA. Thus, in contrast to uncoupled DA, WCDA is shown to have a more balanced analysis and is effective in improving climate predictions (Zhang et al., 2011).

For SCDA, observations can be directly used to update both atmosphere and ocean domains (represented as the dashed arrows in Figure 1.1 (b)). Theoretically, SCDA is the optimal approach to the coupled model initialization because it considers the coupling dynamics in both the forecast and analysis steps. However, the implementation of SCDA is relatively challenging because its feasibility and

robustness depend on not only the design of the DA algorithm but also the understanding of the air-sea coupling dynamics (Zhang et al., 2020). Although several pioneering approaches to SCDA with simplified models have been proposed in recent years (Sluka et al., 2016; Sluka, 2018; Yoshida and Kalnay, 2018; Frolov et al., 2016), showing the great potential of SCDA in improving the climate predictions, more efforts are still required for SCDA to be implemented in practice and operationally.

Currently, most operational centers use WCDA for their coupled model initialization, including the NCEP operational CFS. The atmosphere and the ocean domains of the CFS are independently initialized by the Gridpoint Statistical Interpolation system (GSI) (Kleist et al., 2009) and the Global Ocean Data Assimilation System (GODAS) (Behringer 2007), respectively. The GSI is a variational DA system (e.g., 3D-Var), while flow-dependent error statistics derived from ensemble forecasts within a 6-hr window are introduced to rescale the static variance used in cost function (Saha et al., 2010; Saha et al., 2014). The GODAS is originally evolved from Derber and Rosati (1989) and uses a 3D-Var scheme to assimilate ocean temperature and salinity. In the CFS, relaxation (nudging) of the surface sea temperature and salinity is taken place at the top level as a stronger constraint to prevent field drifting. A brief introduction to nudging is in Section 1.1.3.

1.1.1 The ensemble-based coupled DA

The ensemble Kalman filter (EnKF, Evensen, 2003) is a widely used and convenient method for CDA applications. It has significant advantages, such as flow-

dependent error statistics, and it doesn't require the use of tangent linear and adjoint models, making it easy to implement on complicated, multi-component coupled models.

Many studies have successfully implemented EnKFs to initialize coupled models, from simple toy models to large operational models. Chen and Zhang (2019) developed a regional fully coupled EnKF system to improve tropical cyclone (TC) prediction. Their OSSE of Hurricane Florence (2018) results show that assimilating SST, sea-surface height anomaly, and sea surface current could provide better TC analysis and forecasts. Zhang et al. (2007) developed an ensemble-based CDA system for GFDL second-generation coupled model (CM2). Sluka et al. (2016) and (2018) implemented LETKF to the intermediate complexity coupled model, SPEEDY-NEMO, and the operational CFSv2. Their study investigated the characteristics of SCDA and WCDA and found that the SCDA of ocean observations can significantly improve the atmosphere. The CFSv2-LETKF proposed by Sluka (2018) then be implemented to improve the seasonal prediction of Indian summer monsoon rainfall and shows impressive effectiveness in enhancing the S2S forecast skills (Gade et al., 2022).

One of the biggest challenges for ensemble-based CDA is the sampling error due to limited ensemble size. When the ensemble size is insufficient, the spurious coupled cross-error correlations would significantly degrade the performance of CDA, especially for SCDA. Han et al. (2013) used a 5-variable simple coupled model and demonstrated that the improvement from SCDA would be more considerable than WCDA only when the ensemble size was increased to ~ 10^4. Therefore, special care,

such as the correlation cutoff method (Yoshida and Kalnay, 2018), would be needed for SCDA to deal with the rank deficiency of the ensemble.

The capability of a coupled model to simulate the coupled effects also substantially influences the precision of the estimated cross-domain error correlation. If the forecast model does not resolve the coupling dynamic correctly, it would be difficult for the background error covariance to reflect the realistic coupling features. Furthermore, the multi-timescale effect also raises the difficulty of CDA development. For example, it is found that the observations of slow-varying components (e.g., ocean) provide limited increments to fast-varying variables (e.g., atmosphere), while it is easier for the slow-varying states to be improved by observations from fast-varying components (Han et al., 2013; Sluka et al., 2016). Therefore, new strategies for optimizing the use of observation and advanced approaches for CDA to be better adapted to practical, operational Earth system models become important research topics for the development of CDA.

1.1.2 The Local Ensemble Transform Kalman Filter (LETKF)

The LETKF (Hunt et al., 2007) is one of the most popular ensemble-based DA schemes. Its analysis is derived independently at each model grid by combining the local information from the ensemble backgrounds and the observations. At each analysis time, the analysis equations are expressed as:

$$\overline{\boldsymbol{x_a}} = \overline{\boldsymbol{x_b}} + \mathbf{X}_\mathrm{b}\,\widetilde{\mathbf{P}_a}\,(\mathbf{H}\mathbf{X}_\mathrm{b})^\mathrm{T}\mathbf{R}^{-1}[\boldsymbol{y}_o - \mathbf{H}\overline{\boldsymbol{x_b}}]\,, \tag{1.1}$$

$$\boldsymbol{X}_a = \boldsymbol{X}_b\left[(k-1)\,\widetilde{\mathbf{P}_\mathrm{a}}\right]^{\frac{1}{2}}, \tag{1.2}$$

$$\widetilde{\mathbf{P}_{\mathrm{a}}} = [(k-1)\mathbf{I}_{k\times k} + (\mathbf{H}\boldsymbol{X}_b)^T\mathbf{R}^{-1}(\mathbf{H}\boldsymbol{X}_b)]^{-1}, \quad (1.3)$$

where subscript letters *a* and *b* denote the analysis and background, respectively. The $\boldsymbol{X}_{(.)}$ represents the matrix of ensemble perturbations where each column is the vector of the deviations from the mean state $\overline{\boldsymbol{x}_{(.)}}$, namely $\boldsymbol{X}_{(.)} = \{(\boldsymbol{x}^i_{(.)} - \overline{\boldsymbol{x}_{(.)}}) \,|\ldots|\, (\boldsymbol{x}^k_{(.)} - \overline{\boldsymbol{x}_{(.)}})\}$ and $\boldsymbol{x}^i_{(.)}$ is the state vector of the *ith* ensemble with an ensemble size k . **H** is the linear observation operator that converts information from model space to observation space. $\boldsymbol{y}_o$ denotes the local observations, and **R** is the corresponding observation error covariance. $\widetilde{\mathbf{P}_{\mathrm{a}}}$ is the analysis error covariance in a k-dimensional ensemble space spanned by the local ensembles. This attribute avoids the direct calculation of the error covariance in the M-dimensional model space (given that usually M >> k in NWP applications), and thus, the analysis can be obtained very efficiently.

Since the background error covariance $\mathbf{P}_{\mathrm{b}}$ in LETKF is derived in spanned ensemble space, it is impossible to implement the localization function directly on the background error covariance through the Schur product in physical space like Hamill et al. (2001). Instead, Hunt et al. (2007) proposed another brilliant way to implement localization for LETKF by simply multiplying the elements of **R** by an appropriate localization weight range from zero to one. This feature, where the localization function works at the R matrix, is also known as R localization. The characteristics of R localization and its differences from B localization were discussed by Greybush et al. (2011).

1.1.3 Nudging in the coupled model

The nudging scheme (Hoke and Anthes, 1976), also known as Newtonian relaxation, is a simple and widely used DA method for constraining the model state. The idea of nudging is adding small corrections to the state during the model integration, which would gradually push the model state toward the given field (e.g., observations or analysis). For example, the nudging process for the zonal wind velocity forecast equation can be written as

$$\frac{\partial u}{\partial t} = -v \cdot \nabla u + fv - \frac{\partial \phi}{\partial x} + \frac{u_{obs} - u}{\tau}. \quad (1.4)$$

The final term is the correction given by the nudging, where τ is the nudging timescale that determines the rate of state convergence. Traditionally, the timescale τ is chosen based on empirical considerations and may vary with variables.

Nudging can be applied as a WCDA for coupled model initialization. Chen et al. (1995) used the nudging of wind stress observations for the coupled model initialization and showed significant improvements in El Niño forecasting. Keenlyside et al. (2008) implemented nudging to restore the SST field for the coupled model and successfully enhanced the forecast skill on North Atlantic and tropical Pacific oceans. For the operational CFSv2, the sea surface temperature and salinity at the top level of the ocean model are relaxed to the NOAA OISST and climatological data to avoid field drifting (Saha et al., 2010).

1.2 Localization

1.2.1 Why do we need localization for ensemble-based DA?

One of the notable features of ensemble-based DA (e.g., EnKFs) is its flow-dependent background error covariance derived from the background ensembles (e.g., forecasts initialized at the last analysis time), which involves the time-evolving error statistics for the model state. The implied background and observation error covariances determine how much observation information should be used to generate a new analysis. Therefore, the accuracy of the background error covariance estimates is one of the most critical keys toward an optimal analysis for EnKF.

Houtekamer and Mitchell (1998) noticed that the background error covariance estimated by too few ensembles would introduce spurious error correlations in the assimilation. Incorrect error correlations are harmful to the analysis and could lead to a filter divergence. Hamill et al. (2001) then performed conceptual experiments demonstrating how existing noises in the background error covariance influence the EnKF analysis. Their results showed that the relative error, also known as the noise-to-signal ratio, significantly increases when the ensemble size is reduced, and a large relative error would consequently degrade the analysis accuracy. These early studies concluded that a sufficient ensemble size is essential for EnKF to obtain reliable background error estimates and generate accurate analysis.

Having large ensembles is computationally expensive, especially for high-resolution models. Hence, finding a balance between accuracy and computational cost becomes an inevitable challenge for modern EnKF applications. Recent EnKF studies usually limit their ensemble size to about 100 members, and the ensemble size

employed in operational NWPs is even less due to the consideration of computational efficiency (Houtekamer and Zhang, 2016; Kondo and Miyoshi, 2016). Therefore, in order to reduce the sampling errors induced by limited ensembles, covariance localization becomes one of the essential techniques for EnKF applications.

1.2.2 The distance-dependent localization

The most common and operationally used localization strategy is limiting the effects from distant observations (Houtekamer and Mitchell, 1998; Hamill et al., 2001). A straightforward way to implement that is to apply a Schur product, where each element in the ensemble-based error covariance is multiplied by an element from a prescribed correlation function (Houtekamer and Mitchell, 2001). The most widely used prescribed correlation function is the Gaussian-like, distance-dependent function proposed by Gaspari and Cohn 1999 (hereafter GC99). For a B localization (e.g., the Schur product is applied on the background error covariance B), a compact support of the GC99 function can be represented as,

$$\rho_{ij} = exp\left[\frac{-d(i,j)^2}{2L^2}\right], \tag{1.5}$$

where ρ_{ij} is the localization weight, and *d(i,j)* is the distance between the *ith* analysis grid and the *jth* observation. L is the localization length which is usually manually defined. The GC99 function assumes that the observations farther from the analysis grid are less correlated (and even uncorrelated beyond a finite distance); as a result, the impact from distant observations would be suppressed on the analysis during assimilation.

1.2.3 The issues of distance-dependent localization

However, the employment of distance-dependent localization also brings issues and concerns, such as losing distant information and inducing unbalanced analysis (Miyoshi et al., 2014; Lorenc, 2003; Kepert, 2009). Miyoshi et al. (2014) utilized a 10240-member EnKF to investigate the true error correlations of atmospheric variables, and they confirmed that continental-scale, even planetary-scale, error correlations certainly exist in atmospheric variables. That means the use of distance-dependent localization could artificially remove the real long-range signals from the analysis increments. Another follow-up experiment with the 10240-member EnKF demonstrated that removing localization could significantly improve the analysis and its subsequent 7-day forecasts, and the key component for these improvements is the long-range correlations between distant locations (Kondo and Miyoshi, 2016).

The imbalance analysis is another noteworthy issue for localization (Cohn et al., 1998; Lorenc, 2003; Kepert, 2009). An excellent paper from Greybush et al. (2011) summarized the unbalanced problem induced by localization. They argued that the imbalance analysis could happen for either B or R localizations, and the EnKF analysis accuracy could be consequently affected by the manually defined localization length in GC99. The B and R localizations indicate whether the localization function is applied on the background error covariance B or the observation error covariance R. They also found that the B localization has a longer optimal localization length with respect to the analysis accuracy. In contrast, the R

localization is more balanced than the B localization underlying the same localization length, and the balance of the analysis is enhanced when the localization length increases. A similar conclusion is mentioned in Lorenc (2003) that the unbalance induced by localization would relax with longer localization length and significantly minimized when the length is larger than 3000 km.

1.2.4 Approaches of non-distance-dependent localizations

In addition to defining the localization by geographical distance, the empirical localization method (Anderson, 2007; Anderson and Lei, 2013; hereafter AL13) derives a static and flow-dependent localization from posterior ensembles. The core concept of AL13 is to find a localization weight that performs minimum analysis error, where a cost function is solved iteratively with subset ensembles and observations under a series of OSSEs. The AL13 method shows comparable analysis accuracy to the optimally tunned traditional localization (GC99) on the 40-variable Lorenz model.

Bishop and Hodyss (2009) proposed an adaptive localization strategy using ensemble correlations raised to a power (ECO-RAP). This method calculates the ensemble error correlation within each DA window and raises the collected correlation matrix to qth power (q=6 in their experiment) to eliminate small correlations and emphasize large ones. Then, a smooth operator is applied to the correlation matrix to ensure the consistency of the correlation length scale. In contrast to the correlation cutoff method (YK18, Yoshida and Kalnay, 2018) we used in this study, ECO-RAP estimates the localization function adaptively with simultaneous

ensemble correlations, while YK18 tends to estimate a non-adaptive localization function with climatological correlation. Both the methods need an additional function to filter out small correlations and emphasize larger ones, while ECO-RAP would require a more robust filter to identify and deal with small correlations (like noises) and sampling errors in the simultaneous background ensembles, which is the most critical part for the feasibility of ECO-RAP.

1.3 Estimating the observation impacts in DA

As more observations become available for DA in the past decades, it greatly benefits the quality of analyses and forecasts with advanced model state estimations. However, the surge in observation amounts from various platforms also heightens the need for techniques capable of monitoring the impact of individual observation in DA.

The Forecast Sensitivity to Observation (FSO) proposed by Langland and Baker (2004, hereafter LB04) is one of the most advanced methods of estimating the observation impact without data-denial (with and without observations) experiments. The method attributes the differences between the forecasts initialized with and without DA to individual observations using the adjoint formulation and has been applied to variational DA systems in operational and research centers (Zhu & Gelaro, 2008; Cardinali, 2009; Lorenc & Marriott, 2014; Zhang et al., 2015).

For the ensemble-based DA, Liu and Kalnay (2008) and Li et al. (2010) first introduced the Ensemble Forecast Sensitivity to Observation (EFSO) for the LETKF. EFSO tends to utilize ensemble information from the DA rather than the

adjoint/tangent linear models, allowing it to apply to models without a tangent-linear model. Kalnay et al. (2012, hereafter K12) further refined the EFSO formulation, making it simpler and applicable to all ensemble Kalman filters (EnKFs). A brief introduction to the EFSO proposed by K12 can be found in section 1.3.1.

1.3.1 The Ensemble Forecast Sensitivity to observation (EFSO)

This section briefly introduces the EFSO formulation following K12. Here, $\overline{\mathbf{x}_t^g}$ and $\overline{\mathbf{x}_t^f}$ are the forecasts at time *t* initialized from the background and the analysis at *t* = 0, respectively. Their forecast errors verified against the truth $\mathbf{x}_t^v$ at time *t* can be represented as:

$$\boldsymbol{e}_t^f = \overline{\mathbf{x}_t^f} - \mathbf{x}_t^v\,,\ \boldsymbol{e}_t^g = \overline{\mathbf{x}_t^g} - \mathbf{x}_t^v. \tag{1.6}$$

The verifying truth $\mathbf{x}_t^v$ can be its own analyses or independent data. Following LB04, the forecast error reduction (or increase) Δe^2 is defined as:

$$\Delta e^2 = \boldsymbol{e}_t^{f\mathrm{T}} \boldsymbol{C}\, \boldsymbol{e}_t^f - \boldsymbol{e}_t^{g\mathrm{T}} \boldsymbol{C}\, \boldsymbol{e}_t^g = (\boldsymbol{e}_t^f - \boldsymbol{e}_t^g)^{\mathrm{T}} \boldsymbol{C}\,(\boldsymbol{e}_t^f + \boldsymbol{e}_t^g)$$

$$\approx \left[\mathbf{M}\left(\overline{\mathbf{x}_0^a} - \overline{\mathbf{x}_0^b}\right)\right]^{\mathrm{T}} \boldsymbol{C}\left(\boldsymbol{e}_t^f + \boldsymbol{e}_t^g\right) = [\mathbf{M}\mathbf{K}\delta\boldsymbol{y}_o]^{\mathrm{T}} \boldsymbol{C}\left(\boldsymbol{e}_t^f + \boldsymbol{e}_t^g\right). \tag{1.7}$$

The upper letters *a* and *b* represent the analysis and background, respectively. $\overline{\mathbf{x}^{(\cdot)}}$ is the mean state. $\delta\boldsymbol{y}_o = \boldsymbol{y}_0 - \mathbf{H}(\overline{\mathbf{x}_0^b})$, where $\boldsymbol{y}_o$ denotes observations, and **H** is the observation operator. **M** represents the tangent linear forecast model, and **K** is the Kalman gain. ***C*** is the forecast error norm. A schematic figure for the idea of EFSO is shown in Figure 1.2.

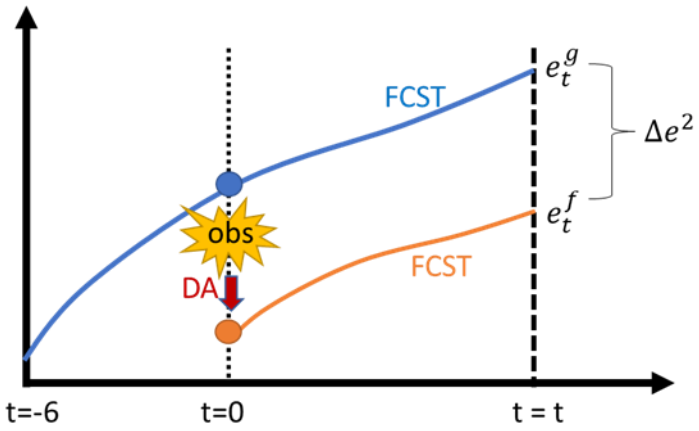

Figure 1.2 Schematic of the EFSO (Δe^2). The terms e_t^g and e_t^f represent the forecast errors (t = t) of the forecasts initialized from the background (t = 0) and analysis (t = 0), respectively.

K12 extended the adjoint-based cost function (Eq (1.7)) to a simpler, tangent-linear-model-free form for ensemble-based DA by generally assuming $\mathbf{K} = \left(\frac{1}{k-1}\right)\mathbf{X}_0^a\mathbf{X}_0^{a\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$ and $\mathbf{X}_t^b \approx \mathbf{M}\mathbf{X}_0^a$, where $\mathbf{X}^{(\cdot)}$ represents the perturbation departure from the mean state, and k is the ensemble size. Thus, Eq (1.7) becomes (Eq (6) in K12):

$$(\Delta e^2)^{EFSO} = \left(\frac{1}{k-1}\right)\delta \boldsymbol{y}_o^T\,[\rho \circ \mathbf{R}^{-1}(\mathbf{H}\mathbf{X}_0^a)\mathbf{X}_t^{b\,T}\boldsymbol{C}\left(\boldsymbol{e}_t^f + \boldsymbol{e}_t^g\right)], \tag{1.8}$$

where ρ is the localization matrix whose row is the weighting for each observation, and the symbol $\circ$ denotes the element-wise multiplication (Schur product). **R** is the observation error covariance. $(\Delta e^2)^{EFSO}$ is the EFSO estimation, where a negative (positive) sign represents the beneficial (detrimental) observation for the forecast verified at time *t*.

1.3.2 Applications of EFSO

Ota et al. (2013) successfully implemented EFSO on the NCEP GFS-EnKF system and demonstrated the effectiveness of EFSO in quantifying the impacts of observations from different platforms. Lien et al. (2018) designed a novel data selection strategy based on EFSO for assimilating new measurements. Chawang and Kutty (2022) applied EFSO on the WRF model to monitor the impact of satellite atmospheric motion vectors (AMV) during the intensification period of tropical cyclons Hudhud (2014) and Phailin (2013) formed over the Bay of Bengal. Yamazaki et al. (2021) implemented EFSO in an atmospheric general circulation model and examined the feasibility and spatial characteristics of EFSO with different ranges of forecasts. They found that EFSO can be useful in estimating the observation impacts in short-range forecasts at all latitudes and in medium-range forecasts for the Arctic. A proactive quality control (PQC) system that removes instantaneous detrimental observations based on EFSO within a single DA cycle was shown to be effective in enhancing the analyses and forecast quality (Hotta et al., 2017; Chen et al., 2020). All the above studies have shown how powerful and effective EFSO is in investigating the observation impacts and guiding the observation use for DAs.

1.4 Outline of this research

Pioneering efforts on CDA have been increasingly proposed in recent years, while challenging issues related to optimizing the ensemble-based CDA still require

deeper investigations. For example, what are the observation impacts in the coupled system? How to optimize the use of observations for CDA and improve the coupled forecasting? This dissertation aims to propose three advanced approaches from different aspects using LETKF. The ultimate goal of this dissertation research is to investigate the observation impacts in a coupled system, explore potential approaches for improving CDA, and gain more insights into enhancing the predictability of Earth system models.

1. **Applying prior correlations for ensemble-based spatial localization (Chap 2)**

This study introduces a non-adaptive, correlation-dependent localization scheme evolved from the correlation cutoff method (Yoshida and Kalnay, 2018; hereafter, YK18). The key idea is to "localize" the information from observation to analysis according to their square background error correlations estimated from a preceding offline run. Although YK18 was initially proposed as a variable localization strategy for coupled systems, it can be further utilized as a spatial localization by appropriately employing the cutoff function (Yoshida 2019). Similar to AL13, YK18 provides a static and flow-dependent localization function from posterior ensembles. However, YK18 does not need a truth value for OSSEs nor run iteratively like AL13, while an additional cutoff function described in Section 2.3 is required to filter out small perturbations in the error correlations. In this study, we applied the YK18 as spatial localization and investigated its feasibility using the Lorenz 96 model and LETKF. We also extensively explored the potential of integrative localization strategies.

2. Correcting the existing SST biases with SST WCDA (Chap 3)

The presence of significant SST biases in the CFS and their potential influences on degrading the accuracy of seasonal predictions have been much discussed (Kumar et al., 2012; Jin and Kinter, 2009; Wang et al., 2010; Chaudhari et al., 2013; KBRR et al., 2020). Traditionally, the SST and surface sea salinity (SSS) in the CFSv2 are mainly constrained by relaxing (nudging) to the "Reynolds SST" (Reynolds et al., 2002, 2007) and to the annual salinity climatology (Conkright et al., 1999), respectively. However, we noticed that SST nudging is not very effective in reducing existing SST biases (Sluka, personal communication, (2019)). To address this issue, we propose to replace nudging with the WCDA of level 2 (L2) and level 4 (L4) SST products to reduce the existing SST biases. Several conceptual experiments were carried out on the operational-like CFSv2-LETKF system (Sluka et al., 2018) with real observations. Given the more effective reduction in the SST biases with WCDA, we expect the CFSv2 analysis and forecast to be beneficially improved.

3. Estimating the oceanic observation impacts on CDA using EFSO (Chap 4)

While the EFSO has become increasingly popular for atmospheric DAs, it has not yet been applied to any ocean or coupled systems to the best of our knowledge. The primary challenge for the oceanic EFSO development is the lack of a proper error norm to simultaneously include the forecast errors in salinity and sea temperature, the two most important variables that govern the dynamical characteristics of the ocean, such as heat transport and circulations. The error norm used in the EFSO formulation tends to incorporate comprehensive impacts into a single scalar, allowing the forecast

errors of different variables to be quantified under the same unit. For atmospheric EFSO, the dry and moist energy norms (Ehrendorfer et al., 1999) are commonly used to measure the forecast errors of essential meteorological variables, such as wind, temperature, and surface pressure. However, salinity is not directly attached to a known energy term as atmospheric variables, which leads to a significant obstacle in the oceanic EFSO development.

This study first proposes the oceanic EFSO by introducing a novel ocean density-based error norm that simultaneously includes sea temperature and salinity. The oceanic EFSO was implemented on the atmosphere-ocean coupled system, the CFSv2-LETKF, and used to investigate the impacts of ocean observations under the WCDA framework.

Finally, the results of this dissertation research are briefly summarized in Chapter 5. Chapter 5 also discusses the ongoing works, potential applications, and future studies inspired by current research.

Chapter 2: Applying prior correlations for ensemble-based spatial localization

2.1 Introduction

This chapter investigates the feasibility of the correlation cutoff method (Yoshida and Kalnay, 2018; hereafter, YK18) as a spatial localization using the LETKF and Lorenz 96 model. The key idea of YK18 is to "localize" the information from observation to analysis according to their square background error correlations estimated from a preceding offline run. Although YK18 was initially proposed as a variable localization strategy for coupled systems, it could be further utilized as a spatial localization by appropriately employing the cutoff function (Yoshida 2019). We compared the YK18 with the well-explored traditional distance-dependent localization GC99 and further explored the potential of the hybrid use of GC99 and YK18 under different configurations, aiming to gain more insights into integrative localization applications.

2.2 Methodology and Model

2.2.1 Model

The classic Lorenz model (hereafter L96 model; Lorenz and Emanuel, 1998) is a one-dimensional, univariate simplified atmospheric model that consists of a nonlinear term (e.g., representing advection), a linear term (e.g., representing mechanical or thermal dissipation), and an external forcing. The governing equations are:

$$\frac{dX_i}{dt} = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F\ (+f_i), \tag{2.1}$$

where the model variable is denoted by X_i , $i = 1, \dots, M$, and $M = 40$. F is the constant external forcing and is set to be 8 here. The variables form a cyclic chain, where $X_{-1} = X_{M-1}$ and $X_0 = X_M$. The varying forcing term f_i is neglected for the classic L96 model. The model is integrated with the fourth-order Runge-Kutta scheme with a time step of 0.0125 units (four steps correspond to 6 hours) and was initialized by adding a single random perturbation onto the rest state and integrating for 90 days to remove the model spin-up.

A variant L96 model with a spatially varying forcing f_i appending to the L96 model is proposed here to mimic a more sophisticated model dynamic. We constrained the total external forcing ($F + f_i$) with a value range of 6 to 10, ensuring the model dynamic remains chaotic and has a wavenumber of 8. This additional forcing characterizes a land-ocean pattern (Figure 2.1 (a)), where the land region has an irregular and more substantial forcing (e.g., source), and the ocean region has a uniform and smaller forcing (e.g., sink). The primary influence of the changes in F is on its error growth rate. That said, increasing F has only a little effect on the qualitative appearance of the wave curves, while the error doubling time has an observable decrease (Lorenz and Emanuel, 1998).

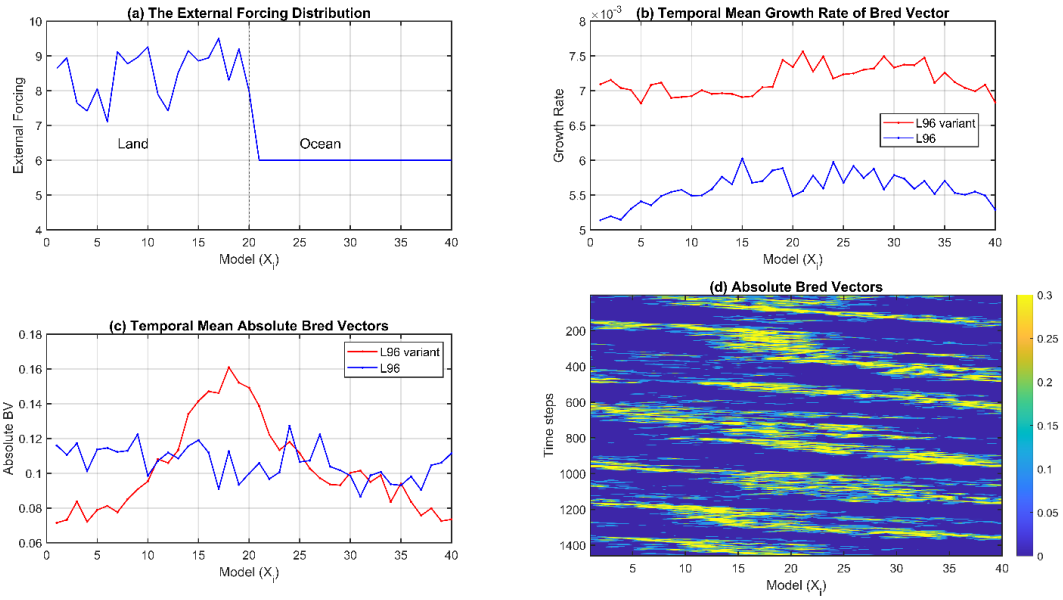

Figure 2.1 (a) The external forcing ($F + f_i$) used in the variant L96 model. The temporal mean of (b) the growth rate and (c) absolute bred vectors. (d) The time evolution of the absolute bred vectors for the variant L96 models. The breeding rescale cycle is 4 steps ($n = 4, \Delta t = 0.0125$), which equals our DA window length. The breeding rescale amplitude is 1.0.

To understand the fundamental properties of the variant L96 model, we examined the bred vectors (BV, Toth and Kalnay, 1993, 1997; Kalnay et al., 2002) of the two models. BV is a nonlinear generalization of the leading Lyapunov vectors (see Toth and Kalnay, 1993 and 1997, for a more detailed exposition). Their growth rate is calculated as $\frac{1}{n\Delta t} \ln \left(\left\| \delta x^f \right\| / \left\| \delta x^0 \right\| \right)$, where δx^f and δx^0 are the final and initial perturbations within the breeding window, respectively. n is the window size and Δt is the integration step. The growth rate can be seen as a measure of the local instability of the flow. Figure 2.1 (b) shows the temporal mean BVs growth rate for the classic and variant L96 models. The variant L96 model has an overall higher

growth rate than the classic L96 model (Figure 2.1 (b)), which agrees with the statement of Lorenz and Emanuel (1998). Moreover, the perturbations tend to grow on the land-sea interface (Figure 2.1 (c)) and propagate eastward with the group velocity (Figure 2.1 (d)). In summary, we expect the variant L96 model to offer more complicated dynamics than the L96 model, and its more rapid error growth would let the corrections from DA be more quickly lost.

2.2.2 The distance-dependent localization

Following Hunt et al. (2007), we use the positive exponential function as the localization function:

$$\rho_{i_j} = exp\left[\frac{d(i,j)^2}{2L^2}\right], \tag{2.2}$$

where ρ_{i_j} is the localization weight and *d(i,j)* is the distance between the *ith* analysis grid and the *jth* observation. L is the localization length which is usually manually defined. Eq. (2.2) is a smooth and static Gaussian-like function that offers the same localization effect as the GC99 when applied to LETKF. Since the observation errors are assumed to be uncorrelated in our experiments (R is diagonal), the localization weight would be independently assigned for the assimilated observation j and analysis grid i. So, when the distance ($d(i,j)$ in Eq (2.2)) increases, a larger value of ρ_{i_j} would be multiplied to R, inflating observation error for the jth observation. That would lead to a smaller value in the corresponding rows of the Kalman gain ($\boldsymbol{X}_b[(k-1)\mathbf{I} + (\mathbf{H}\boldsymbol{X}_b)^T\mathbf{R}^{-1}(\mathbf{H}\boldsymbol{X}_b)]^{-1}(\mathbf{H}\boldsymbol{X}_b)^T\mathbf{R}^{-1}$) of LETKF, down-weighting the observation on updating the background; thus, the impact of distant observations

would be suppressed on the analysis. When the compact support is presented with the localization function, the observations located beyond a certain distance (this study uses 3.65 times L) from the analysis grid would be discarded by assuming $\rho_{i\,j} = 0$.

2.2.3 The correlation cutoff method

The correlation cutoff method (Yoshida and Kalnay, 2018; Yoshida, 2019), a pioneering localization approach for coupled systems, localizes the information from observation to analysis according to their ***square background error correlations***. This method is carried out with two steps:

Step 1. Obtaining the square error correlation from an independent run

The prior square error correlations are collected from a preceding independent DA cycling run. At each analysis time t, an instantaneous background ensemble correlation between the *ith* analysis grid and the *jth* observation is computed as:

$$corr_{i\,j}(t) = \frac{\sum_{k=1}^{K}[x_{k\,i}\,(t) - \overline{x_i(t)}][h_j(x_k(t)) - \overline{h_j(x_k(t))}]}{\sqrt{\sum_{k=1}^{K}\left[x_{k\,i}(t) - \overline{x_i(t)}\right]^2}\sqrt{\sum_{k=1}^{K}\left[h_j(x_k(t)) - \overline{h_j(x_k(t))}\right]^2}} \tag{2.3}$$

where $x_{k\,i}(t)$ is the state vector of the *kth* ensemble at the ith analysis grid at time *t*. $h_j(x_k(t))$ is the linear interpolation to the background state $x_k(t)$ from the analysis grid to the *jth* observation location. The symbol $\overline{(\,)}$ denotes the ensemble mean of a given vector. K is the total ensemble size.

Then, the temporal mean of the squared correlation is computed by:

$$< corr_{i\,j}^2 >= \frac{1}{\mathrm{T}} \sum_{t=1}^{T} corr_{i\,j}^2(t) \tag{2.4}$$

T is the total analysis cycles in the independent run. In the original YK18, this prior error correlation is used as a criterion for variable localization in the coupled DA, by which only those highly correlated observations would be assimilated. For the spatial localization approach, the value $< corr_{i_j}^2 >$ will serve as the "prior error correlation" to estimate the localization function as described in Step 2.

Step 2. Converting the prior error correlation into the localization weighting

The localization function is derived by substituting the prior error correlation obtained in Step 1 with a chosen cutoff function. The purpose of using the cutoff function is to generally smooth out small perturbations and ensure the weight range is between 0 and 1. Here, we followed Yoshida (2019) using the quadratic function (Eq (2.5)) as our choice of the cutoff function. The localization weight ρ_{i_j} assigned for the *j*th observation at the *i*th analysis grid can be written as:

$$\rho_{i_j} = \begin{cases} 0 & (x \le c), \\ 1 - \left(\frac{1-x}{1-c}\right)^2 & (c < x \le 1), \\ 1 & (x > 1) \end{cases} \tag{2.5}$$

where $x =< corr_{i_j}^2 >$ and c is a tunable parameter that defines the slope for the function. We set c equal to 0.05 and 0.01 for the classic and variant L96 experiments, respectively.

An additional threshold was applied to exclude observations with a square error correlation smaller than 1/(K-1). This threshold is chosen because the squared sample

correlation estimated by K random samples extracted from an uncorrelated distribution would converge to 1/(K-1) (Pitman, 1937). So, any value not much larger than 1/(K-1) is assumed to be unreliable (Yoshida, 2019).

2.3 Experimental settings

2.3.1 Localization methods

In this study, we investigated four types of localization strategies:

- **GDL**: Distance-dependent localization introduced in Section 2.2.2. The localization length used for each experiment is experimentally tuned for a minimum temporal mean analysis RMSE. The cutoff radius is set to be 3.65 times the localization length.
- **YK18**: Correlation-dependent localization, in which the weighting function is derived from the correlation cutoff method introduced in Section 2.2.3.
- **Hybrid**: a hybrid application of GDL and YK18, in which the localization weighting is equal to $\alpha\ GDL + (1-\alpha)YK18$. The combination ratio α is 0.5 for our experiment. This method was tested for the classic L96 model experiment.
- **Hybrid II:** Combination use of GDL and YK18, in which YK18 is employed for the first 80 DA cycles for shortening the DA spin-up, and GDL is subsequently applied for the rest of the DA cycles. This method is used only for the variant L96 model experiment.

For YK18, an independent offline run with sequential DA cycling is conducted to acquire the prior error correlation before running the DA experiments. The running period for the offline run is three years with a 6-hr analysis window. The first four months are assumed to be the DA spin-up period and were removed. We used 10 ensembles for the offline run with configurations the same as the GDL experiments in Sections 2.4.2 and 2.4.3. Offline runs were performed respectively for the classic and variant L96 models.

Theoretically, the best localization length for GDL is directly proportional to the ensemble size, and there must exist an optimal combination of the localization length and the inflation factor (Hamill et al., 2001). This study applied the multiplicative covariance inflation (Anderson, 2001), and its best combination with localizations is experimentally defined based on the minimum averaged analysis error for each experiment. The parameters used in the experiments for the L96 and L96 variant models are listed in Table 2.1 and Table 2.3.

2.3.2 Truth and Observations

The truth was obtained from the model free-run, and the observations were generated by adding random Gaussian errors with a variance of 1.0 onto the truth state every 6 hours. The initial ensembles are obtained from the perturbed model states and integrated for 75 days until the ensemble trajectories converge to the model attractor. The total experiment period is one year.

The analysis result is evaluated by the root-mean-square error (RMSE) with the true state. For each variable, the RMSE can be represented as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(\overline{x_i^a} - x_i^e)^2}\,, \tag{2.6}$$

where M is the number of model grids, which equals 40 for the L96 model. The $\overline{x_i^a}$ and x_i^e are the analysis ensemble mean and the verified state, respectively.

2.4 Results

2.4.1 The characteristics of the YK18 function

The squared error correlation estimated from the independent background ensembles is the core of the YK18 localization function. Before moving to DA experiments, we first discussed:

(1) How do different factors (ensemble and observation) in the offline run impact the corresponding error correlation estimation (Eq (2.4))?

(2) What are the main differences in the localization functions (e.g., GDL and YK18)?

First, we examined the temporal mean squared correlation (Eq (2.4)) estimated by different observations and ensemble sizes of the offline runs. Trials with observation sizes of 40, 20, and 13 (representing uniform coverages of 100%, 50%, and 30%, respectively) are carried out on the classic L96 model with 40 ensembles. We found that the squared correlation estimation (Eq (2.4)) is not very sensitive to the observation size changes (Figure 2.2 (a)) as long as the analysis of the offline run is well-constrained. Moreover, the minor differences in the estimated squared error

correlation (Figure 2.2 (a)) would be ultimately smoothed out by the cutoff function (Eq (2.5)) in practice. Therefore, the final localization weights derived from the offline runs with different observation sizes will be almost identical. This characteristic, in other words, provides clear evidence to use past data to estimate the error correlations for newly added observations, which is a significant advantage for the applicability of YK18 in modern DA.

Figure 2.2 (b) shows how the offline run period (e.g., the number of samples) would affect the prior error correlation estimation (Eq (2.4)). The mean-square error (MSE) was verified with the result estimated by large ensembles (ens =100) and a long period (3 years). It is noticeable that the required number of samples (e.g., length of offline run) for the estimated error correlation to converge to climatology is associated with the ensemble size and model complexity. When the model is more complicated (blue and red lines), or the ensemble size is smaller, the required cycles for the convergence of estimated error correlations would accordingly increase (Figure 2.2 (b)). Therefore, a longer offline run or past data may be required when using fewer ensembles or a more complicated model.

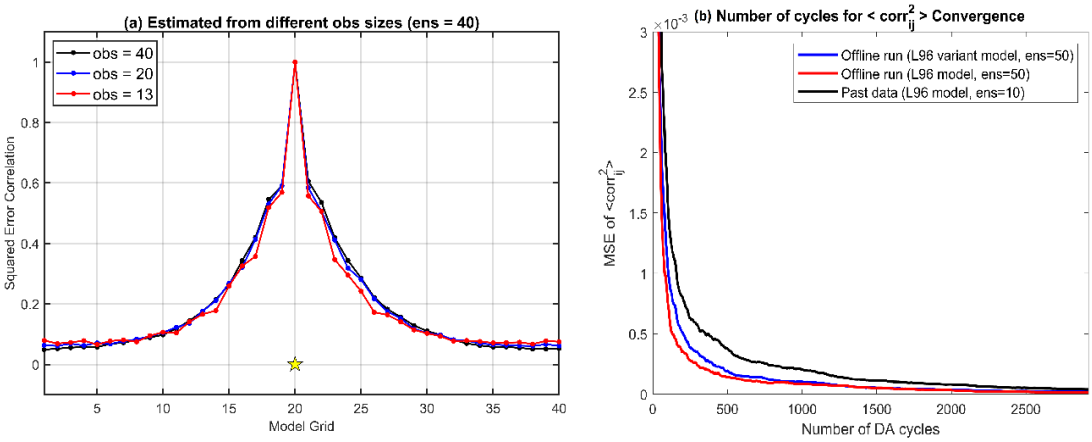

Figure 2.2 (a) The temporal-mean squared error correlations estimated from different observation amounts. The yellow star represents the correlated observation location. (b) The MSE of Eq (2.4) estimated by the past data with 10 ensembles (black) and by the ideal offline runs with the L96 model (red) and the L96 variant model (blue).

The localization functions of GDL and YK18 applied for our DA experiments are shown in Figure 2.3. The optimal localization length for GDL is associated with multiple factors like ensemble size, observation distributions, and model dynamics. For example, when the ensemble size shrinks, the optimal localization length would correspondingly decrease so that a more substantial suppression effect can be performed on those spurious correlations in distant regions (Ying et al., 2018). In contrast, the YK18 localization function, once it is defined, is independent of the ensemble size changes. Unlike GDL, which provides a fixed function for every observation, YK18 offers customized localization functions for each observation based on their prior error correlations; for example, the different asymmetric features of the YK18 function (red line) in Figure 2.3 (b) and (c).

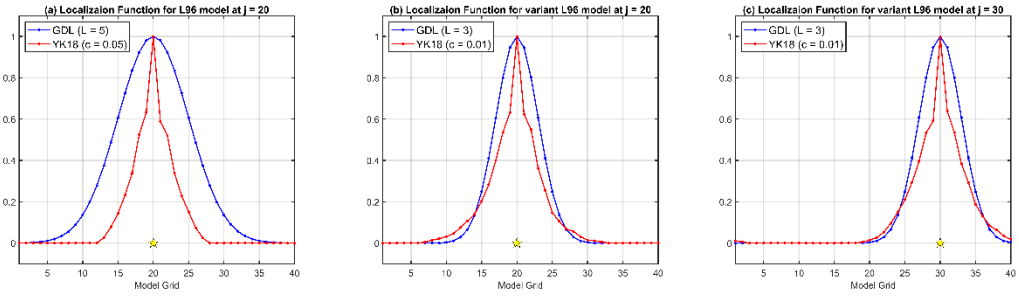

Figure 2.3 The localization functions of GDL (blue) and YK18 (red) for the (a) classic L96 model and (b)(c) the variant L96 model but for the different observation sites. The yellow stars represent the corresponding observation sites. The results presented here are for the case of 10 ensembles and 40 observations.

2.4.2 DA Experiment Scenario I: classic L96 model

This section utilized the classic L96 model to investigate the impacts of GDL, YK18, and Hybrid. The total experiment period is one year (after the first 60 cycles of spin-up) with a DA window of 6 hours. The tested ensemble sizes are 8 and 10. Observations are uniformly distributed with a total number of 20 and 40. The parameters for the localization and inflation for the experiments are shown in Table **2.1**.

Table 2.1 The parameters used in the class L96 model experiments given in Eq (2.2) and Eq (2.5). The symbol α represents the multiplicative inflation parameter.

		GDL	YK18	Hybrid
ens =10	obs = 40	L = 5, α = 1.04	c = 0.05, α = 1.03	L = 7, α = 1.03
	obs = 20	L = 4, α = 1.03	c = 0.05, α = 1.03	L = 6, α = 1.03
ens = 8	obs = 40	L = 3, α = 1.04	c = 0.05, α = 1.04	L = 7, α = 1.06
	obs = 20	L = 3, α = 1.07	c = 0.05, α = 1.04	L = 7, α = 1.06

Figure 2.4 shows the analysis RMSE of GDL, YK18, and Hybrid. The YK18 presented the lowest RMSEs among all the methods during the DA spin-up period (Figure 2.4), particularly when the ensemble size and observations were reduced (Figure 2.4 (d)). This result shows that YK18 can shorten the DA spin-up and perform a comparable analysis as GDL. The DA spin-up means the required period for the ensemble-based DA system to build a reliable background error covariance, and the analysis error reaches convergence. The phrase “spin-up” used in the following sections refers to the DA spin-up.

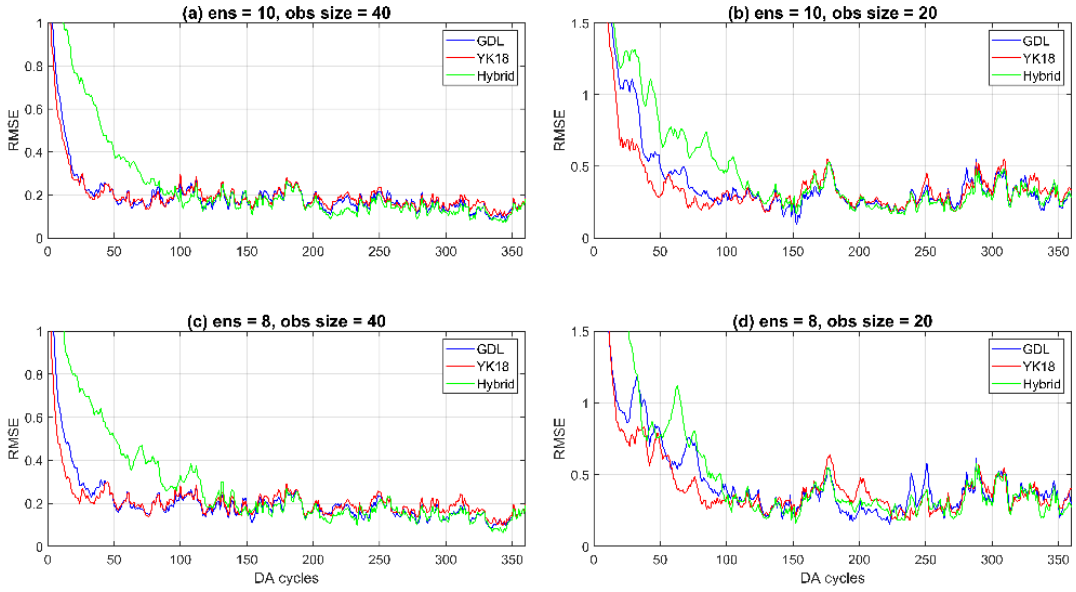

Figure 2.4 The time series of the analysis RMSE for GDL (blue line), YK18 (red line), and Hybrid (green line) for the cases of 10 ensembles with (a) 40 and (b) 20 observations; and cases of 8 ensembles with (c) 40 and (d) 20 observations.

The capability of YK18 in accelerating the spin-up mainly comes from its more precise interpretation of the error correlations derived from the independent (or past) ensembles. Figure 2.5 shows the localized background error covariance ($\rho \mathbf{X}_b \mathbf{X}_b^T$) of GDL (blue line) and YK18 (red line) at the first (Figure 2.5 (a)) and the second (Figure 2.5 (b)) DA cycles. The true covariance (black line) was obtained by perturbing the truth state and evolving through the corresponding DA window (6 hours) with a large ensemble size of 5000, which can be seen as an optimal estimation without sampling errors. At the first DA cycle, where GDL and YK18 were initialized with the same ensembles, it is apparent that the localized error covariance

of YK18 is significantly closer to the true covariance, showing less spurious than GDL, especially for distant covariances. (Figure 2.5 (a)). With a better estimate of the background error covariance, YK18 performed a superior analysis at the initial cycle and subsequently improved the background error estimation for the next cycle (Figure 2.5 (b)). Thus, with prior knowledge of the error correlations, YK18 can optimize the use of observations, inducing more "on-point" corrections for the analysis and reducing the required number of cycles for the DA system's spin-up. This advantage of YK18 is also present in the variant L96 model (Figure 2.5 (c)(d)).

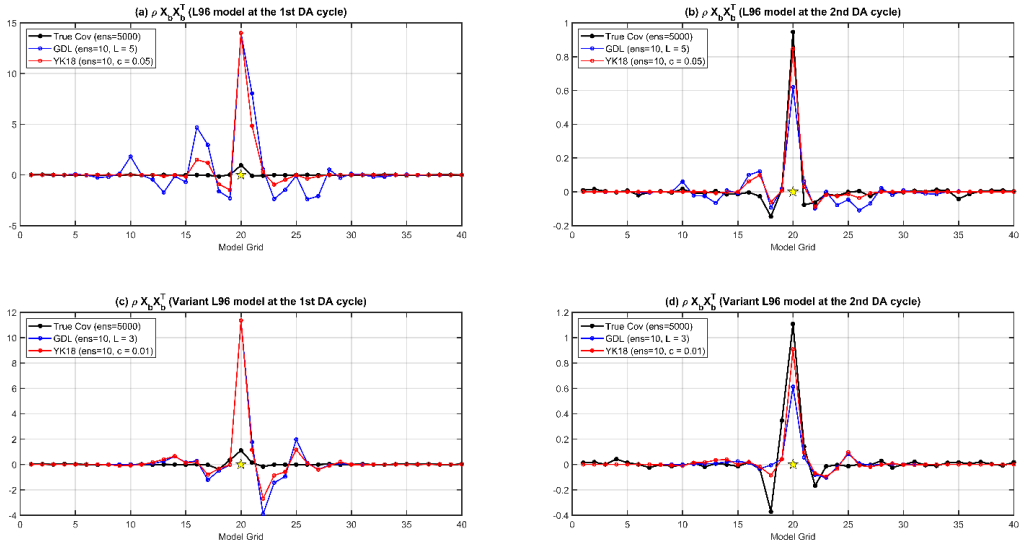

Figure 2.5 The true (black) and localized background error covariances ($\rho \boldsymbol{X}_b \boldsymbol{X}_b^T$) of GDL (blue) and YK18 (red) for the L96 model at (a) the first and (b) the second DA cycles, and for the variant L96 model at (c) the first and (d) the second DA cycles. The localization functions and configurations are the same as in Figure 2.3.

Table 2.2 shows the 1-yr mean analysis RMSE without the spin-up period (first 100 cycles). Generally, the long-term averaged performance of the three localizations is very similar, while Hybrid is slightly better than the other two (Table 2.2). The best localization length for Hybrid is longer than pure GDL, which allows it to gain more observation information after the DA convergence. Note that it is unlikely for GDL to apply such long localization length at the beginning because it needs a relatively shorter localization length to constrain the spurious error covariances during the spin-up. In our experiments, the GDL went through filter divergence at the early stage when using localization lengths larger than 7. In contrast, by averaging with the tighter function from YK18, the Hybrid was able to get through the spin-up with a longer localization length. However, on the other hand, it requires a significantly longer spin-up period than the other two methods due to its weaker constraint in the early stage.

Table 2.2 The long-term mean analysis RMSE for the classic L96 model.

	Observation = 40		**Observation = 20**	
	Ens = 8	Ens = 10	Ens = 8	Ens = 10
GDL	0.178	0.175	0.292	0.245
YK18	0.192	0.185	0.302	0.280
Hybrid	0.176	0.163	0.271	0.253

2.4.3 DA Experiment Scenario II: the variant L96 model

Considering the classic L96 model is favorable for GDL due to its simple model dynamics, the variant L96 model that offers a more complicated model dynamic was employed here. We used 10 ensembles with different observation sizes of 40, 30, and 20. The 20 and 40 observations are uniformly distributed. The 30 observations are distributed densely on the land (20 observations) and coarsely in the ocean area (10 observations). Three localization methods were tested here: GDL, YK18, and Hybrid II. Hybrid II uses YK18 for the first 80 DA cycles to accelerate the spin-up, then GDL for the rest of the cycles. Since the parameters are respectively tuned for each method, the localization length used in GDL and Hybrid II may differ. The parameters used for this section are listed in Table 2.3.

Table 2.3 The parameters used in the variant L96 model experiments given in Eq (2.2) and Eq (2.5). The symbol α represents the multiplicative inflation parameter.

		GDL	**YK18**	**Hybrid II**
ens =10	obs = 40	L = 5, α = 1.06	c = 0.005, α = 1.03	L = 5 , α = 1.03
	obs = 30	L = 4, α = 1.06	c = 0.005, α = 1.04	L = 5, α = 1.04
	obs = 20	L =3, α = 1.03	c = 0.005, α = 1.05	L = 5, α = 1.06

Figure 2.6 shows the analysis RMSE of the three methods on the variant L96 model. Since Hybrid II is identical to YK18 for the initial 100 DA cycles, the green overlaps with the red line in Figure 2.6. As expected, GDL requires a significantly longer spin-up for the more complicated model, especially when fewer observations

are available (Figure 2.6 (b) and (c)). The YK18, again, showed impressive efficiency in accelerating the spin-up, particularly with fewer observations, and generated a better analysis than GDL at the early stage. Nevertheless, this advantage of YK18 became more pronounced with a more complicated model and fewer observations.

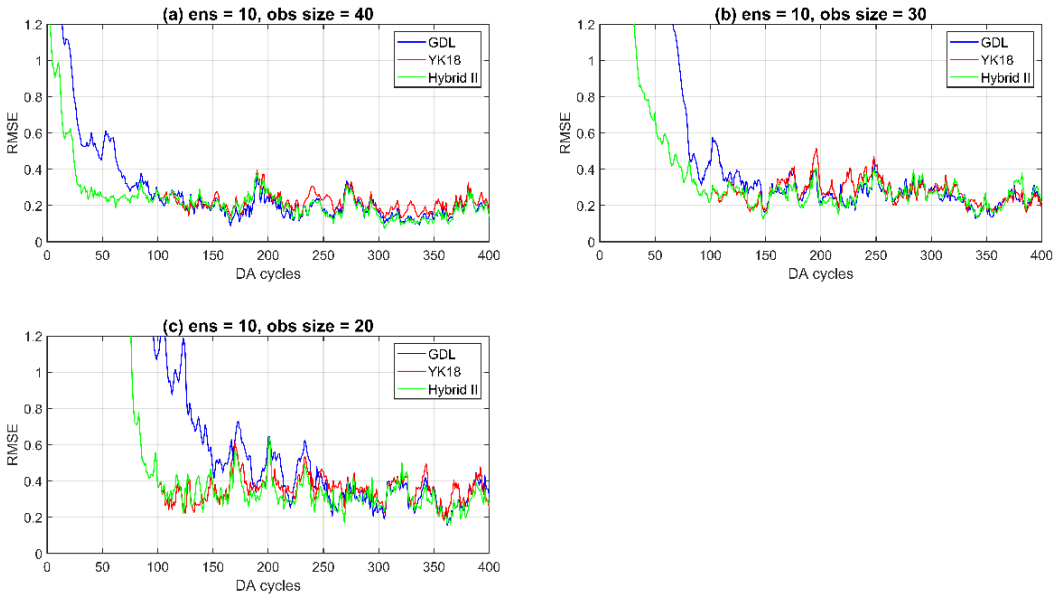

Figure 2.6 The analysis RMSE of the GDL (blue), YK18(red), and Hybrid II (green) with observations of (a) 40, (b) 30, and (c) 20 for the variant 96 model experiment.

Table 2.4 The long-term mean analysis RMSE for the variant L96 model (10 ensembles).

	obs = 40	obs = 30	obs = 20
GDL	0.185	0.254	0.317
YK18	0.210	0.255	0.319
Hybrid II	0.178	0.234	0.312

Table 2.4 is the 1-yr average of the analysis RMSE after the first 100 spin-up cycles. After the system's spun-up, the averaged analysis RMSEs of all methods are similar, while Hybrid II is slightly better than the other two methods (Table 2.4). We found that the mixed use of YK18 and GDL Hybrid II is superior to solely using YK18 or GDL. Hybrid II inherits the benefit of YK18 of accelerating spin-up and outperforms GDL after the system convergence, presenting the best performance among all methods. That is, possibly, because Hybrid II has a longer optimal localization length than GDL, allowing it to acquire more observation information during the assimilation and provide a more accurate analysis. Moreover, Hybrid II has a significantly shorter spin-up than Hybrid I, making it a better hybrid strategy for cases requiring DA spin-ups.

Finally, it is important to highlight that YK18 is an exceptionally efficient localization method. In practice, using GDL requires multiple preceding trials to find an optimal length for the experiments of interest, which may consume considerable computational resources and time. Moreover, when the ensemble size or observation amount changes, the optimal localization length may vary accordingly, so additional tuning for the localization length might be needed for GDL. In contrast, YK18 only needs one offline run to determine the error correlations, whereas it performs a comparable analysis as GDL, even with a faster spin-up. Although an initial tunning for the parameter c in Eq (2.5) is necessary at the beginning, once it is tuned, it can adapt to future ensemble or observation size changes since it is not sensitive to the variation of those factors. This feature, on the other hand, allows YK18 to avoid further trial-and-error tunings and be more efficient than GDL.

2.5 Summary and discussion

This study explored the feasibility of using the correlation cutoff method (YK18, Yoshida and Kalnay 2018; Yoshida 2019) as a spatial localization and compared the accuracy of the two types of localization, the correlation-dependent (YK18) and distance-dependent (GDL), preliminarily on the Lorenz (1996) model with the LETKF. We also proposed and explored the potential of the two types of hybrid localization applications (Hybrid and Hybrid II). Our results showed that YK18 performs a similar analysis as GDL but with a significantly shorter spin-up, especially when fewer ensembles and observations are presented. YK18 can accelerate the spin-up by optimizing the use of observations with its prior knowledge of the actual error correlations, effectively reducing the required number of cycles toward the analysis convergence. In our experiments with the variant L96 model, we demonstrated that these advantages of YK18 would become even more pronounced under a more complicated dynamic.

It is worth highlighting that YK18 is more efficient and economical than GDL. Traditionally, the use of GDL requires multiple trial-and-errors to define the optimal localization length for the experiments of interest. In contrast, YK18 only needs one offline run to obtain the prior error correlations, whereas it provides a comparable analysis as GDL, even with a faster spin-up. For operational or research centers that have plentiful archives of historical ensemble datasets, it is possible to directly obtain the required prior error correlation for YK18 from the past data (e.g., historical ensemble forecasts) without executing additional offline runs.

We found that the hybrid methods, the combination uses of YK18 and GDL, generated a more accurate analysis than that solely using GDL or YK18. Hybrid II has the same advantages as YK18 in accelerating the spin-up and a larger optimal localization length than GDL. These features allow Hybrid II to spin up quicker, obtain more observation information after the system convergence, and generate a slightly better analysis than GDL and YK18. Since the analysis unbalances would be relaxed by a larger localization length, we expect the hybrid methods would deliver a more balanced analysis than GDL with a multivariate model. Further investigation of this advantage will be part of our future work.

We would like to emphasize that the L96 model used in this study is highly advantageous to GDL because of its univariate and simple dynamic without teleconnection features. So, the two known problems in GDL, unbalanced analysis and losing long-range signals, would not appear here to degrade its performance. Despite that, this model is still an excellent testbed for preliminary DA studies because it offers a simple and ideal environment for first exploring the fundamental characteristics of new methods. With that in mind, it is encouraging that YK18 performed a comparable analysis to GDL (even with a shorter spin-up) under such an environment that is particularly advantageous to GDL. We believe YK18 has a great potential to generate a relatively accurate and balanced analysis than GDL in a more sophisticated, multivariate model. More studies with a multivariate and more realistic model would be required and will be conducted as our future works.

Another future work will be extending the use of YK18 to location-varying observations. One potential solution is to use neural networks to estimate

corresponding error correlations for YK18 applications (Yoshida, 2019). Yoshida (2019)'s experiments proved that neural networks could estimate the background error correlations for observation at arbitrary locations. Although high computational costs and numerous samples are inevitable for training neural networks, once the network is developed, it can provide significant advantages in estimating the error correlations for location-varying observations such as satellite data.

Chapter 3: Reducing the SST biases in the CFSv2 analysis under the WCDA framework

3.1 Introduction

This chapter explores the impact of alternatively using WCDA instead of the traditional relaxation (nudging) scheme to the CFSv2 on constraining the surface and upper sea temperature. We investigated how WCDA and nudging affect the ocean temperature and atmosphere analyses and forecasts with conceptual experiments on the operational-like CFSv2-LETKF and real observations.

3.2 Methods

3.2.1 The NCEP Climate Forecast System version 2 (CFSv2)

The second version of the NCEP Climate Forecast System (CFSv2) (Saha et al. 2010, 2014) is a fully coupled atmosphere-ocean-land model that provides retrospective forecasts, reanalysis, and operational predictions of climates on a global scale. The CFSv2 includes an atmospheric component of the Global Forecasting System (GFS) and an ocean component from the GFDL Modular Ocean Model version 4 (MOM4). The land component is from the 4-level Noah land surface model (Ek, 2003), which runs as part of the GFS model with dynamic vegetation. A

three-layer interactive sea ice model (Winton, 2000) is utilized as a subcomponent performing ice dynamics, vertical thermodynamics, ice transport, and surface albedo.

The atmospheric model is integrated with a spectral resolution of T126 in horizontal (~ 100 km) and 64 sigma-pressure hybrid vertical levels. The ocean model is run with a horizontal resolution of 0.25 degrees near the equator (10°S to 10°N latitude band) for better capturing the equatorial dynamics and 0.5 degrees elsewhere. There are 40 vertical layers for the MOM4, and above the ocean surface, a coupler is employed between the atmosphere and ocean components, interpreting the air-sea interactions and flux exchanges.

3.2.2 The CFSv2-LETKF

The CFSv2-LETKF (Sluka et al., 2018) is built based on the CFSv2 and the LETKF and can update the CFSv2 ensembles under WCDA and SCDA. For WCDA, the atmosphere and ocean components are updated independently with observations in their respective domains. The background error covariance of LETKF is estimated from CFS forecast ensembles, delivering flow-dependent features in its error statistics. For a coupled system, the state vector $\mathbf{x}$ includes the atmosphere and ocean components. Namely,

$$\mathbf{x} = [\mathbf{x}^{\mathrm{ATM}} \ \mathbf{x}^{\mathrm{OCN}}] .$$

In the CFSv2-LETKF, the atmospheric analysis variables ($\mathbf{x}^{\mathrm{ATM}}$) are winds (U, V), air temperature (T_{air}), moisture (Q), and surface pressure (Ps). The oceanic analysis variables ($\mathbf{x}^{\mathrm{OCN}}$) are sea temperature (T) and salinity (S). Our experiments were

conducted under WCDA, so the $\mathbf{x}^{ATM}$ and $\mathbf{x}^{OCN}$ are updated independently with observations in their respective domains. Although no cross-domain increment is involved in DA, the increments made in one domain (e.g., ocean) can still affect another domain (e.g., atmosphere) through the CFSv2 coupled forecasting.

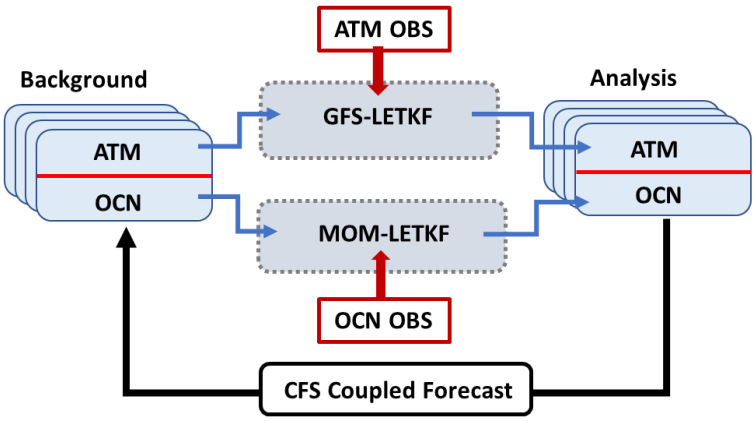

Figure 3.1 Schematic of the CFSv2-LETKF under WCDA framework.

3.3 Experiment Design

This study conducted a series of 1-month experiments to investigate the impact of WCDA and nudging on constraining the CFSv2 SST. The effect of different types of SST observations was also explored and compared.

3.3.1 Observations

A subset of the NCEP PREBUFR data described in Table 3.1 were assimilated for all experiments for the atmosphere components. We used the Canadian

Meteorological Center (CMC) daily mean foundation sea temperature as the Level 4 (L4) SST observations (Brasnett B., 2008) and NOAA Advanced Very High-Resolution Radiometer (AVHRR) infrared satellite SST data as the Level 2 (L2) SST in our WCDA experiments. The NOAA Optimum Interpolation Daily SST version 2 (OISSTv2, Reynolds et al., 2007), also known as the Reynolds SST, was applied in the nudging experiments for imitating the operational CFSv2 SST relaxation. L4 SST is a global, gridded, blended, and gap-free SST analyzed field generated by satellite-retrieved SST observations, and L2 SST is reprocessed, satellite-retrieved SST at the observed pixels. A detailed description of these three SST data is listed in Table 3.2. The observation distributions of atmosphere and SST observations are shown in Figure 3.2 and Figure 3.3, respectively.

The L2/L4 SST data can be downloaded from the Group for High-Resolution Sea Surface Temperature (GHRSST, https://www.ghrsst.org/).

Table 3.1 Observations used in this study

Types of observation	Description	Experiment used
Atmospheric Observation		
ADPUPA	Upper-air (RAOB, PIBAL, RECCO, DROPS) reports	All experiments
ADPSFC	Surface land (synoptic, metar) reports	
SFCSHP	Surface marine (ship, buoy, c-man/tige gauge platform) reports	
AIRCFT	AIREP, PIREP, AMDAR, TAMDAR aircraft reports	
PROFLR	Wind profiler and acoustic sounder reports	
SATWND	Satellite-derived wind reports	
VADWND	VAD (NEXRAD) wind reports	
WDSATR	Windsat scatterometer wind data (reprocessed)	
Ocean Observation		
CMC L4 SST	CMC High-resolution Blended Analysis of Daily Foundation sea temperature	WCDA & Nudging
NOAA OISSTv2	NOAA Optimum Interpolation Daily SST	Nudging
NOAA L2 SST	NOAA AVHRR L2 SST	WCDA

Table 3.2 SST observations used in this study

	OISSTv2	CMC L4SST	NOAA AVHRR L2SST
Resolution	Level 4 (0.25 degrees)	Level 4 (0.25 degrees)	Level 2
Platforms	NOAA AVHRR NASA EOS AMSR-E In-situ (ships, buoys)	TRMM TMI CORIOLIS WINDSAT NOAA AVHRR NASA EOS AMSR-E ERS ATSR Buoys ENVISAT AATSR	NOAA AVHRR

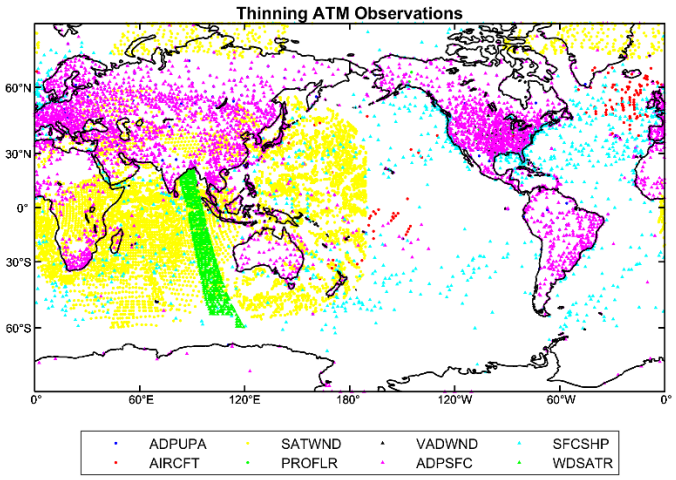

Figure 3.2 A snapshot of the location of assimilated atmospheric observations.

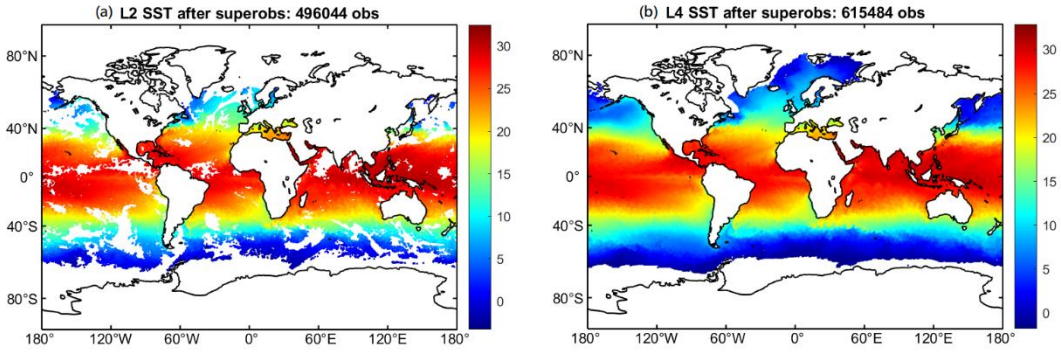

Figure 3.3 The location of assimilation (a) L2 SST and (b) L4 SST.

3.3.2 WCDA and nudging experiments

Four experiments listed in Table 3.3 were carried out on the CFSv2-LETKF for one month. The main difference between the nudging and the WCDA experiments is how they use the gridded SST data to constrain the SST. For nudging, SST and SSS are directly relaxed to the selected data to constrain the fields at the interface domain (between the atmosphere and ocean) of CFSv2. In contrast, WCDA tends to update the subsurface sea temperature in the MOM4 by the assimilation of SST data. With an improved sea temperature analysis at the upper ocean (e.g., 5m depth), the model can better estimate the interface SST for CFSv2.

For the nudging experiments, the SST was relaxed to the OISSTv2 (Reynolds et al., 2002, 2007) and L4SST with a time scale of 5 days, and the SSS was relaxed to the annual salinity climatology (Conkright et al., 1999) with a time scale of 30 days. This configuration mimics the operational global ocean data assimilation system at NCEP (GODAS). For the WCDA cases, the ocean temperature was updated daily at 12 UTC. The SST relaxation was suspended while we kept the SSS relaxation the same as in the nudging case to avoid field drifting. A comparison between the nudging case, GODAS, and WCDA case is shown in Table 3.4.

Table 3.3 The SST WCDA and nudging experiments

EXP name	**Nudge OISSTv2**	**Nudge L4SST**	**WCDA L4SST**	**WCDA L2SST**
ATM DA	**NCEP prebufr** / DA cycle = 6 hours			
OCN DA	x	x	**CMC L4SST** DA cycle : 1 day	**AVHRR L2SST** DA cycle: 1 day
SST Nudging	**OISSTv2** Tscale: 5 days	**CMC L4SST** Tscale: 5 days	x	x
SSS Nudging	Climatology (Tscale: 30 days)			

Table 3.4 Comparison of GODAS, WCDA, and nudging experiments

	Operational GODAS	**Nudging EXPs**	**WCDA EXPs**
Sea Surface Temperature	Relax to NOAA OISST (Reynolds et al. 2002) with a time scale of 5 days	Relax to gridded SST field (OISSTv2 and CMC L4 SST) with a time scale of 5 days	No relaxation
Salinity	Relax to annual salinity climatology (Conkright et al., 1999) with a time scale of 10 days	Same as GODAS but with a time scale of 30 days	
Sub-surface Temperature	- Top level (5m): Relaxation to gridded daily mean SST (every 6hr) - Subsurface: 3DVAR assimilation with surface and profile data	Free run	WCDA high-resolution SST at 10m (CMC L4 SST) and 0m (L2 SST)
Salinity		Free run	Free run

For all experiments, the meteorological variables of wind field (U, V), temperature (T), specific humidity (Q), and surface pressure (Ps) were updated by assimilating the conventional data (e.g., no radiance, GPS, and precipitation data) of NCEP PREBUFR every 6 hours at 00, 06, 12, 18 UTC.

The experiment period is one month and initialized from 00 UTC, June 01, 2006, with 40 analysis ensembles provided by IITM. The covariance inflation is the relaxation to prior spread (RTPS) method (Whitaker and Hamill, 2012). The RTPS rate is 0.9 for both atmosphere and the ocean. The localization (Gaspari and Cohn (1999)) length scale is 500 km horizontally and 0.4 scale height vertically for the atmosphere domain. For ocean localization, the horizontal length scale varies from 80 km to 300 km with respect to the latitude of the analysis grid, and the length scale in vertical is 50 meters.

3.4 Results of the CFSv2 analysis

3.4.1 The impacts of WCDA and nudging

Before comparing the impacts of WCDA and nudging on the SST constraint, we would like first to examine the fundamental effects of the two methods on the CFS SST field. It is known that CFS has a significant warm bias in SST simulation which may degrade its accuracy of seasonal predictions (Kumar et al., 2012; Jin and Kinter, 2009; Wang et al., 2010; Chaudhari et al., 2013; KBRR et al., 2020). We found the same issue in our CFSv2-LETKF, as shown in Figure 3.4, that the initial state of all

experiments (the first background) had noticeable warm biases, especially in the tropical region and the northern Pacific.

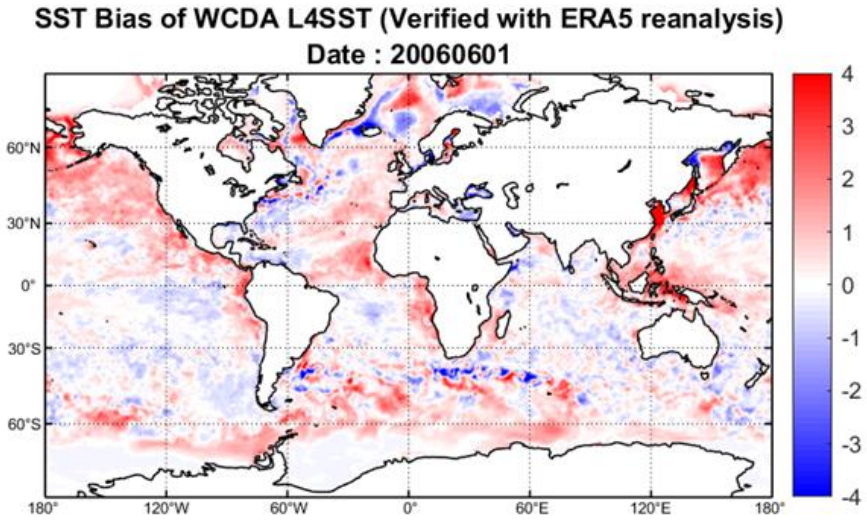

Figure 3.4 The SST bias (WCDA minus ERA5 reanalysis) at 20060601 (first background).

To understand the effects given by SST DA, we compared the analysis of 5m and 15m sea temperatures with and without SST DA. Figure 3.5 shows that the SST DA gave an overall cooling effect on the SST, which mitigated the SST warm bias in the CFS (Figure 3.5 (c)(d)).

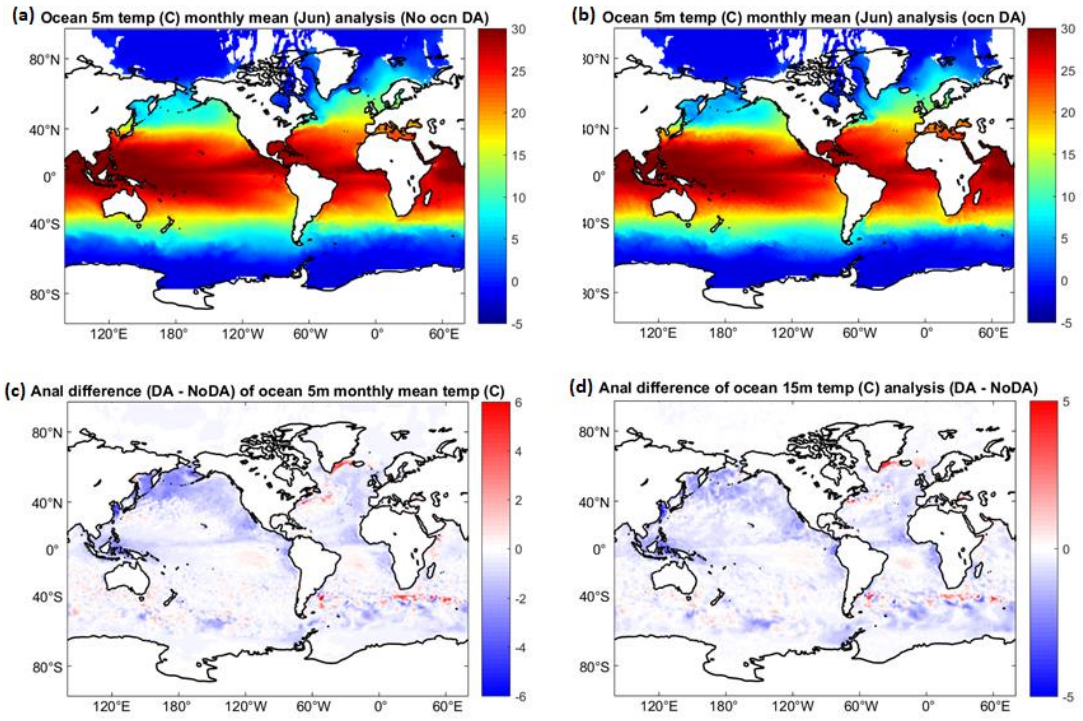

Figure 3.5 The ocean 5m depth temperature of (a) noDA and (b) with WCDA. The temperature differences between WCDA and no DA for (c) 5m and (d) 15m depth temperatures.

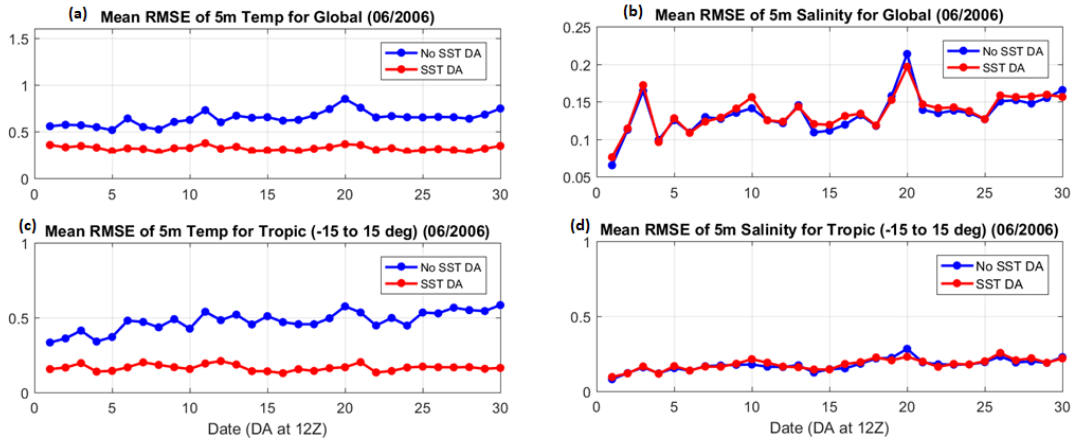

Figure 3.6 The time series of the global and monthly mean RMSE of 5m (a) sea temperature and (b) salinity for no DA (blue) and WCDA L4SST (red) experiments.

(c)(d) the same as (a)(b) but are for the Tropics (-15 to 15 degrees). The RMSEs were verified with ocean profiles.

Next, we examined the RMSE reduction of WCDA in upper sea temperature and salinity. Figure 3.6 shows the time series of the mean RMSEs of 5m sea temperature and salinity with (red) and without SST DA (blue). It is clear that with SST DA, the 5m sea temperature RMSE was significantly lower than without DA, indicating that SST DA is effective in constraining the upper ocean temperature. In contrast, the SST DA didn't greatly impact the salinity RMSEs. This result is reasonable because sea temperature is not highly correlated to salinity in the upper ocean (e.g., mixed layer). That said, the increment given by assimilating SST observations on salinity is very limited. Therefore, the salinity field at the surface and subsurface were mainly constrained by the salinity relaxation applied at the ocean surface. As a result, the salinity analysis from the nudging and WCDA experiments will be comparable under the current framework, where salinity assimilation was not yet conducted.

Vertical localization plays a vital role in determining the observation impact from the observed level to other levels. The vertical cutoff radius (182 meters) we used in this study makes the corrections from the SST observations mainly happened at the surface and the mixed layer. Although some increments could be brought to the thermocline and deep ocean through the vertical mixing and ocean current, the amount was minimal within the 1-month experiment.

In conclusion, we confirmed that WCDA and nudging both could reduce the warm SST bias in the CFSv2 based on the above results. We expected the difference in salinity between the two methods might not be significant, and the increments from

SST observations would be basically presented within the mixed layer. Therefore, the discussions in the following sections will focus on the SST and sea temperature in the mixed layer.

3.4.2 Impact on the SST constraints

This section investigates the impact of the four experiments (Table 3.1) on the CFSv2 SST and upper sea temperature analysis. Figure 3.7 is the time series of the mean SST analysis RMSE within the regions of Tropics (24 ºS ~ 24 ºN), Extratropic in the North hemisphere (NH, 25 ºN ~ 65 ºN), and Extratropic in the South hemisphere (SH, 25 ºS ~ 65 ºS) for the four experiments. Here, we used ERA5 reanalysis as the independent reference to evaluate the RMSEs for our experiments. Since we only assimilated/nudged SST into MOM4, the ECMWF reanalysis (e.g., ERA5) and Operational Sea Temperature and Ice Analysis (OSTIA), which assimilates satellite-retrieved SST, ice data, and ocean profiles for its ocean, is expected to be more accurate than our ocean analysis, making it an appropriate independent reference for our conceptual experiments.

First, the WCDA cases noticeably outperform the nudging cases, especially for Tropics and SH, where WCDA shows ~ 60% improvements in the SST analysis. For the NH, WCDA performed a superior SST analysis for the first 15 days and a comparable performance as the nudging cases in the later stage, where the RMSEs of the WCDA cases in the NH slightly increased at the end of the experiment period. That is because of the shrinking spread after several DA cycles. The inflation we applied was adequate for Tropic and SH; however, it might be too small for the

summer NH due to its comparatively lower SST variability. Figure 3.8 shows the map of the monthly mean SST analysis biases, which provides geographical information on the presence of improvements. It is very impressive that WCDA significantly mitigated the SST biases almost everywhere.

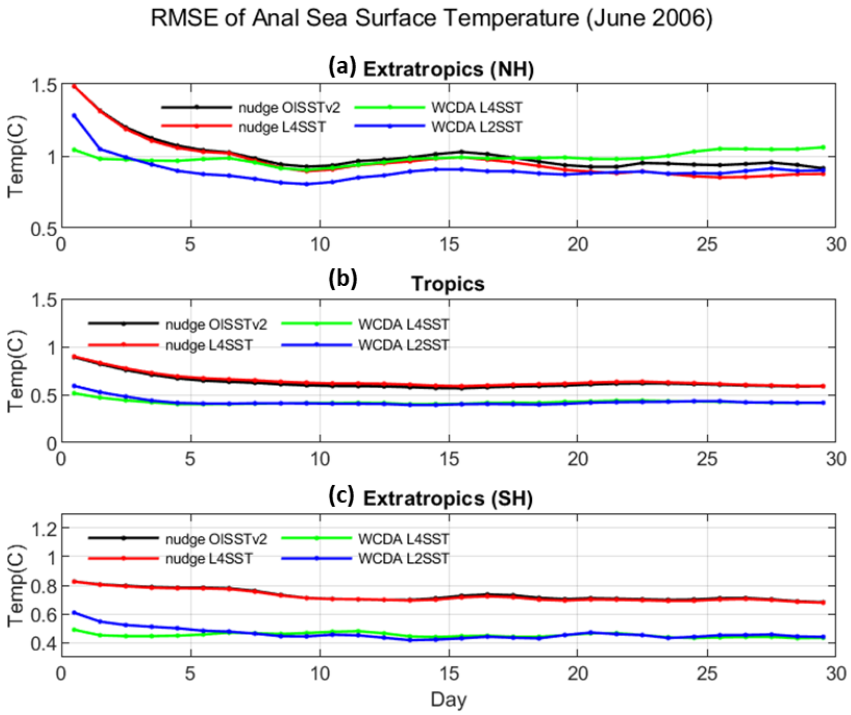

Figure 3.7 The time series of the analysis RMSE for the four experiments at the region of (a) NH extratropics, (b) Tropics, and (c) SH extratropics.

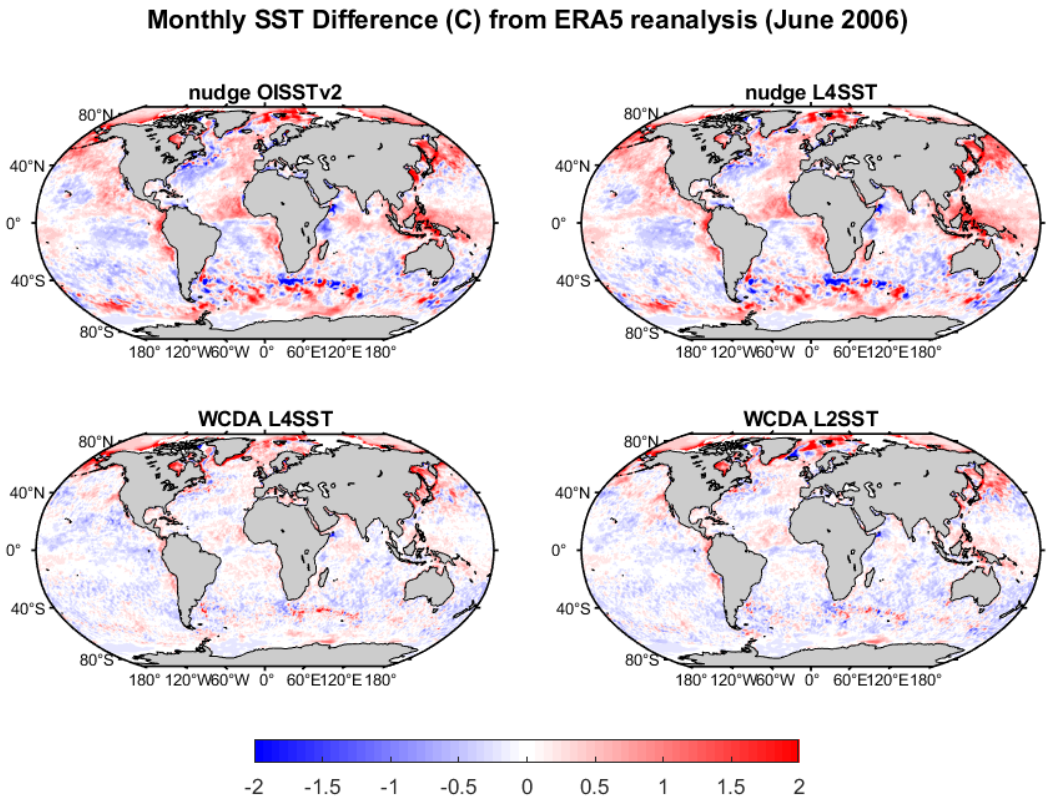

Figure 3.8 The map of monthly-mean SST bias (experiment minus ERA5 reanalysis) for four experiments.

We found that the method plays a more critical role than the choice of data in SST constraining. For example, nudging to the OISSTv2 and L4SST gives virtually identical SST results. Similar to WCDA cases, assimilating L4 SST or L2 SST generally provides similar global biases patterns. To be more specific, L4 SST improved the Tropics more, and L2 SST represented a smaller error in the high-latitude regions, though the differences were tiny. A shred of more clear evidence to back this argument is the comparison of WCDA L4 SST and the nudge L4 SST shown in

Figure 3.9. These two experiments used the same observations but in different ways. We found that the SST warm/cold biases shown in nudge L4SST were significantly reduced in WCDA L4SST. This result demonstrates how the observations being digested into the CFSv2 substantially impact the effectiveness of SST constrain and the consequent analysis, and apparently, WCDA has superior performance than nudging.

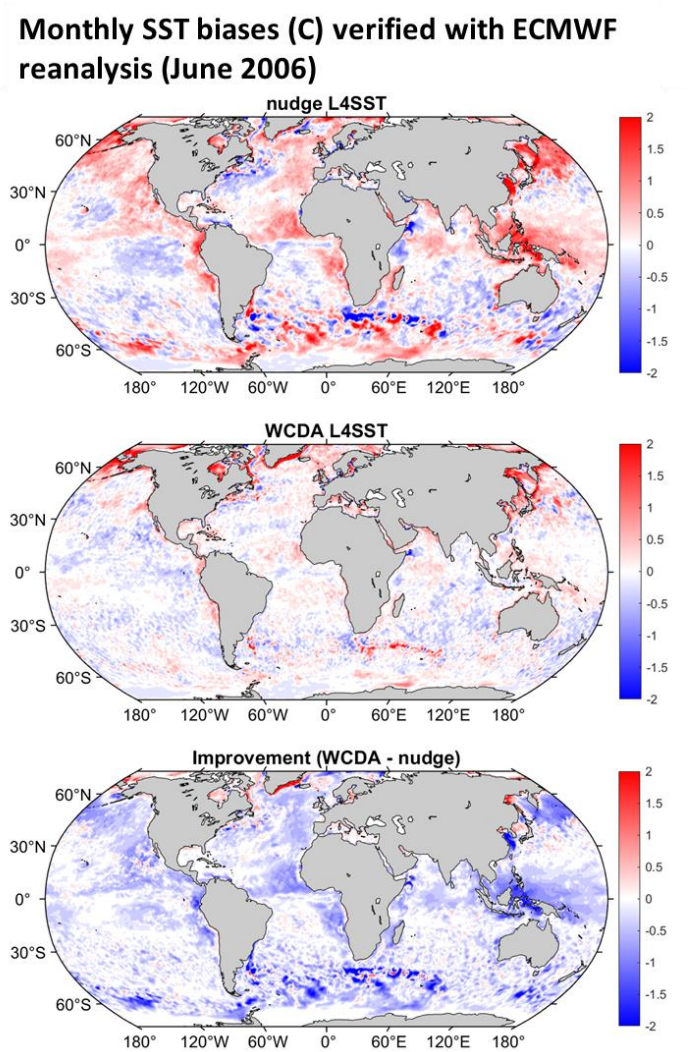

Figure 3.9 The monthly-mean SST bias map of nudge L4SST (upper) and WCDA L4SST (middle), and the SST bias difference (WCDA minus nudge) (bottom). The SST bias is verified with ERA5 reanalysis. For the improvement figure (bottom), the blue (red) color means WCDA (nudge) is better.

3.4.3 Impacts on the ocean temperature

The SST corrections made by either nudging or WCDA at the ocean surface would gradually impact the subsurface or deeper layers through the vertical mixing or ocean current during the model integrations, so it can be expected that SST WCDA and nudging could consequently affect the interior ocean temperature after a period of time. Figure 3.10 is the vertical profile of the monthly mean analysis RMSE of the sea temperature (ST) for all experiments. The black line represents the RMSE of the free-run, where the ocean state was the pure forecast from initial ensembles (its atmosphere state was updated by DA the same as other experiments). The free-run result can be seen as a bottom line of the experiments, which gives us an idea of how much improvement each experiment provided against the unadjusted state.

In general, both WCDA and nudging could reduce the RMSE of SST and effectively constrain the SST error within 1 ºC. We found that the WCDA provides more improvements than the nudging within the mixed layer, especially in the Tropics and SH (Figure 3.10). The presence of a deeper and more extensive RMSE reduction in the SH is due to a thicker ocean mixed layer in the local wintertime, where the improvements could be brought down to a deeper depth through active vertical mixing.

We also found that the corrections from WCDA can reach a deeper depth more quickly than nudging. For example, the ocean temperature at 150 meters depth can be simultaneously corrected in the first WCDA cycle, while it needs ~ 7 to 10 days (depending on regions) for the model to convey the increments from the surface to the same depth in the nudging experiments. The same conclusion is also mentioned by Counillon et al., 2016, who found assimilating SST observations can improve the interior ocean simultaneously by the ensemble-derived cross-covariance. This feature makes DA a more efficient method for conveying observation information from the surface to deeper layers to update the subsurface and interior ocean.

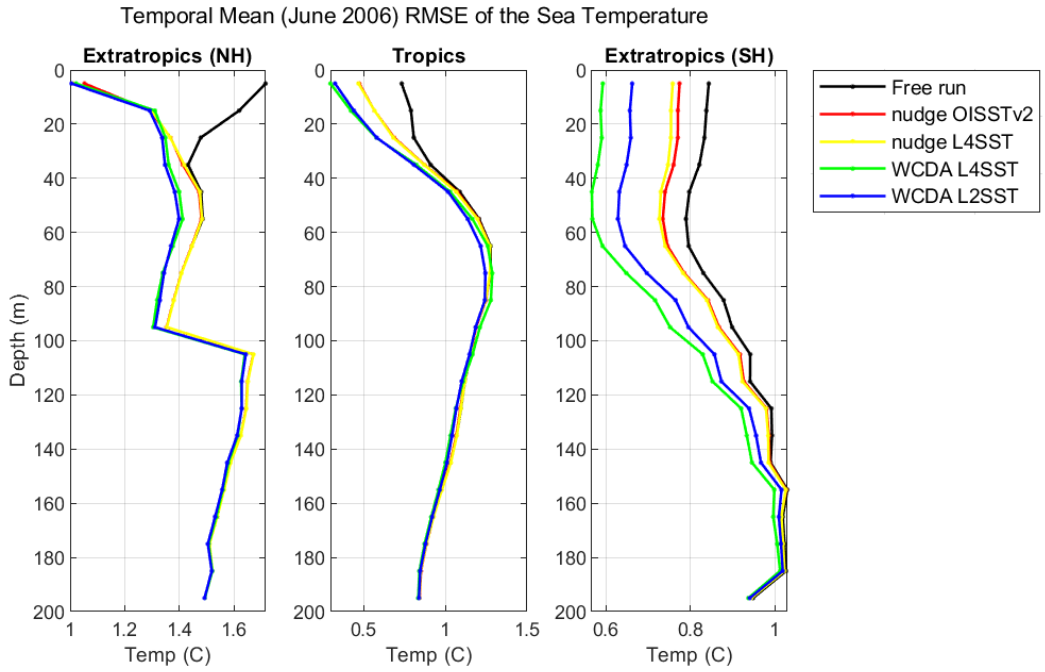

Figure 3.10 The vertical profile of the sea temperature monthly-mean analysis RMSE of control (black), nudge OISSTv2 (red), nudge L4SST (yellow), WCDA L4SST (green), and WCDA L2SST (blue) for NH (left), Tropics (center), and SH (right).

3.4.4 Impacts on the atmospheric 2m temperature

The WCDA updates the atmosphere and ocean states independently. Namely, there will be no cross-domain increment from SST observation to the atmospheric variables. However, the corrections in the SST fields made by WCDA or nudging could impact the lower-level atmosphere through the air-sea flux exchanges during coupled forecasting. One of the representative variables for examing the air-sea coupled effect is the 2-meter air temperature (2mT). When the SST varies, the variations will impact the heat flux exchange in the coupler of CFSv2 and gradually influence the atmosphere above the sea surface, such as skin temperature and 2mT. Note that the 2mT field is a diagnosis variable estimated by other model variables. That means the SST variation affects the 2mT in an indirect way.

Figure 3.11 is the monthly-mean 2mT RMSE for the four cases. We found that the 2mT fields were significantly improved by replacing the nudging with WCDA (Figure 3.11 (e) (f)), especially for the high-latitude and tropical regions. This result clearly shows that the improvements in the SST field could further improve the atmosphere through coupled forecasting, and WCDA could provide a more accurate 2mT estimation due to its more significant improvements on the SST analysis.

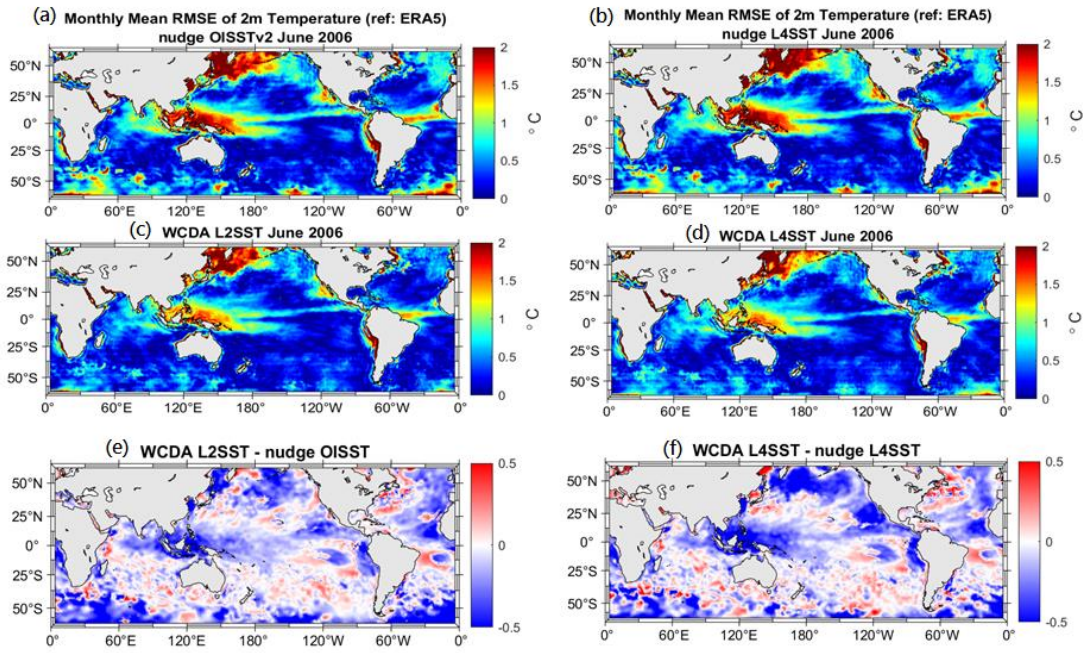

Figure 3.11 The monthly-mean analysis RMSE (verified with ERA5) of 2m temperature for (a) nudge OISSTv2, (b) nudge L4SST, (c) WCDA L4SST, and (d) WCDA L2SST. (e) The RMSE differences between (a) and (c). (f) The same as (e), but for (b) and (d).

3.5 Impacts on the CFSv2 SST forecasts

This section explored the impacts of WCDA and nudging on the CFSv2 forecasts. A series of 10-day ensemble forecasts initialized from five analyses in a row from June 15 to 19, 2006, at 12Z were conducted (Figure 3.12). The forecast results were verified with ERA5 reanalysis (SST) and ocean profiles (depth > 5m).

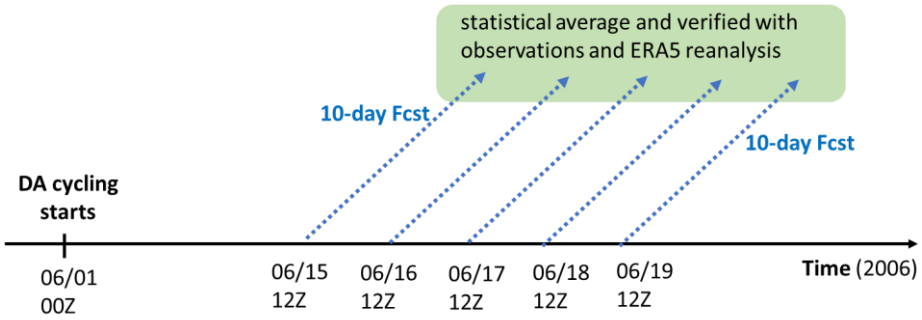

Figure 3.12 Schematic of the CFSv2 forecast experiments.

Figure 3.13 shows the 10-day forecast RMSE of SST on the global and regional scales. The result indicates that the forecasts initialized from WCDA analyses were significantly better than those from nudgings. We found that the WCDA L2SST has a superior forecast than WCDA L4SST at the upper ocean, while WCDA L4SST performed slightly better predictions in a deeper depth (Figure not shown). Since L4SST is the foundation temperature at 10m depth, and L2SST is the surface temperature at 0m, L2SST could give more precise information about the sea surface, such as the diurnal heating and cooling, and the increment from L4SST could go slightly deeper than L2SST. However, these differences were insignificant and generally provided similar forecast results.

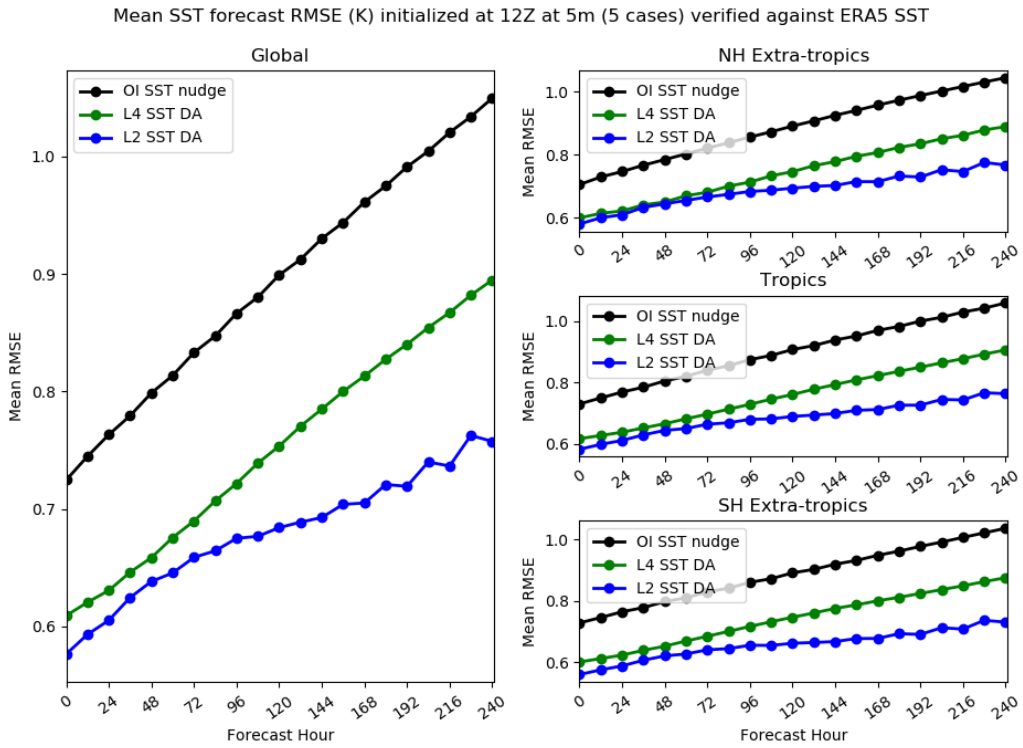

Figure 3.13 The time series of the 10-day forecast RMSE of SST for nudge OISSTv2 (black), WCDA L4SST (green), and WCDA L2SST (blue) for global and different regions. Figure provided by Dr. Kriti Bhargava.

3.6 Summary

This chapter discussed the impacts of SST WCDA and nudging on the CFSv2 ocean analysis and forecasts. We found that SST WCDA, either with L4SST or L2SST, outperforms the nudging, especially in Tropics and SH. The WCDA reduces the SST biases almost everywhere, significantly enhancing the SST analysis accuracy. Another advantage of WCDA over nudging is its efficiency in bringing the observation information to other unobserved ocean layers. WCDA can simultaneously correct the ocean state at different unobserved layers with the surface observations based on their error correlations during assimilation. That makes WCDA

capable of improving the ocean mixed-layer temperature more efficiently and effectively than nudging, thus performing a better temperature analysis in the ocean sub-layers.

Although the atmosphere and ocean states were updated independently under the WCDA framework, the forecast of the lower surface atmosphere (e.g., 2m temperature) and its subsequent analysis could also benefit from a better SST field through the coupled forecasting. We found that the 2m air temperature analysis was significantly improved by replacing the nudging with WCDA, especially for the tropical region.

Finally, we compared the accuracy of the ensemble forecasts initialized from WCDA and nudging cases. Results show that WCDA performed better ocean forecasts than nudging, indicating that WCDA effectively enhanced the CFSv2 forecasts.

This conceptual experiment shows the effectiveness of using WCDA instead of traditional nudging in the CFSv2 SST constraining. A possible future direction would be exploring the SST SCDA on the ocean and atmosphere states.

Chapter 4: Estimating the ocean observation impacts on coupled DA using EFSO

4.1 Introduction

In Chapter 3, we investigated the role of WCDA in enhancing the predictability of the coupled system and how assimilating high-resolution SST products could benefit the reduction of SST bias in the CFS. Since a large number of oceanic observations from different platforms have become available for WCDA, it is essential to explore techniques capable of monitoring the impact of individual observation in DA.

One of the powerful tools for efficiently identifying the beneficial/detrimental impacts of every observation in ensemble-based DA is the Ensemble Forecast Sensitivity to observation (EFSO). EFSO has been widely used in atmospheric DA but has not yet been applied to any ocean or coupled DA due to the lack of a proper error norm for oceanic variables. This chapter aims first to introduce a novel density-based error norm that simultaneously includes sea temperature and salinity forecast errors, by which EFSO becomes available to ocean DA. We then implemented the oceanic EFSO on the CFSv2-LETKF for quantifying the individual impact of ocean observations under the weakly coupled DA framework.

4.2 Model and Methodologies

4.2.1 The error norms for EFSO on the CFSv2

The EFSO can be estimated by different error norms. A common choice for atmospheric EFSO is the dry and moist energy norm (Ehrendorfer et al., 1999).

$$\boldsymbol{e}^T \boldsymbol{C}\boldsymbol{e} = \frac{1}{2}\frac{1}{S}\int_S \int_0^1 \left(u'^2 + v'^2 + \frac{C_p}{T_r} T'^2_{air} + w_q \frac{L^2}{C_p T_r} q'^2 + \frac{R_d T_r}{P_r^2} p'^2_s \right) d\sigma \, dS$$

Here, u', v', T', p'_s and q' are the forecast error of prognostic meteorological variables of U, V, Tair, Ps, and Q, respectively. S represents the target region, σ is the vertical sigma coordinate. C_p, R_d, and L are the specific heat of the air at constant pressure, the gas constant of the dry air, and the latent heat of condensation per unit mass, respectively. Tr and Pr are the reference temperature and surface pressure, for which we use constant values of 280K and 1000 hPa, respectively. w_q is 1 for moist energy norm and 0 for dry energy norm.

For the oceanic EFSO, two error norms are proposed: the ocean temperature/salinity Euclidean norms (T/S L2 norms) and the ocean density norm, which includes the dominant prognostic ocean variables T and S. The details of each error norm are listed in Table 1. For the density norm, the density conversion can be conducted by the Thermodynamic Equation of Seawaters-2010 (TEOS-10, McDougall and Barker, 2011), where a function of ocean T, S, latitude (θ), and vertical ocean depth (z) was applied to estimate the density perturbation. A more detailed description of the conversion can be found in Appendix A.

Table 1. The oceanic EFSO norms. T', S' , and d' denote the perturbation of the ocean T, S, and converted density, respectively. S represents the target region, z is the vertical ocean depth, and θ is the latitude.

	T L2 norm	S L2 norm	Density norm
Variable	Ocean T	Ocean S	Ocean T and S
Unit	K	PSU	kg/m^3
Error norm	$e^T Ce = \frac{1}{2}\frac{1}{S}\int_S \int_z (T'^2)\, dzdS$	$e^T Ce = \frac{1}{2}\frac{1}{S}\int_S \int_z (S'^2)\, dzdS$	$e^T Ce = \frac{1}{2}\frac{1}{S}\int_S \int_z d'(T,S,z,\theta)^2\, dzdS$

4.3 Experimental Settings

In this study, we followed the same configuration of CFSv2-LETKF described in section 3.2.2. The CFSv2 atmosphere model resolution is spectral T126 (~ 1 degree) with 64 vertical levels (up to 0.02hPa). The ocean model has 40 vertical subsurface layers and a horizontal resolution of 0.25 degrees near the equator (10°S to 10°N latitude band) and 0.5 degrees elsewhere. The 40 ensemble members were initially chosen from the NCEP CFS Reanalysis (CFSR) data and freely ran for four months to let the ocean ensemble grow enough spread. Then, we proceeded with a 2-month DA cycling run as the spin-up and started our 1-month experiments from 00 UTC 01 March 2010.

We assimilated the conventional data (e.g., no radiance, GPSRO, and precipitation data) in the NCEP PREPBUFR every 6 hours (00, 06, 12, 18Z) for the atmosphere. For the ocean, the high-resolution NOAA Geo-Polar Blended level-4 Sea Surface

Temperature (L4SST) (Harris and Maturi, 2012) and the ocean T and S profiles from the World Ocean Database (Boyer et al., 2013) were assimilated daily at 12Z. We suspended the SST relaxation (nudging) and instead assimilated the NOAA blended SST to constrain the near-surface ocean temperature, which shows superior results in improving the CFS SST analysis, as discussed in Chap 3.

The localizations (Gaspari and Cohn, 1999) were applied with a length scale of 500 km horizontally and an e-folding scale of 0.4 scale height vertically for the atmospheric observations. For the ocean, the vertical length scale is 50 m, and the horizontal localization length varies from 80 km to 300 km according to latitudes, which is longer at the equator and shorter in high-latitude regions. The latitude-dependent localization aims to include the impact of the Rossby radius of deformation. If the localization radius is too small, such as less than the Rossby radius of deformation, it may lead to an imbalanced analysis and induce gravity waves in the forecast (Sluka, 2018).

The covariance inflation is the relaxation to the prior spread (RTPS) method (Whitaker and Hamill, 2012) with a rate of 0.9 for the atmosphere and 0.95 for the ocean. Since the RTPS inflates analysis ensembles after assimilation, it would magnify the Kalman gain in EnKF (Eq (5) in Ota et al., 2013), which, in turn, would lead to an overestimation of the forecast error reduction, magnifying the EFSO value compared to the original formulation but without changing the sign of EFSO (Kotsuki et al., 2019).

The atmospheric and oceanic EFSOs were evaluated with their own analyses daily at 12Z. For the atmospheric EFSO, we followed the same configuration as Hotta et al.

(2017). Our results of the atmospheric EFSO on the CFSv2 were consistent with Hotta et al., 2017. For the oceanic EFSO, since the evaluation forecast length of 24 hours is a short timescale for the ocean current motions (e.g., on the order of 10 cm/s), the use of moving localization in K12 is not necessary here.

4.4 Results

4.4.1 Comparison of ocean T/S norm and density norm

Since the density forecast performance is directly related to the T and S field, we expect the two types of norms would deliver similar EFSO patterns but with different units. Here, we compared the mean EFSO estimated by the two norms with respect to the ocean depth (Figure 4.1). We found that the two norms performed similar EFSO patterns along with the depth, and both show that assimilating the T and S profiles is generally beneficial for the ocean state at all levels. More specifically, the EFSO impacts are most notable in the mixed layer and are relatively small in the deep ocean. It is because of the abundance of observations and the more substantial forecast error changes due to the active dynamics at the upper ocean. Additionally, due to the assimilation of the high-resolution L4SST, the impact of T profiles near the surface would be diluted, so it is relatively small (but not zero) near the surface (Figure 4.1 (a)).

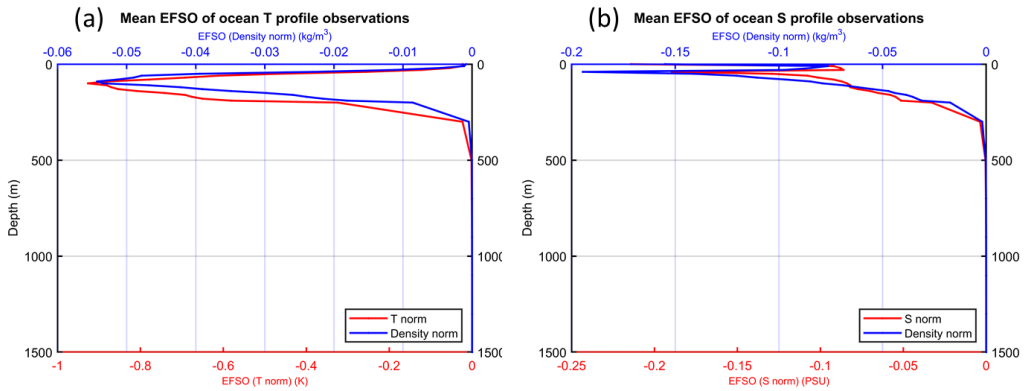

Figure 4.1 The mean EFSO estimated by the T/S norms (blue) and density norm (red) for (a) the T profile and (b) the S profile observations. The value of the blue (red) line corresponds to the blue (red) axis.

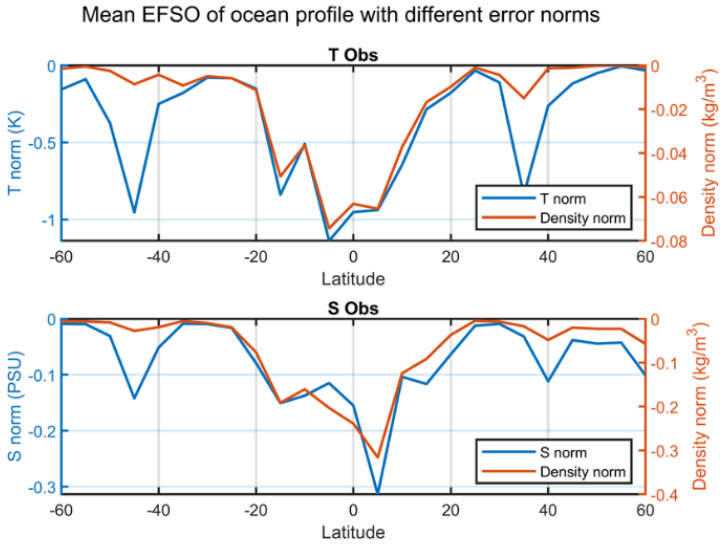

Figure 4.2 The mean EFSO estimated by the T/S norms (blue) and density norm (red) for the T profile (upper panel) and S profile (lower panel).

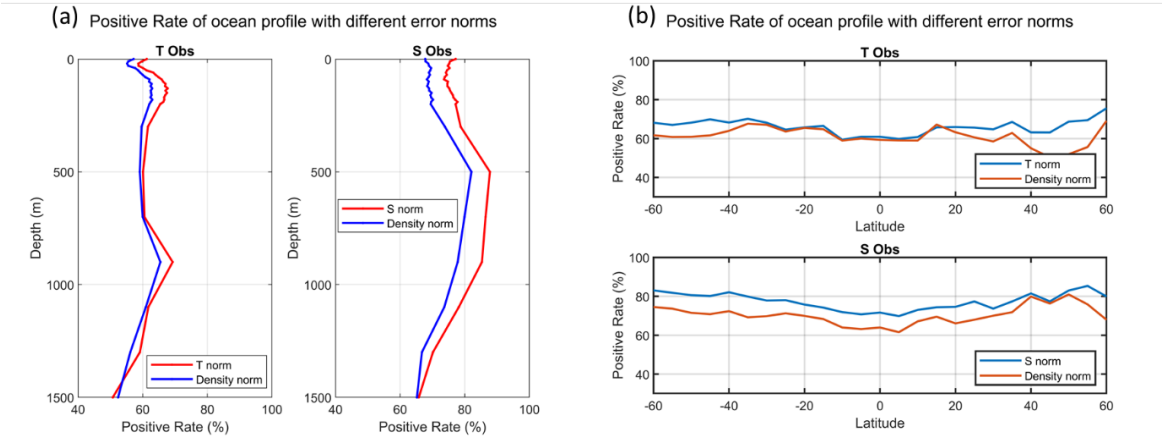

Figure 4.3 The positive rate of ocean profile observations for different EFSO norms with (a) the ocean depth and (b) the latitude changes.

Figure 4.2 and Figure 4.3 (b) are the zonal mean EFSOs and the zonal mean positive rates estimated by the two norms. The positive rate is calculated as

$$positive\ rate = \frac{number\ of\ beneficial\ observations}{total\ number\ of\ observations} \times 100\%.$$

We found that the EFSOs and positive rates are consistent in the Tropics but not in the high-latitude regions. That is because the relative importance of changes in temperature and salinity in determining seawater density varies with water temperature. For the density conversion, temperature variations are more dominant in warm ocean waters (e.g., Tropics), whereas salinity variations are more important in cold ocean waters (e.g., high latitude regions). Therefore, the consistency of the two norms is more elevated for temperature in the Tropics and salinity in the high-latitude regions. Moreover, the mixed water area of Kuroshio Current and the Atlantic meridional overturning circulation (AMOC) around 35ºN to 40ºN also increases the variability in the temperature and salinity, together resulting in a more significant

mismatch between the T L2 norm and density norm at 35ºN to 40ºN (Figure 4.2). Figure 4.4 shows the annual mean state of salinity, temperature, and density between 80°W and 20°E. Noticeably, the temperature and salinity perform more substantial gradients at the overturning region than the potential density.

The positive rates estimated by the two EFSO norms generally agree with each other with varying depth (Figure 4.3 (a)) and latitude (Figure 4.3 (b)). Overall, the positive rate of density norm was slightly smaller than the T/S norms. That is because the density norm includes the multivariate impacts, and the forecast errors from another variable would influence the overall density estimation. Therefore, it is not surprising that the positive rate estimated by the density norm is lower than by the T/S L2 norm at higher latitudes (Figure 4.3).

In summary, this section compares the EFSOs estimated by the T/S L2 norms and the density norm. The results show that the two norms offer similar EFSO patterns, and the density norm could effectively provide a comprehensively incorporated EFSO for ocean temperature and salinity. In the later sections, we will discuss the oceanic observation impacts primarily using the EFSO with density norm.

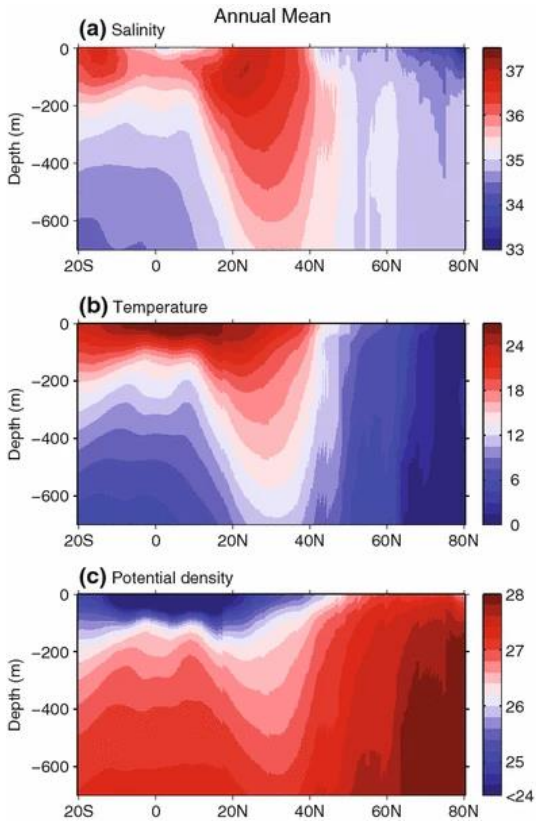

Figure 4.4 The zonally averaged (between 80°W and 20°E) annual mean (a) salinity (psu), (b) temperature (°C), and (c) potential density (kg/m3) as a function of depth and latitude. Figure adapted by Wang et al., 2010.

4.4.2 Oceanic EFSO dependency with the lead-time

This section aims to investigate the oceanic EFSO impacts with respect to the ocean depth and evaluation forecast lead time. Figure 4.5 (a) and Figure 4.6 (a) are the mean EFSO impacts of ocean profiles estimated with different depth intervals and forecast lead times for sea temperature and salinity, respectively.

First, assimilating ocean profiles benefits the ocean state at all depths. Since the forecast error grows nonlinearly with time, the magnitude of the EFSO impact would generally increase with respect to a longer forecast lead time.

Secondly, it is noticeable that the EFSOs within the mixed layer and at the surface are more significant than in the deep ocean. That is because of the more active dynamics and the shorter timescales at the upper ocean. The dynamical timescale of the ocean substantially varies with depth. For example, the timescale at the ocean surface could be as rapid as hourly to daily, while in the deep ocean (> 3000m depth), the timescale increases to hundreds of years. Therefore, the magnitude of the forecast error difference, evaluated as the EFSO impacts, would be more significant at the top of the ocean while very small in the deep ocean within the same evaluation forecast lead time. Thus, the smaller EFSO at a deeper depth doesn't mean that the observation there is less important. Note that the relatively smaller EFSO for sea temperature (Figure 4.5) at the surface is due to the assimilation of dense L4SST, where the observation impacts of temperature profiles would be significantly diluted.

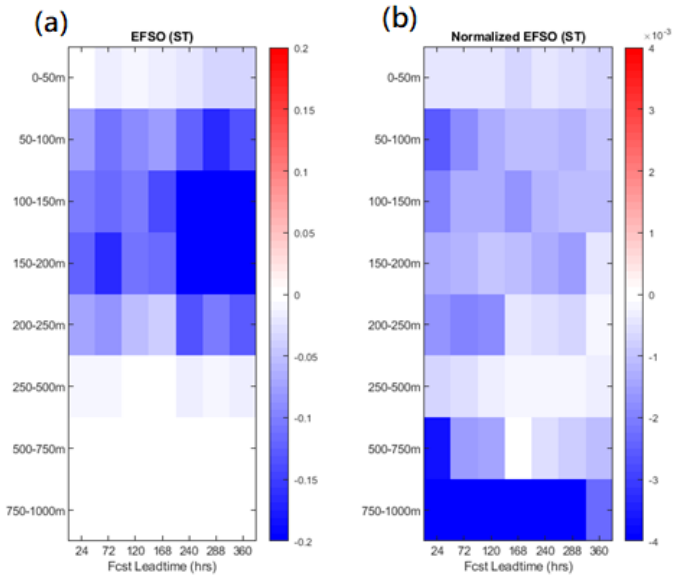

Figure 4.5 The EFSO impact of the temperature profile with respect to depth and forecast lead time for (a) the typical EFSO and (b) the normalized EFSO. Blue (red) color represents the beneficial (detrimental) impact.

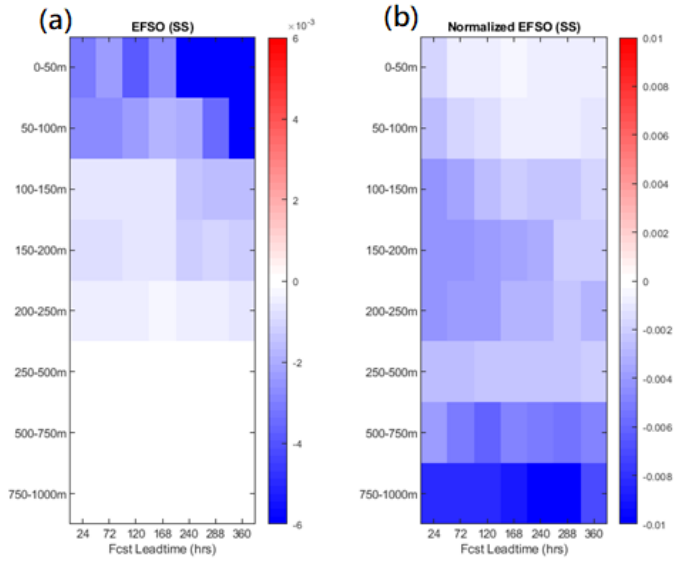

Figure 4.6 Same as Figure 4.5, but for the salinity profile.

Since the EFSO is estimated by the error differences between the two forecasts, it would generally be proportional to the evaluation forecast lead time due to the nonlinear error growth. To better understand the actual dependence of EFSO on the evaluation forecast lead time and depth, we followed Chen (2018) calculating the normalized mean EFSO, which can be represented as $\widetilde{e_{ij}} = \overline{e_{ij}}/\sqrt{\sum_j \overline{e_{ij}}}$, where $\overline{e_{ij}}$ denotes the EFSO of the *jth* depth interval with lead time *i*. By normalizing the EFSOs, the impact of the nonlinear error growth can be generally included, allowing us to have a more fair comparison of EFSOs between different depths and lead times. Figure 4.5 (b) and Figure 4.6 (b) are the normalized mean EFSOs for sea temperature and salinity. We found that

- The impacts of the deep ocean observations are actually very pronounced, and the high impacts can last at least half a month.
- Near the surface (0 - 50m depth), the observation impacts fade away quickly with time. The impacts continuously decrease after 24hrs.
- The observation impacts can last longer in the thermocline (300 – 1000m depth). For salinity, it can last up to three days.
- The mixed layer (0 – 250m depth) has relatively larger EFSO impacts, which peak in one to three days, and drop after five days.

Figure 4.7 shows the positive rate of L4SST and ocean profiles with forecast lead time. We found that the salinity profile obviously has a higher positive rate than temperature observations, especially within 500m to 700m depths. The positive rates of the three types of observation all decreased with a longer forecast lead time. This

result indicates that some observations benefit short-term forecasts but might be detrimental for longer forecasts.

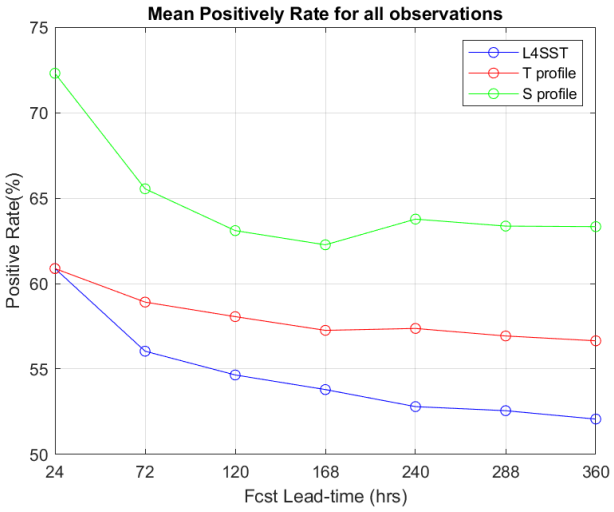

Figure 4.7 The mean positive rate of L4SST (blue), T profile (red), and S profile (green) with respect to forecast lead time.

4.4.3 Geographical impacts of satellite SST and ocean profiles

This section investigates the geographical significance and relative importance of the oceanic observations using the EFSOs of L4SST and ocean profiles. The L4SST are global foundation SST uniformly distributed with a resolution of 0.25 degrees, making its EFSO a perfect illustration of the geographical distribution of ocean observation impacts near the surface. In contrast, the ocean profiles consist of various in-situ measurements such as moored buoys, Argo floats, and ships, which are generally more abundant in the Tropics, coastal regions, and following the ocean currents. Their distributions and EFSOs typically reflect the relative abundance of

observations over the ocean and represent the geographical observation impact on entire water columns.

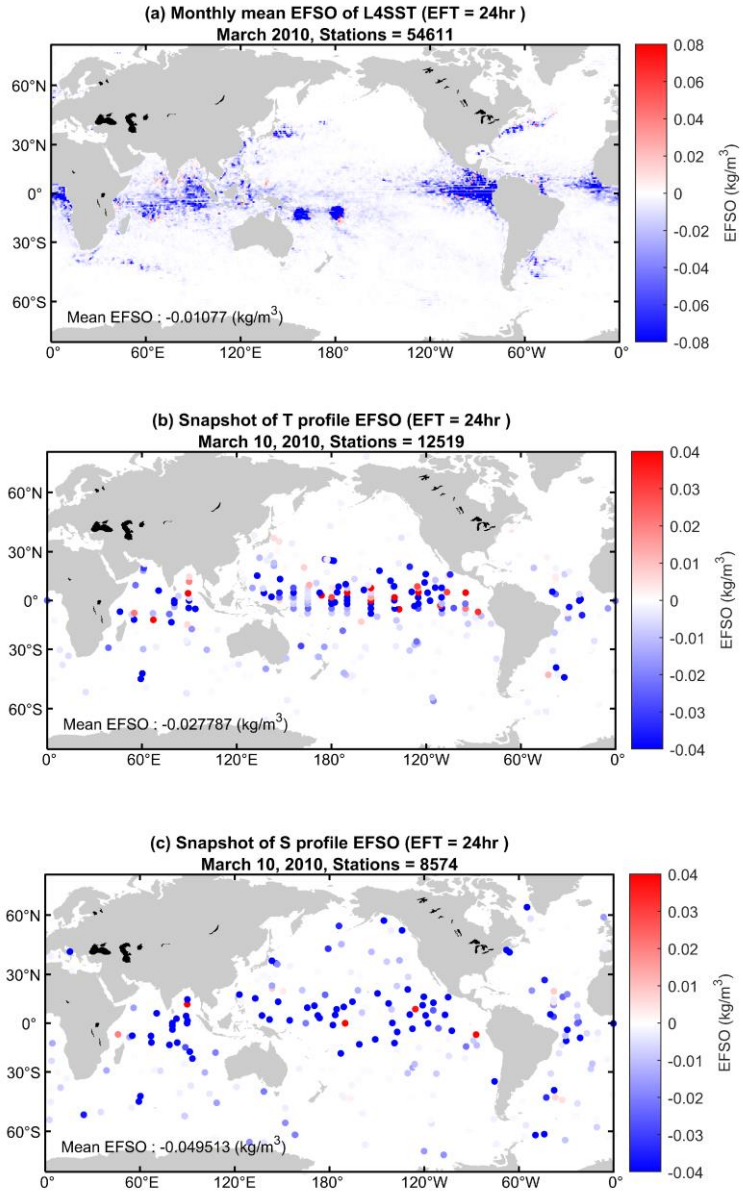

Figure 4.8 Geographical distribution of (a) the monthly-mean EFSO of L4SST and the snapshots of EFSOs of the (b) T and (c) S profiles on March 10, 2010. The

corresponding EFSO for each observation is shown with colors, where blue (red) represents beneficial (detrimental) impacts.

Overall, assimilating L4SST and profiles provides positive impacts on the CFS forecast. Figure 4.8 shows the global map of the monthly-mean EFSO of L4SST (Figure 4.8 (a)) and the snapshots of the EFSOs of T (Figure 4.8 (b)) and S (Figure 4.8 (c)) profiles, where each dot on the plots represent one assimilated observation. For both observations, relatively more significant impacts are seen in areas with larger climate variabilities, such as the eastern equatorial Pacific Ocean, the Gulf Stream in the Atlantic Ocean, and the Kuroshio Current in the North Pacific (Figure 4.8). We also found two cyclonic features with substantial EFSOs (Figure 4.8 (a)) in the southwestern Pacific Ocean. That was related to the air-sea interactions between the upper ocean and the Tropical Cyclones (TCs) Tomas and Ului (2010). The storm activities would thermally (e.g., latent heat fluxes from evaporations and precipitations) and dynamically (e.g., turbulent mixing by surface winds) perturb the upper ocean temperature and salinity, leading to significant error variations in the ocean state (Emanuel, 2001; Bueti et al., 2014; Sriver et al., 2010). For example, the heavy rainfall the storm brings may dilute the salinity, and the strong wind stress may induce a cooling effect on the upper sea temperature. In other words, those perturbations would introduce more considerable uncertainties in the background, gaining the weighting of observation information during DA and thus resulting in larger EFSOs along the storm tracks. Moreover, the quantified EFSO result shows that the influences from TCs could be more considerable than that from its underlying SST variability on ocean DA. All these results demonstrate that the ocean

observations located at storm-track and high SST variability areas would be more influential and valuable for DA, and the appearance of TC should be evaluated for ocean DA when designing future observing systems.

4.4.4 The vertical distributions of ocean observation impacts

The ocean T and S profiles are multi-platforms, three-dimensional observations measuring ocean temperature and salinity at various depths and locations. The profiles contain observations from the surface to as deep as 1000m depth, but over 75% of the profile data were within the mixed layer (Figure 4.9 (a) and (d)). Note that we are simply referring to the mixed layer as the top 200 meters of the ocean, as a globally averaged thickness, but the mixed layer thickness could vary based on seasons and latitude.

The observation impacts could vary with latitudes and ocean depths. Figure 4.9 (b) and (e) are the zonal mean EFSOs per observation of the T and S profiles. The overall negative EFSOs (blue-shaded color) show that assimilating profiles benefit all ocean levels. The impacts are most significant in the tropical mixed layer (Figure 4.9 (b) and (e)), where more active dynamics, such as stronger flux exchanges and mixings, wind-driven turbulence, and active storm activities, are taken place and lead to more considerable forecast error variations. Note that due to the assimilation of dense L4SST and the CFS surface salinity relaxation, the impact of profiles near the surface would be diluted, so their EFSOs would be relatively small at the surface.

The deep-water formation and mixed water areas around 40ºN also have noticeable EFSOs (Figure 4.9 (b) and (e)). These areas, which are related to the

mixed water area of Kuroshio Current and the Atlantic meridional overturning circulation (AMOC), have considerable temperature and salinity gradients (Wang et al., 2010), which would result in more considerable density variations and, thereby, more significant EFSOs. Finally, we found slight degradations in the upper ocean above 65 ºN (red region in Figure 4.9 (b) and (e)). That is possibly due to the appearance of sea ice. When sea ice is involved, the density conversion will become very complicated and challenging. Since the density norm applied here is based on seawater, further investigations on sea ice EFSOs would be needed.

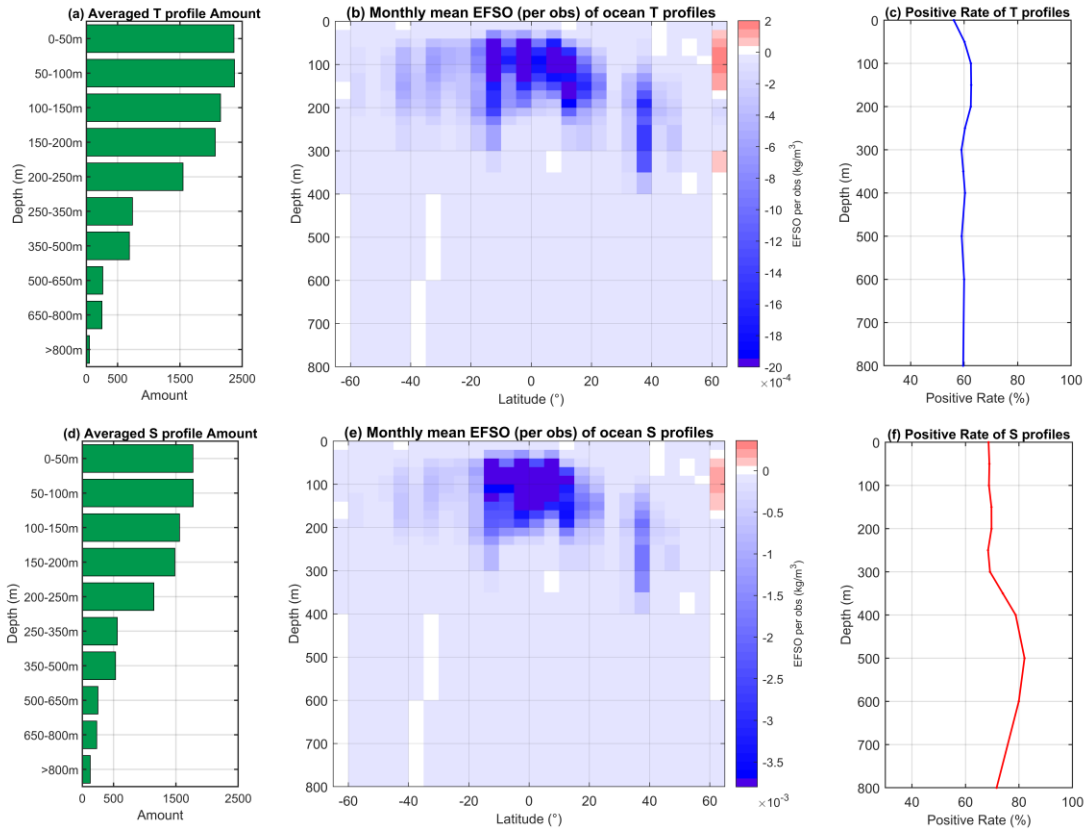

Figure 4.9 The monthly mean of the (a)(d) observation amounts, (b)(e) the zonal-

mean EFSOs, and (c)(f) the positive rate with respect to ocean depths. Variables shown here are the ocean T (a,b, and c) and S profiles (d, e, and f).

Figure 4.9 (c) and (f) are the positive rates of T and S profiles, respectively. The positive rate is calculated as the percentage of beneficial observation amount among all assimilated observations. We found that the positive rates are ~60% for the T profile and ~70% for the S profile, and both are not sensitive to depth changes.

4.4.5 The data-denial experiments

To validate the oceanic EFSO estimation and explore its potential in improving CFSv2 forecasts, we conducted a 1-month offline (e.g., non-cycled) data-denial experiment. At each ocean analysis time (e.g., daily at 12Z), the analysis was generated again by only assimilating those beneficial ocean observations identified by the EFSO. Then, we compared its subsequent forecasts with the original forecasts. The examined variables are the ocean T, S, and the atmospheric 2-meter temperature (2mT). The ocean T and S forecasts were verified with ocean profiles, and the atmospheric 2mT forecast was evaluated with the ECMWF reanalysis. Since nearly 90% of the rejected ocean observations were above 250m depth (Figure 4.10 (a) and (b)), we expected that most forecast improvements would occur in the upper ocean.

Figure 4.10 (c) shows the monthly-mean MSE differences between the EFSO-refined and original forecasts. As expected, most improvements are within the mixed layer. The T forecasts for all levels were significantly improved by removing the detrimental observations (Figure 4.10 (c)). In contrast, the S forecast was notably

improved in the mixed layer but slightly degraded in the thermocline. A possible cause is too few S observations after the data removal. Hotta et al. (2017) found that removing all the detrimental observations does not always guarantee the best result. That said, removing too many observations simultaneously in the thermocline may negatively impact the salinity analysis, especially since S observations are minimal there. Moreover, removing one observation could be beneficial for one variable but detrimental for another. This effect would be more pronounced in the thermocline than in the mixed layer because T and S are strongly correlated there. Despite the slight degradation in the thermocline salinity, it is noticeable that the mean S forecast for the entire water column was still improved.

Finally, we extended the CFSv2 forecast to five days and evaluated the relative forecast improvement (hereafter RFI), which is calculated as:

$$relative\ forecast\ improvement\ (RFI) = \frac{e^f_{original} - e^f_{new}}{e^f_{original}} \times 100\%,$$

where the e^f_{new} and $e^f_{original}$ are the error of the forecasts initialized from the EFSO-refined and original analyses. So, a positive RFI means the EFSO-refined forecast is more accurate than the original, and vice versa. Figure 4.10 (d) shows the monthly-mean RFI of the mixed layer T and S and the atmospheric 2mT of the CFSv2 5-day forecasts. It is impressive that the EFSO-refined improvements could last for at least five days for CFSv2 ocean forecasts and two days for the low-level atmosphere 2mT forecast. This result demonstrates the importance and significant advantage of optimizing the oceanic observation use in a coupled system. As the ocean state is improved, the corrections could instantly be delivered to the low-level atmosphere

through the air-sea flux exchanges, consequently improving the entire CFSv2 forecasts.

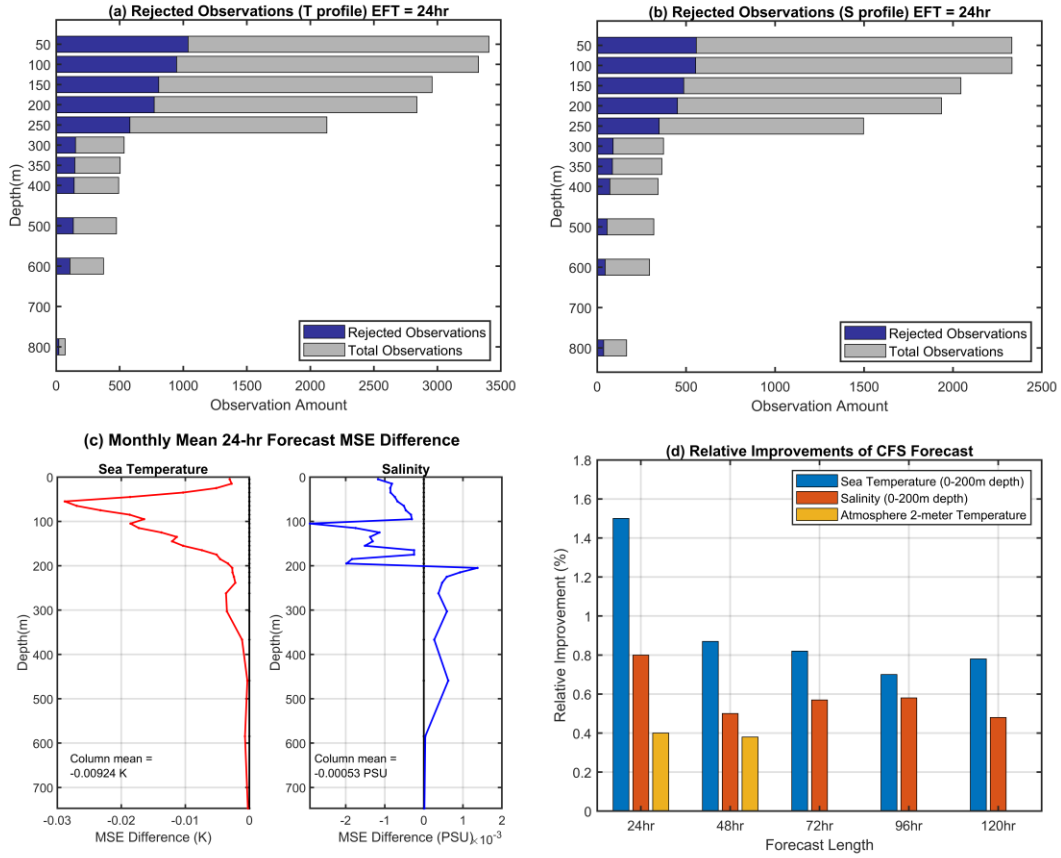

Figure 4.10 The averaged total and removed observation amount for the (a) T and (b) S profiles. (c) The monthly-mean differences (EFSO-refined minus original) in mean-squared errors (MSE) of the 24-hr forecasts. (d) The monthly-mean relative forecast improvement (%) of the ocean T (blue), S (red) in the mixed layer, and the atmosphere 2mT (yellow) forecasts.

4.5 Summary and discussion

This study proposes a new approach to oceanic EFSO with a novel density-based norm. We implemented the oceanic EFSOs on the operational-like CFSv2-LETKF and investigated the ocean observation impacts under the WCDA framework. Then, a 1-month data-denial experiment was conducted to validate the impact estimation and the feasibility of using oceanic EFSO to improve the CFSv2 forecasts.

Our results show that the oceanic EFSO can effectively identify each ocean observation impact, including the L4SST and ocean profiles. Our main findings include:

1. The ocean observations distributed over regions of higher variabilities, ocean current overturning, mixed water area, and storm track have more significant impacts and could be more useful for DA.
2. The CFSv2 forecasts were significantly improved by removing the detrimental ocean observations detected by EFSO. The improvements in the forecasts can last at least five days for the ocean and two days for the atmosphere 2mT.

Potential future directions include extending the ocean EFSO/PQC from current offline to cycling so that the improvements due to PQC can be accumulated and brought to the next DA cycle. Another future study is exploring a universal ocean energy norm that shares the same unit as the atmospheric energy norm, so it can be used to evaluate the impacts of sea surface observations on fully coupled DA.

Chapter 5: Conclusion

5.1 Summary

In summary, this dissertation proposes several advanced approaches to exploring the observation impacts on DA and enhancing the predictability of the coupled system from different aspects.

First, we examined the feasibility of the correlation cutoff method (YK18, Yoshida and Kalnay, 2018) as a spatial localization on the Lorenz 96 model. We found that the YK18 provides a similar analysis to the traditional distant-dependent localization (e.g., GC99) but with a shorter DA spin-up period. The advantage of YK18 in accelerating DA spin-up becomes more significant when the observation and ensemble sizes are reduced and under a more complex model. We also found that the integrated localization offers an even better analysis than solely using YK18 or GC99. The result of this study has been published in Chang and Kalnay (2022).

Second, we attempt to replace the traditional SST relaxation with the WCDA of satellite-retrieved SST products in the CFSv2. A series of conceptual experiments with real observations were conducted on the operational-like CFSv2-LETKF system. We found that the SST WCDA reduced the existing SST biases almost everywhere and significantly improved the CFSv2 analysis and forecast. Our result indicates that the WCDA would be a more effective scheme than the relaxation to constrain the surface ocean state in the CFSv2.

Finally, we proposed a novel ocean density norm for the EFSO to be applied to the ocean state for the first time. We implemented the oceanic EFSO on the CFSv2-LETKF and investigated the ocean observation impacts on the coupled system. Our results demonstrated that the ocean EFSO is effective and efficient in identifying the beneficial/detrimental impact of each observation. Our data-denial experiment also shows that the oceanic EFSO has great potential to be used as a quality control method (e.g., Proactive Quality Control (PQC), Hotta et al., 2017) to enhance the predictability of the coupled system.

5.2 Possible future works

5.2.1 Implementing YK18 on the EMARS

One of the ongoing and future works is implementing YK18 on the Ensemble Mars Atmosphere Reanalysis System (EMARS, Greybush et al., 2012, and 2019) to improve the Mars analysis and forecast. The EMARS is the first ensemble-based system that integrates the GFDL Mars Global Climate Model (MGCM) and LETKF DA system, providing Mars analysis and prediction.

One of the significant advantages of YK18 is that it localizes the information based on error correlations rather than distance so that the distant signals can be preserved during DA. This advantage is especially important for Mars DA to resolve the large-synoptic scale weather dynamics and teleconnection features, such as the transient eddies and dust storms that could range to half the hemisphere (Gillespie et al., 2020). Current EMARS applies the traditional GC99 localization with a 600km Gaussian e-

folding distance. This radius length is much shorter than the scale of large-scale weather features. Namely, using GC99 may cause a loss of distant information for Mars DA.

As a first step, we built a new observation operator for the surface station in EMARS. We proposed to conduct a series of OSSE experimentsThe simulated surface observation is preliminarily set to be at the lowest model level 28, corresponding to a reference pressure of 607.9 hPa. A detailed reference of MGCM model levels, pressure, and heights can be found in Greybush (2011) Table 2.1.

To better illustrate the impact of GC99 and YK18 on Mars DA, we calculated the error correlations between the simulated surface temperature observation and model state from 16 ensemble forecasts during MY29 Ls 210 to 240, early in dust storm season. Figure 5.1 compares the long-term averaged error correlation of the simulated surface station and model state between daytime and nighttime. We found that the error correlation of Mars temperature has significant diurnal variation. It is evident that the error correlation has a horizontal and vertically more significant extent in the area. Therefore, using GC99 with a fixed localization length may not be an ideal strategy for Mars DA.

Figure 5.2 demonstrates the correlation and tentative localization maps using YK18 and GC99. It is noticeable that the YK18 successfully captures the main error correlation with long-range features. In contrast, the GC99 performs a flow-independent feature in weighting the observation impacts and intentionally discards the distant signals. Thus, based on the above evidence, we expect the YK18 can

further improve the Mars analysis with its advantages of better adapting to dynamic variations and preserving more distant information.

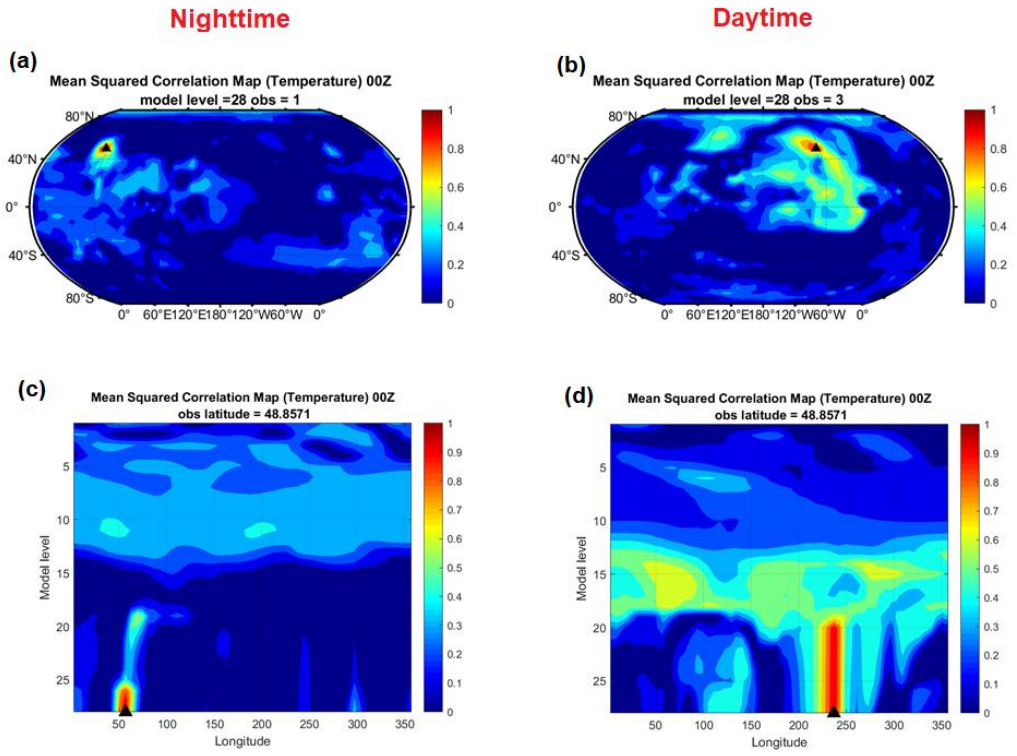

Figure 5.1 The mean squared error correlation map between the temperature at model level 28 and the observation at (a) (longitude, latitude) = (57º, 48.85ºN), and (b) (longitude, latitude) = (23º, 48.85ºN). (c)(d) is the same observation as (a)(b) but presented vertically. The observation site is plotted as a black triangle.

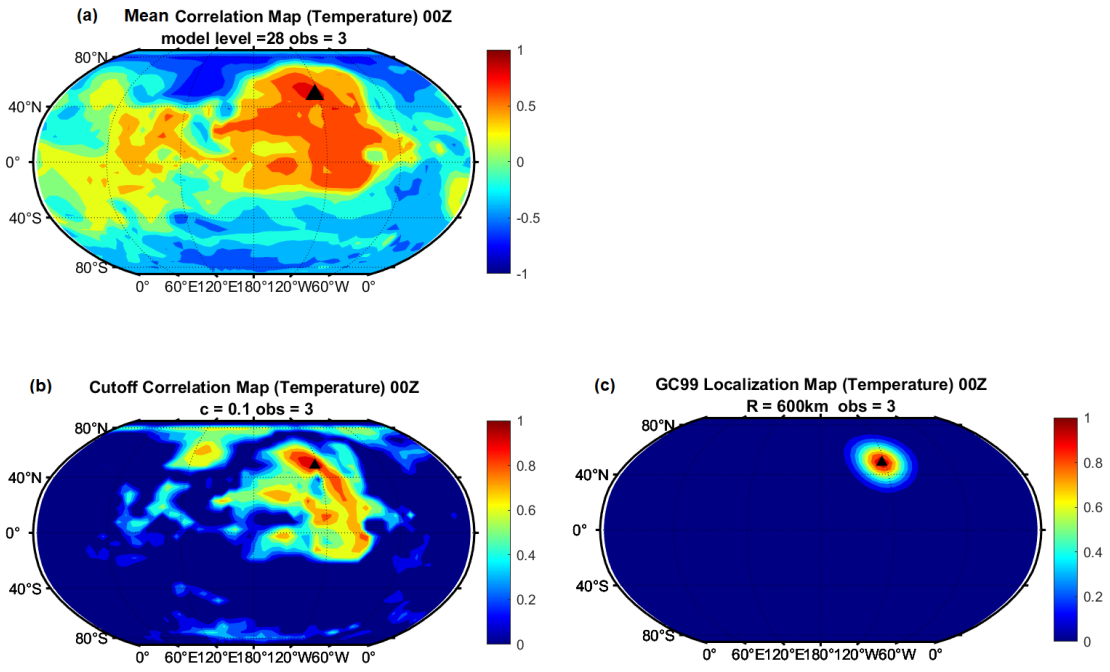

Figure 5.2 (a) The mean error correlation map between the surface temperature observation and temperature at model level 28. (b) and (c) are the localization weighting maps for YK18 and GC99, respectively. The observation site is plotted as the black triangle.

5.2.2 Applications of the oceanic EFSO

Inspired by the encouraging results demonstrated in Chapter 4, there are two potential directions for future oceanic EFSO applications. First, it is promising to extensively apply the oceanic EFSO as a data selection strategy (e.g., PQC, Hotta et al., 2017) to improve the CFS analysis and forecasts. A flowchart of the cycling PQC is shown in Figure 5.3.

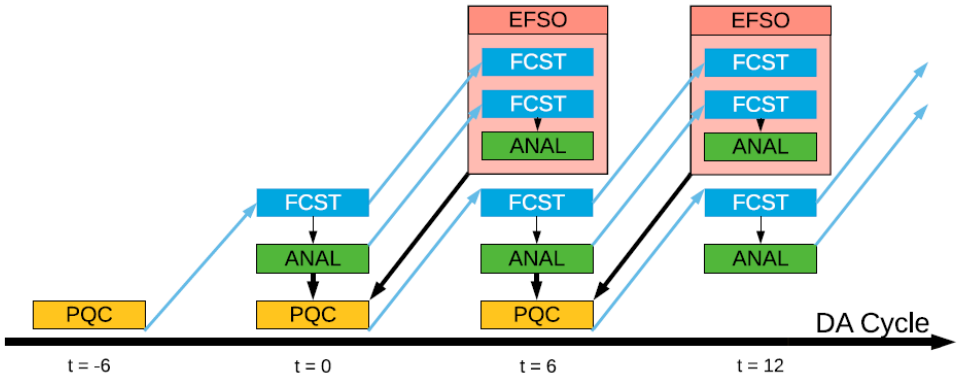

Figure 5.3 Flowchart of cycling PQC (Chen and Kalnay, 2019).

The second is the development of EFSO for the SCDA system. A possible direction is extending the density norm to an ocean energy norm. This energy norm can be derived by multiplying the height variations with the inversion of the density norm (Appendix A). Since the ocean energy norm has the same unit as the atmospheric energy norm, it would be possible to be used for evaluating the EFSOs in SCDA. In addition to SCDA application, this energy norm also can be used to evaluate the impact of sea surface height observations through the forecast error variations of the height in the proposed energy norm.

Appendix A: The ocean density norm conversion and its extension to energy norm

I. The density operator for the oceanic EFSO

The density operator is developed based on the Thermodynamic Equations of Seawater 2010 (TEOS-10) tool. The goal of this operator is to derive the in-situ ocean density from the CFS temperature and salinity fields. There are three steps included in the operator:

(1) Convert the practical salinity (PSU) into absolute salinity (g/kg)

The conversion follows the algorithm of McDougall et al., 2012. Here, the required ocean pressure field can be either obtained from the CFS outputs or estimated from the ocean depth with the computationally-efficient 75-term expression (gsw_p_from_z) in TEOS-10. This step can be achieved by the function "gsw_sa_from_sp."

(2) Convert the in-situ temperature into the conservative temperature

Next, we obtain the conservative temperature of seawater from CFS in-situ temperature. This step can be achieved by the function "gsw_ct_from_t" in TEOS-10.

(3) Derived the in-situ density (kg/m^3)

Finally, we calculate the in-situ density from the absolute salinity and conservative temperature derived from steps 1 and 2, using the computationally-efficient expression for specific volume in terms of conservation temperature, absolute salinity, and pressure (Roquet et al., 2015). This step can be achieved by the function "gsw_rho" in TEOS-10.

II. Converting the density norm to the energy norm

It is possible to extensively convert the density norm into an energy norm. That can be achieved by reversing the density term and multiplying it by a pressure term. Namely, $norm = (Pressure\ (Pa)) * \left(\frac{1}{density(kg/m^3)}\right)$, so that the unit will be $Pa * m^3/kg = J/kg$, which is the same unit as the atmosphere EFSO. That means the ocean and atmosphere can share the same universal unit, which opens the door to the "strongly coupled" EFSO for future applications.

The energy norm is not yet applied to our current configuration. Since we only assimilate T and S and our evaluation forecast length is short, the pressure forecast error difference would be extremely small, leading to an approximate zero EFSO value.

Appendix B: Handling the Chaotic-Oscillatory System with the "online" approaches of the EnOC and RIP schemes

1. Introduction

Chaotic-oscillatory modes commonly exist in phenomena forced by cyclical forcings, such as Madden-Julian oscillations (MJO), El Niño/Southern Oscillation (ENSO), and diurnal/annual carbon cycles. Oscillations are more predictable than chaotic processes due to their periodic nature and lower sensitivity to the initial condition. Inspired by this advantage of oscillation forecasting, Bach et al. (2020) proposed the Ensemble Oscillation Correction (EnOC) method that optimizes the initial ensembles and consequently enhances the predictability of chaotic-oscillatory systems (e.g., systems that consist of both chaotic and oscillation processes). The core concept of EnOC is to purposely select “good” ensembles based on their data-driven forecasts in the oscillatory mode, which results in a minimum forecast error in both the oscillation and the physical spaces.

Unlike EnOC, which improves the background ensembles by purposely selecting ensembles, the Running-In-Place method (RIP, Kalnay and Yang, 2010) refines the background by reusing observations within the DA window. RIP can effectively handle strong nonlinearity, accelerates the spin-up, and enhances the predictability of a chaotic system (Yang et al., 2012).

EnOC is an offline method in which a preceding offline run conducts the ensemble selection procedure with a known truth. However, the use of offline runs could limit the feasibility and flexibility of EnOC in practical DA applications. In this thesis, we proposed an "online" approach for EnOC, in which the ensembles are optimized at each analysis cycle based on their data-driven forecast and ***innovation*** (e.g., observation minus background). This approach allows EnOC to adapt to practical DA systems without any offline runs and can effectively refine the mean state of the background. We further investigated the characteristics and feasibility of these two types of ensemble-refining strategies on the forced Lorenz 63 model.

2. The forced Lorenz 63 (L63) model

The forced Lorenz 63 model is an advanced version of the Lorenz (1963) model with a sinusoidal forcing in the x component. This model includes both oscillatory and chaotic behaviors, making it an ideal testbed for methods employed in chaos-oscillatory systems. In this thesis, we follow the version used by Bach et al., 2020 with governing equations:

$$\frac{dx}{dt} = \sigma(y - x) + cu \tag{B.1}$$

$$\frac{dy}{dt} = x(\rho - z) - y \tag{B.2}$$

$$\frac{dz}{dt} = xy - \beta z \tag{B.3}$$

$$\frac{du}{dt} = v\,, \quad \frac{dv}{dt} = -\Omega^2 u \tag{B.4}$$

Here, equations (B.1) to (B.3) are the dynamical core of the L63 model, and equation (4) represents the periodic forcing and Ω = 0.3. The L63 parameters are the classic $(\sigma, \rho, \beta) = (10,\ 28,\ \frac{8}{3})$ and c = 5. The model is integrated by the 4th-order Runge-Kutta scheme with a time step of 0.05 units.

3. Running-In-Place (RIP) method

The main idea of RIP (Kalnay and Yang, 2010) is to improve the background ensembles by using the information from "future" observations. At each analysis cycle, RIP utilizes the no-cost smoother (Kalnay et al., 2007), which is the central core of its algorithm, iteratively to maximize the use of the observations within the assimilation window. The iterations will stop when no additional helpful information is estimated to be extracted. For a typical RIP of ith iteration, the updated analysis ensemble $\widetilde{x^a}^{i+1}{}_{n-1}$ at time t_{n-1} is given by:

$$\widetilde{\boldsymbol{x}^a}^{i+1}{}_{n-1} = \overline{\boldsymbol{x}}_{n-1}{}^{a,i} + \mathbf{X}_{n-1}^{a,i}\,\overline{\mathbf{w}_n^{a,i}}\ ,$$

where the updated perturbations are derived as:

$$\widetilde{\mathbf{X}^{a,i+1}}_{n-1} = \mathbf{X}_{n-1}^{a,i}\,\mathbf{W}_n^{a,i} + \mathbf{E}_{n-1}^{i+1}\ ,$$

Where E_{n-1}^{i+1} are small Gaussian perturbations added before integrating the smoothed analysis from t_{n-1} to t_n during the iteration. It is added to avoid the same analysis ensemble as previously derived under a linear condition (Yang et al., 2012).

The RIP method is more capable of handling the system's nonlinearity than the pure LETKF and can significantly accelerate the spin-up of LETKF.

4. Ensemble Oscillation Correction (EnOC)

The EnOC (Bach et al., 2020) is an ensemble correction method for chaos-oscillatory systems to enhance their predictability. The core concept is to project the model variables from physical space to oscillation space, where the oscillatory signals can be more easily predicted with few modes, then define optimal ensembles based on their forecasts in the oscillation space. It includes three steps:

(1) Extract the oscillation modes from the historical data using the Multi-channel singular spectrum analysis (M-SSA, Ghil et al., 2002). The trajectory of the historical data in the oscillation space is assumed to be the "truth" in the oscillation mode.

(2) Projecting the state in the physical space onto the oscillation space and running data-driven forecasting in the oscillation mode. Data-driven forecasting can be achieved by the analog method (Krishnamurthy and Sharma, 2017).

(3) Evaluate the forecast result with the truth state and select the ensembles that result in the smallest error. These selected ensembles are then used as the corrected initial condition to improve the model prediction. One can find more details in Bach et al., 2020.

The "online" approach we proposed for EnOC is to identify "good" ensembles based on their data-driven forecast in the oscillation mode and the direct measurement of the oscillatory variables, instead of evaluating with the known truth as in the original EnOC. With this new approach, the preceding offline run with known truth is

no longer needed, so the online EnOC can be implemented on DA systems to obtain optimal ensembles within a given window.

5. Experimental Setup

Experiments with the pure LETKF, the LETKF-EnOC, and the LETKF-RIP were conducted with the forced Lorenz 63 model. A random Gaussian error with a variance of 2.0 is added as the model error for each variable every 15 steps.

6. Preliminary Results

Figure B.*1* shows the mean analysis RMSE of the three methods with different ensemble sizes. In general, LETKF and LETKF-RIP, like most EnKFs, performed better with a larger ensemble size. As demonstrated by Yang et al., 2012, it is difficult for LETKF to handle the strong nonlinearity within a long DA window (e.g., 25 steps), especially when the ensemble is limited (e,g., ens=5). In contrast, RIP can significantly improve the LETKF analysis and avoid potential filter divergence. We found that LETKF-EnOC is not sensitive to the ensemble size and performs similar analyses no matter how the ensemble size changes. This characteristic of LETKF-EnOC makes it outperform the other two methods when having a small ensemble size.

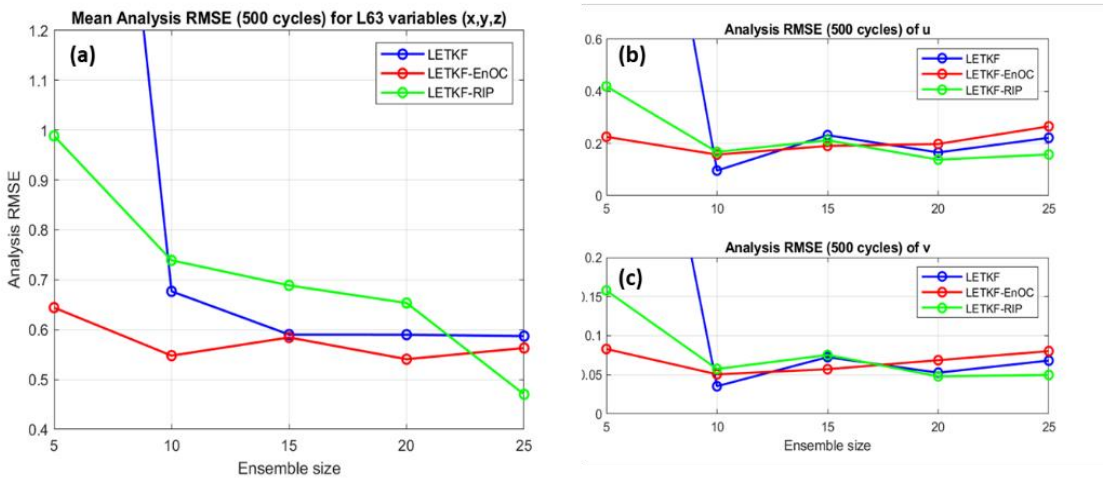

Figure B.1 Mean analysis RMSE of LETKF (blue), LETKF-EnOC (red), and LETKF-RIP (green) for (a) chaotic variables (x,y,z) and oscillatory variables (b) u and (c) v with respect to different ensemble sizes.

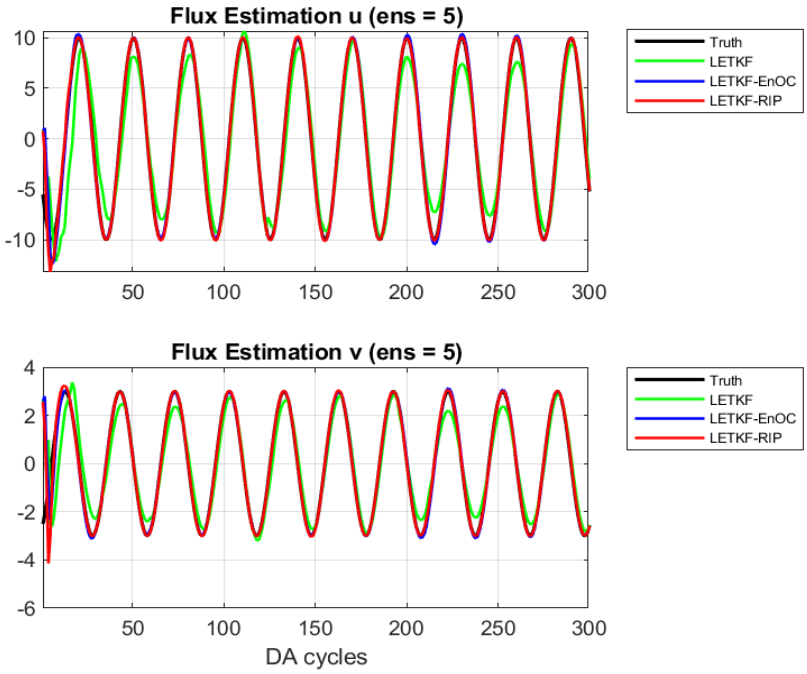

Figure B.2 The truth state (black) and the analysis of LETKF (blue), LETKF-EnOC (red), and LETKF-RIP (green) for oscillatory variables (a) u and (b) v.

We further examined the feasibility of the three methods in flux estimation. In the flux estimation problem, the flux variables (u,v) are unobserved and estimated by other prognostic variables during DA. Figure B.2 represents the flux analysis result of the three methods. We found that both the LETKF-EnOC and LETKF-RIP can capture the signal of unobserved fluxes more quickly than the pure LETKF. The LETKF-EnOC has the best analysis with very few ensembles. However, it is the most expensive method that requires nearly five times more CPU time than the pure LETKF. The primary computational cost for LETKF-EnOC is the data-driven process (e.g., analog forecasting) at each DA cycle. In contrast, the costliest part for LETKF-RIP is its iterations, particularly during the spin-up period. After the DA system convergence, the number of required iterations for RIP would accordingly drop, and the computational cost would be gradually reduced.

So, which method is better, EnOC or RIP? The answer depends on users' needs. For a low-dimensional model with an oscillatory feature, EnOC is an ideal tool since it can offer better analysis and forecasts with limited ensembles and adapt to a longer observation window. However, EnOC may not be feasible for a pure chaotic system because its algorithm mainly depends on the forecasts in oscillation mode. In contrast, RIP is more flexible to different types of systems and is effective in accelerating DA spin-up. Moreover, it can significantly improve the analysis of highly nonlinear, chaotic systems.

Bibliography

Anderson, J. L. (2007). Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, *230*(1-2), 99-111.

Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation, *Monthly weather review*, *129*(12), 2884-2903, https://doi.org/10.1175/1520-0493(2001)129%3C2884:AEAKFF%3E2.0.CO;2.

Anderson, J., & Lei, L. (2013). Empirical Localization of Observation Impact in Ensemble Kalman Filters, *Monthly Weather Review*, *141*(11), 4140-4153. doi: https://doi.org/10.1175/MWR-D-12-00330.1

Bach, E., Mote, S., Krishnamurthy, V., Sharma, A. S., Ghil, M., & Kalnay, E. (2021). Ensemble Oscillation Correction (EnOC): Leveraging oscillatory modes to improve forecasts of chaotic systems. *Journal of Climate, 34*, 5673–5686.

Behringer, D. W. (2007). 3.3 the global ocean data assimilation system (GODAS) at NCEP. In *Proceedings of the 11th symposium on integrated observing and assimilation systems for the atmosphere, oceans, and land surface.*

Bender, M. A., Ginis, I., & Kurihara, Y. (1993). Numerical simulations of tropical cyclone-ocean interaction with a high-resolution coupled model. *J. Geophys. Res., 98*, 23245.

Bishop, C., & Hodyss, D. (2009). Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus A: Dynamic Meteorology and Oceanography*, *61*(1), 84-96.

Boyer, T.P., J.I. Antonov, O.K. Baranova, C. Coleman, H.E. Garcia, A. Grodsky, D.R. Johnson, R.A. Locarnini, A.V. Mishonov, T.D. O'Brien, C.R. Paver, J.R. Reagan, D. Seidov, I.V. Smolyar, and M.M. Zweng, (2013), World Ocean Database 2013., S. Levitus, Ed., A. Mishonov, Technical Ed.; NOAA Atlas NESDIS 72, 209 pp., doi:10.7289/V5NZ85MT

Brasnett, B. (2008). The impact of satellite retrievals in a global sea-surface-temperature analysis. *Quarterly Journal of the Royal Meteorological Society*, *134*(636), 1745-1760.

Bueti, M. R., I. Ginis, L. M. Rothstein, and S. M. Griffies. (2014) Tropical cyclone–induced thermocline warming and its regional and global impacts. *J. Climate*, **27**, 6978–6999, doi:10.1175/JCLI-D-14-00152.1

Chang, C. C., & Kalnay, E. (2022). Applying prior correlations for ensemble-based spatial localization. *Nonlinear Processes in Geophysics*, *29*(3), 317-327.

Chaudhari, H. S., Pokhrel, S., Saha, S. K., Dhakate, A., & Hazra, A. (2015, August). Improved depiction of Indian summer monsoon in latest high resolution NCEP climate forecast system reanalysis. *Int. J. Climatol.*, *35*, 3102–3119.

Chaudhari, H. S., Pokhrel, S., Saha, S. K., Dhakate, A., Yadav, R. K., Salunke, K., . . . Rao, S. A. (2013, April). Model biases in long coupled runs of NCEP CFS in the context of Indian summer monsoon. *Int. J. Climatol.*, *33*, 1057–1069.

Chawang, N., & Kutty, G. (2022, December). Ensemble-based forecast sensitivity approach to estimate the impact of satellite-derived atmospheric motion vectors in a limited area model. *J. Earth Syst. Sci.*, *131*.

Chen, Dake, Stephen E. Zebiak, Antonio J. Busalacchi, and Mark A. Cane. (1995) "An improved procedure for EI Nino forecasting: Implications for predictability." *Science* 269, no. 5231, 1699-1702.

Chen, T. C. (2018). *Applications of ensemble forecast sensitivity to observations for improving numerical weather prediction (Doctoral dissertation).* University of Maryland, College Park.

Chen, T. C., & Kalnay, E. (2019). Proactive quality control: Observing system simulation experiments with the Lorenz'96 model. *Monthly Weather Review*, *147*, 53–67.

Chen, T. C., Kalnay, E., & Hotta, D. (2017). Use of EFSO for online data assimilation quality monitoring and proactive quality control. In *CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling.*

Chen, T.-C., & Kalnay, E. (2020, September). Proactive quality control: Observing system experiments using the NCEP global forecast system. *Mon. Weather Rev.*, *148*, 3911–3931.

Chen, X., & Zhang, F. (2019, November). Development of a convection-permitting air-sea-coupled ensemble data assimilation system for tropical cyclone prediction. *J. Adv. Model. Earth Syst.*, *11*, 3474–3496.

Cohn, S. E., Da Silva, A., Guo, J., Sienkiewicz, M., & Lamich, D. (1998). Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Monthly Weather Review*, *126*(11), 2913-2926.

Conkright, M. E., Levitus, S., O'brien, T. D., Boyer, T. P., Stephens, C., Johnson, D. R., & Forgy. (1999). *World Ocean Database 1998: CD-ROM data set documentation.*

Counillon, F., Keenlyside, N., Bethke, I., Wang, Y., Billeau, S., Shen, M. L., & Bentsen, M. (2016). Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian Climate Prediction Model. *Tellus A: Dynamic Meteorology and Oceanography*, *68*(1), 32437.

Derber, J., & Rosati, A. (1989). A global oceanic data assimilation system. *Journal of physical oceanography*, *19*(9), 1333-1347.

Ehrendorfer, M., Errico, R. M., & Raeder, K. D. (1999, June). Singular-vector perturbation growth in a primitive equation model with moist physics. *J. Atmos. Sci., 56*, 1627–1648.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., . . . Tarpley, J. D. (2003, November). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res., 108*.

Emanuel, K. (2001), Contribution of tropical cyclones to meridional heat transport by the oceans, *J. Geophys. Res.*, 106 (D14), 14771– 14781, doi:10.1029/2000JD900641.

Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, *53*(4), 343-367.

Frolov, S., Bishop, C. H., Holt, T., Cummings, J., & Kuhl, D. (2016). Facilitating strongly coupled ocean–atmosphere data assimilation with an interface solver. *Monthly Weather Review*, *144*(1), 3-20.

Gade, S. V., Sreenivas, P., Rao, S. A., Srivastava, A., & Pradhan, M. (2022, August). Impact of the ensemble Kalman filter based coupled data assimilation system on seasonal prediction of Indian summer monsoon rainfall. *Geophys. Res. Lett., 49*.

Gaspari, G., & Cohn, S. E. (1999, January). Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc., 125*, 723–757.

Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., . . . Yiou, P. (2002, February). Advanced spectral methods for climatic time series. *Rev. Geophys., 40*, 3–1–3–41.

Griffies, S. M., Harrison, M. J., Pacanowski, R. C., & Rosati, A. (2004). A technical guide to MOM4. *GFDL Ocean Group Tech. Rep*, *5*(5), 371.

Gillespie, H. E., Greybush, S. J., & Wilson, R. J. (2020). An investigation of the encirclement of Mars by dust in the 2018 global dust storm using EMARS. *Journal of Geophysical Research: Planets*, *125*, e2019JE006106. https://doi.org/10.1029/2019JE006106

Greybush, S. J. (2011). *Mars weather and predictability: Modeling and ensemble data assimilation of spacecraft observations*. University of Maryland, College Park.

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., & Hunt, B. R. (2011). Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review*, *139*(2), 511-522.

Greybush, S. J., Kalnay, E., Ide, K., Miyoshi, T., Mcconnochie, T., Hoffman, M. J., . . . Wilson, R. J. (2012). Ensemble Kalman Filter Data Assimilation of Thermal Emission Spectrometer (TES) Profiles into a Mars Global Circulation Model. *J. Geophys. Res. Planets, 117*.

Greybush, S. J., Kalnay, E., Wilson, R. J., Hoffman, R. N., Nehrkorn, T., Leidner, M., . . . Miyoshi, T. (2019, November). The Ensemble Mars Atmosphere Reanalysis System (EMARS) version 1.0. *Geosci. Data J., 6*, 137–150.

Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review, 129*(11), 2776-2790.

Harris, A. and Maturi, E. (2012), Geo-Polar Blended SST Algorithm Theoretical Basis Document, Version 2.1. NOAA NESDIS Center for Satellite Applications and Research.

Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review, 126*(3), 796-811.

Houtekamer, P. L., and Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation, *Monthly Weather Review*, *129*(1), 123-137, https://doi.org/10.1175/1520-0493(2001)129%3C0123:ASEKFF%3E2.0.CO;2.

Houtekamer, P. L., & Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *144*(12), 4489-4532.

Han, G., Wu, X., Zhang, S., Liu, Z., & Li, W. (2013, December). Error covariance estimation for coupled data assimilation using a Lorenz atmosphere and a simple pycnocline ocean model. *J. Clim., 26*, 10218–10231.

Hoke, J. E., & Anthes, R. A. (1976, December). The initialization of numerical models by a dynamic-initialization technique. *Mon. Weather Rev., 104*, 1551–1556.

Hotta, D., Chen, T.-C., Kalnay, E., Ota, Y., & Miyoshi, T. (2017, August). Proactive QC: A fully flow-dependent quality control scheme based on EFSO. *Mon. Weather Rev., 145*, 3331–3354.

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007, June). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D, 230*, 112–126.

Jin, E. K., & Kinter, J. L. (2009, April). Characteristics of tropical Pacific SST predictability in coupled GCM forecasts using the NCEP CFS. *Clim. Dyn., 32*, 675–691.

Kalnay, E., & Yang, S.-C. (2010, July). Accelerating the spin-up of ensemble Kalman filtering. *Q. J. R. Meteorol. Soc., 136*, 1644–1651.

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., & Ballabrera-Poy, J. (2007, October). 4-D-Var or ensemble Kalman filter? *Tellus A*.

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., & Ballabrera-Poy, J. (2007, January). Response to the discussion on ``4-D-Var or EnKF?" by Nils Gustafsson. *Tellus A, 59*, 778.

Kalnay, E., Ota, Y., Miyoshi, T., & Liu, J. (2012, December). A simpler formulation of forecast sensitivity to observations: application to ensemble Kalman filters. *Tellus A, 64*, 18462.

Kalnay, E., Corazza, M., & Cai, M. (2002). Are bred vectors the same as Lyapunov vectors?. In *EGS general assembly conference abstracts* (p. 6820).

Kbrr, H. P., Pentakota, S., Anguluri, S. R., George, G., Salunke, K., Krishna, P. M., . . . Rao, D. N. (2020, September). Impact of horizontal resolution on sea surface temperature bias and air–sea interactions over the tropical Indian Ocean in CFSv2 coupled model. *Int. J. Climatol., 40*, 4903–4921.

Keenlyside, N. S., Latif, M., Jungclaus, J., Kornblueh, L., & Roeckner, E. (2008, May). Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature, 453*, 84–88.

Kepert, J. D. (2009). Covariance localisation and balance in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *135*(642), 1157-1176.

Kleist, D. T., Parrish, D. F., Derber, J. C., Treadon, R., Wu, W.-S., & Lord, S. (2009, December). Introduction of the GSI into the NCEP Global Data Assimilation System. *Weather Forecast., 24*, 1691–1705.

Kondo, K., & Miyoshi, T. (2016). Impact of removing covariance localization in an ensemble Kalman filter: Experiments with 10 240 members using an intermediate AGCM. *Monthly Weather Review*, *144*(12), 4849-4865.

Kotsuki, S., Kurosawa, K., & Miyoshi, T. (2019), On the properties of ensemble forecast sensitivity to observations. *Q.J.R. Meteorol. Soc.*, *145*(722), 1897-1914. https://doi.org/10.1002/qj.3534

Krishnamurthy, V., & Sharma, A. S. (2017, August). Predictability at intraseasonal time scale. *Geophys. Res. Lett., 44*, 8530–8537.

Kumar, A., Chen, M., Zhang, L., Wang, W., Xue, Y., Wen, C., ... & Huang, B. (2012). An analysis of the nonstationarity in the bias of sea surface temperature forecasts for the NCEP Climate Forecast System (CFS) version 2. *Monthly weather review*, *140*(9), 3003-3016.

Langland, R. H., & Baker, N. L. (2004), Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus A: Dynamic Meteorology and Oceanography*, *56*(3), 189-201. doi:10.3402/tellusa.v56i3.14413

Li, H., Liu, J. and Kalnay, E. (2010), Correction of 'Estimating observation impact without adjoint model in an ensemble Kalman filter'. Q.J.R. Meteorol. Soc., 136: 1652-1654. https://doi.org/10.1002/qj.658

Lien, G. Y., Hotta, D., Kalnay, E., Miyoshi, T., & Chen, T. C. (2018), Accelerating assimilation development for new observing systems using EFSO. *Nonlinear*

Processes in Geophysics, *25*(1), 129-143. https://doi.org/10.5194/npg-25-129-2018

Liu, J. and Kalnay, E. (2008), Estimating observation impact without adjoint model in an ensemble Kalman filter. *Q.J.R. Meteorol. Soc.*, 134: 1327-1335. https://doi.org/10.1002/qj.280

Lorenc, A.C. & Marriott, R.T. (2014), Forecast sensitivity to observations in the Met Office Global numerical weather prediction system. *Q.J.R. Meteorol. Soc.*, 140: 209-224. https://doi.org/10.1002/qj.2122

Lorenc, A. C. (2003). The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *129*(595), 3183-3203.

Lorenz, S., Grieger, B., Helbig, P., & Herterich, K. (1996). Investigating the sensitivity of the atmospheric general circulation model ECHAM 3 to paleoclimatic boundary conditions. *Geologische Rundschau*, *85*(3), 513-524.

Lorenz, E. N., & Emanuel, K. A. (1998, February). Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, *55*, 399–414.

McDougall, T.J. & P.M. Barker (2011), Getting started with TEOS-10 and the Gibbs Seawater (GSW) Oceanographic Toolbox, 28pp., SCOR/IAPSO WG127, ISBN 978-0-646-55621-5.

Miyoshi, T., Kondo, K., & Imamura, T. (2014, July). The 10,240-member ensemble Kalman filtering with an intermediate AGCM. *Geophys. Res. Lett.*, *41*, 5264–5271.

Moore, A. M., & Kleeman, R. (1996, July). The dynamics of error growth and predictability in a coupled model of ENSO. *Q. J. R. Meteorol. Soc.*, *122*, 1405–1446.

Mulholland, D. P., Laloyaux, P., Haines, K., & Balmaseda, M. A. (2015). Origin and impact of initialization shocks in coupled atmosphere–ocean forecasts. *Mon. Weather Rev.*, *143*, 4631–4644.

Ota, Y., Derber, J. C., Kalnay, E., & Miyoshi, T. (2013, December). Ensemble-based observation impact estimates using the NCEP GFS. *Tellus A*, *65*, 20038.

Penny, S. G., & Hamill, T. M. (2017, July). Coupled data assimilation for integrated earth system analysis and prediction. *Bull. Am. Meteorol. Soc.*, *98*, ES169–ES172.

Pitman, E. J. (1938, February). Significance tests which may be applied to samples from any populations iii. The analysis of variance test. *Biometrika*, *29*, 322–335.

Rai, A., & Saha, S. K. (2018, January). Evaluation of energy fluxes in the NCEP climate forecast system version 2.0 (CFSv2). *Clim. Dyn.*, *50*, 101–114.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002, July). An improved in situ and satellite SST analysis for climate. *J. Clim., 15*, 1609–1625.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007, November). Daily high-resolution-blended analyses for sea surface temperature. *J. Clim., 20*, 5473–5496.

Roquet, F., Madec, G., McDougall, T. J., & Barker, P. M. (2015). Accurate polynomial expressions for the density and specific volume of seawater using the TEOS-10 standard. *Ocean Modelling*, *90*, 29-43.

Rosati, A., Miyakoda, K., & Gudgel, R. (1997, May). The impact of ocean initial conditions on ENSO forecasting with a coupled model. *Mon. Weather Rev., 125*, 754–772.

Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., ... & Goldberg, M. (2010). The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, *91*(8), 1015-1058.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y., Chuang, H., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., & Becker, E. (2014). The NCEP Climate Forecast System Version 2, *Journal of Climate*, *27*(6), 2185-2208. doi: https://doi.org/10.1175/JCLI-D-12-00823.1

Sivareddy, S., Paul, A., Sluka, T., Ravichandran, M., & Kalnay, E. (2017), The pre-Argo ocean reanalyses may be seriously affected by the spatial coverage of moored buoys. *Scientific Reports*, *7*(1), 1-8. https://doi.org/10.1038/srep46685

Sluka, T. C. (2018). *Strongly coupled ocean-atmosphere data assimilation with the local ensemble transform Kalman filter (Doctoral dissertation)*. University of Maryland, College Park.

Sluka, T. C., Penny, S. G., Kalnay, E., & Miyoshi, T. (2016, January). Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation. *Geophys. Res. Lett., 43*, 752–759.

Sriver, R. L., Goes, M., Mann, M. E., & Keller, K. (2010), Climate response to tropical cyclone-induced ocean mixing in an Earth system model of intermediate complexity, *J. Geophys. Res.*, 115, C10042, doi:10.1029/2010JC006106.

Toth, Z., & Kalnay, E. (1993). Ensemble forecasting at NMC: The generation of perturbations. *Bulletin of the american meteorological society*, *74*(12), 2317-2330.

Toth, Z., & Kalnay, E. (1997). Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review*, *125*(12), 3297-3319.

Wang, B. (2005). Fundamental challenge in simulation and prediction of summer monsoon rainfall. *Geophys. Res. Lett., 32*.

Wang, W., Chen, M., & Kumar, A. (2010, June). An assessment of the CFS real-time seasonal forecasts. *Weather Forecast., 25*, 950–969.

Wang, C., Dong, S., & Munoz, E. (2010), Seawater density variations in the North Atlantic and the Atlantic meridional overturning circulation. *Climate dynamics*, *34*(7), 953-968. https://doi.org/10.1007/s00382-009-0560-5

Winton, M. (2000), A reformulated three-layer sea ice model. *Journal of atmospheric and oceanic technology*, *17*(4), 525-531. https://doi.org/10.1175/1520-0426(2000)017<0525:ARTLSI>2.0.CO;2

Whitaker, J. S., & Hamill, T. M. (2012, September). Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Weather Rev., 140*, 3078–3089.

Woolnough, S. J., Vitart, F., & Balmaseda, M. A. (2007, January). The role of the ocean in the Madden–Julian Oscillation: Implications for MJO prediction. *Q. J. R. Meteorol. Soc., 133*, 117–128.

Yamazaki, A., Miyoshi, T., Inoue, J., Enomoto, T., & Komori, N. (2021, April). EFSO at different geographical locations verified with observing-system experiments. *Weather Forecast.*

Yang, S.-C., Kalnay, E., & Hunt, B. (2012, August). Handling nonlinearity in an ensemble Kalman filter: Experiments with the three-variable Lorenz model. *Mon. Weather Rev., 140*, 2628–2646.

Ying, Y., Zhang, F., and Anderson, J. L. (2018). On the selection of localization radius in ensemble filtering for multiscale quasigeostrophic dynamics, *Monthly Weather Review*, *146*(2), 543-560, https://doi.org/10.1175/MWR-D-17-0336.1.

Yoshida, T. (2019). *Covariance Localization in Strongly Coupled Data Assimilation (Doctoral dissertation, University of Maryland, College Park).*

Yoshida, T., & Kalnay, E. (2018). Correlation-cutoff method for covariance localization in strongly coupled data assimilation. *Monthly Weather Review*, *146*(9), 2881-2889.

Zhang, X., Wang, H., Huang, X. Y., Gao, F., & Jacobs, N. A. (2015), Using adjoint-based forecast sensitivity method to evaluate TAMDAR data impacts on regional forecasts. *Advances in Meteorology*, 2015. https://doi.org/10.1155/2015/427616

Zhang, S., Harrison, M. J., Rosati, A., & Wittenberg, A. (2007). System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather Review*, *135*(10), 3541-3564.

Zhang, R.-H., Yu, Y., Song, Z., Ren, H.-L., Tang, Y., Qiao, F., . . . Wang, L. (2020, July). A review of progress in coupled ocean-atmosphere model developments for ENSO studies in China. *J. Oceanol. Limnol., 38*, 930–961.

Zhang, S. (2011, December). A study of impacts of coupled model initial shocks and state–parameter optimization on climate predictions using a simple pycnocline prediction model. *J. Clim., 24*, 6210–6226.

Zhang, S., Liu, Z., Zhang, X., Wu, X., Han, G., Zhao, Y., . . . Deng, X. (2020, June). Coupled data assimilation and parameter estimation in coupled ocean–atmosphere models: a review. *Clim. Dyn., 54*, 5127–5144.

Zhu, Y., & Gelaro, R. (2008), Observation sensitivity calculations using the adjoint of the Gridpoint Statistical Interpolation (GSI) analysis system. *Monthly Weather Review*, *136*(1), 335-351. https://doi.org/10.1175/MWR3525.1