# ABSTRACT

Title of dissertation: DEVELOPMENTS IN LAGRANGIAN DATA ASSIMILATION AND COUPLED DATA ASSIMILATION TO SUPPORT EARTH SYSTEM MODEL INITIALIZATION

Luyu SUN
Doctor of Philosophy, 2019

Dissertation directed by: Professor James Carton and Stephen Penny
Department of Atmosphere and Oceanic Science

The air-sea interface is one of the most physically active interfaces of the Earth's environments and significantly impacts the dynamics in both the atmosphere and ocean. In this doctoral dissertation, developments are made to two types of Data Assimilation (DA) applied to this interface: Lagrangian Data Assimilation (LaDA) and Coupled Data Assimilation (CDA).

LaDA is a DA method that specifically assimilates position information measured from Lagrangian instruments such as Argo floats and surface drifters. To make a better use of this Lagrangian information, an augmented-state LaDA method is proposed using Local Ensemble Transform Kalman Filter (LETKF), which is intended to update the ocean state (T/S/U/V) at both the surface and at depth by directly assimilating the drifter locations. The algorithm is first tested using "identical twin" Observing System Simulation Experiments (OSSEs) in a simple double gyre configuration with the Geophysical Fluid Dynamics Laboratory (GFDL) Mod-

ular Ocean Model version 4.1 (MOM4p1). Results from these experiments show that with a proper choice of localization radius, the estimation of the state is improved not only at the surface, but throughout the upper 1000m. The impact of localization radius and model error in estimating accuracy of both fluid and drifter states are investigated.

Next, the algorithm is applied to a realistic eddy-resolving model of the Gulf of Mexico (GoM) using Modular Ocean Model version 6 (MOM6) numerics, which is related to the 1/4-degree resolution configuration in transition to operational use at NOAA/NCEP. Atmospheric forcing is first used to produce the nature run simulation with forcing ensembles created using the spread provided by the 20 Century Reanalysis version 3 (20CRv3). In order to assist the examination on the proposed LaDA algorithm, an updated online drifter module adapted to MOM6 is developed, which resolves software issues present in the older MOM4p1 and MOM5 versions of MOM. In addition, new attributions are added, such as: the output of the intermediate trajectories and the interpolated variables: temperature, salinity, and velocity. The twin experiments with the GoM also show that the proposed algorithm provides positive impacts on estimating the ocean state variables when assimilating the drifter position together with surface temperature and salinity.

Lastly, an investigation of CDA explores the influence of different couplings on improving the simultaneous estimation of atmosphere and ocean state variables. Synchronization theory of the drive-response system is applied together with the determination of Lyapunov Exponents (LEs) as an indication of the error convergence within the system. A demonstration is presented using the Ensemble Trans-

form Kalman Filter on the simplified Modular Arbitrary-Order Ocean-Atmosphere Model, a three-layer truncated quasi-geostrophic model. Results show that strongly coupled data assimilation is robust in producing more accurate state estimates and forecasts than traditional approaches of data assimilation.

# Developments in Lagrangian Data Assimilation and Coupled Data Assimilation to Support Earth System Model Initialization

by

Luyu SUN

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
2019

Advisory Committee:
Professor James Carton, Chair/Advisor
Professor Stephen G. Penny, Co-Advisor
Professor Eugenia Kalnay
Professor Brian Hunt
Professor Eric Slud

# Acknowledgments

I acknowledge the support from the National Oceanographic and Atmospheric Administration (NOAA) National Environmental Satellite, Data, and Information Service (NESDIS) awarded via the National Aeronautics and Space Administration (NASA) Research Opportunities in Earth and Space Science (ROSES) program. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center and the University of Maryland High-Performance Computing (HPC) Program.

# Table of Contents

# List of Tables

# List of Figures

# List of Abbreviations

| | |
|---|---|
| 20CR | 20th Century Reanalysis |
| 20CRv3 | 20th Century Reanalysis version 3 |
| 3D-Var | Three-dimensional Variational Assimilation |
| 4D-Var | Four-dimensional Variational Assimilation |
| ADT | Absolute Dynamic Height |
| ALE | Arbitrary Lagrangian-Eulerican |
| AVISO | the Archiving, Validation, and Interpretation of Satellite Oceanographic data |
| DA | Data Assimilation |
| ECMWF | European Centre for Medium-Range Weather Forecasts |
| GDP | Global Drifter Program |
| GEBCO | General Bathymetric Chart of the Oceans |
| GLAD | Grand LAgrangian Deployment |
| GoM | Gulf of Mexico |
| HYCOM | Hybrid Coordinates Ocean Model |
| LaDA | Lagrangrian Data Assimilation |
| LCE | Loop Current Eddy |
| LETKF | Local Ensemble Transform Kalman Filter |
| LE | Lyapunov Exponent |
| MAOOAM | Modular Arbitrary-Order Ocean-Atmosphere Model |
| MOM | Modular Ocean Model |
| MOM1 | Modular Ocean Model Version 1 |
| MOM4p1 | Modular Ocean Model Version 4.1 |
| MOM5 | Modular Ocean Model Version 5 |
| MOM6 | Modular Ocean Model Version 6 |
| NCOM | Navy Coastal Ocean Model |
| NWP | Numerical Weather Prediction |
| OSCAR | Ocean Surface Current Analyses Realtime |
| pdf | probability distribution function |
| RMSE | Root Mean Square Error |
| SODA | Simple Ocean Data Assimilation |
| SS | Skill Score |
| SSH | Sea Surface Height |
| SSS | Sea Surface Salinity |
| SST | Sea Surface Temperature |
| TLM | Tangent Linear Model |
| TS | Temperature and Salinity |
| WOA05 | World Ocean Atlas 2005 |

# Chapter 1: Introduction

## 1.1 Surface Ocean Currents

Surface ocean currents are primarily driven by atmosphere surface winds that influence the uppermost ocean "mixed layer", which can extend typically to depths of about 100-300m depth (Apel, 1987). The scales of ocean currents at the sea surface span from the basin-wide circulation, to mesoscale eddies and fast narrow currents, to sub-mesoscale features, to smaller scales such as turbulence. The differences in these scales result from a combination of different forcing and different underlying geophysics intertwined with the motion of the deep ocean which is difficult to measure and monitor. A detailed knowledge of the surface flow field is expected to be of high value for the prediction of Earth system model forecasts at a variety of timescales. Nevertheless, fundamental studies about the structure and impact of surface currents at different scales are currently insufficient and many scientific problems still remain unsolved in accurately estimating the ocean surface currents due to the lack of the measurements or mature data assimilation strategies. A recent European Centre for Medium-Range Weather Forecasts (ECMWF) Newsletter article (Mogensen et al., 2018) showed that sea-surface temperature (SST) coupling is important for the prediction of not only the large-scale tropical cyclones but also

the 2-meter temperature as well, while follow-on work is still necessary to improve the understanding of the processes involved in interactions between the atmosphere and the ocean (including sea ice) in accordance with ECMWFs Strategy.

In order to have a better understanding and then determine a more accurate prediction of the surface currents in different temporal and spatial scales, we need to improve in three main perspectives: (1) launching a large number of measurement devices to monitor the surface currents; (2) constructing realistic global and local circulation models to be capable of detecting not only the large-scale of circulation but also sub-mesoscales flows (100 m to tens of kilometers, hours to days); and (3) proposing robust data assimilation methods to provide accurate prediction by combining all the relevant ocean measurements with a high-dimensional numerical forecast model.

## 1.2 Surface Drifters

In general, a direct measurement of the instantaneous surface current is notoriously difficult. The current state-of-the-art approach is to use remotely sensed satellite altimetry and ocean vector winds to derive gridded sea level anomalies and therefore determined the flow field of the sea surface. For example, the Ocean Surface Current Analyses Realtime (OSCAR) product (Bonjean and Lagerloef, 2002) determines the near-surface ocean by using sea surface height, surface vector wind and sea surface temperature through quasi-linear and steady flow momentum equations. The resulting velocity map is of $1/3^\circ$ resolution and updated every 5 days.

More advanced approaches including the use of synthetic aperture radar (SAR) for high-resolution imaging of surface (Romeiser et al., 2009) are of growing interest.

Surface drifters/surface drifting floats, in recent years, have a strongly influential usage in monitoring the ocean currents by tracking a preponderance of flow-following instruments, which are of low cost and transmit nearly instantaneous data with high frequency. The trajectories of drifters transported by ocean currents provide information about the underlying dynamics and physical properties of the ocean. Poje et al. (2014) pointed out that these Lagrangian experiments are the most feasible means of simultaneously measuring the submesoscale of ocean surface structure. The Global Drifter Program (GDP) sponsored by the National Oceanic and Atmospheric Administration (NOAA) is a part in situ, part remote-sensing program (Figure 1.1) that provides the capacity to produce a rough estimate of near surface currents by tracking a large number of surface drifters drogued at 15-m depth throughout the global ocean Lumpkin et al. (2007).

The Grand Lagrangian Deployment (GLAD) is another surface drifter program, conducted by the Consortium for Advanced Research on Transport of Hydrocarbon in the Environment (CARTHE), aimed at in studying the Loop Current Eddy (LCE) in the Gulf of Mexico (GoM). Nearly 300 surface drifters were deployed in late July 2012 and drogued at 1-m depth. Armed with GPS transmitter, the GLAD drifters are capable of updating their own instantaneous locations every 15 mins. Historical data of GLAD drifter trajectories are collected from July 20, 2012 through October 22, 2012. Prior to the GLAD drifter deployments, the loop current has already shed an eddy located near the center of GoM where the majority

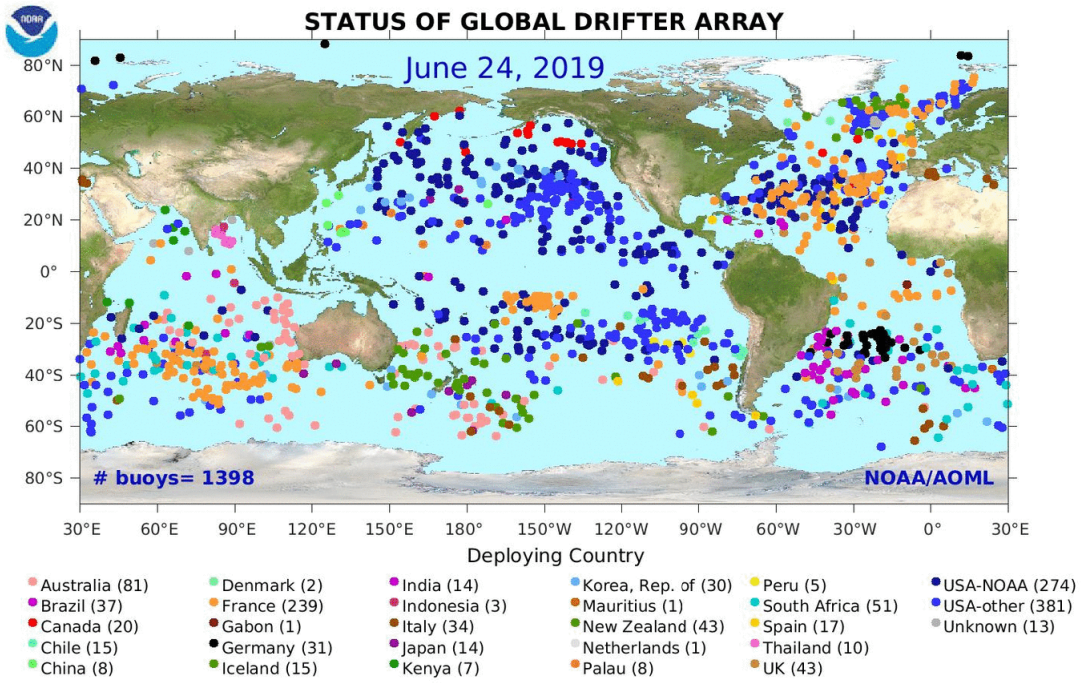


Figure 1.1: GPS locations of the Global Drifter Program (GDP) on June 24, 2019. Colored marks represent different deploying country. This is generated by `www.aoml.noaa.gov/phod/gdp/index.php`.

of drifters were spotted (Coelho et al., 2015).

Ocean data assimilation systems have typically used temperature and salinity measurements from surface drifters indirectly via calibration of satellite-based SST and sea surface salinity (SSS) retrieval products (Bitterman and Hansen, 1993; Melnichenko et al., 2016), though they are sometimes assimilated directly (Dong et al., 2017). The position data of surface drifters are rarely used in any way, though there have been some pioneering investigations in reconstructing the mesoscale Eulerian velocity fields based on observations of the Lagrangian positions (Nilsson et al., 2012; Taillandier et al., 2006). Therefore, how to fully utilize the information pro-

vided by the drifter positions and improve more information in monitoring of the high-frequency motions in the ocean becomes an interesting research topic.

## 1.3 Studies of Data Assimilation

Despite modern advances in modeling the ocean and atmosphere, there are still unresolved processes, inaccurate parameterizations, and inevitable errors in boundary and initial conditions. If these errors are not corrected over time during long model integrations, the aggregated errors will result in a divergence from a reliable forecasting trajectory.

Data assimilation (DA) is a mathematical discipline applied in Numerical Weather Prediction (NWP) to efficiently solve this objective, which combines observations with "prior knowledge" (e.g. a skillful forecast generated from a numerical model) to provide an improved estimate of the true state of the physical system and the corresponding uncertainty of that estimate. In this dissertation, we focus on the following two perspectives of DA: (1) utilizing Lagrangian observations (e.g. drifters, subsurface floats etc.) to improve the synoptic states in a General Circulation Model (GCM); (2) studying the impact of various degrees of coupling in simultaneously assimilating both atmosphere and ocean observations.

### 1.3.1 Lagrangian Data Assimilation (LaDA)

LaDA is a DA method that specifically assimilates the position information measured from Lagrangian instruments such as surface drifters and deep Argo floats

that move along with the flow of the ocean currents. There are three main challenges for all LaDA methods: (1) the indirect representation of Lagrangian observations in terms of general forecast model variables; (2) the inherently nonlinear evolution of drifter trajectories; and (3) the efficiency of application to high-dimensional models.

One DA approach uses a mathematical and statistical approximation to transform from a sequence of geophysical locations to an instantaneous Eulerian velocity and assimilates these proxy measurements to update the prognostic model state variables (Molcard et al., 2003; Nilsson et al., 2012; Taillandier et al., 2006). While this approach is easy to extend to the realistic model with a completion of an advanced observation operator, the approximation of the velocity measurements in general simplifies the motion and ignores the local instabilities in the underlying geophysics.

An alternative approach (Ide et al., 2002; Kuznetsov et al., 2003) extends the original fluid model state to include the drifter positions in an augmented-state vector $\mathbf{x} = (\mathbf{x}_F, \mathbf{x}_D)^T$ and evolves this extended dynamical system. Given a sufficiently short analysis cycle, this approach is able to update the model states with high accuracy by utilizing the quasi-linear error growth in the simulated drifter position and the correlations of errors between the ocean tracer and velocity fields. This Lagrangian approach may be more effective than the assimilation of transformed Eulerian velocity proxy observations in strongly nonlinear dynamic models (Kuznetsov et al., 2003; Vernieres et al., 2011), though it requires an effective drifter simulator inside the corresponding forecast model.

Under the framework of the augmented-state approach, Apte et al. (2008)

showed that LaDA can be applied with the ensemble Kalman filter (EnKF) (Evensen, 2004; Ide et al., 2002) with high-dimensional model states, though it fails to capture dynamics with the high-order nonlinearity because of the resulting non-Gaussian distribution of error from the nonlinear trajectory model. An alternative assimilation method that could be used in the presence of this nonlinearity is the particle filter (PF) (Salman et al., 2008; Santitissadeekorn et al., 2014). However, the canonical PF does not scale well to high dimensions (Snyder et al., 2008). While new types of PF have recently been proposed for high dimensional geophysical systems (Penny and Miyoshi, 2016; Poterjoy, 2016; van Leeuwen, 2003), none have yet been shown viable for operational applications. A hybrid method (Slivinski et al., 2015) has been proposed to address the issues due to nonlinearity and high-dimensionality at the same time by updating the flow states using an EnKF while updating the drifter states using a PF. The algorithm of Slivinski et al. (2015) was examined using a framework of perfect twin experiments with the linear shallow water equations. Results indicated that compared to the EnKF, this hybrid approach produced improved estimates of the Bayesian posterior and better tracked the true state.

### 1.3.2 Coupled Data Assimilation (CDA)

The ultimate goal of studying the ocean surface is to have a better understanding on both of the ocean and atmosphere dynamics at different depths and heights respectively. Assimilation approaches related to updating the ocean and atmosphere states are generally categorized based on the choice of forecast models and

the analyzing states: (1) uncoupled DA ; (2) weakly coupled DA; and (3) strongly coupled DA.

Uncoupled DA (UCDA) is the most commonly used approach that take advantage of the solo ocean model forced by atmosphere data (e.g. atmosphere reanalyses) and solo atmosphere model forced by ocean data (e.g. a SST objective analysis, or ocean reanalyses). Different DA strategies are applied to these forced forecast models independently and update the ocean and atmosphere states separately.

Weakly coupled DA (WCDA) and strongly coupled DA (SCDA) together regard the coupled model as the forecast model, but differ in the way of updating analysis states. In WCDA, the observation in one medium does not influence the determination of the analysis in the other medium. In this approach, the interrelation between the ocean and atmosphere states are ignored during the analysis update and are only realized during the forecast model integration. Whereas, SCDA takes into account the connection between data of the two media by assimilating observations of the two media together into one coupled system.

## 1.4 Study Objectives

The content of this thesis is mainly separated into to two parts: LaDA and CDA. The primary goal of studying LaDA is to propose a feasible DA approach to assimilate the direct drifter locations and then to provide corrections with high accuracy not only for the surface ocean currents but also in the deep ocean as well. Specifically, we incorporate the Local Ensemble Transform Kalman Filter (LETKF)

with the LaDA augmented-state framework, which is denoted as LETKF-LaDA. The design of this method aims to be readily extended to realistic models of refined resolution and implemented efficiently using parallel computing. We investigate the proposed method from the following perspectives:

1. What is the optimal localization radius of LETKF-LaDA in an eddy-permitting forecast model and, more importantly, in an eddy-resolving model in terms of Baroclinic Rossby Radius of Deformation?

2. How deep does the information from the surface drifters give beneficial impacts on the LaDA analysis?

3. What is the influence of assimilating surface drifters along with in situ T/S, comparing with the traditional DA method to assimilate purely T/S observational data?

4. How does the model resolution impact the performance of LETKF-LaDA?

5. Is LETKF-LaDA capable of recovering the sub-mesoscale currents that can neither be accurately predicted by the forecast model nor completely recovered from the satellite altimetry?

In order to benefit the future study on LaDA, we also build up an online drifter module to simulate drifter motions for the latest version of the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model version 6 (MOM6). The drifter trajectory is obtained by integrating the velocity with the same model time step as the associated ocean model. In addition, the drifter module is further

improved by emulating the temperature, salinity and even velocity measurements at the corresponding drifter positions.

As for the study related to CDA, we compare UCDA, WCDA and SCDA from an idealized theoretic perspective. In this thesis, we consider an example coupled quasi-geostrophic model and its application on the aforementioned three CDA methods. We apply mathematical knowledge from dynamical system and synchronization theories to provide a theoretical analysis on the estimation performances. The following problems are discussed in the corresponding chapter:

1. Can forced ocean-atmosphere models synchronize with the coupled model? What would be the difference when the surface forcing is applied at different frequencies? Is it necessary to provide extra information or mechanisms for the forced system to synchronize with the coupled system?

2. What are the impacts of different coupling strategies on the DA in terms of the error convergence? Will there be any difference in the atmosphere and ocean components?

3. What are the effects of the observations from different medium on the performance of SCDA?

4. What is the performance of UCDA, WCDA and SCDA over a range of various forecast lead times?

## 1.5 Outline

Chapter 2 introduces the detailed algorithm of LETKF-LaDA. We commence our discussion with the verifying experiments using the 3D ocean model with the double-gyre configuration of the Modular Ocean Model version 4.1 (MOM4p1), in order to study its impact on correcting the ocean fields from surface currents to deep ocean. We then compare it with the traditional DA method assimilating temperature and salinity at the drifter locations. In Chapter 3, we extend the application of LETKF-LaDA to an eddy-resolving model with a GoM configuration in MOM6. The creation of an online drifter module adapted to MOM6 is also introduced in this chapter. In Chapter 4, we study the differences in UCDA, WCDA and SCDA from the perspective of synchronization. We employ the tool of Lyapunov Exponents (LEs) to analyze the performance of different DA methods. Chapter 5 summarizes the conclusions and the follow-on work.

# Chapter 2: LETKF-LaDA on Double-Gyre Model of MOM4p1

## 2.1 Overview

### 2.1.1 Formulation of augmented-state LaDA

The augmented-state approach (Ide et al., 2002) extracts information carried by Lagrangian tracers by extending the model states as a combination of fluid and drifter states,

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_F \\ \\ \mathbf{x}_D \end{bmatrix},$$

where $\mathbf{x}_D$ contains the information measured by drifters, such as: longitude, latitude, depth, as well as temperature and salinity when these are available in the model forecast states. In this study, we mainly consider the case of drifter data consisting of longitude and latitude measurements and assume drifters maintain a constant depth. Thus if there are $N_D$ drifters, then $\mathbf{x}_D$ is a $2N_D$-component vector and $\mathbf{x}_F$ is of $N_F$ components, where $N_F$ is the dimension of fluid model states. The dimension of the vector $\mathbf{x}_D$ can exceed $2N_D$ if extra fields are included in the drifter prognostic states (e.g. temperature and salinity).

The drifter advection equation is then added to the original fluid dynamical

system as,

$$\begin{cases} \dfrac{d\mathbf{x}_F^f}{dt} = M_F(\mathbf{x}_F^f, t) \\ \dfrac{d\mathbf{x}_D^f}{dt} = M_D(\mathbf{x}_F^f, \mathbf{x}_D^f, t) \end{cases}$$

As a clarifying example, suppose that only the position data carried by Lagrangian drifters are observed. Then the dimension of the observation space is $2N_D$ and the observations can be represented in terms of the observation operator $\mathbf{H}$ and true states $\mathbf{x}^t$ as,

$$\mathbf{y}^o = \begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}_F^t \\ \\ \mathbf{x}_D^t \end{bmatrix} + \varepsilon = \mathbf{H}\mathbf{x}^t + \varepsilon,$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix}, \qquad \varepsilon \sim N(0, \mathbf{R}).$$

The matrix $\mathbf{I}$ in the above formula is an identity matrix of dimension $2N_D \times 2N_D$. The quantity $\varepsilon$ is a Gaussian random error, with the $2N_D \times 2N_D$ observation error covariance matrix $\mathbf{R}$. The forecast time length must be chosen sufficiently small to maintain approximately Gaussian error statistics. The augmented-state LaDA provides an estimator of $\mathbf{x}_{i+1} = (\mathbf{x}_F, \mathbf{x}_D)^T$ at time $t_{i+1}$ given the observations $\mathbf{y}_{i+1}^o$. With the Gaussian assumption, the Kalman filter based methods attempt to provide the best linear unbiased estimator by taking advantage of the estimated observation error covariance matrix $\mathbf{R}$ and the estimated background error covariance matrix $\mathbf{P}$. Our background error covariance matrix is composed of the combined fluid and drifter states defined as,

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{FF} & \mathbf{P}_{FD} \\ \\ \mathbf{P}_{FD}^T & \mathbf{P}_{DD} \end{bmatrix},$$

where $\mathbf{P}_{FF}$, $\mathbf{P}_{FD}$, $\mathbf{P}_{DD}$, denote the background error covariance matrices of the fluid state vector, the cross-covariance between the fluid and drifters state vectors, and the drifters state vector, respectively. The detail definition of the components of these background error covariance matrices will be specified in the later section. The ensemble Kalman filter (EnKF) updates the prior error covariance matrix by computing the sample error covariance matrix from ensemble perturbations around the ensemble forecast mean.

### 2.1.2 Review of LETKF

The LETKF is an Ensemble Square Root Filter (EnSRF) proposed by Hunt et al. (2007) as an extension of works by Bishop et al. (2001), Halliwell et al. (2014) and Houtekamer and Mitchell (1998), using the localization approach of Ott et al. (2004). There are generally two kinds of localization approach: in observation space (R-localization), and in model space (B-localization) (Greybush et al., 2011). The LETKF uses R-localization, which selects and weights local observations in a prescribed region around each grid point while excluding observations outside this region.Salman et al. (2006) shows a proper selection of the localization region is beneficial in using EnKF to assimilate the drifter positions within the shallow-water system. In order to preserve vertically consistent dynamics, no localization is applied in the vertical (Penny et al., 2015). As a consequence, surface observations impact the analysis of the entire water column. Because previous studies have found superior results using this approach to vertical localization when applying LETKF

in the ocean (Penny et al., 2015; Sluka et al., 2016), we only consider experiments applying variations in the horizontal localization radius.

After the localized region is determined, we compute an analysis update to the center grid point at all depths. In general, the EnKFs assume Gaussian error statistics, which are estimated from the perturbations of the ensemble forecast around the ensemble forecast mean state. The analysis solution is thus confined to a maximum $(K-1)$-dimensional linear space defined by the ensemble states, where $K$ is the number of ensemble members. By applying the localization technique, we allow the global analysis to be formed from a larger dimensional space, though the localized solution is still formed within a linear space limited by the ensemble size. The localization makes it possible to approximate the solution of spatially extended high-dimensional nonlinear problems with better accuracy. The linear transform applied as part of the LETKF is an efficient technique to update the ensemble deterministically by transforming the prior distribution to the posterior distribution. The LETKF algorithm is easily parallelized by calculating the analysis independently at each grid point, thus facilitating the technical transition to more realistic applications.

## 2.2 Algorithm of LETKF-LaDA

Combining sections 2.1.1 and 2.1.2, we denote the LETKF method applied to solve the Lagrangian problem as LETKF-LaDA. To simplify the description, we first consider the case when drifter locations are observed and the dimension of

observation space is $2N_D$. A more general case considering the surface temperature and salinity is elaborated at the end of this section.

1. Run the dynamical model to obtain the global ensemble forecast states $\mathbf{x}^{f(k)}_{[g]} = (\mathbf{x}^{f(k)}_{F[g]},\ \mathbf{x}^{f(k)}_{D[g]})^T$ $(k = 1, 2, ..., K)$, then obtain the corresponding global ensemble mean $\overline{\mathbf{x}}^f_{[g]} = (\overline{\mathbf{x}}^f_{F[g]},\ \overline{\mathbf{x}}^f_{D[g]})^T$ and the forecast error perturbation matrix $\mathbf{X}^f_{[g]}$ , whose $k$th column is $\mathbf{x}^{f(k)} - \overline{\mathbf{x}}^f_{[g]}$. The subscript [g] indicates the global state vector.

2. Apply the specialized observation operator $\mathbf{H}$ defined in section 2.1.1 to the augmented model states in order to form the ensemble $\mathbf{y}^{f(k)}_{[g]}$ of the forecast observation vectors by $\mathbf{y}^{f(k)}_{[g]} = \mathbf{H}\mathbf{x}^{f(k)}_{[g]}$. Compute the corresponding mean, $\overline{\mathbf{y}}^f_{[g]}$, and error perturbation matrix, $\mathbf{Y}^f_{[g]}$, defined in observation space. In this case, we assume all observation data is carried by the drifters so that $\mathbf{y}^{f(k)}_{[g]} = \mathbf{x}^{f(k)}_{D[g]}$, the mean vector $\overline{\mathbf{y}}^f_{[g]} = \overline{\mathbf{x}}_{D[g]}$ and $2N_D \times K$ matrix $\mathbf{Y}^f_{[g]} = \mathbf{X}^f_{D[g]}$. For our experiments, the global observation vector is denoted as $\mathbf{y}^o_{[g]}$ and the observation error covariance matrix $\mathbf{R}_{[g]}$ is defined as a diagonal matrix with diagonal entries specified as the observation variance $\sigma^2$, which will be discussed in section 2.3.

3. Determine the local analysis of the LETKF-LaDA system, using the local arrays of the global terms: $\overline{\mathbf{x}}^f_{[l]}$, $\mathbf{X}^f_{[l]}$, $\mathbf{Y}^f_{[l]}$, $\mathbf{R}_{[l]}$, $\mathbf{y}^o_{[l]}$ and $\overline{\mathbf{y}}^f_{[l]}$. The selection approach of the localization region depends on the type of the analysis variables to be updated. For the fluid variables we apply the localization region as a cylinder centered at a horizontal grid point $(i, j)$ with horizontal localization

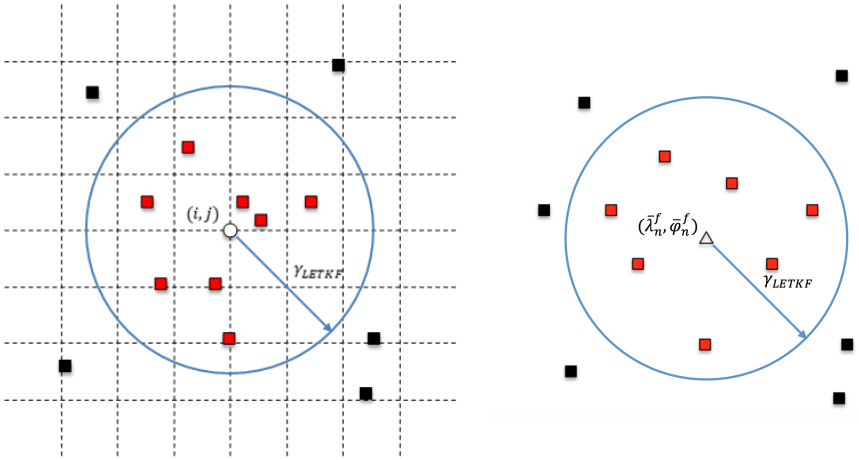


Figure 2.1: (a) the localization region defined to update the fluid states $\overline{\mathbf{x}}^f_{F[l]}$ at grid point $(i, j)$, the white circle at the center of the figure. The background dashed lines represent the mesh grids of the forecast model. (b) The localization region defined to update the drifter states $\overline{\mathbf{x}}^f_{D[l]}$ associated to a forecasting drifter ID $n$. The center white triangle is located at the forecast ensemble mean position of drifter No. $n$. In both figures, $\gamma_{LETKF}$ denotes the radius of the localization region and the squares represents the observation drifter locations. Those observation drifters inside the circles (red squares) are marked by their IDs as the localized observation.

radius $\gamma_{LETKF}$ (see Figure 2.1 (a)), which will be specified in section 2.3. For global model state variables $\overline{\mathbf{x}}^f_{[g]}$ and $\mathbf{X}^f_{[g]}$, rows associated with the fluid variables at this grid point $(i, j)$ from all the depth levels are chosen to formulate their corresponding local variables $\overline{\mathbf{x}}^f_{F[l]}$ and $\mathbf{X}^f_{F[l]}$. For example, suppose the

prognostic flow states are gridded velocities $u_F$ and $v_F$ , then we have,

$$\overline{\mathbf{x}}^f_{F[l]} = \begin{bmatrix} \overline{\mathbf{u}}^f_F(i,j,1) \\ \overline{\mathbf{v}}^f_F(i,j,1) \\ \vdots \\ \overline{\mathbf{u}}^f_F(i,j,h) \\ \overline{\mathbf{v}}^f_F(i,j,h) \end{bmatrix}_{2h\times 1} \tag{2.1}$$

$$\mathbf{X}^f_{F[l]} = \begin{bmatrix} \mathbf{u}^{f(1)}_F(i,j,1) - \overline{\mathbf{u}}^f_F(i,j,1) & \cdots & \mathbf{u}^{f(K)}_F(i,j,1) - \overline{\mathbf{u}}^f_F(i,j,1) \\ \mathbf{v}^{f(1)}_F(i,j,1) - \overline{\mathbf{v}}^f_F(i,j,1) & \cdots & \mathbf{v}^{f(K)}_F(i,j,1) - \overline{\mathbf{v}}^f_F(i,j,1) \\ \vdots & & \vdots \\ \mathbf{u}^{f(1)}_F(i,j,h) - \overline{\mathbf{u}}^f_F(i,j,h) & \cdots & \mathbf{u}^{f(K)}_F(i,j,h) - \overline{\mathbf{u}}^f_F(i,j,h) \\ \mathbf{v}^{f(1)}_F(i,j,h) - \overline{\mathbf{v}}^f_F(i,j,h) & \cdots & \mathbf{v}^{f(K)}_F(i,j,h) - \overline{\mathbf{v}}^f_F(i,j,h) \end{bmatrix}_{2h\times K} \tag{2.2}$$

where $h$ is the bottom level of the model. In Figure 2.1 (a), all the observed drifters located in this localized region are marked by their IDs and we define the number of localized observed drifters as $N_{D[l]}$. We then choose the rows of $\mathbf{Y}^f_{[g]}$, $\mathbf{y}^o_{[g]}$ and $\overline{\mathbf{y}}^f_{[g]}$ related to these marked drifter IDs to create $2N_{D[l]} \times K$ matrix $\mathbf{Y}^f_{[l]}$, and $2N_{D[l]} \times 1$ vectors $\mathbf{y}^o_{[l]}$ and $\overline{\mathbf{y}}^f_{[l]}$. Similarly, rows and columns are chosen to formulate $2N_{D[l]} \times 2N_{D[l]}$ diagonal matrix $\mathbf{R}_{[l]}$. Figure 2.1 (b) illustrates the approach to define the localization region in order to update the local state variables of the simulated drifters. As shown in Figure 2.1 (b), the selection of the localization region is associated with each drifter ID rather than each model grid point. For each simulated drifter ID $n$, its forecast ensemble mean location $(\overline{\lambda}^f_n, \overline{\varphi}^f_n)$ is defined as the center of the corresponding

localization region. Localized model state vector $\overline{\mathbf{x}}^f_{D[l]}$ and error perturbation matrix $\mathbf{X}^f_{D[l]}$ include all the entries associated to this ID $n$,

$$\overline{\mathbf{x}}^f_{D[l]} = \begin{bmatrix} \overline{\lambda}^f_{D,n} \\ \overline{\varphi}^f_{D,n} \end{bmatrix}, \qquad \mathbf{X}^f_{D[l]} = \begin{bmatrix} \lambda^{f(1)}_{D,n} - \overline{\lambda}^f_{D,n} & \cdots & \lambda^{f(K)}_{D,n} - \overline{\lambda}^f_{D,n} \\ \varphi^{f(1)}_{D,n} - \overline{\varphi}^f_{D,n} & \cdots & \varphi^{f(K)}_{D,n} - \overline{\varphi}^f_{D,n} \end{bmatrix}. \tag{2.3}$$

Similarly to the approach used to update the fluid variable, $\overline{\mathbf{y}}^f_{[l]}$, $\mathbf{Y}^f_{[l]}$, $\mathbf{y}^o_{[l]}$ and $\mathbf{R}_{[l]}$ can be defined accordingly based on the localization region. Among all the marked IDs, the observation associated to its own drifter ID $n$ is included. To simplify the notation, the subscriptions "$F[l]$" in (2.1,2.2) and "$D[l]$" in (2.3) of the model variables (i.e. $\overline{\mathbf{x}}^f$ and $\mathbf{X}^f$) are all replaced by "$[l]$" in the next step.

4. The remaining steps follow the original LETKF (Hunt et al., 2007), repeated here for completeness. Form the local analysis error covariance and weight matrix as,

$$\tilde{\mathbf{P}}^a_{[l]} = \left[(k-1)\mathbf{I}/\rho + (\mathbf{Y}^f_{[l]})^T \mathbf{R}^{-1}_{[l]} \mathbf{Y}^f_{[l]}\right]^{-1}, \qquad \mathbf{W}^a_{[l]} = \left[(k-1)\tilde{\mathbf{P}}^a_{[l]}\right]^{1/2},$$

where $\rho$ is the covariance inflation factor. The ensemble weight vector used to compute the local mean analysis is,

$$\mathbf{x}^{a(k)}_{[l]} = \overline{\mathbf{x}}^f_{[l]} + \mathbf{X}^f_{[l]} \mathbf{w}^{a(k)}_{[l]}.$$

5. Collect all the local analysis states at each grid point and each drifter ID to form the global analysis state vector,

$$\mathbf{x}^{a(k)}_{[g]} = \begin{bmatrix} \mathbf{x}^{a(k)}_{F[g]} \\ \mathbf{x}^{a(k)}_{D[g]} \end{bmatrix}$$

This algorithm can be extended to a more general case in which additional measurements of the fluid variables are associated with each drifter. For example, if additional temperature and salinity measurements are made for each drifter ID $n$, $\overline{\mathbf{x}}^f_{D[l]}$ and $\mathbf{X}^f_{D[l]}$ in (2.3) at step 3 can be extended as,

$$\overline{\mathbf{x}}^f_{D[l]} = \begin{bmatrix} \overline{\lambda}^f_{D,n} \\ \overline{\varphi}^f_{D,n} \\ \overline{T}^f_{D,n} \\ \overline{S}^f_{D,n} \end{bmatrix}, \qquad \mathbf{X}^f_{D[l]} = \begin{bmatrix} \lambda^{f(1)}_{D,n} - \overline{\lambda}^f_{D,n} & \cdots & \lambda^{f(K)}_{D,n} - \overline{\lambda}^f_{D,n} \\ \varphi^{f(1)}_{D,n} - \overline{\varphi}^f_{D,n} & \cdots & \varphi^{f(K)}_{D,n} - \overline{\varphi}^f_{D,n} \\ T^{f(1)}_{D,n} - \overline{T}^f_{D,n} & \cdots & T^{f(K)}_{D,n} - \overline{T}^f_{D,n} \\ S^{f(1)}_{D,n} - \overline{S}^f_{D,n} & \cdots & S^{f(K)}_{D,n} - \overline{S}^f_{D,n} \end{bmatrix}.$$

The remaining steps in the above algorithm are the same.

The above further extension of the system allows a direct differencing of the observed drifter measurements with the modeled drifter measurements, i.e. $\mathbf{y}^o - H\mathbf{x}^f$. We assume the interpolation operator is automatically embedded in the drifter dynamic model $M_D$ rather than explicitly relying on an observation operator $H$ to map the gridded temperature and salinity states to the drifter positions at the appropriate time.

## 2.3 Experiments Setup

### 2.3.1 The numerical ocean model and spin-up procedure

We use the "identical twin" approach in OSSEs to evaluate the impact of LETKF-LaDA. The nature run and the forecast model in all experiments use the GFDL subtropical double gyre configuration of the B-grid hydrostatic non-Boussinesq ocean model MOM4p1 (Griffies, 2008). We work in a rectangular, closed basin

on a beta plane with longitude ranging from 0°E to 10°E and latitude ranging from 15°N to 35°N. Ocean circulation is driven by the zonal wind stress defined as $F_\lambda(\varphi) = 0.1 \cdot \sin(\pi \cdot (\varphi - 20°N)/10°\ N)\ N/m^2$ , where $\varphi$ is the latitude. Fifty constant $z$-level coordinates are used extending down to 5500m.

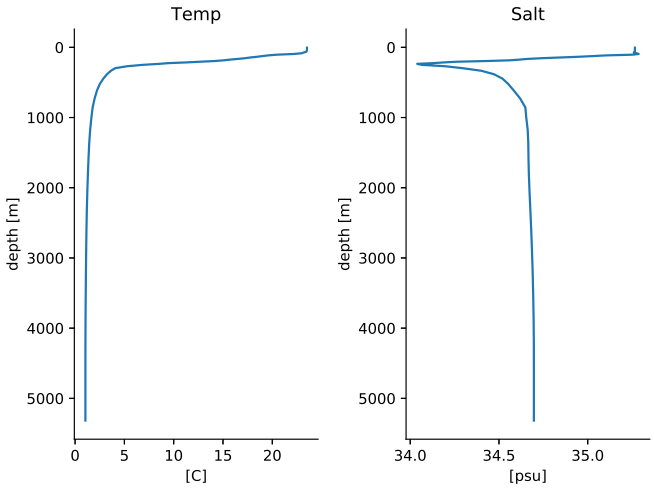


Figure 2.2: Temperature and salinity initial conditions provided by SODA on 2 Jan 1981 at location 175°E 25°N varying from 5m to 5316m. The thermocline is located below 75m and above 235m.

The nature run is generated using a 1/4-degree horizontal resolution with integration time step $\Delta t = 1800$ seconds. We specify the initial conditions of the temperature and salinity fields by replicating a point profile of temperature and salinity state estimates from the Simple Ocean Data Assimilation (SODA) on 2 Jan 1981 (Figure 2.2) and then initializing the model with this horizontally uniform stratification. The true drifter locations are simulated in the nature run, randomly initialized in the energetic region $2.5°E \sim 5°E$ and $20°N \sim 27°N$, as shown in Figure 2.3(a). The deployment depth, 15m, is the same as the official fixed depth

assigned for the NOAA GDP surface drifters.

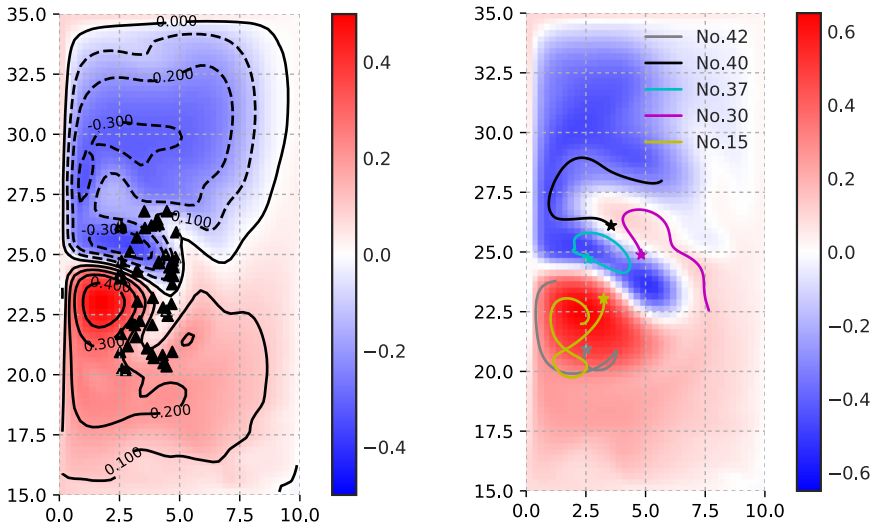


Figure 2.3: (a) The sea surface height (SSH) of the nature run at the end of the 6-month spin-up procedure (contour interval is 0.1m); (b) sampled drifter trajectories in 91 days are shown with SSH (m) at the terminal time. Initial drifter locations are marked as triangles in (a) and asterisks in (b).

We construct an ensemble of wind stress fields by adding a constant to the zonal component of the true wind field with the magnitude, which is randomly chosen by $\theta \sim N(0, 0.1)N/m^2$ for each ensemble member. This addition of the constant is applied uniformly for all grid points and time steps, i.e. $F_\lambda^{(k)}(\varphi) = F(\varphi) + \theta^{(k)}$ $(k = 1, 2, \cdots, K)$. To initialize the ensemble, we first spin up each ensemble ocean member for six months with the perturbed atmospheric conditions $F_\lambda^{(k)}$. All experiments use an ensemble size $K = 40$. We next populate 50 drifters ($N_D = 50$) simulated at 15m depth, initialized at the 'true' positions identically for each en-

semble member. The ensemble members are then integrated for another 16 days to generate the initial ensemble spread of drifter positions. We take advantage of an embedded drifter module in MOM4p1 that provides output of drifter positions as well as temperature and salinity at the drifter locations. Within this module, the drifter positions are calculated by integrating the corresponding velocities obtained after applying a bilinear interpolation to the gridded ocean velocity fields at the corresponding depth. The temperature and salinity data are determined directly by bilinear interpolation of the gridded temperature and salinity fields to the corresponding drifter locations.

The LETKF-LaDA system uses a daily analysis cycle following Jacobs et al. (2014). The optimal choice of analysis cycle window depends on the error doubling time of the modeled drifter positions, and remains an open question. The observations are assimilated daily with positioning errors drawn from a normal distribution using a prescribed standard deviation of $\sigma = 0.1$ degree in both longitude and latitude. The parameter $\sigma$ is determined by averaging the daily error growth of all the 50 drifter locations deployed in 40-member flow ensemble generated by the first step of spin-up process as previously described. We investigate the influence of the horizontal localization radius and model resolution on the performance of the LETKF-LaDA. We define the horizontal localization radius, $\gamma_{LETKF}$, as a multiple of the Rossby radius of deformation (RRD) at the respective latitudes (Chelton et al., 1998). The choice of $\gamma_{LETKF}$ in different experiments will be specified in the next section. None of the experiments use multiplicative inflation (i.e. $\rho = 1$). The control experiment is initialized using the ensemble mean ocean and drifter states

at the end of the two-step spin-up process.

### 2.3.2 Error metrics

The locations of drifters in this work are based on longitude-latitude coordinates on a spheroid surface instead of the Cartesian coordinates on a flat surface as used in previous augmented-state LaDA studies (Salman et al., 2006). In this section, we utilize the measurements of geophysical distance to evaluate the difference between two locations of interest. The errors in drifter states are computed by taking the average distance between ensemble mean positions and the true positions of all the drifters,

$$RMSE_D = \frac{1}{N_D}\sum_{n=1}^{N_D} d\Big((\lambda^t_{D,n},\ \varphi^t_{D,n}), (\overline{\lambda}_{D,n},\ \overline{\varphi}_{D,n})\Big),$$

where $d((\cdot,\cdot),(\cdot,\cdot))$ is the function to compute the geophysical distance between two locations based on their latitude and longitude coordinates. Salman et al. (2008) defined a dimensionless fluid field norm in the horizontal direction that evaluates the error in terms of a percentage of the true fluid state. The error corresponding to the velocity field is formulated as a type of kinetic energy field. Because the double-gyre model used for our experiments has the additional complexity of including multiple vertical layers, we sum the horizontal errors from the top depth level $h_o$ to level $h$,

$$|T| = \left[\frac{\sum_{m=h_o}^{h}\sum_{i,j}\big(T_F(i,j,m)-T^t_F(i,j,m)\big)^2}{\sum_{m=h_o}^{h}\sum_{i,j}T^t_F(i,j,m)^2}\right]^{1/2}, \quad |S| = \left[\frac{\sum_{m=h_o}^{h}\sum_{i,j}\big(S_F(i,j,m)-S^t_F(i,j,m)\big)^2}{\sum_{m=h_o}^{h}\sum_{i,j}S^t_F(i,j,m)^2}\right]^{1/2}$$

$$|KE| = \left[\frac{\sum_{m=h_o}^{h}\sum_{i,j}\big(u_F(i,j,m)-u^t_F(i,j,m)\big)^2+\big(v_F(i,j,m)-v^t_F(i,j,m)\big)^2}{\sum_{m=h_o}^{h}\sum_{i,j}u^t_F(i,j,m)^2+v^t_F(i,j,m)^2}\right]^{1/2}, \tag{2.4}$$

where $T_F(i,j,m)$, $S_F(i,j,m)$, $u_F(i,j,m)$ and $v_F(i,j,m)$ indicate the flow states at the grid point $(i,j,m)$, in order to verify the convergence of the proposed LETKF-LaDA approach. We use the $RMSE_D$ defined above to quantify the norm of the drifter states.

Due to the relationship between geostrophic currents and sea surface height (SSH), we are also interested in the forecast error correlation between the drifter states and the surrounding flow fields. For a specific drifter with ID $n$, we denote its true position as $(\lambda^t_{D,n}, \varphi^t_{D,n})$ and the drifter state corresponding to this ID as $\mathbf{x}^t_{D,n}$. The fluid state at the horizontal grid point $(i,j)$ with depth $h$ is specified as the vector $\mathbf{x}^f_F(i,j,h)$. The error correlation between the drifter state and its closest vertical ocean field is denoted as,

$$\begin{aligned} r_n(i,j,h) &= Corr(\varepsilon^f_F(i,j,h), \varepsilon^f_{D,n}) \\ &= \frac{E\Big[(\mathbf{x}^f_F(i,j,h) - \mathbf{x}^t_F)(\mathbf{x}^f_{D,n} - \mathbf{x}^t_{D,n})\Big]}{\sqrt{E\Big[(\mathbf{x}^f_F(i,j,h) - \mathbf{x}^t_F)^2\Big]} \cdot \sqrt{E\Big[(\mathbf{x}^f_{D,n} - \mathbf{x}^t_{D,n})^2\Big]}}, \end{aligned} \tag{2.5}$$

where $E(\cdot)$ denotes the mean. Similarly, the error correlation corresponding to SSH is

$$\begin{aligned} r_{SSH,n}(i,j,h) &= Corr(\varepsilon^f_F(i,j,h), \varepsilon^f_{SSH}) \\ &= \frac{E\Big[(\mathbf{x}^f_F(i,j,h) - \mathbf{x}^t_F)(\mathbf{x}^f_{SSH} - \mathbf{x}^t_{SSH})\Big]}{\sqrt{E\Big[(\mathbf{x}^f_F(i,j,h) - \mathbf{x}^t_F)^2\Big]} \cdot \sqrt{E\Big[(\mathbf{x}^f_{SSH} - \mathbf{x}^t_{SSH})^2\Big]}}, \end{aligned} \tag{2.6}$$

where $\mathbf{x}_{SSH}$ is selected at the closest grid point position to the $n$th drifter location.

## 2.4 Results and Discussion

We first set the localization radius of LETKF-LaDA as $\gamma_{LETKF} = 3L_R$, where $L_R$ stands for the approximate RRD at the corresponding latitude, and implement the LETKF-LaDA for 91 days (about 3 months). Within this time period, some of the true drifters complete one circuit around the gyre (see Figure 2.3(b)). Figure 2.4 depicts contours of absolute error averaged from the top layer 5m to the 235m. Since the initial drifter positions are all close to the gyre, a large reduction in error occurs within the western-central region for all prognostic model variables. Figure 2.5 shows that the error is reduced not only in the surface fields but also at deeper levels. The relatively small error reduction in the northern and southern boundary regions is due to sparse observation coverage (Figure 2.4). Figure 2.6(a) highlights the time variation of the analysis RMSEs and ensemble spreads in velocities at confluence region (0.125°E$\sim$ 2.875°E and 23.625°N$\sim$ 26.625°N). The exponential decays in both of the RMSEs and ensemble spread of flow velocities at the confluence region are able to guarantee the ensemble drifters at this region distributed around the truth with small ensemble spreads and therefore provide accurate analysis states. Salman et al. (2008) shows that different drifter deployments can affect the convergence of the error in drifter displacements. In Figure 2.6(b), we display the time variation of analysis RMSE for sampled drifters labeled in Figure 2.3(b). The RMSE of most of the drifters are below the prescribed observation error standard deviation (i.e. 0.1 degree), except for the drifter No.37 circuiting near the confluence. It is observed that this drifter has a sudden increase in the RMSE as it is about to complete one

loop and then the error gradually decays in the end.

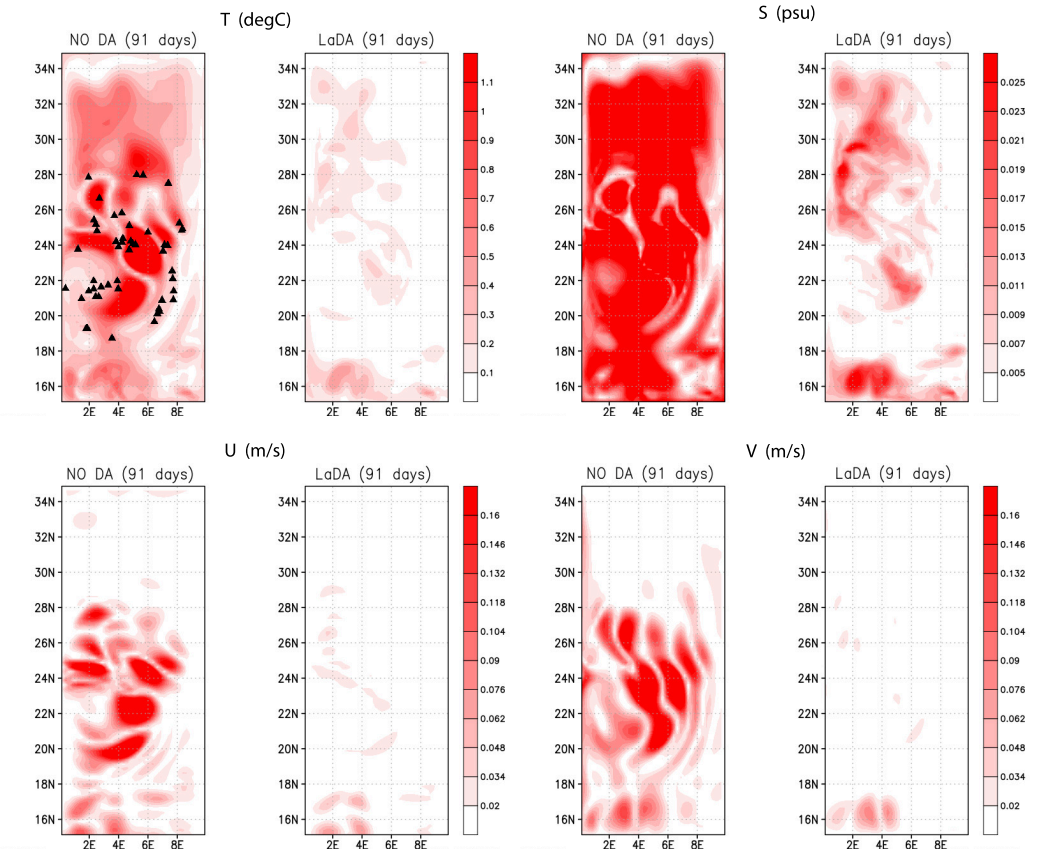


Figure 2.4: Absolute error comparisons between the control run (i.e. $|\mathbf{x}^C - \mathbf{x}^t|$) and the LETKF-LaDA (i.e. $\mathbf{x}^a - \mathbf{x}^t$) assimilating only the drifter position to analyze ocean flow fields temperature (T), salinity (S) and velocities (U, V) at end of 91 days. The errors are shown in the longitude-latitude plane and averaged from the top layer 5m to the 235m. The true drifter locations at the end of 91 days are marked by closed triangles in the first contour figure of the left column.

In general, the performance of the LETKF-LaDA depends on the localization radius, which affects both the extent of influence by observations and the reduction of nonlinear error in the prognostic model states. In the following discussions, we

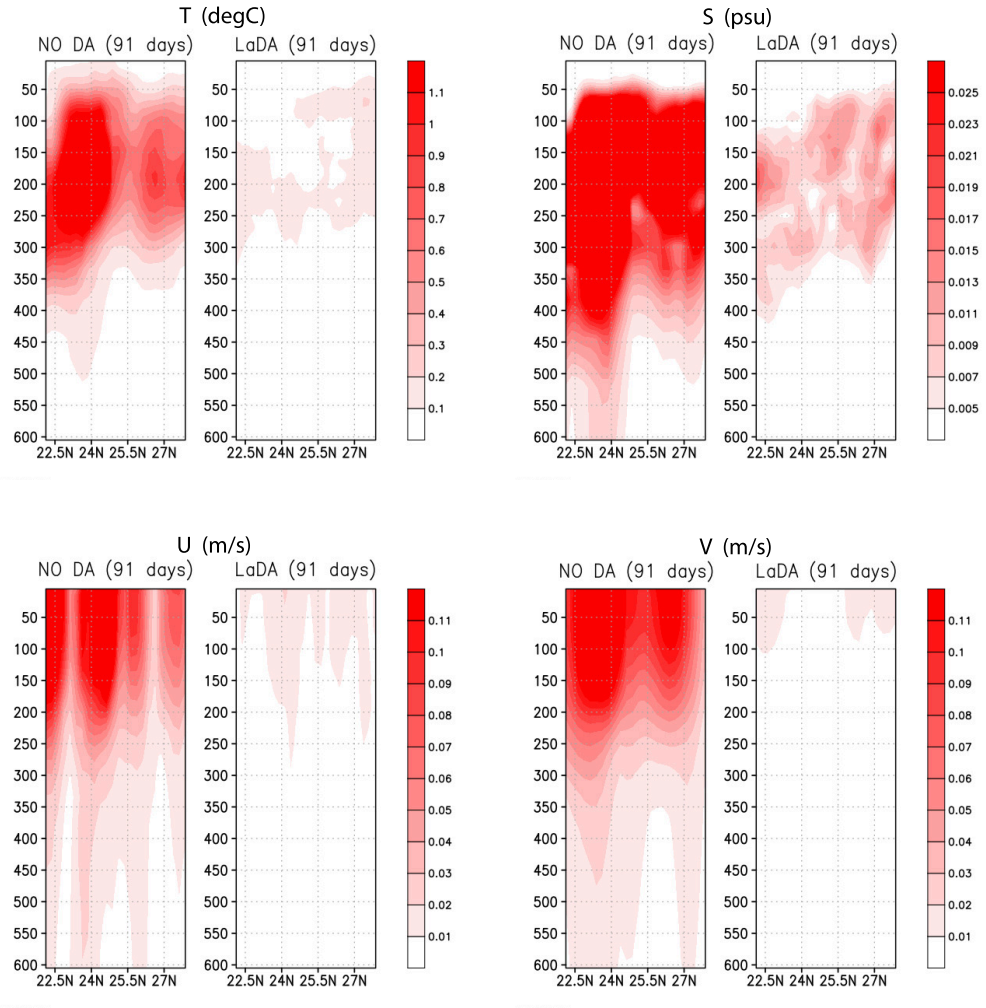


Figure 2.5: Absolute error comparisons between the control run (i.e. $|\mathbf{x}^C - \mathbf{x}^t|$) and the LETKF-LaDA (i.e. $|\mathbf{x}^a - \mathbf{x}^t|$) assimilating only the drifter position to analyze ocean flow fields temperature (T), salinity (S) and velocities (U, V). The errors are shown in the latitude-depth plane and averaged along longitude from 0.625°E to 7.125°E. The x-axis stands for the latitude interval 22.125°N$\sim$ 27.875°N, while the y-axis represents the depth level from 5m to 600m.

first determine an effective localization radius for the given system and then apply this radius for the remainder of the experiments.

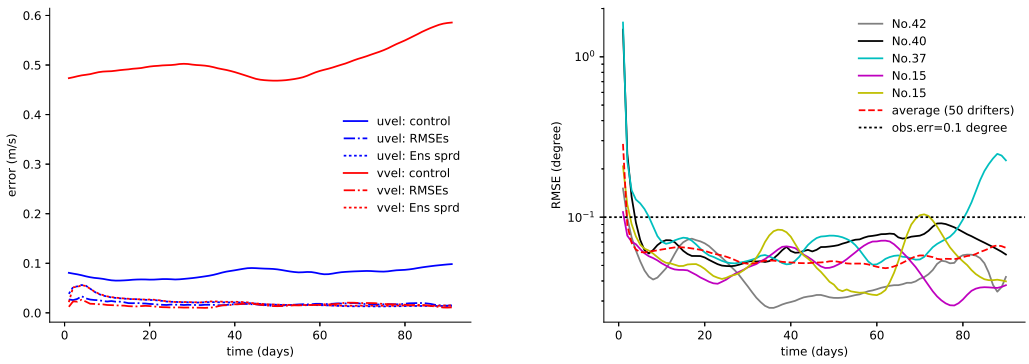


Figure 2.6: a) Time variation of analysis RMSE and ensemble spread in velocities at confluence region (0.125°E∼ 2.875°E and 23.625°N∼ 26.625°N) averaged through all the depth levels; b) time variation of analysis RMSE (in degree) in drifter displacements.

### 2.4.1 Impacts of varying localization radius $\gamma_{LETKF}$

Using the 1/4-degree horizontal resolution, we vary the localization radius: $\gamma_{LETKF} = 5L_R, 4L_R, 3L_R, 2L_R$ and $L_R$. Each experiment applies the LETKF-LaDA over 91 analysis cycles. We compare the error variations of the cycled analysis states to the control run. Figure 2.7 shows the results of percentage error norm defined in equation (2.4). We compare the errors of the analysis mean with the control run. For the temperature, salinity, and kinetic energy fields, we sum the squared errors in percentage norm from 5m to 1000m meter depth. Cases using $3L_R$ and $2L_R$ produce the lowest errors over the course of the experiments. The $2L_R$ localization radius produces slightly lower errors than $3L_R$ radius in controlling the mean drifter error distance below 3km within the 1/4° meshgrid map. The experiment with

$5L_R$ radius generates the most accurate analysis in model prognostic fields at end of the first cycle, however it begins diverging after 60 cycles due to continuous error growth in drifter states beginning after 5 cycles. Using the smallest attempted localization radius $L_R$, the error decays relatively slowly in both the prognostic fields and estimated drifter positions in the first 60 cycles. The errors in prognostic fields reach a low value comparing with other radii at the end of this time period. With localization radius $L_R$, the LETKF-LaDA produces in a sudden reduction in error between 60 to 80 days, however this error reduction does not appear stable. The sudden "shock" in error after the 1st cycle of the LETKF-LaDA with the various localization radii may be the consequence of an instability generated by a large change in the state variables.

In order to monitor the performance of the LETKF-LaDA in the vertical using varying localization radii, we consider the analysis states at the end of the 91st analysis cycle and compute the error in each horizontal layer from 5m to 1000m depth (see Figure 2.8). In the temperature field, all errors have a similar shape in the vertical direction with a maximum at approximately 270m depth. Both of the $3L_R$ and $2L_R$ localization radii produce smaller errors at all depth levels than all the rest of the localization radii. The experiment with localization radius behaves similarly to those with radii $3L_R$ and $2L_R$ below the thermocline, though a marginally larger error is produced in the top layer (5m$\sim$200m). For salinity, experiments using $3L_R$ and $2L_R$ radii have the lowest error above 270m depth, while errors in all experiments decay to almost the same value at deeper levels. All LETKF-LaDA cases have difficulty in reducing error in kinetic energy below 300m

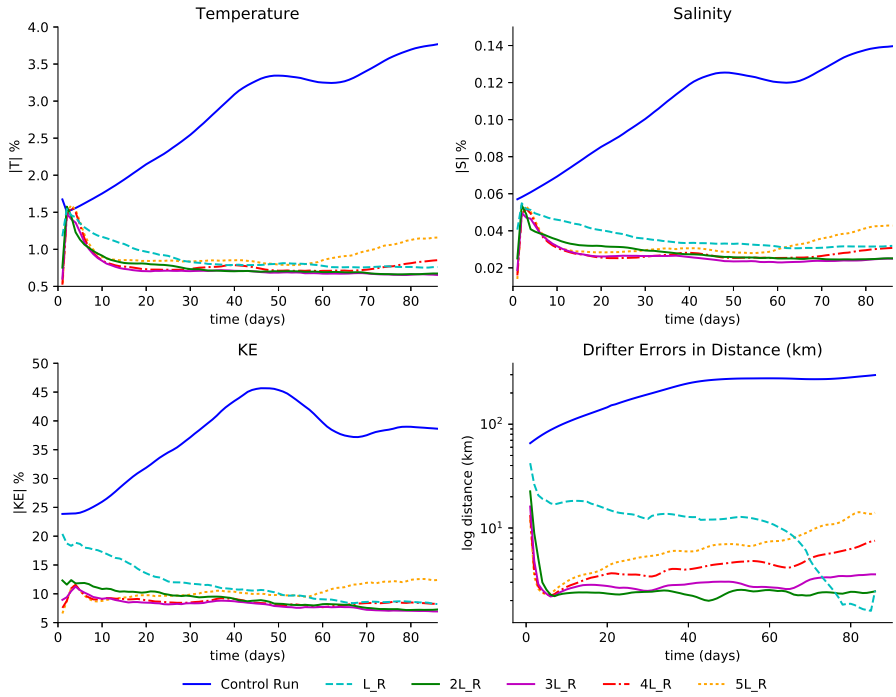


Figure 2.7: Error norms of temperature, salinity, kinetic energy and drifter states ($RMSE_D$) in the given period [0, 91] using formula (2.4). The error norms of the ocean flow fields are aggregated from 5m depth to 1000m depth (i.e. $h_o = 5$m and $h = 1000$m). In each subplot, we show the experiment results of control run (blue solid line), $5L_R$ (orange dotted line), $4L_R$ (red dash-dotted line), $3L_R$ (purple solid line), $2L_R$ (green solid line) and $L_R$ (cyan dashed line).

depth, though the case using a radius of $3L_R$ produces the smallest error in all depth levels.

The LETKF-LaDA requires the growth in drifter position error to grow quasi-linearly in time for the duration of the analysis cycle. The degree of nonlinearity depends on the stability of the dynamics, being greatest in the gyre region, and for the purpose of LaDA depends upon the time between observations. In order to

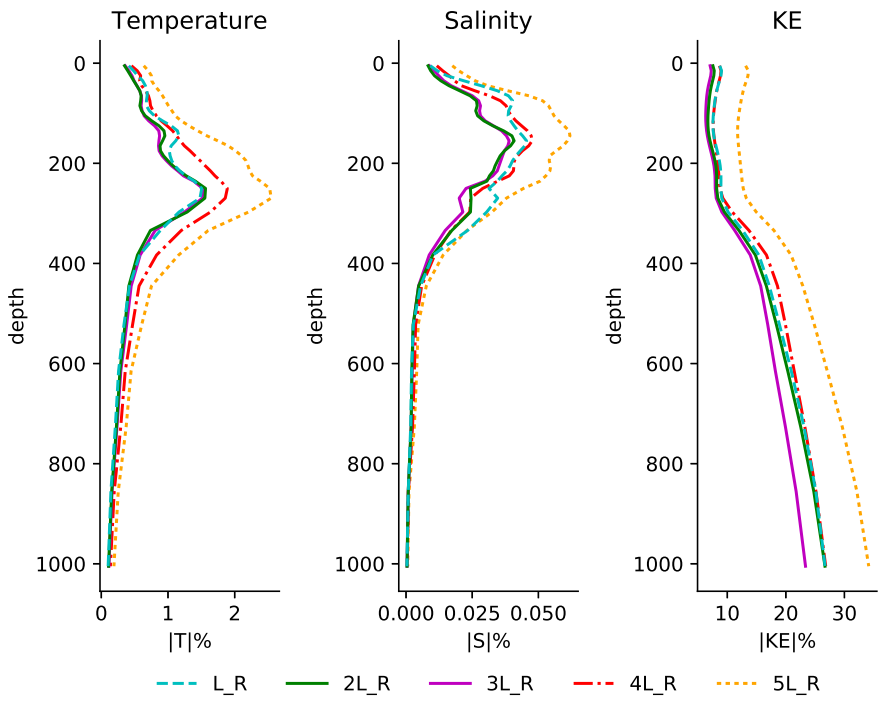


Figure 2.8: Error in temperature, salinity and kinetic energy at the end of 91th DA cycle along the vertical direction with experiments of localization radius $5L_R$, $4L_R$, $3L_R$, $2L_R$ and $L_R$. These quantities are evaluated by formula (2.4) at each individual vertical level with $h_o = h$. The control run is not shown in these figures because its errors largely exceed the scale of the given results in all fields.

evaluate the performance of the LETKF-LaDA in updating the state in the presence of varying degrees of nonlinearity, we examine trajectories for two different drifter IDs (Figure 2.9). We select one drifter in the gyre (No. 44) and another in a region with relatively slow approximately linear flow (No. 40). For both the linear and nonlinear trajectories, the LETKF-LaDA is able to track the true drifter trajectories.

In summary, with the quantitative and qualitative comparison in all the states and drifter trajectories, we find that the LETKF-LaDA with localization radii $3L_R$

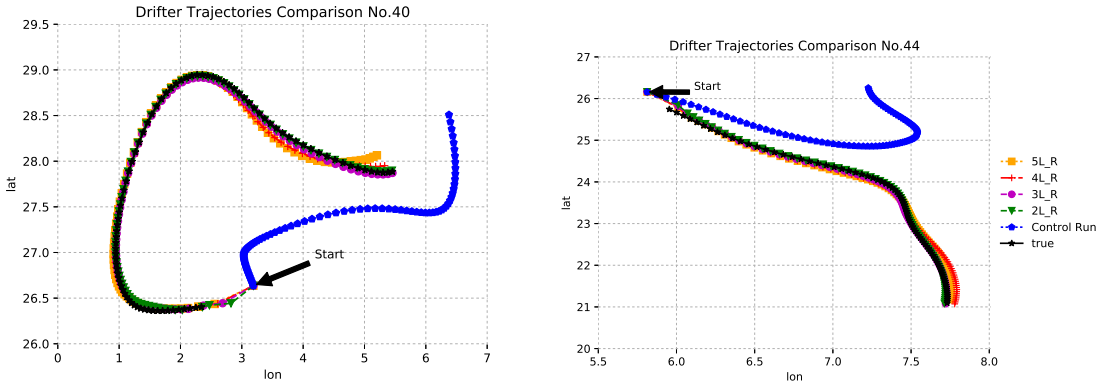


Figure 2.9: Drifter trajectories of nature run (black) compared to the control run (blue) and the LETKF-LaDA varying the localization radius: $5L_R$, $4L_R$, $3L_R$ and $2L_R$ in 91 days. All cases use identical drifter starting positions. The green line ($2L_R$) is the closest trajectory to the black line (obs) at the end for both of the drifters.

and $2L_R$ result in the lowest errors in the state estimate.

### 2.4.2 Comparing to conventional assimilation of in situ temperature and salinity

Drifters may be equipped not only with a GPS locator but also sensors providing in situ measurements (Lumpkin et al., 2007). In this section, we add temperature and salinity observations to the simulated Lagrangian drifters and compare the performance of three observing strategies: (1) using the LETKF to assimilate only the in situ temperature and salinity observations at the surface (no position information), (2) using the LETKF-LaDA to assimilate only drifter positions, and (3) using the LETKF-LaDA to assimilate both drifter locations and in situ temperature and

salinity observations with observation errors $\sigma_T = 0.1°C$ and $\sigma_S = 0.1$psu. All experiments use the localization radius $\gamma_{LETKF} = 3L_R$, as determined in the previous section.

Figure 2.10 summarizes the time variation of error in the estimated ocean state. The LETKF-LaDA experiments provide more accurate estimates of the prognostic model state variables than that of the conventional T/S DA after passing through the "shock" period in all the flow fields. The addition of the in situ data to some extent reduces the "shock" in temperature and salinity error for the LETKF-LaDA in the first few cycles. In the vertical comparison at the terminal time (see Figure 2.11), the LETKF-LaDA outperforms assimilation of conventional observations at all depths using the metric given by formula (2.4). Assimilating the combination of in situ and Lagrangian position data further improves the accuracy of the LETKF-LaDA salinity estimates in all layers and temperature in most of the layers except for the levels between 100m and 200m, though this results in a smaller improvement of the estimations in kinetic energy blow 450m.

### 2.4.3 Model resolution

One factor that could have an impact on the convergence of the LETKF-LaDA is errors in drifter trajectories caused by the model grid resolution. The insufficient modeling flow caused by a coarse grid resolution can possibly result in a relatively large discrepancy between the forecast drifter and the truth. This is a particular concern for transitioning to the assimilation of real observational data. In order to

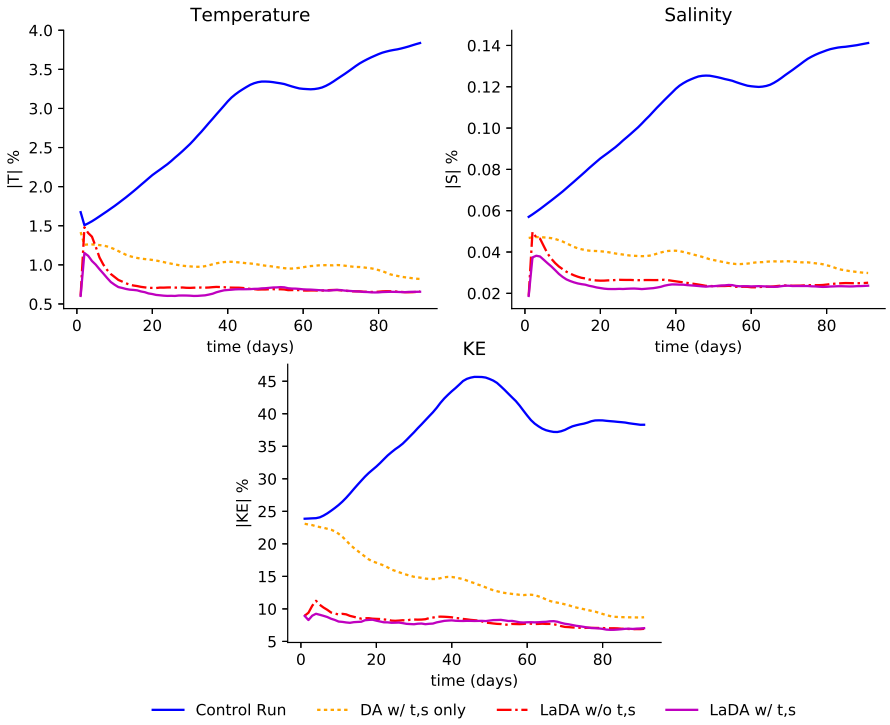


Figure 2.10: Error norms of temperature, salinity and kinetic energy in the given period [0, 91] using formula (2.4). The error norms of the ocean flow fields are aggregated from 5m to 1000m depth (i.e. with $h_o$ = 5m and $h$ =1000m). In each subplot, we show the experiment results of control run (blue solid line), assimilation of surface drifter measured T and S (orange dotted line), LETKF-LaDA assimilating only drifter positions (red dashdotted line) and LETKF-LaDA assimilating both drifter positions and surface drifter measured T and S (purple solid line).

examine sensitivity to resolution, model forecasts are obtained using configurations of 1/3° and 1/2° horizontal grid resolutions. The initial coarser ensemble fields are generated by applying bilinear interpolation to the previous initial ensemble fields defined in 1/4° gridding system as in section 2.3. Figure 2.12 shows a comparison of the SSH fields in one-day forecast generated by models of different resolutions

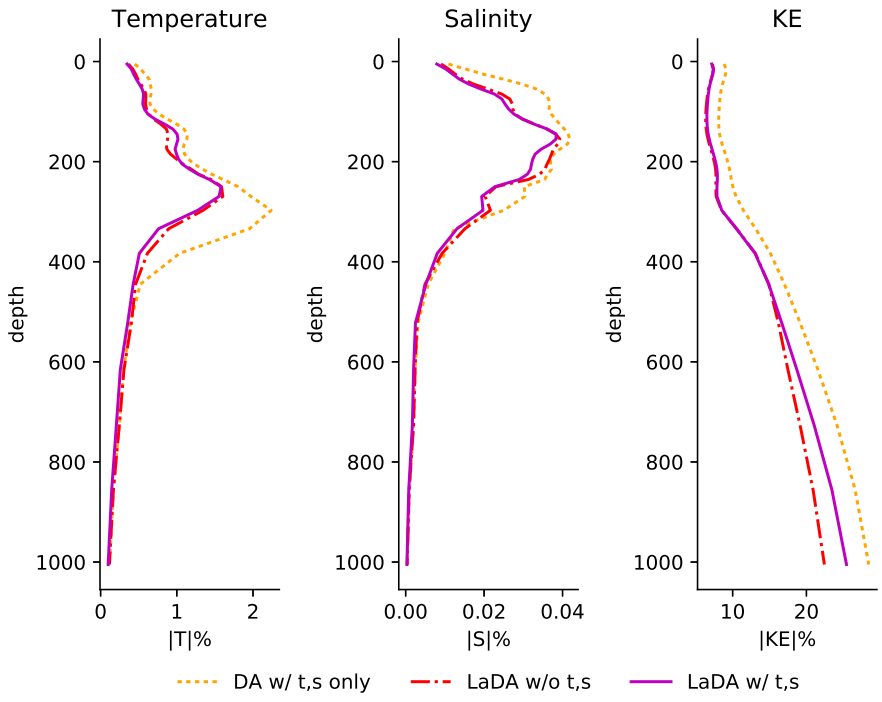


Figure 2.11: Vertical profile of change in error for temperature, salinity and kinetic energy at the end of 91th DA cycle with experiments of traditional DA assimilating T and S, LETKF-LaDA with and without assimilation of T and S. These quantities are evaluated by formula (2.4) at each individual vertical level with $h_o = h$. The control run is not shown in these figures because its errors largely exceed the scale of the given results in all fields.

before we start applying the LETKF-LaDA. Forecast models of coarser grids fail to represent the scale of SSH at the center of the gyres and create larger errors at eastern edge of the gyre. The LETKF-LaDA is performed with localization radius $3L_R$.

We observe in Figure 2.13 that assimilating Lagrangian data is possible with a forecast model resolution that resolves the dynamics present in the observations

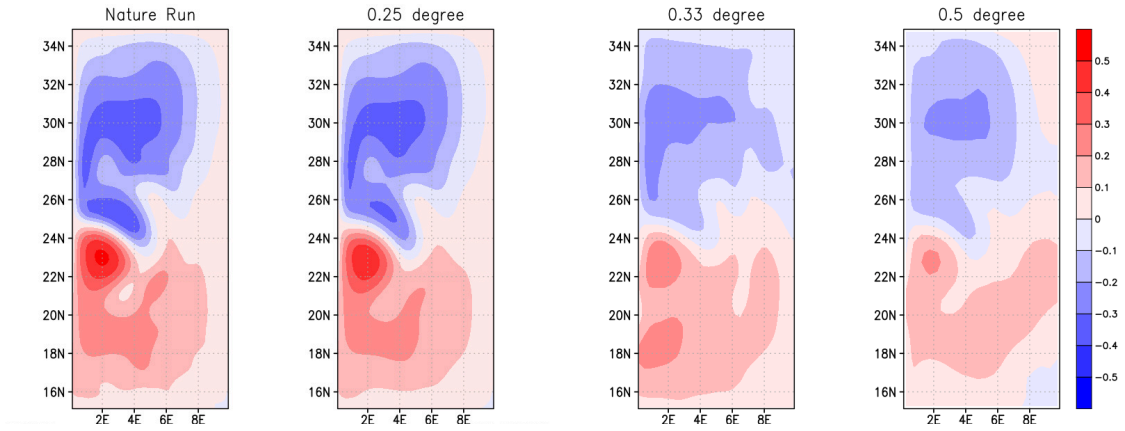


Figure 2.12: SSH contour comparisons between the nature run and one-day forecast ensemble means generated by forecast models of 1/4°, 1/3° and 1/2° resolutions before we start applying the LETKF-LaDA.

(e.g. in this case with a perfect model). Experiments using both coarser resolutions fail to stabilize the error growth in ocean and drifter states due to the exponential growth in the drifter position errors.

The above results show that the accuracy of the forecasting model influences the performance of the LETKF-LaDA. The relation between the convergence of the LETKF-LaDA and the prognostic flow scale of the forecasting model must be examined in the future with a high-resolution model.

### 2.4.4 Relation between the Lagrangian states and SSH

Due to the geostrophic relationship between the surface currents and pressure in the midlatitudes, and because SSH can be used as a proxy for integrated pressure in the upper ocean column, we expect some redundancy in the information provided by surface drifter position data and satellite altimeter measurements. For example,

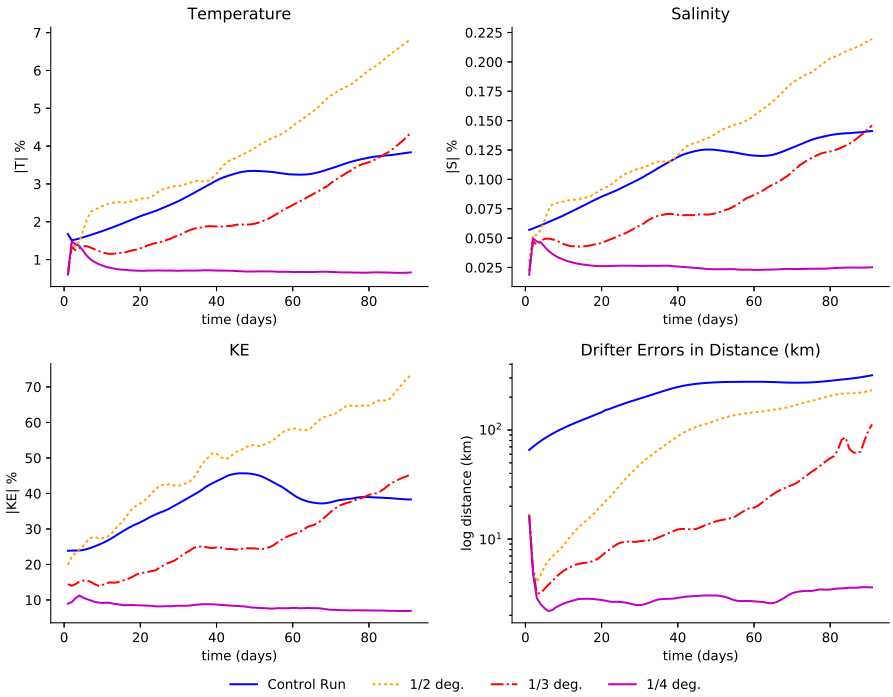


Figure 2.13: Error norms of temperature, salinity, kinetic energy and drifter states (RMSED) in the given period [0, 91] using formula (2.4). The error norms of the ocean flow fields are aggregated from 5m depth to 1000m depth (i.e. $h_o = 5$m and $h = 1000$m). In each subplot, we show the experiment results of control run (blue solid line), LETKFLaDA with forecast model of 1/2° (orange dotted line), 1/3° (red dashdotted line) and 1/4° (purple solid line) resolutions.

Carrier et al. (2016) showed that assimilating the combination of the along-track SSH and the Eulerian flow velocity approximated by drifter locations can provide improved forecasts of SSH, compared to using either the observations of the along-track SSH or the Eulerian velocity alone. Similarly, we expect the LETKF-LaDA will be a viable method to improve forecasts of SSH by taking advantage of drifter position data.

We use the error correlation between the drifter state variables (e.g. longitude and latitude) and the ocean prognostic fields (e.g. temperature, salinity and velocity) to extract information from the Lagrangian position data that can be used to update the subsurface state. Because the general quasi-geostrophic balance between the surface currents and dynamic ocean height fields indicates that a passive drifter will move following the contour of SSH, we infer that the correlation between the SSH forecast error and the Eulerian velocity forecast error is intertwined with the forecast errors corresponding to the drifter positions. To verify this, we spin up the ensemble as introduced in section 2.3 and compute the SSH ($r_{SSH}$) and drifter forecast error correlations ($r_{lon}$ and $r_{lat}$) using the formula (2.6) and (2.5) defined in section 2.3.2.

We examine error correlations between the drifter positions and the 3D fluid and compare in the longitude-depth plane. For example, the $r_{lon}$ and $r_{lat}$ associated with drifter No.14 share similar patterns with the contour profile of $r_{SSH}$ for all the flow fields (Figure 2.14). Continuing the experiments for all the other drifters, we summarize the connection between $r_{SSH}$ and $r_{lon}$, $r_{lat}$ in longitude-depth plane as following: for a specific drifter ID $n$,

$$r_{SSH,n} \approx \alpha \cdot r_{lon,n} + \beta \cdot r_{lat,n}, \quad \alpha \propto -u_n, \qquad \beta \propto sign(u_n) \cdot v_n,$$

where ($u_n$, $v_n$) is the drifter velocity of drifter ID $n$. For those drifters located in unstable portions of the flow, a large ensemble size is needed to verify the above relation numerically. In the last row of Figure 2.14, the estimations of $r_{SSH}$ are determined by this formula for drifter No. 14, using the coefficients: $\alpha = -u_n/(|u_n|+$

$|v_n|)$ and $\beta = sign(u_n) \cdot v_n/(|u_n| + |v_n|)$. Comparing the first three rows with the last row in Figures 14, we observe that with this approximation formula, $r_{lon}$ and $r_{lat}$ partially recover the positivity and negativity of the correlations $r_{SSH}$ in the given region, though the magnitude is different. The strongly linear relation between $r_{lon}/r_{lat}$ and $r_{SSH}$ suggests a similar relation in their corresponding error covariance matrices.

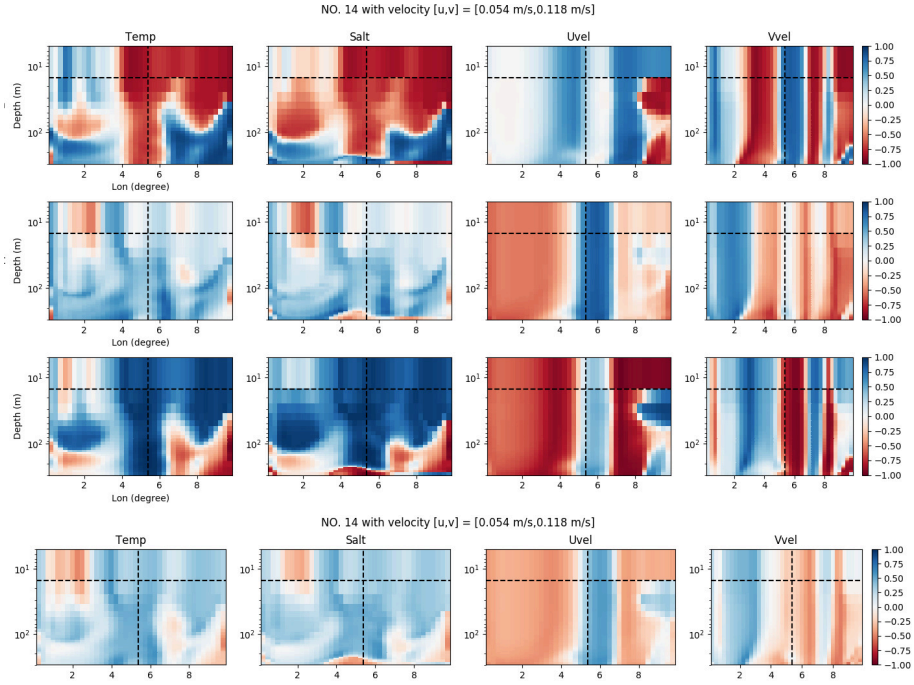


Figure 2.14: Correlation comparisons among $r_{lon}$ (1st row), $r_{lat}$ (2nd row) and $r_{SSH}$ (3rd row) associated with drifter No. 14 in the longitude-depth plane. The 4th row shows the correlation estimation of $r_{SSH}$ using the linear combination of $r_{lon}$ and $r_{lat}$ associated with drifter No. 14 in the longitude-depth plane using coefficients: $\alpha = -u_n/(|u_n| + |v_n|)$ and $\beta = sign(u_n) \cdot v_n/(|u_n| + |v_n|)$ .

## 2.5 Conclusion

In this chapter, we introduced a localized augmented-state Lagrangian Data Assimilation implemented with the Local Ensemble Transform Kalman Filter (LETKF-LaDA). Extending previous work with augmented-state LaDA methods, we applied this method to an ocean model with 3D dynamics and representation of temperature and salinity tracers. The LETKF-LaDA was found to provide more accurate analyses throughout the ocean column from the surface to 1000m depth by assimilating drifter position observations, compared to the conventional assimilation of in situ drifter measurements of temperature and salinity measurements.

In the context of the LETKF-LaDA, we studied the impact of localization radius on stabilizing the error growth over time. We found that with radius $5L_R$, five times the Rossby radius of deformation, the error decays fast in the first DA cycle but the nonlinear error grows unchecked over longer time. We also found that if the localization radius is too small, e.g. $L_R$, error often grows unconstrained in regions far away from the observations. The best performing localization radius in our experiments with the LETKF-LaDA assimilating 50 drifters deployed randomly around the gyre was between $2L_R \sim 3L_R$.

Correlations between errors in the drifter locations and ocean states ($r_{lon}$ and $r_{lat}$) were also studied. For a specific drifter, $r_{lon}$ and $r_{lat}$ had an approximately linear relation with the SSH error correlation ($r_{SSH}$) at the drifter location. Since the error correlation is an influential factor in the performance of DA methods, future research will examine the use of the LETKF-LaDA to simultaneously assimilate both drifter

position and along-track SSH observations to improve the forecast accuracy of the SSH field and the ocean surface currents, especially where there are gaps in the altimeter observations.

The LETKF-LaDA was implemented using a parallel computing framework in order to accelerate the computation of the analysis in anticipation of scaling to an operational global ocean data assimilation system. An additional factor affecting the stability of the LaDA filter is the number of drifter observations. A software limitation in MOM4p1 prevents us from studying the effect of this parameter because the built-in drifter module of MOM4p1 does not support running a large group of drifters. As the number of simulated drifters increases, the drifters become more likely to be lost or repeated during the model integration. The primary cause of this issue is an error in communication across different processors.

Conducting an OSSE using the "identical twin" approach is an important first step for validation of any new DA method. However this type of experiment ignores important concerns that are necessary to determine the viability of the LETKF-LaDA for operational use, such as the presence of systematic model errors. For example, in our experiments varying model resolution, results indicated that the LETKF-LaDA is sensitive to degraded model resolution. In future research, we will apply the LETKF-LaDA using a more realistic model configuration of the latest GFDL MOM6, configured with various resolutions that span a range from eddy permitting to eddy resolving and forced with time-variant near-surface wind forcing, and transit the whole system into assimilating real historical drifter and hydrographic profile data.

# Chapter 3: LETKF-LaDA on Gulf of Mexico of MOM6

## 3.1 Overview

Modern ocean models are capable of resolving the oceanic large-scale flow, including western boundary currents such as the Gulf Stream to the Kuroshio. Nevertheless, at mesoscales and sub-mesoscales, the current ocean models inadequately predict the flow with high accuracy even using high-resolution model. Individual tracers or drifters, in contrast, are essential observations to provide reliable information for the search and rescue operations, and hydrocarbon/chemical spill simulations (Poje et al., 2014). Therefore, it is important to take a full advantage of the information carried by actual drifters/floats in order to properly constrain the model simulation with the prescribed resolution. Here, we summarize the major studies of applying different DA methods to multiple GoM forecast models assimilating drifter measurements.

Vernieres et al. (2011) used a multi-layer reduced gravity model of the GoM in a perfect model scenario to investigate the impact of three kinds of synthetic measurements on the performance of DA: Eulerian velocity at the fixed stations, 2D horizontal Lagrangian drifter locations, and 3D Lagrangian drifter locations. The investigation was conducted using an EnKF under the framework of augmented-

state method (Ide et al., 2002). Results indicated that the two LaDA methods were capable of reproducing the eddy-shedding process by one drifter as long as the drifter is launched in a proper position. In addition, the two LaDA methods also had larger region of influence compared with the Eulerian method, with the LaDA in 3D scenario being the largest.

During the summer 2012, the Consortium for Advanced Research on Transport of Hydrocarbon in the Environment (CARTHE) conducted experiments of the Grand Lagrangian Deployment (GLAD) by deploying more than 300 drifters at the GoM and recording their displacement every 15mins for three months. The database was released later in a combination of estimated instantaneous velocities derived from drifter trajectories, which motivates the assimilation of the drifter measurements in more realistic scenarios. Drifters in GLAD program were gradually launched since July 20, 2012 and the launching process took coupled of days to complete. Figure 3.1 shows the 297 drifter GPS locations on Aug 3rd, 2012, when the launching process for all the drifters was already completed. Note that most of the drifters were initially deployed in the northeastern region. In addition, Hurricane Issac passed through the GoM area, which inflicts damage on the observations in late August. Details on the data processing of all the drifters are shown in Olascoaga et al. (2013) and Coelho et al. (2015)

Jacobs et al. (2014) and Carrier et al. (2014, 2016) all utilized the Navy Coastal Ocean Model (NCOM), capable of producing very accurate ocean simulations for regional and global applications at mesoscales and sub-mesoscales. Variational DA methods including 3D-Var and 4D-Var were attempted in this series of research

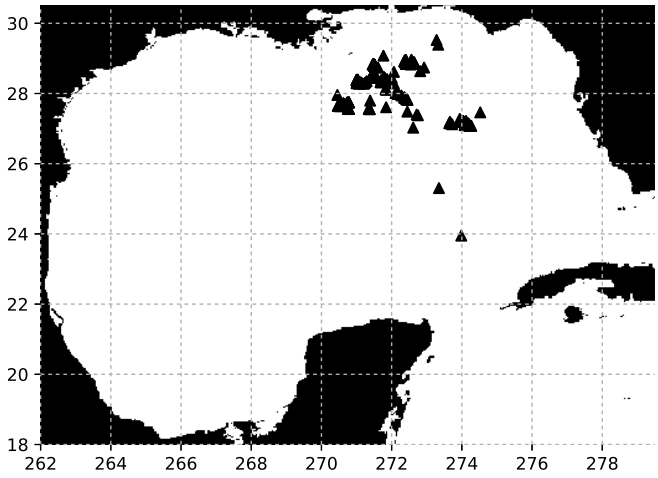


Figure 3.1: Horizontal distribution of 297 drifter locations recorded in GLAD database on Aug 3rd, 2012.

to study the impact of the "proxy velocity" entries of the GLAD on the analysis and the forecast state variables. Results show that: (1) the GLAD velocity data significantly contribute to improvements in characterizing of the circulation; and (2) with a combination of along-track SSH observations and the GLAD velocity data, the NCOM-4DVAR system was able to provide a better SSH forecast when compared to the experiments of assimilating either the aforementioned two observations alone.

Coelho et al. (2015) proposed a non-intrusive cycling method named as Multi-Model Ensemble Kalman Filter, where they generated the members of the ensemble forecast using a set of realistic ocean models with different horizontal resolutions, which included the NCOM at 1 km and 3 km resolutions, the US Navy operational NCOM at 3 km resolution (AMSEAS), and two versions of the Hybrid Coordinates Ocean Model (HYCOM) set at 4 km. In this work, the GLAD velocity data

was assimilated to improve the water velocity prediction along the observed drifter trajectories.

So far, among the applications to the realistic prediction models, only the Eulerian velocity entries of the GLAD dataset were used and assimilated into the DA system. There has not been many attempts to apply the augmented-state assimilation (Ide et al., 2002) of the direct drifter positions into a real-world scenario, possibly due to the absence or underdevelopment of drifter module in the ocean model, or the curse of dimensionality faced by most of the current augmented-state methods.

In this chapter, we extend the application of the LETKF-LaDA method proposed in section 2.1.2 to a high-resolution eddy-resolving regional ocean model, the Modular Ocean Model version 6 (MOM6) configured for GoM. Unlike the constant surface wind field used in the previous chapter, the GoM model is forced by a time-variant surface wind field and related near-surface atmosphere fields, such as, downward shortwave/longwave radiation flux and precipitation. With the utilization of MOM6 in transition to operational use for ocean monitoring at NCEP, and as the ocean component of coupled models being developed for Numerical Weather Prediction (NWP), we expect the LETKF-LaDA will be able to contribute to the whole DA system by successfully enhancing the prediction of ocean circulation by assimilating the drifter positions directly. The work is built using three main procedures: (1) developing necessary functionality in an online drifter module for MOM6; (2) using the regional GoM model with the realistic atmosphere surface forcing; and (3) the application of the LETKF-LaDA assimilating drifter positions.

## 3.2 Model Description

In the previous developments of MOM, from MOM1 in 1991 to MOM5 in 2012, the upgrade in each version of the model makes tremendous strides, such as improvements in the numerical methods, physical parameterizations and computational infrastructure. MOM has become an evolutionary tool to assist oceanographers to have a better understanding of ocean dynamics. Despite the valuable advancement in the previous versions of MOM, there were still severe issues in the mesoscale eddy equation, which caused the spurious diapycnal mixing associate with numerical advection (Griffies et al., 2000; Ilcak et al., 2012).

MOM6 is the latest ocean modeling project in the effort to make revolutionary advances while solving the remaining issues from all previous versions of MOM. The main improvements include the general-coordination formation, conservative representation of wetting and drying and the novel parameterization of sub-grid scale physics (Griffies et al., 2015). The application of the Arakawa C-grid, instead of the the Arakawa B-grid in MOM5, enables a more accurate simulation of mesoscale eddies. In addition, the parameterization of mesoscale and submesoscale eddies has also been updated by incorporating the cutting-edge theories, such as the mesoscale eddy kinetic energy methods of Marshall and Adcroft (2010), the resolution function of Hallberg (2013), the improved energy budget-based parameterization for the mesoscale eddy diffusivity of Jansen et al. (2015), and the newly updated algorithm in parameterizing the submesoscale eddy of Fox-Kemper et al. (2011, 2008). This advantage of MOM6 shows a great potential in the application of LETKF-LaDA

with drifter observations, targeting to forecast the mesoscale and submesoscale currents with high accuracy.

In this section, we will explain in detail about building up the necessary ingredients for the later use of combing LETKF-LaDA and MOM6 as the forecast model to improve the prediction of surface currents in the GoM.

### 3.2.1 Construction of a Gulf of Mexico regional model configuration

The model domain is defined in a region extending 18°N to 30.5°N and 262°E to 279.5°E using spherical coordinates and a horizontal resolution 1/24°. The horizontal grid is generated by interpolating from the gridded bathymetric data set of General Bathymetric Chart of the Oceans (GEBCO). The vertical coordinates used in the current study are defined as 75-level $z^*$ coordinates, though the Arbitrary Lagrangian-Eulerican (ALE) vertical coordinate is available in MOM6. We will take account of the impact of time-variant ALE in the future research.

The surface atmospheric forcing is provided by the 20th Century Reanalysis version 3 (20CRv3) for winds, temperature, specific heat, precipitation, pressure and downward shortwave and longwave radiation. The 20th Century Reanalysis (20CR) is the pioneer project producing an ensemble of sub-daily global atmospheric conditions spanning over 100 years. According to Slivinski et al. (2019b), the previous version 2c suffers from significant issues including inaccurate estimates of confidence and a global sea level pressure bias in the mid-19th century. The latest version 20CRv3 utilizes improved DA methods and a more advanced forecast model,

in order to attenuate the significant issues involving estimates of confidence and a global sea level pressure in the mid-19th century (Slivinski et al., 2019a).

All the atmospheric boundary forcing used in this chapter is time-variant, with an update frequency of 3 hours. In addition, the ensemble forcing for each member is selected from the 80-member 20CRv3 ensemble data set. Figure 3.2 shows the monthly ensemble spread of the 20CRv3 eastward/northward wind fields at a height of 10 meters in 2012.

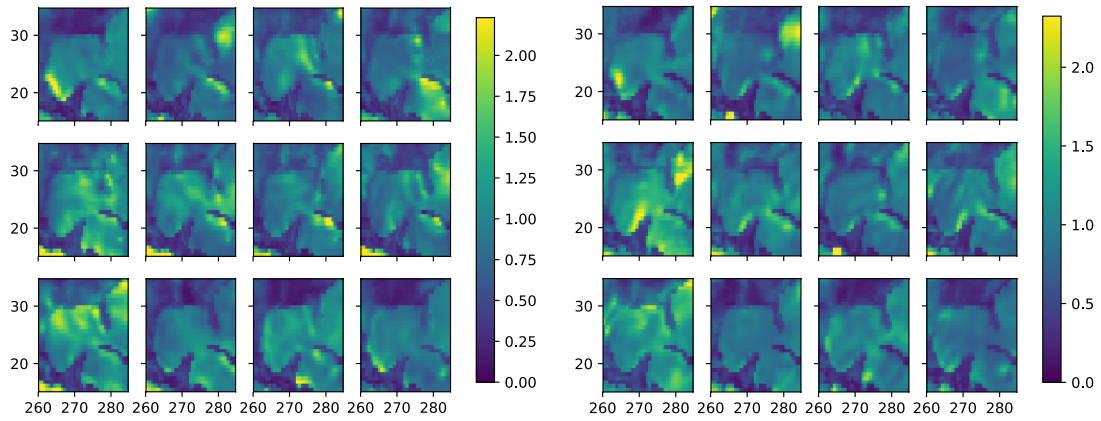


Figure 3.2: Monthly ensemble spread of eastward (left) and northward (right) wind fields at 10 meter height in 2012. Each panel represents the ensemble spread at the beginning of each month.

The initial data of the potential temperature and salinity is mapped from the objectively analyzed (1° grid) climatological fields from the World Ocean Atlas 2005 (WOA05). As a preliminary stage of the extension to a realistic scenario, we utilize the default "wall" boundary condition, where no flux exists at the eastern and southern boundaries. Therefore, with this experiment setting, no loop current

is expected to appear during the experimenting time period. An examination of a more accurate boundary condition based on the SODA data set will be conducted in the future research.

Figure 3.3 shows snapshots of SSH and sea surface temperature (SST) after 215 days (around 7 months) free run on the GoM model with 1/24° resolution using the 20CRv3 atmosphere forcing of ensemble member No.001.

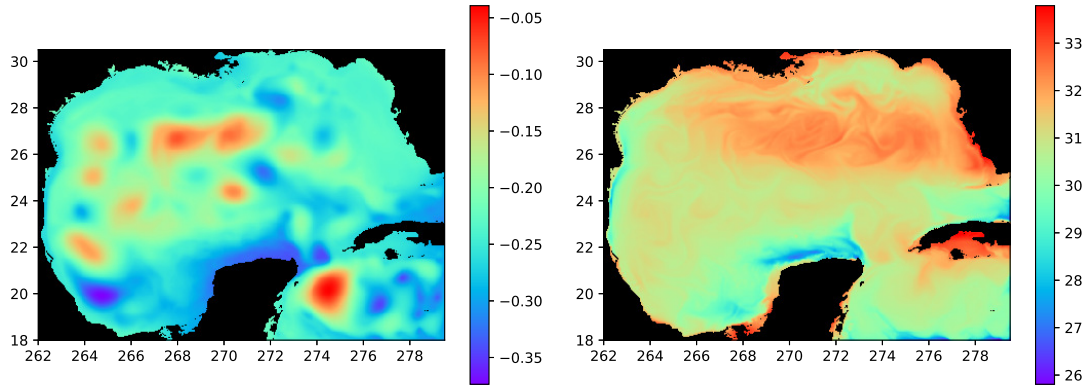


Figure 3.3: Screenshots of SSH (left; unit: m) and SST (right; unit: °C) at the end of 215-day free run of GoM with resolution 1/24° and "wall" boundary condition. This free run is initialized with the WOA05 climatological fields and forced with 20CRv3 atmosphere forcing for ensemble member No.001.

### 3.2.2 Construction of drifter module for MOM6

As shown in the previous chapter, one of the reasons to extend the application of LETKF-LaDA to MOM6 is the deficiency of the drifter simulation module embedded in the MOM4p1 and MOM5. The online drifter code in previous versions of

MOM mainly suffers from the complexities of a distributed computing environment, resulting in serious issues related to the repetition and the loss of drifters during a free model run. This prevents us from deploying a number of drifters in a specific region and employing the historical database, such as the GDP and the GLAD, including hundreds and even thousands deployments.

As such, many offline drifter simulation operators and analysis toolboxes (Delandmeter and van Sebille, 2019; Lange and van Sebille, 2017) have been invented as remedies to attenuate the inconvenience brought by the flawed online drifter simulation modules. However, because of their offline setup, the background flow field can be only input by a sequence of the general ocean model output. This can consume extra storage space and excessive I/O when implementing the drifter simulator, specially when using a high-resolution model. More importantly, since the output interval is in general larger than the model time step, the drifter trajectory generated by these offline toolboxes is less accurate because it cannot be obtained by integrating the instantaneous interpolated flow velocity at each model time step.

Here, we build an online drifter simulation module within MOM6. This will benefit not only future research studying the impact of different LaDA methods involving the MOM6, but also other applications related to the drifter/particle simulation within MOM6. The main idea in constructing the drifter module is to regard the complete processor-communication functions of the MOM6 iceberg module as a template, in order to create a well distributed background flow environment for the drifter behaviors. At each model time step, flow velocities are defined on the C-grid (as shown in the green diamond labels of Figure 3.4) in each data domain, therefore

we first shift the C-grid velocities to the B-grid (the corner of each cell). Specifically, we shift all the zonal velocity northward and all the meridional velocity westward. We then focus on each local cell (Figure 3.4(a)) to implement the bilinear interpolation of velocities into the drifter location in this local rectangle cell. A rotation map (i.e. the Jacobian matrix) defined on each tracer cell (i.e. blue circle marks in Figure 3.4) is multiplied by the interpolation coefficient vector to convert interpolating relations within the local rectilinear cell to those within an actual curvilinear cell (see Figure 3.4(b)). Then the drifter velocity is obtained within spherical coordinates. Using this instantaneous velocity, the drifter location at the end of each model time step is then determined by using the 4th order of Runge Kutta method.

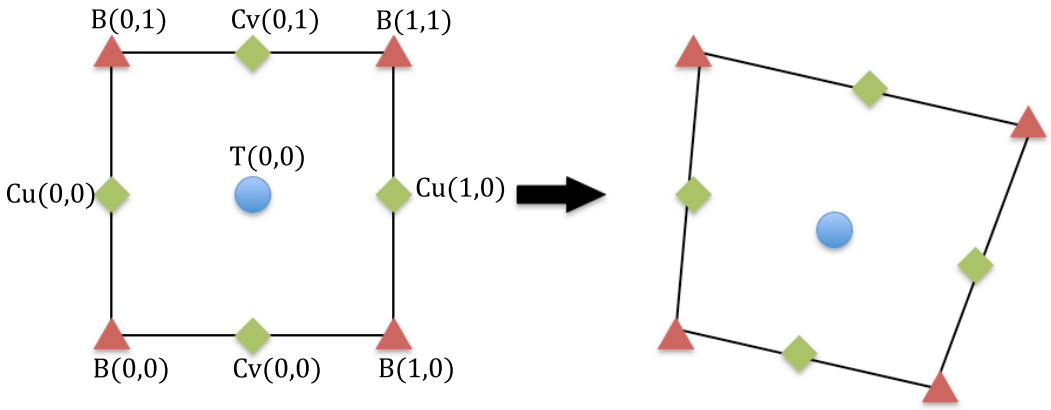


Figure 3.4: (letf) Different grids defined on the local unit rectlinear cell with B-grid (brown triangle labels), C-grid (green diamond labels; "Cu" for zonal velocity and "Cv" for meridional velocity) and tracer grid (blue circle label). (right) Representation of different types of grid in the actual curvilinear cell.

Figure 3.5 shows a drifter trajectory for one-month model run with 1/24° resolution for the background flow field. In the enlarged panel, an inertial oscillation

can be observed through the drifter trajectory.

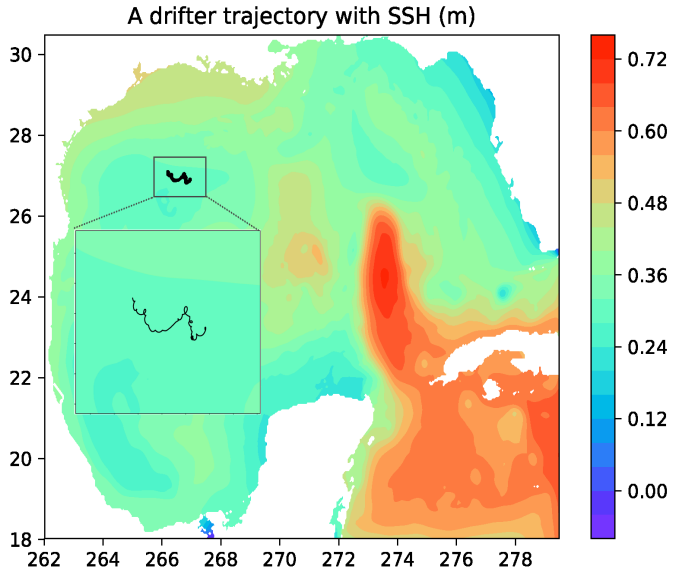


Figure 3.5: The trajectory of a drifter initially deployed at (266.96, 26.89) from 01/01/2012-01/31/2012 with the background SSH (unit: m) contour. The Gulf of Mexico model is of 1/24 degree horizontal resolution and is forced by the surface fields of 20CRv3. The initial condition and the boundary condition are both provided by the SODA Dataset (instead of the "wall" boundary condition). The zoomed-in figure of the drifter trajectory is shown in the embedded panel.

Some beneficial attributes of the new online drifter module must be highlighted, in comparison to the previous version:

1. Completely solves the issues of repeating and disappearing drifters.

2. Besides the drifter location, the drifter RESTART file also includes the extra interpolated variable values: temperature, salinity, and zonal and meridional

velocity located at the terminal drifter location.

3. In addition to the drifter RESTART file, the latest online drifter module is also capable of generating the drifter trajectory file with the location update frequency of a time scale of one minute. This will make it possible to fully utilize high-frequency drifter databases, such as GLAD in future research.

Yet, further improvements and upgrades must still be made in the current drifter simulation code:

1. The drifter simulation module is limited to simulate the behaviors of only surface drifters. The selection of drifters on different depth levels should also be included in a future development of the drifter module, which will be a promising feature to simulate the trajectory of the drifting platforms in the deep ocean, such as the Argo floats.

2. The current drifter module only takes account of the 2D dynamics. The consideration of 3D dynamics in the future improvement would make the drifter module more realistic.

## 3.3 Experiments Setup

We use the "identical twin" approach of OSSEs as a preliminary stage of testing the impact of LETKF-LaDA, where we use the same model configuration for both the nature run and the forecast model, i.e. 1/24° resolution and the "wall" boundary condition. Since 20CRv3 contains the atmosphere surface forcing of an 80-member

ensemble, the nature run is defined to be forced by the mean forcing of the 20CRv3 80-member ensemble in year 2012. The initial conditions of the 3D temperature and salinity are obtained by using the World Ocean Atlas 2005 climatology database. The nature run starts on Jan 1st, 2012, and 300 drifters are deployed randomly in the whole region on August 3rd, 2012 (Figure 3.6) at 1m depth in order to emulate the GLAD setup. Figure 3.7 shows a point profile of temperature and salinity in the nature run on August 3rd, 2012. The drifter observations are collected every day with observation error 0.08° in both longitude and latitude. We emulate the temperature and salinity measurement of the drifters by bilinearly interpolating the "true" underlying flow fields to the observed drifter locations with observation error 0.4 °C for temperature and 0.2 psu for salinity. Assuming that the dominant errors in the observations are errors of representativeness, the choices of the above observation errors all depend on the averaging error growth rate in the first 24h forecast. The observation windows are defined as one day for all observation types, which is the same setup as Jacobs et al. (2014).

We set the ensemble size as $K = 26$ and the ensemble atmosphere surface forcing selected as the first 26 members in the ensemble of 20CRv3. To initialize the ensemble for LETKF-LaDA, we first spin up each ensemble ocean member for 215 days (around 7 months) without drifters. Figure 3.8 shows the time variation of the ensemble spread in SSH, SST, and sea surface zonal and meridional velocity (SSU/SSV) during the 215 days, which shows that the ensemble errors are saturated enough at the end of this time period. Identically to the nature run, we populate the same 300 drifters ($N_D = 300$) drogued at 1m depth as shown in Figure 3.6 on August

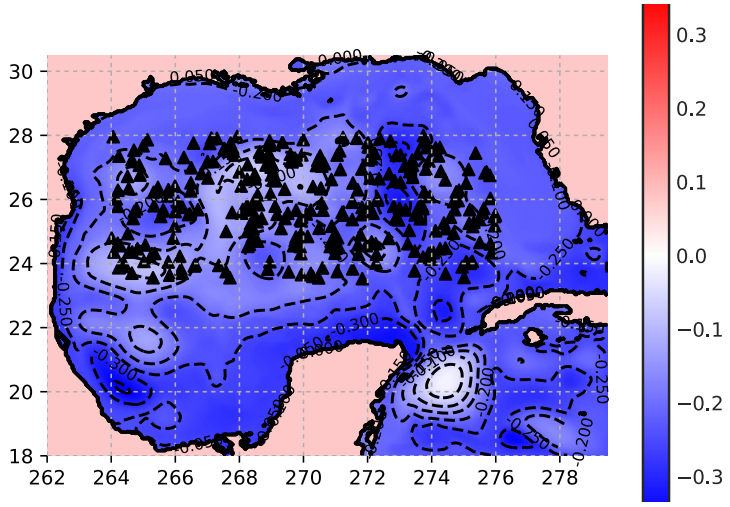


Figure 3.6: Deployments of the initial 300 drifters randomly distributed in the region 264°E∼ 276°E, 23.5°N∼ 28°N with background contour (contour interval is 0.05m) as the true SSH (m) at the end of the first 215 days free model run.

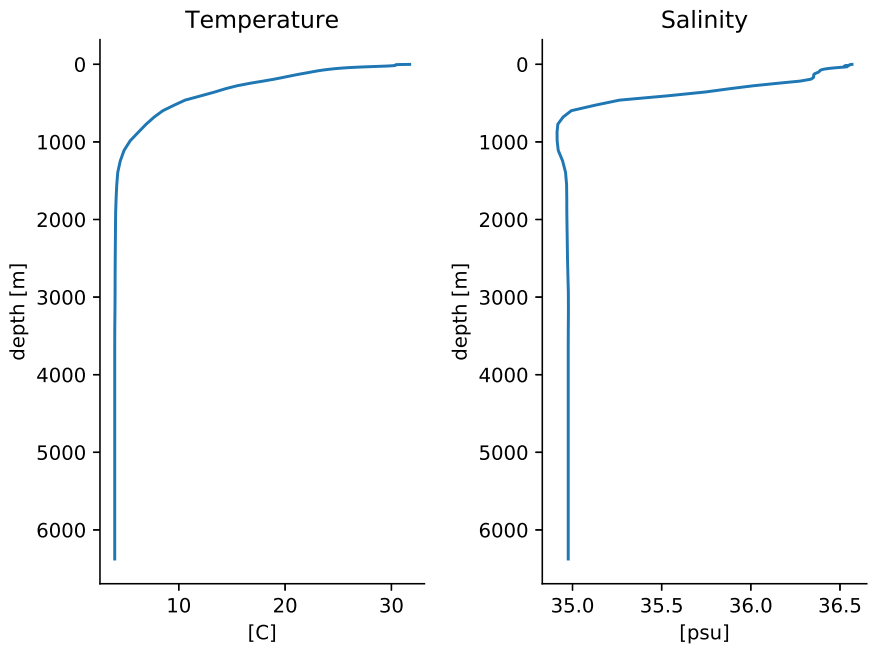


Figure 3.7: A point profile of temperature and salinity on August 3rd, 2012, locating at 266.271°E 24.3125°N .

3rd, 2012. The drifters are advected by the prescribed ensemble surface currents for one day to generate the initial ensemble spread for drifter state variables. The choice of the relatively short one-day spin-up process for the drifter state variables is due to the high refinement within the horizontal grid of the forecast model. Like the experiment setting in Chapter 2, and matching the observation frequency, the LETKF-LaDA uses a daily analysis cycle following Jacobs et al. (2014). The localization radius is tuned to be as long as the baroclinic Rossby radius of deformation $\gamma_{LETKF} = L_R$, which is related to the dominant length scale of unstable waves in a stratified shear flow (Chelton et al., 1998).

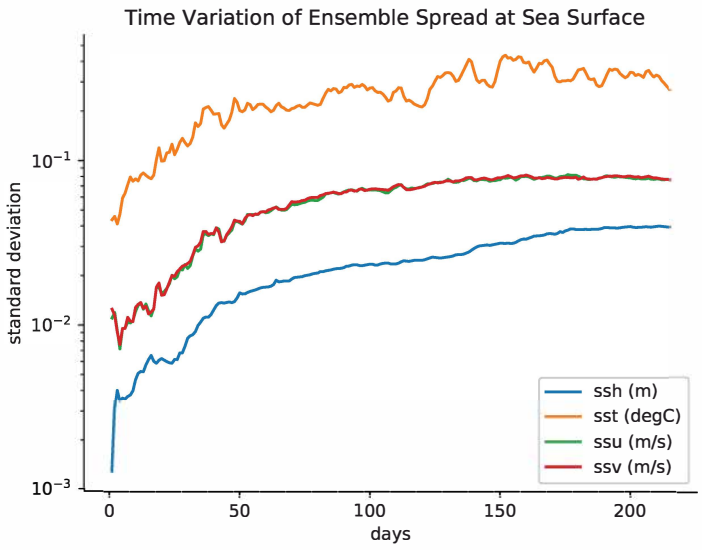


Figure 3.8: Time variation of the area-averaged ensemble spread during the first step of spin-up process, i.e. 215 days.

The free run is initialized as the ensemble mean flow (among the 26 members) obtained on August 4th, 2012, the end of the spin-up process for both of the flow and drifter state variables. The atmospheric surface forcing is defined by the 26-member

ensemble mean, which can be regarded as a combination of the "true" atmosphere surface forcing and a time-variant marginal error. No extra condition or mechanism is used to constrain the free model run in the whole process. In this chapter, we also use the same percentage norm (2.4) as a metric to quantify the estimation error.

## 3.4 Results and Discussion

### 3.4.1 LETKF-LaDA assimilating drifter locations only

Among all the previous work in the augmented-state LaDA approach, the assimilation of the direct drifter positions is only used for analyzing flow velocity and surface velocity related fields (e.g. the mean height field in Salman et al. (2006); Slivinski et al. (2015)). Chapter 2 showed a promising result for improving the 3D temperature/salinity using the covariance between the temperature/salinity of the whole local water column and the surface drifter states using the eddy-permitting model. Two experiments are conducted in order to examine the impact of LaDA on the temperature and salinity: (1) a baseline of the free run experiment; and (2) using LaDA assimilating drifter positions to analyze all ocean state variables including temperature, salinity, and zonal and meridional velocity.

Figure 3.9 shows the time variation of the accumulated forecast error from $h = 1$m to $h = 1200$m in temperature, salinity, and kinetic energy defined in (2.4). Among all the basic 3D ocean fields, the LETKF-LaDA performs better than the free run, except for the salinity. Different from chapter 2, we find that assimilating only the drifter locations has difficulties in constraining the salinity field, which could

be caused by the involvement of the atmosphere surface fields (e.g. precipitation, or temperature; Lima et al. (2019)) in this more realistic scenario. This indicates that additional observations related to ocean salinity are necessary for improving estimation in future research assimilating historical drifter data. A significant error reduction in kinetic energy appears in the first 6 days, while the error slowly increases after. This could be due to the time varying wind field and surface temperature for the prescribed nature run, which provides a different error growth rate at each time step. Figure 3.10 shows the time variation of the area-averaged RMSEs of surface temperature and winds, where it shows an increase of the error growth in the wind fields starting from day 6. For temperature, the LETKF-LaDA outperforms the free run and the variation of the ocean temperature error is strongly influenced by the error variation of the atmosphere surface temperature (see the blue line in Figure 3.10(b)).

We next consider the vertical influence of the LETKF-LaDA in estimating the accuracy at the terminal time step, i.e. the end of 15th day (see Figure 3.11). For temperature, LETKF-LaDA is less accurate than the free run in the top 10m depth, while the estimation is improved below. This implies that the temperature observation is needed to improve the estimation of surface temperature. For salinity field, assimilating only the surface drifter position only provides a better estimation in the surface layer of about 10m to 20m depth. For the estimation of kinetic energy, the positive influence of the surface drifter observation can be extended to around 800m depth.

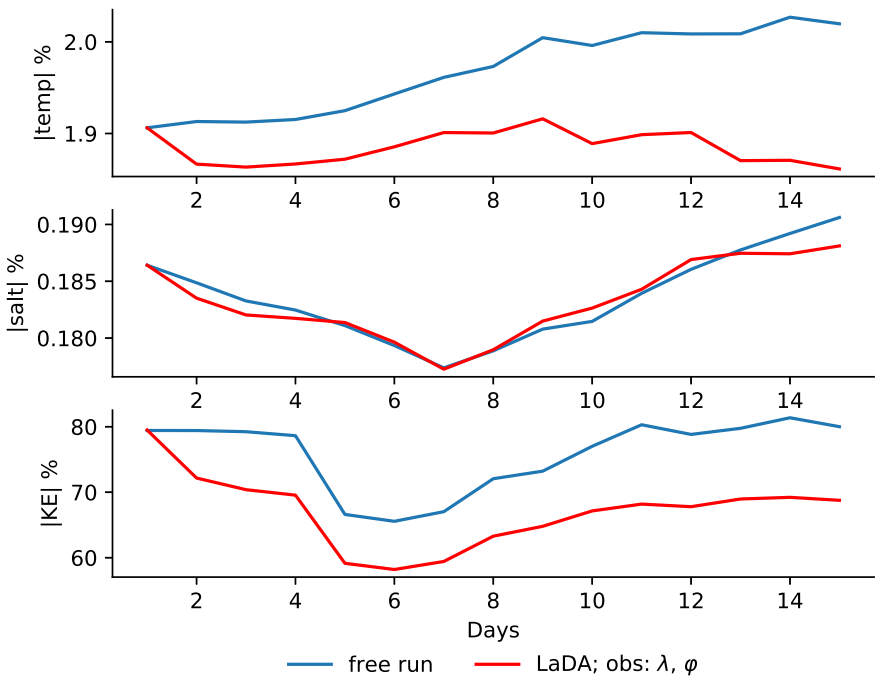


Figure 3.9: Forecast error norms of temperature, salinity, kinetic energy in the given period [0, 15] using formula (2.4) for free run (blue line) and LETKF-LaDA (red line) with localization radius $L_R$. The error norms of the ocean flow fields are aggregated from 1m depth to 1200m depth (i.e. $h_o = 1$m and $h = 1200$m). $\lambda$ and $\varphi$ stand for drifter longitude and latitude.

### 3.4.2 The impact of LETKF-LaDA on assimilating drifter locations and TS measurements

Carrier et al. (2014) conducted two experiments using 4D-Var: (1) assimilating temperature and salinity (hereafter referred as TS) only; and (2) assimilating approximated velocity at the sampled drifter positions together with TS. The TS observation used in their work included both remotely sensed and in situ ocean

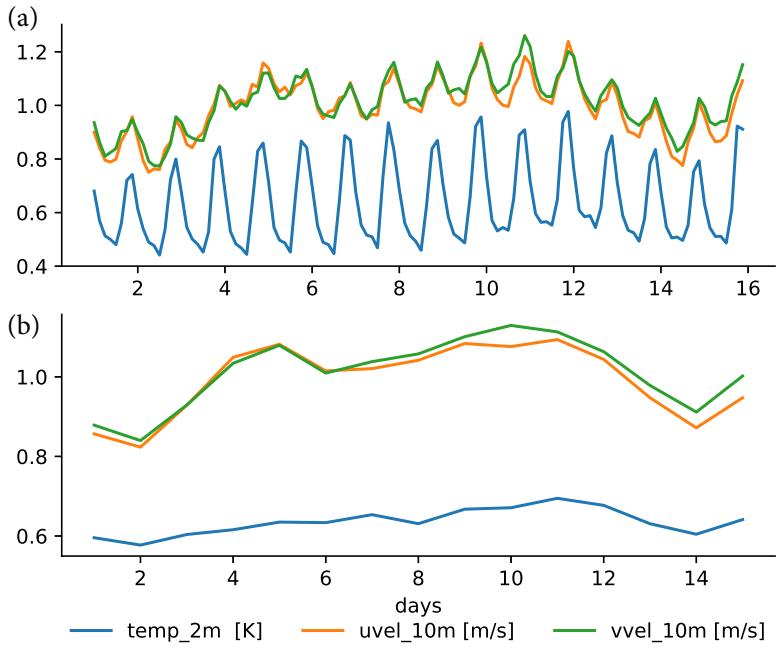


Figure 3.10: Area-averaged RMSEs of the surface temperature (blue, unit: K), zonal wind field (orange, unit: m/s), and meridional wind field (green, unite: m/s) with, (a) the original update requency of 3hrs (upper panel), and (b) the daily average (lower panel).

observation: (Geostationary Operational Environmental Satellite) GOES-East sea surface temperatures (SSTs), Argo profiling floats, expendable bathythermographs (XBT) and drifter buoys. By first comparing the free run and the assimilation of TS alone, they concluded that it is difficult for the TS observations to sufficiently constrain the error in the model solution of the surface velocity field. The addition of the "proxy" drifter velocity observations not only match the result in temperature and salinity states, but also further improve the estimation of the velocity field.

In this subsection, we consider the involvement of the surface TS measurement in LaDA, which are emulated by bilinearly interpolating the corresponding flow fields

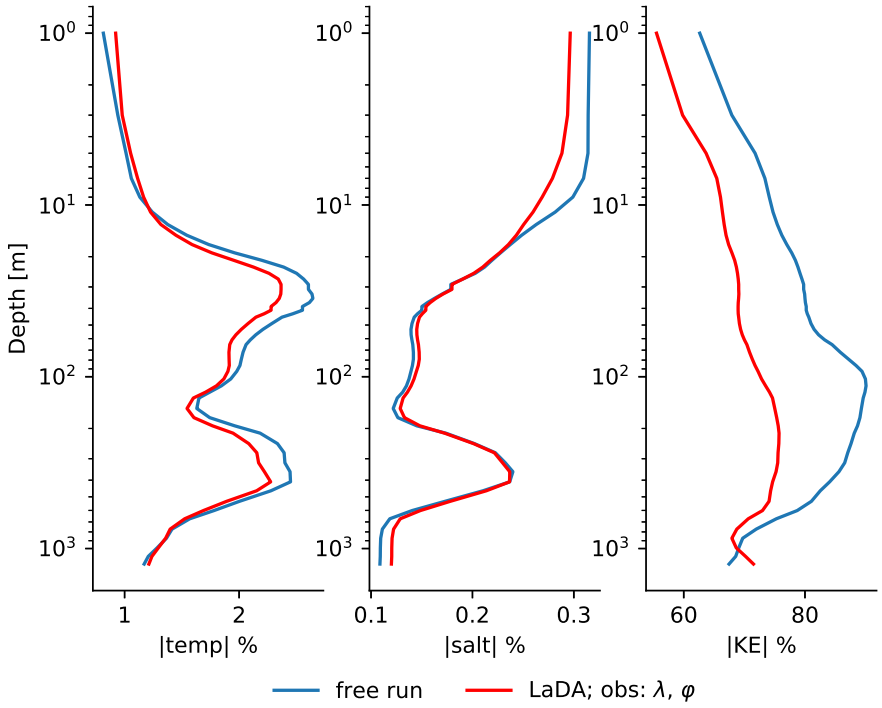


Figure 3.11: Forecast error in temperature, salinity and kinetic energy at the end of 15th DA cycle along the vertical direction with experiments of free run (blue line) and LETKF-LaDA using $L_R$ (red line). These quantities are evaluated by formula (2.4) at each individual vertical level with $h_o = h$. $\lambda$ and $\varphi$ stand for drifter longitude and latitude.

to the drifter positions. Four experiments are conducted: (1) a free run; (2) a control run obtained by using LETKF to assimilate only the TS carried by drifters with the same ensemble size (i.e. $K = 26$) and localization radius (i.e $\gamma_{LETKF} = L_R$); (3) using LETKF-LaDA to assimilate only the drifter position; and (4) using LETKF-LaDA to assimilate surface TS together with the drifter positions.

Similar to Carrier et al. (2014), we first compare the free run with the control run in Figure 3.12. It shows that assimilating the surface TS measurements can

significantly improve the salinity estimation, and slightly correct the temperature forecast after the 9th cycle. Different from the results in Carrier et al. (2014), we notice that the surface TS is indeed capable of constraining the velocity field using LETKF, though the error reduction is marginal. We then compare the results between the control run and the two experiments using the LETKF-LaDAs (Figure 3.12). By adding the TS measurement, the LETKF-LaDA can outperform the control run in salinity and further improve the prediction in temperature, though the estimation in velocity is comparable to the case without TS observations.

Figure 3.13 shows the vertical error variation at the terminal time step horizontally averaged at each level. We conclude that by adding the surface TS observations, LETKF-LaDA is largely improved in the estimation of surface salinity and the impact can be extended to around 40m depth. Moreover, the addition of surface TS observations improves the estimation of the surface temperature compared to using the LETKF-LaDA assimilating the drifter positions only. The two LETKF-LaDA cases exhibit significantly reduction in the velocity error above 800m, while the benefit fades away in the deeper ocean. This can be due to the fact that the velocity error in the deeper ocean grows more linearly than the surface velocity error does and over longer timescales, which implies a potential value of including the velocity-related information in deep ocean from Argo float trajectories for the future study of LaDA.

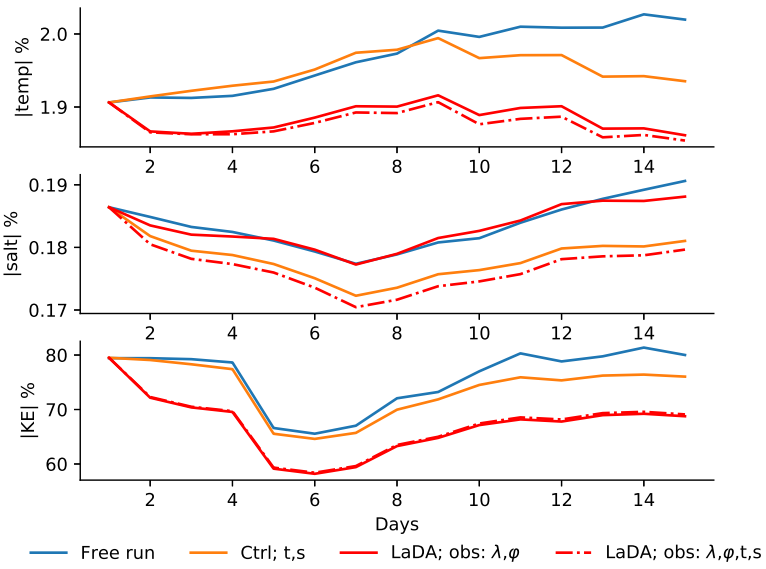


Figure 3.12: Forecast error norms of temperature, salinity, kinetic energy in the given period [0, 15] using formula (2.4) for experiments: (1) free run (blue line); (2) control run (orange line); (3) using LETKF-LaDA to assimilatie drifter locations (red solid line); and (4) using LETKF-LaDA to assimilate both drifter locations and TS measurements on drifters (red dotdash line). Localization radius of all the experiments are defined as $L_R$. The error norms of the ocean flow fields are aggregated from 1m depth to 1200m depth (i.e. $h_o = 1$m and $h = 1200$m). $\lambda$ and $\varphi$ stand for drifter longitude and latitude.

### 3.4.3 Comparing LaDA with EuDA

We now compare the performance of LaDA versus EuDA. The velocity observations are simulated by directly interpolating the true surface ocean velocity fields to the corresponding drifter locations. Notice that in the operational use where the true gridded flow is unknown, the instantaneous velocity observations are in general

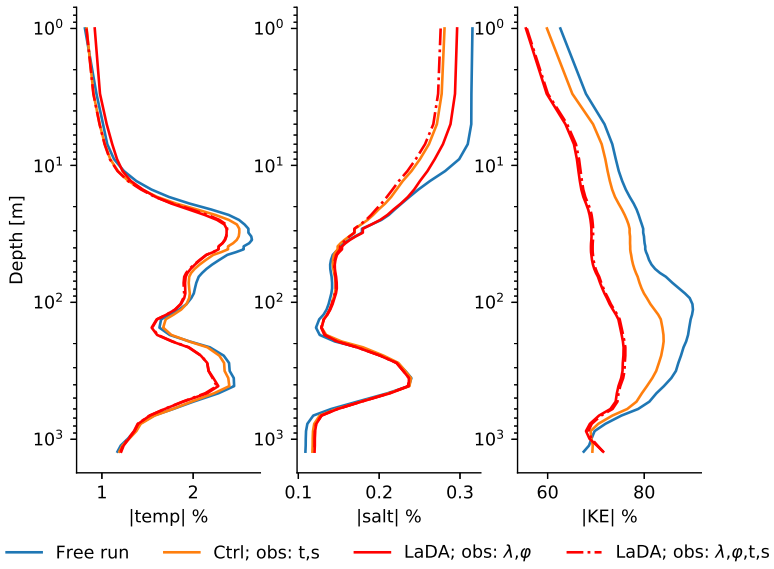


Figure 3.13: Error in temperature, salinity and kinetic energy at the end of 10th DA cycle along the vertical direction with experiments: (1) free run (blue line) ;control run (orange line); (2) using the LETKF-LaDA to assimilatie drifter locations (red solid line); and (3) using the LETKF-LaDA to assimilate both drifter locations and TS measurements on drifters (red dotdash line). Localization radius of all the experiments are defined as $L_R$. These quantities are evaluated by formula (2.4) at each individual vertical level with $h_o = h$. $\lambda$ and $\varphi$ stand for drifter longitude and latitude.

approximated by a sequence of the drifter locations and known with much less accuracy than the observations we have simulated in this subsection. Therefore, the EuDA that we present here is under an idealized scenario and the simulated velocity observations contains enough information of nonlinear dynamics in the surface flow. The observation error for velocity is chosen as 0.08m/s, which is comparable to the observation error of the drifter location as 0.08° per day. The justification of this

is based on the fact that $1/24^\circ$ is approximately equivalent to 3.75km within the region of GoM and the details are shown as following,

$$1^\circ = 3750 \text{ m}/(1/24^\circ) = 9 \times 10^4 \text{ m}.$$

$$1 \text{ day} = 24 \times 3600 \text{ s}.$$

Therefore,

$$0.08^\circ/\text{day} \approx 0.08 \times 9 \times 10^4 \text{ m}/(24 \times 3600 \text{ s}) \approx 0.08 \text{ m/s}$$

Four experiments are conducted to compare the performance between the LaDA and the ideal EuDA using LETKF: (1) the free run; (2) the same control run defined in the previous subsection; (3) using the LETKF-LaDA to assimilate surface TS in addition to the drifter locations; and (4) using the traditional LETKF to assimilate Eulerian flow velocity and TS together. The time variation of the forecast error is shown in Figure 3.14. It shows that the velocity error of the LETKF-LaDA is marginally smaller than the one of the ideal EuDA. Furthermore, the LETKF-LaDA can further improve the temperature estimation than the ideal EuDA, though the salinity forecast obtained by the LETKF-LaDA is less accurate.

In the vertical comparison at the terminal time step (Figure 3.15), we observe that the vertical forecast error in both of the LaDA and the ideal EuDA are comparable, and they are both better than the control run and the free run in predicting ocean temperature and kinetic energy above around 800m. In addition, the ideal EuDA performs slightly better than LaDA in estimating the velocity below 800m. In order to determine the vertical comparison in details between the LaDA and the EuDA, we utilize the following index to describe their performance.

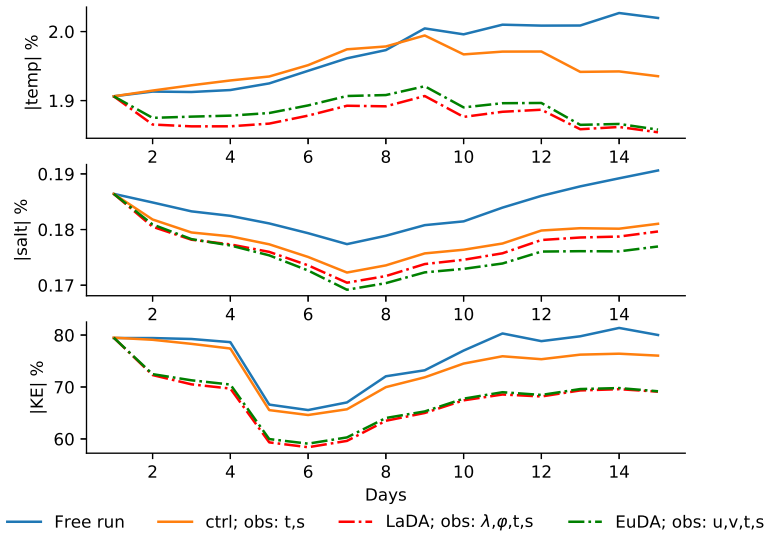


Figure 3.14: Forecast error norms of temperature, salinity, kinetic energy in the given period [0, 15] using formula (2.4) for experiments: (1) free run (blue line); (2) control run (orange line); (3) using LETKF-LaDA to assimilatie both drifter locations and TS measurements on drifters (red dotdash line); and (4) using LETKF-EuDA to assimilate "proxy" velocity and TS measurements on drifters (green dotdash line). Localization radius of all the experiments are defined as $L_R$. The error norms of the ocean flow fields are aggregated from 1m depth to 1200m depth (i.e. $h_o = 1$m and $h = 1200$m). $\lambda$ and $\varphi$ stand for drifter longitude and latitude.

Carrier et al. (2014) introduced a metric, called skill score (SS), intended for comparing the results of two experiments that have marginal difference against each other. The skill score of the experiment 1 based on the performance of the experiment 2 is shown as following,

$$\mathrm{SS} = 1.0 - \frac{\mathrm{RMSE}_1}{\mathrm{RMSE}_2}, \tag{3.1}$$

where experiment 2 can be regarded as the baseline of the comparison. This for-

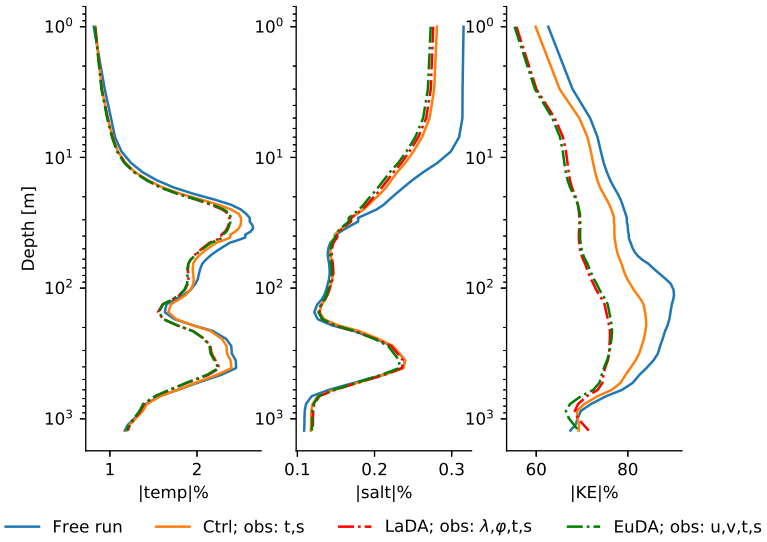


Figure 3.15: Error in temperature, salinity and kinetic energy at the end of 15th DA cycle along the vertical direction with experiments: (1) free run (blue line); (2) control run (orange line); (3) using LETKF-LaDA to assimilatie both drifter locations and TS measurements on drifters (red dotdash line); and (4) using LETKF-EuDA to assimilate "proxy" velocity and TS measurements on drifters (green dotdash line). Localization radius of all the experiments are defined as $L_R$. These quantities are evaluated by formula (2.4) at each individual vertical level with $h_o = h$. $\lambda$ and $\varphi$ stand for drifter longitude and latitude.

mula indicates that if the experiment 1 performs better (worse) than the baseline experiment 2 in terms of a lower (higher) error, then the SS metric will be positive (negative).

Figure 3.16 shows the SS of the LaDA based on the performance of the EuDA. The error reduction of the LaDA is relatively less than the one from the ideal EuDA in salinity for almost all the levels, while it is better in improving the velocity field

in the depth level between 20m and 200m.

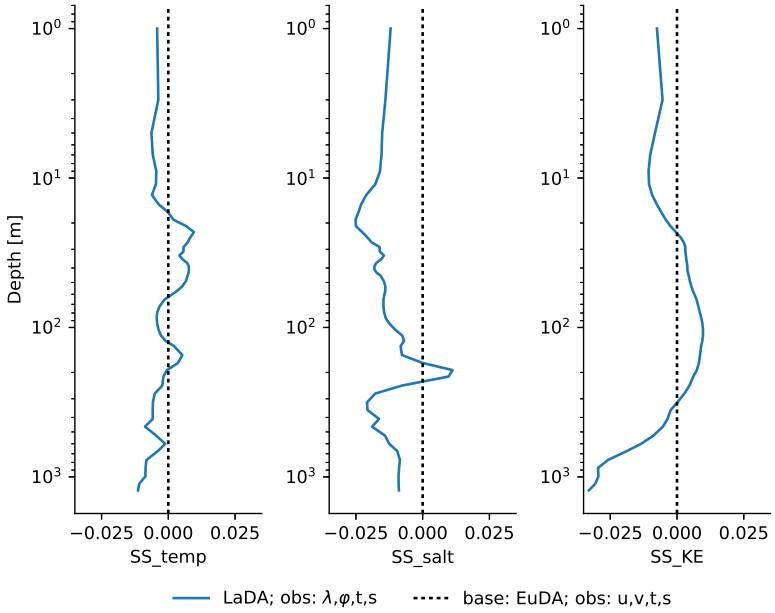


Figure 3.16: Vertical SS scores at the end of 10th DA cycle along the vertical direction of using LETKF-LaDA to assimilatie both drifter locations and TS measurements on drifters (blue line) with the base of using LETKF-EuDA to assimilate "proxy" velocity and TS measurements on drifters (black dashed line). Localization radius of all the experiments are defined as $L_R$. $\lambda$ and $\varphi$ stand for drifter longitude and latitude.

### 3.4.4 The impact of variable localization on LETKF-LaDA

The reason why the LaDA doesn't perform as well as the EuDA in forecasting the salinity after including the salinity observation can be manifold, such as the length of the assimilation window and unbalanced setup on the observation error. One possible reason can be due to the high dependence of the LaDA on the drifter

forecast error. The addition of the TS observations results in the misfitting of the the drifter position, which provides a less accurate estimation in the drifter positions and therefore provides less accurate covariance matrix in cost function at the next step (see Figure 3.17). As such, we consider of employing a "variable localization" technique (Kang et al., 2011) to constrain the error growth corresponding to the drifter position by zeroing out the covariance components between the drifter positions (i.e. longitude and latitude) and the drifter TS measurements.

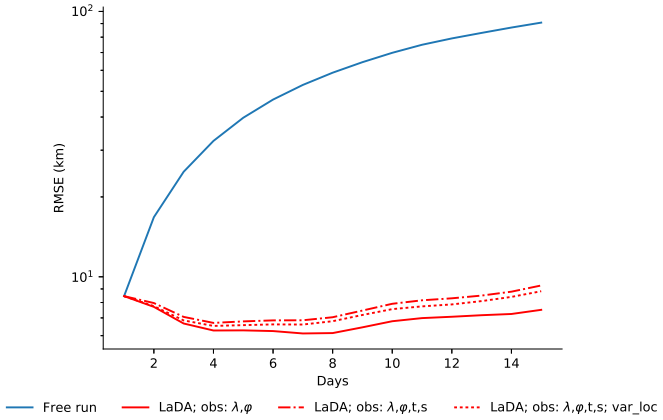


Figure 3.17: RMSEs of the forecast drifter locations generated by (1) free run (blue line); (2) LETKF-LaDA assimiating on the drifter locations (red solid line); (3) LETKF-LaDA assimilating all types of the drifter measurements (red dotdash line); and (4) LETKF-LaDA assimilating all types of the drifter observation using variable localization (red dotted line). $\lambda$ and $\varphi$ stand for drifter longitude and latitude.

Figure 3.18 shows the time variation of the SS for the experiment using the aforementioned variable localization against the base of the one without the variable localization. We can observe that the variable localization does show a positive impact in estimating the salinity field and this has eliminated around 10% advantage

of EuDA with respect to the salinity (compared with Figure 3.16), while this results in the slightly inaccurate prediction of velocity in the first 9 DA cycles and of temperature between the 7th and 11th day.

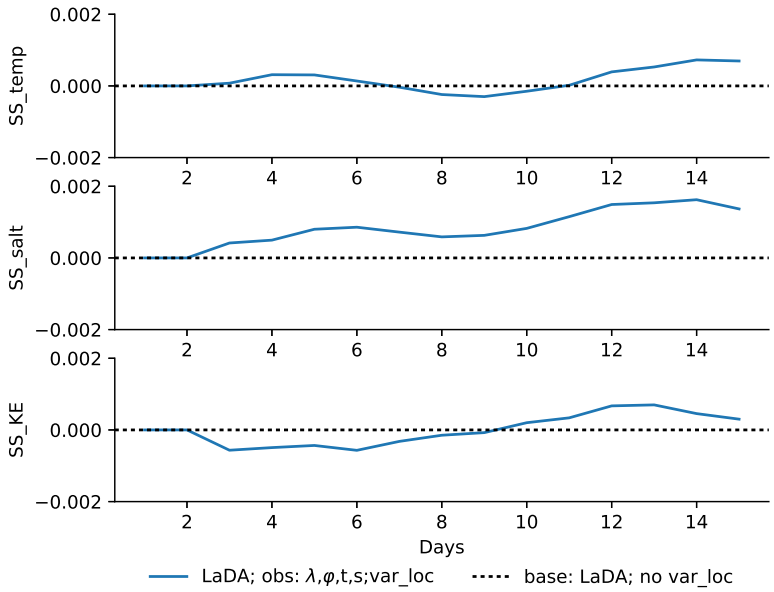


Figure 3.18: SS score of temperature, salinity, kinetic energy and drifter states in the given period [0, 15] for the experiment of LETKF-LaDA to assimilatie ALL measurements on drifters using variable localization (blue line) based on the LETKF-LaDA without variable localization (black dashed line). Localization radius of all the experiments are defined as $L_R$. The error norms of the ocean flow fields are aggregated from 1m depth to 1200m depth (i.e. $h_o = 1$m and $h = 1200$m). $\lambda$ and $\varphi$ stand for drifter longitude and latitude.

## 3.5 Conclusion

In this chapter, we upgrade the examination of the LETKF-LaDA into a more realistic scenario where we build up an eddy-resolving model with the configuration of GoM and it is forced with the historical surface conditions sampled from 20CRv3. In addition, we build up an online drifter module adaptive to MOM6, which solves the deficiency of the drifter module appearing in the previous versions of MOM, such as the loss and repetition issues in simulating a large number of drifters. By implementing the drifter module within the high-resolution gridded flow, we can also observe the phenomena of the inertial oscillation through the drifter trajectory. However, the current drifter module is still limited in simulating only the surface drifters at the first depth level of the model, i.e. 1m in the given model configuration and the release of this constraint will be tackled in the follow-on research.

As in the previous chapter, we use the "identical twin" experiments of OSSEs to test the performance of LETKF-LaDA. Compared with the free run, we conclude that the LETKF-LaDA improves the forecast error in not only the velocity but also the temperature as well, even when the observations only include the drifter positions. The misfitting issue in the salinity field could be due to the additional effect in the surface precipitation and 2-meter temperature error. The positive impact of the error reduction of kinetic energy can be extended to the deep ocean around 800m depth. We then investigate the impact of the drifter location observations on the LETKF-LaDA by defining the control run as the one simply assimilating the surface TS. We notice that without temperature observations, LETKF-LaDA is still

able to further reduce the error in temperature and velocity versus the experiment assimilating TS alone. Even though assimilating the drifter locations alone cannot constrain the salinity as well as the control run and free run, the addition of the TS measurements is able to make up this deficiency in LETKF-LaDA.

We then compare the performance of the LaDA and the idealized EuDA when assimilating the surface TS measurements together with the drifter positions. Results show that the LaDA outperforms the EuDA in improving the estimation of temperature and velocity, while it is less efficient in improving the salinity. We propose a possible remedy for this issue to apply the variable localization strategy by zeroing out the covariance components between the drifter positions and the TS measurement.

In the follow-on research, we will open the eastern and southern boundary of the given model and will use the historical database as the open boundary condition. The drifter observations will be replaced by the GLAD database recorded from August to October to examine the impact of LETKF-LaDA on the real data. Note that the GLAD database does not include any additional measurements besides the drifter locations and approximated velocity. Based on our previous work, we conclude that the TS measurements are necessary for LETKF-LaDA to further improve the estimation of all ocean state variables. Therefore, we need to add the remotely sensed and in situ TS observation together with the drifter position when later examining the LETKF-LaDA using real historical data.

# Chapter 4: Transition from the Uncoupled Media to the Coupled Media: the Dynamical Model and Data Assimilation

## 4.1 Introduction

In the previous two chapters, all the dynamical models and the data assimilation methods are demonstrated only within the ocean, though the ocean models are constrained by different atmosphere surface forcing. We observe, in chapter 3, that it is difficult to continuously reduce the error in ocean states when the ocean model is forced by the time-variant and high-frequency surface forcing. One of the reasons is possibly due to the single-direction fluxes from the atmosphere to the surface ocean and the high dependency of the given ocean models on the downward surface flux data. At each DA cycle, while we provide analysis temperature states with higher accuracy than the forecasts, there is neither any correcting feedback to the surface forcing nor the upward radiation fluxes to the atmosphere, which drives the forecast trajectory of the next cycle still inevitably veering away from a well initialized prediction. As such, a mutual communication between atmosphere and ocean at the interface-layer is a significant factor influencing the performance of DA methods.

In pioneering researches, the forecast models are evolving from multi-models in different independent media to a single model in coupled media, where downward and upward radiation fluxes are considered in the same model at the interface layer. Different DA method are emerging to improve the multi-media forecast states based on the choice of the forecast model and the definition of the covariances between states from different media. Uncoupled DA (UCDA) is a traditional DA strategy used in operational prediction centers, in which the atmosphere and ocean are predicted and analyzed separately. In recent research, for example, coupled models were applied as forecast models within the DA system in National Centers for Environment Prediction (NCEP; Saha et al. (2006)) and Geophysical Fluid Dynamics Laboratory (GFDL; Zhang et al. (2007)). In both of the aforementioned research studies, independent DA methods are applied to analyze the atmosphere and ocean states, though the bilateral communication on the interface layer is considered in the forecast model. This DA system is known as weakly coupled DA (WCDA). The latest development (Penny and coauthors, 2008; Penny and Hamill, 2017) in studying CDA is to treat the atmosphere and ocean as one coupled media in order to forecast and analyze states together, i.e. strongly coupled DA (SCDA).

Pecora and Carroll (1991, 2015) conclude that a necessary condition for a response system to synchronize with the driver system is the negative values of all the Lyapunov exponents (LEs), or also known as conditional LEs, for the response system. Suppose the nature dynamics is governed by a coupled driving system, then the forced ocean/atmosphere models can be ideally summarized as different response systems, with the surface forcing as the driving information provided by

the "nature" driving coupled system. In addition, the system of a forecast model and a DA method can be also viewed as a response system to the nature driving system, with the observation as the driving information. In our work, we utilize the idea of the drive/response system and the theory of synchronization from the perspectives of the LEs to discuss the convergence of different couplings in models and DA systems.

All demonstrations in this chapter are carried out with the simplified Modular Arbitrary-Order Ocean-Atmosphere Model (MAOOAM) developed by De Cruz et al. (2016), which is a three-layer (including two layers of atmosphere and one layer of ocean) truncated quasi-geostrophic model. The configuration and the choice of parameters are the same as the examination by Vannitsem and Lucarini (2016), where the state vectors are of dimension $N = 36$ with $N_a = 20$ for atmosphere and $N_o = 16$ for the ocean. As for the DA method, we focus on the performance using of Ensemble Kalman Transform Filter (EKTF) in this thesis while more results corresponding to variational methods are shown in Penny et al. (2019).

In section 4.2, we first derive the forced ocean/atmosphere model under the framework of drive-response system. The corresponding tangent linear model (TLM) related to these forced models are further determined for the computation of LEs. Different discussions based on LEs are elucidated in section 4.3, including: the comparisons between the uncoupled and coupled models, different couplings in DA and impacts of observations on SCDA. The examination of the forecast lead time inherited from the analyses generated by different coupling approaches is conducted at the end of the same section.

## 4.2 Synchronization and Lyapunov Exponents

### 4.2.1 Derivation of forced ocean/atmoshphere model

We define the coupled model governing the dynamics of both atmosphere and ocean states as following,

$$\frac{d\mathbf{x}^c}{dt} = \mathcal{M}(\mathbf{x}^c). \tag{4.1}$$

Categorizing the components inside $\mathbf{x}^c$ into atmosphere states $\mathbf{x}^a$ and ocean states $\mathbf{x}^o$, the above dynamical model can then be separated into two sub-models associated to atmosphere and ocean respectively,

$$\begin{cases} \dfrac{d\mathbf{x}_a^c}{dt} &= \mathcal{M}_a(\mathbf{x}_a^c, \mathbf{x}_o^c) \\ \dfrac{d\mathbf{x}_o^c}{dt} &= \mathcal{M}_o(\mathbf{x}_a^c, \mathbf{x}_o^c) \end{cases}$$

Inheriting the typical use of uncoupled forced models in operational NWP and AMIP/OMIP-type climate studies, we decomposed the above coupled system into the following uncoupled independent systems forced by the boundary information of a sequence prescribed states $\{\tilde{\mathbf{x}}_a^c\}$ and $\{\tilde{\mathbf{x}}_o^c\}$,

$$\begin{cases} \dfrac{d\mathbf{x}_a}{dt} &= \mathcal{M}_a(\mathbf{x}_a, \tilde{\mathbf{x}}_o^c) \\ \dfrac{d\mathbf{x}_o}{dt} &= \mathcal{M}_o(\tilde{\mathbf{x}}_a^c, \mathbf{x}_o) \end{cases}.$$

In the perfect forcing scenario, $\{\tilde{\mathbf{x}}_a^c\}$ and $\{\tilde{\mathbf{x}}_o^c\}$ are collected from the "true" trajectory produced by the coupled model. In this work, we exert the boundary forcing to the uncoupled models by using a set of step-wise functions with updating frequency

$t_b$,

$$\begin{cases} \tilde{\mathbf{x}}_a^c(t) = \mathbf{x}_a^c(t_i), \\ \tilde{\mathbf{x}}_o^c(t) = \mathbf{x}_o^c(t_i), \end{cases} \quad \text{for } t \in [t_i, t_i + t_b],$$

where $t_b$ is generally defined no less than the model integration time step. Using the same numerical integration method used for the original coupled model, the forced ocean/atmosphere models also use the 2nd order Runge-Kutta (RK2) integration method to obtain their trajectories. Here, we take the forced atmosphere model as an example to illustrate the detailed steps of RK2,

$$\begin{cases} k_1 & = \mathcal{M}_a(\mathbf{x}_a(t), \tilde{\mathbf{x}}_o(t)) \\ k_2 & = \mathcal{M}_a(\mathbf{x}_a(t) + \Delta t k_1, \tilde{\mathbf{x}}_o(t)) \\ \mathbf{x}_a(t + \Delta t) & = \mathbf{x}_a(t) + \dfrac{\Delta t}{2}(k_1 + k_2) \end{cases} \tag{4.2}$$

### 4.2.2 Derivation of the conditional tangent linear models for forced ocean/atmosphere models

De Cruz et al. (2016) shows that the MAOOAM can be rewritten as the following system of ODEs

$$\frac{d\mathbf{x}_i}{dt} = \mathcal{M}(\mathbf{x}, \mathcal{T}^i) = \mathbf{x}^T \mathcal{T}^i \mathbf{x},$$

where $\mathcal{T}^i$ is the $(N+1) - \text{by} - (N+1)$ constant matrix depending on the model parameters. In order to simplify the derivation of the force atmosphere/ocean model, we decompose the matrix $\mathcal{T}^i$ and the state vector $\mathbf{x}$ as

$$\mathcal{T}^i = \begin{bmatrix} \mathcal{T}^i_{1,1} & \mathcal{T}^i_{1,a} & \mathcal{T}^i_{1,o} \\ \mathcal{T}^i_{a,1} & \mathcal{T}^i_{a,a} & \mathcal{T}^i_{a,o} \\ \mathcal{T}^i_{o,1} & \mathcal{T}^i_{o,a} & \mathcal{T}^i_{o,o} \end{bmatrix} \qquad \text{and} \qquad \mathbf{x} = \begin{bmatrix} 1 \\ \mathbf{x}_a \\ \mathbf{x}_o \end{bmatrix}$$

In addition, we decompose the total dimension of the model state variables $N$ into the dimension of the atmosphere variables $N_a$ and the dimension of the ocean variables $N_o$.

In the formulae of the forced atmospheric model, we first categorize all the components on the right hand side of the dynamic equations into one part associated with atmospheric variables and the other one independent of the atmospheric variables:

$$\begin{aligned} \frac{d\mathbf{x}_{a,i}}{dt} &= \mathcal{M}_{a,i}(\mathbf{x}_a, \tilde{\mathbf{x}}_o, \mathcal{T}_i) \\ &= \mathbf{x}_a^T \mathcal{T}^i_{a,1} \cdot 1 + 1 \cdot \mathcal{T}^i_{1,a} \mathbf{x}_a + \mathbf{x}_a^T \mathcal{T}^i_{a,a} \mathbf{x}_a + \tilde{\mathbf{x}}_o^T \mathcal{T}^i_{o,a} \mathbf{x}_a + \mathbf{x}_a^T \mathcal{T}^i_{a,o} \tilde{\mathbf{x}}_o + c_{a,i} \\ &\qquad\qquad \text{for } i \in \mathcal{I}_a, \end{aligned}$$

and

$$c_{a,i} = \tilde{\mathbf{x}}_o^T \mathcal{T}^i_{o,1} \cdot 1 + 1 \cdot \mathcal{T}^i_{1,o} \tilde{\mathbf{x}}_o + \tilde{\mathbf{x}}_o^T \mathcal{T}^i_{o,o} \tilde{\mathbf{x}}_o + 1 \cdot \mathcal{T}^i_{1,1} \cdot 1$$

where $\mathcal{I}_a$ denotes the indices subset corresponding to the atmosphere. In addition, $c_{a,i}$ is a constant with respect to $\mathbf{x}_a$ for all $i \in \mathcal{I}_a$. Then it is straightforward to compute the Jacobian of the above ODEs by taking the derivatives corresponding to atmospheric variables. The $i$th row of the atmosphere Jacobian matrix (i.e. $\mathcal{J}_{a,i}$),

could be derived as following,

$$\begin{aligned}\mathcal{J}_{a,i} &= \left[\begin{array}{cccc} \dfrac{\partial \mathcal{M}_{a,i}}{\partial \mathbf{x}_{a,1}} & \dfrac{\partial \mathcal{M}_{a,i}}{\partial \mathbf{x}_{a,1}} & \ldots & \dfrac{\partial \mathcal{M}_{a,i}}{\partial \mathbf{x}_{a,N_a}} \end{array}\right] \\ &= ((\mathcal{T}^i_{a,1})^T + \mathcal{T}^i_{1,a}) + \mathbf{x}_a^T(\mathcal{T}^i_{a,a} + (\mathcal{T}^i_{a,a})^T) + \tilde{\mathbf{x}}_o^T(\mathcal{T}^i_{o,a} + (\mathcal{T}^i_{a,o})^T) \\ &\qquad\qquad \text{for } i \in \mathcal{I}_a,\end{aligned}$$

where $\mathcal{J}_{a,i}$ is of $1 \times N_a$ and the other rows of the atmosphere TLM $\mathcal{J}_a$ can be determined using the same manner.

The differential form of the conditional TLM for a small perturbation $\delta\mathbf{x}_a$ could be determined as,

$$\begin{aligned}\frac{d(\delta\mathbf{x}_{a,i})}{dt} &= \mathcal{J}_{a,i}\delta\mathbf{x}_a \\ &= ((\mathcal{T}^i_{a,1})^T + \mathcal{T}^i_{1,a})\delta\mathbf{x}_a + \mathbf{x}_a^T(\mathcal{T}^i_{a,a} + (\mathcal{T}^i_{a,a})^T)\delta\mathbf{x}_a + \tilde{\mathbf{x}}_o^T(\mathcal{T}^i_{o,a} + (\mathcal{T}^i_{a,o})^T)\delta\mathbf{x}_a \\ &\qquad\qquad \text{for } i \in \mathcal{I}_a,\end{aligned}$$

Likewise, to derive the forced ocean TLM, the forced oceanic model can be extended and rewritten as,

$$\begin{aligned}\frac{d\mathbf{x}_{o,i}}{dt} &= \mathcal{M}_{o,i}(\tilde{\mathbf{x}}_a, \mathbf{x}_o, \mathcal{T}_i) \\ &= 1 \cdot \mathcal{T}^i_{1,o}\mathbf{x}_o + \mathbf{x}_o^T\mathcal{T}^i_{o,o}\mathbf{x}_o + \mathbf{x}_o^T\mathcal{T}^i_{o,1} \cdot 1 + \mathbf{x}_o^T\mathcal{T}^i_{o,a}\tilde{\mathbf{x}}_a + \tilde{\mathbf{x}}_a^T\mathcal{T}^i_{a,o}\mathbf{x}_o + c_{o,i} \\ &\qquad\qquad \text{for } i \in \mathcal{I}_o,\end{aligned}$$

and

$$c_{o,i} = 1 \cdot \mathcal{T}^i_{1,a}\tilde{\mathbf{x}}_a + \tilde{\mathbf{x}}_a^T\mathcal{T}^i_{a,1} \cdot 1 + \tilde{\mathbf{x}}_a^T\mathcal{T}^i_{a,a}\tilde{\mathbf{x}}_a + 1 \cdot \mathcal{T}^i_{1,1} \cdot 1$$

where $\mathcal{I}_o$ is the subset of indices associated to the ocean states. The parameter $c_{o,i}$ is a constant with respect to $\mathbf{x}_o$ for all $i \in \mathcal{I}_o$. Then the forced TLM for a small perturbation of ocean state variables $\delta\mathbf{x}_o$ is defined as following,

$$\begin{aligned}\frac{d\delta\mathbf{x}_{o,i}}{dt} &= \mathcal{J}_{o,i}\delta\mathbf{x}_o \\ &= (\mathcal{T}^i_{1,o} + (\mathcal{T}^i_{o,1})^T)\delta\mathbf{x}_o + \mathbf{x}_o^T((\mathcal{T}^i_{o,o})^T + \mathcal{T}^i_{o,o})\delta\mathbf{x}_o + \tilde{\mathbf{x}}_a^T((\mathcal{T}^i_{o,a})^T + \mathcal{T}^i_{a,o})\delta\mathbf{x}_o, \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } i \in \mathcal{I}_o,\end{aligned}$$

where the $i$th row of the ocean Jacobian matrix, $\mathcal{J}_{o,i}$ and is obtained by,

$$\begin{aligned}\mathcal{J}_{o,i} &= \left[\begin{array}{cccc} \frac{\partial\mathcal{M}_{o,i}}{\partial\mathbf{x}_{o,1}} & \frac{\partial\mathcal{M}_{o,i}}{\partial\mathbf{x}_{o,1}} & \ldots & \frac{\partial\mathcal{M}_{o,i}}{\partial\mathbf{x}_{o,N_o}} \end{array}\right] \\ &= (\mathcal{T}^i_{1,o} + (\mathcal{T}^i_{o,1})^T) + \mathbf{x}_o^T((\mathcal{T}^i_{o,o})^T + \mathcal{T}^i_{o,o}) + \tilde{\mathbf{x}}_a^T((\mathcal{T}^i_{o,a})^T + \mathcal{T}^i_{a,o}) \qquad \text{for } i \in \mathcal{I}_o\end{aligned}$$

### 4.2.3 Numerical methods for computing Lyapunov Exponents

Lyapunov exponents (LEs) are defined on the far-past and far-future tangent linear operators and the Oseledec subspaces. Given an $N$-dimension dynamical system, LEs characterize the asymptotic exponential separation rate of trajectories initialized with infinitesimal errors. The LEs are defined as

$$\lambda_j = \lim_{t\to\infty} \frac{1}{t - t_0} \ln \|\mathcal{F}(t, t_0)\mathbf{e}_j\|, \quad j = 1, 2, ..., N, \tag{4.3}$$

where $\mathbf{e}_j$ is a unit vector at time $t_0$ and $\mathcal{F}(t, t_0)$ is the tangent linear propagator or resolvent corresponding to the TLM of the given dynamic system, e.g. the coupled model system (4.1),

$$\frac{d(\delta\mathbf{x})}{dt} = \mathcal{J}(\mathbf{x}, t)\delta\mathbf{x}, \tag{4.4}$$

where $\mathcal{J} : \mathbb{R}^N \times \mathbb{R}^+ \to \mathbb{R}^{N\times N}$ is the Jacobian matrix of $\mathcal{M}$ with respect to $\mathbf{x}$. Theoretically, the tangent linear propagator $\mathcal{F}$ is defined by the fundamental matrix $\mathbf{\Psi} \in \mathbb{R}^{N\times N}$ as

$$\mathcal{F}(t, t_0) = \mathbf{\Psi}(t)\mathbf{\Psi}(t_0)^{-1}. \tag{4.5}$$

It solves the TLM as,

$$\delta\mathbf{x}(t) = \mathcal{F}(t, t_0)\delta\mathbf{x}(0).$$

The positive LEs representing the error growing modes within the dynamical system. The detailed mathematical definitions and properties are reviewed by Kuptsov and Parlitz (2012). In practice, the general form of $\mathcal{F}$ in a larger interval is numerically determined by a multiplication of a sequence of approximated linear propagators defined on infinitesimal time intervals,

$$\begin{aligned}\mathcal{F}(t_0 + n\delta t, t_0) &= \mathcal{F}(t_0 + n\delta t, t_0 + (n-1)\delta t) \cdots \mathcal{F}(t_0 + 2\delta t, t_0 + \delta t)\mathcal{F}(t_0 + \delta t, t_0) \\ &\approx (1 - \delta t \cdot \mathcal{J}_{n-1}) \cdots (1 - \delta t \cdot \mathcal{J}_1)(1 - \delta t \cdot \mathcal{J}_0), \\ &= \hat{\mathcal{F}}_{n-1} \cdots \hat{\mathcal{F}}_1 \hat{\mathcal{F}}_0\end{aligned} \tag{4.6}$$

where $\delta t$ is a small time step and is usually taken the same as the model integration time step. $\mathcal{J}_i$ is the Jacobian matrix defined at time $t_0 + i \cdot \delta t$. Different from the the theoretical definition of $\mathcal{F}$ in (4.5), we use the forward Euler method to obtain the numerical approximation of the intermediate tangent linear propagator from $t_0 + i\delta t$ to $t_0 + (i+1)\delta t$ through the TLM (4.4),

$$\hat{\mathcal{F}}_i := (1 - \delta t \cdot \mathcal{J}_i) \approx \mathcal{F}(t_0 + (i+1)\delta t, t_0 + i\delta t).$$

The most commonly used numerical method approximating the LEs in a given dynamical system is based on a sequential QR factorization. This numerical method is derived from a basic idea of using an orthogonal time-varying change of variables to transform the dynamical system 4.4 to a triangular system, of which the converging theorem (Dieci et al., 2011; Liapounoff, 1907) is introduced as following, "If we have

a regular system with upper triangular coefficient matrix $\mathbf{B}(t)$ : $\dot{\mathbf{y}} = \mathbf{B}(t)\mathbf{y}$, then, for $j = 1, 2, ..., N$, its Lyapunov exponents are given by

$$\lambda_j = \lim_{t\to\infty} \frac{1}{t - t_0} \int_{t_0}^{t} \mathbf{B}_{jj}(s)\, ds.\text{"}$$

Here, $\lambda_j$ is the same as the LE in (4.3) from the TLM 4.4 if the transformation is orthogonal throughout the time.

In fact, the aforementioned orthogonal transformation can be achieved by using the orthonormal matrix from QR factorizing the tangent linear propagator $\mathcal{F}(t, t_0)$ of the original dynamical system at every time $t$,

$$\mathcal{F}(t, t_0) = \tilde{\mathbf{Q}}(t)\tilde{\mathbf{R}}(t), \tag{4.7}$$

with all the diagonal components $\tilde{\mathbf{R}}_{jj}(t)$ positive. If $\mathbf{y}(t) := \tilde{\mathbf{Q}}(t)^T \delta\mathbf{x}(t)$, then its derivative satisfies a triangular system,

$$\frac{d(\mathbf{y}(t))}{dt} = \mathbf{B}(t)\mathbf{y}(t), \tag{4.8}$$

where $\mathbf{B}(t)$ is the upper triangular matrix. We can also conclude that each diagonal component $\tilde{\mathbf{R}}_{jj}(t)$ satisfies a condition,

$$\ln(\tilde{\mathbf{R}}_{jj}(t)) = \int_{t_0}^{t} \mathbf{B}_{jj}(s)\, ds. \quad \text{for } j = 1, 2, ..., N, \tag{4.9}$$

and,

$$\lambda_j = \lim_{t\to\infty} \frac{1}{t - t_0} \ln(\tilde{\mathbf{R}}_{jj}(t)).$$

Note that for the full-rank matrix, the QR factorization is unique if the diagonal components are all positive. The detailed derivations of the triangular system (4.8) and the formula (4.9) are shown in the Appendix A.

In light of the above theory, the QR algorithm of computing the LEs is shown as below,

1. We first initialize the orthonormal matrix $\mathbf{Q}_0$ as the identity matrix, i.e. $\mathbf{Q}_0 = \mathbf{I}$.

2. At each time time step $t_i := t_0 + i\delta t$ $(i = 0, 1, ..., n-1)$, we compute the approximated intermediate linear propagator $\hat{\mathcal{F}}_i$ defined on the time interval $(t_i, t_{i+1}]$. Then we apply the QR factorization on the multiplication $\hat{\mathcal{F}}_i \mathbf{Q}_i$,

$$\mathbf{Q}_{i+1}\mathbf{R}_{i+1} := QR(\hat{\mathcal{F}}_i \mathbf{Q}_i), \tag{4.10}$$

   where $QR(\cdot)$ is the numerical QR factorization operator. At each step, we save the diagonal terms of each $\mathbf{R}_{i+1}$ $(i = 0, 1, ..., n-1)$

3. Then the numerical values for LEs can be estimated as,

$$\hat{\lambda}_j(n) = \frac{1}{n\delta t} \ln \Big| \prod_{i=1}^{n-1} (\mathbf{R}_{n-i})_{jj} \Big|, \quad \text{for } j = 1, 2, ..., N$$

Notice that within the above algorithm $\hat{\mathcal{F}}_i$ at each time step $t_i$ $(i = 0, 1, ..., n-1)$ can be rewritten as

$$\hat{\mathcal{F}}_i = \mathbf{Q}_{i+1}\mathbf{R}_{i+1}\mathbf{Q}_i^T.$$

Therefore the estimation of the long-term linear propagator is shown as,

$$\mathcal{F}(t_n, t_0) \approx \hat{\mathcal{F}}_{n-1} \cdots \hat{\mathcal{F}}_1 \hat{\mathcal{F}}_0 = \mathbf{Q}_n \mathbf{R}_n \mathbf{R}_{n-1} \cdots \mathbf{R}_1$$

and we have the following relations between the numerical QR algorithm and the theoretical QR factorization (4.7) used in the aforementioned convergence theorem,

$$\tilde{\mathbf{R}}_n \approx \mathbf{R}_n \cdots \mathbf{R}_2 \mathbf{R}_1 \quad \text{and} \quad \tilde{\mathbf{Q}}_n \approx \mathbf{Q}_n,$$

or,

$$\lambda_j = \lim_{t\to\infty} \frac{1}{t-t_0} \ln(\tilde{\mathbf{R}}_{jj}(t)) \approx \lim_{n\to\infty} \frac{1}{n\delta t} \ln(\mathbf{R}_n \cdots \mathbf{R}_1) = \lim_{n\to\infty} \hat{\lambda}_j(n).$$

Despite that there is no theoretical justification of the converging rate of $\hat{\lambda}_j(n)$ to $\lambda_j$ based on the choice of $n$, we define $n$ large enough for a relatively accurate estimation.

We next discussed the computation of LEs after we apply the DA method to the given dynamical system. Carrassi et al. (2008) and Penny (2017) describes the effect of the DA analysis update as a contraction operating on the tangent linear propagator. Suppose we operate a DA method at end of time $t_i$ and the updated linear propagator can be written as,

$$\mathcal{F}^+(t_i, t_{i-1}) = [\mathbf{I} - \mathbf{K}_i\mathbf{H}_i]\mathcal{F}^-(t_i, t_{i-1}),$$

where $\mathcal{F}^-$ is the original tangent linear propagator for the forecast model and $\mathcal{F}^+$ is the updated linear propagator after applying DA update. In numerical computation, $\mathcal{F}^-(t_i, t_{i-1})$ is substituted by the approximation $\hat{\mathcal{F}}_{i-1}$. $\mathbf{I} - \mathbf{K}_i\mathbf{H}_i$ is the contraction matrix resulting from DA with $\mathbf{K}_i$ as the gain matrix and $\mathbf{H}_i$ as the observation operator at $t_i$. The updated LEs can be numerically determined by followint the above QR algorithm replacing each $\hat{\mathcal{F}}_{i-1}$ with $\mathcal{F}^+(t_i, t_{i-1})$ in (4.10).

## 4.3 Results and Discussion

In this section, all the models are numerically integrated using RK2 with the time step $\delta t = 0.01$ model time unit (MTU). We use the discrete QR algorithm reviewed in the previous section to determine LEs with $n = 10^6$ loops to obtain

relatively accurate results. We conduct a series of Observing System Simulation Experiments (OSSEs) to assess the impact of DA methods on different coupling approaches. The nature run is generated by the original MAOOAM coupled model and the observation errors are determined by the 10% of climatological variability of the corresponding states. The initial ensemble perturbations are Gaussian with 10% of climatological variance. The choice of the different DA window is specified in details in the following subsections.

### 4.3.1 Forced ocean/atmosphere model and coupled model

We first compare the Lyapunov spectrum of the coupled model to those of the uncoupled models under the ideal scenario, where the perfect forcing is provided. We examine with different forcing interval $t_b = 0.01, 100, 1000$ model time unit (MTU) and $t_b =$ infinity (insert as a constant forcing). The experiments are set up as the following procedures: first, all the uncoupled modes start from the the initial condition slightly perturbed from the initial condition of the coupled model; second, integrate all the models for transient time $10^7$ MTU to reach their own attractors respectively; and at last, the comparison starts after the completion of the transient and continue running for another $10^6$ MTU.

We find that changing the forcing interval $t_b$ does not significantly impact the Lyapunov spectrum of the forced atmosphere and ocean models if the systems are driven by the same set of prescribed forcing. Figure 4.2 (b) shows the LEs of the forced ocean model using different forcing frequencies. Of the near-zero LEs for all

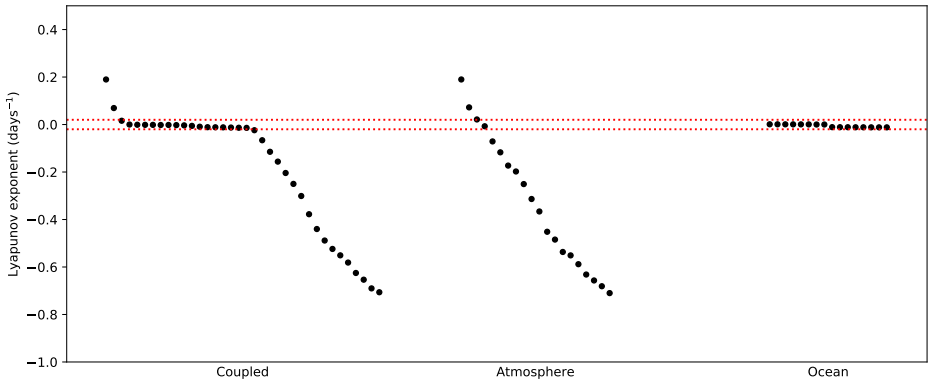


Figure 4.1: The Lyapunov spectrum for the coupled model and forced atmosphere and ocean. For the forced models, forcing time interval is $t_b = 0.01$ MTU. The two dotted red lines are referencing lines with value $\pm 0.01$

the forced ocean, the first three leading LEs are numerically positive and on the scale of $10^{-5}$, the 4th and 5th LEs vary between $10^{-6}$ and $-10^{-6}$, and the 6th to 8th are negative and on the scale of $-10^{-5}$. Meanwhile, of the near-zero LEs for the coupled model, the first four leading LEs are positive with the first LE on the scale of $10^{-4}$ and the 2nd to 4th of $10^{-5}$. The 5th to 8th LEs are numerically negative, among which the 5th and 6th are of the scale $-10^{-5}$, whereas the 7th and 8th are of the scale $-10^{-4}$.

Nevertheless, the choice $t_b$ does impact the degree of synchronization in terms of the RMSEs (see Figure 4.2 (a)). Compared to the near-zero LEs of the corresponding coupled model, we notice that the Lyapunov spectrum for uncoupled models shows a "staircase" profile with a prominent "gap" between the 8th and the 9th LEs, while there is a gradual decay within the Lyapunov spectrum for the

coupled model. This discrepancy implies a loss of phase in the forced ocean model. To illustrate this point, Figure 4.2 (a) shows the relative RMSEs of the uncoupled trajectories against the true coupled trajectory. It shows that within the given testing periods, even for the uncoupled trajectory with $t_b = 0.01$ and the perfect forcing, there is no sign of synchronization in the given time period.

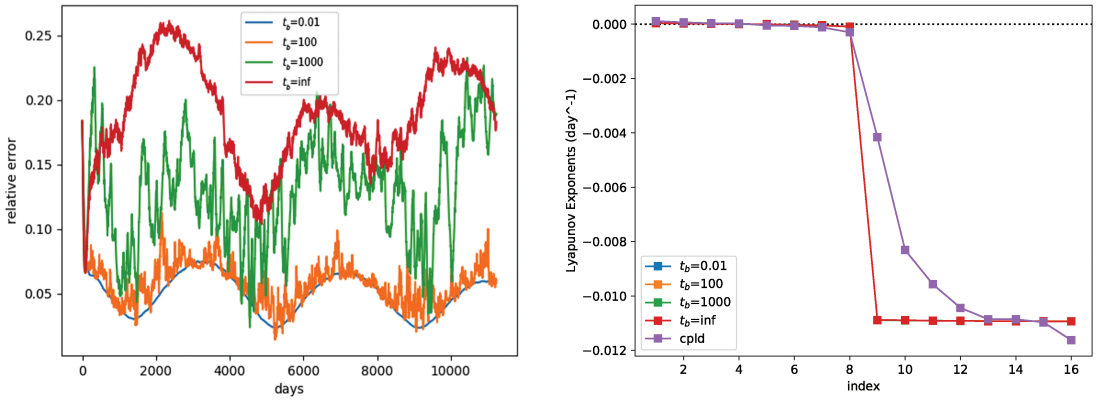


Figure 4.2: (left) The relative error between the forced and coupled trajectory and (right) CLEs of forced ocean model with different forcing window lengths $t_b$. The CLEs associated to $t_b = 0.01$ (blue), $t_b = 100$ (orange), $t_b = 1000$ (green) and $t_b = \text{inf}$ (red) are intertangled with each other and are presented by the red line.

To compare DA applied to the coupled versus forced models, we estimate the minimum ensemble size required for the ETKF to prevent filter divergence. We compare the forced atmosphere model to the coupled atmosphere model, using an observing network restricted to the atmosphere only (Table 4.1). We define divergence as the condition that the root mean square error reaches the same order of magnitude as a free model run starting from perturbed initial conditions. For the forced atmosphere model, the convergence for a given analysis cycle window $\tau$

depends largely on the forcing interval $t_b$. Using an analysis cycle window shorter than the forcing interval causes the ETKF to diverge. However, when the analysis cycle window is too large (i.e. $\tau = 1000$), ETKF diverges for all the given forcing intervals $t_b = 0.01$, 0.1, 1, 10. In addition, when the data assimilation window is of the same length as the forcing interval, the minimum ensemble sizes are identically the same (i.e. 7), which is greater than the number of unstable-neutral modes (i.e. 3-4, shown in Figure 4.4). This can be possibly due to the extraneous error from the observation at each time step.

When using a coupled forecast model, the ETKF converges for the data assimilation windows $\tau = 0.1$, 1 and 10 MTU. The minimum ensemble sizes for convergence are smaller (i.e. 9-12) compared to the its corresponding model dimension - roughly 25% to 33% of the coupled model dimension (36) as compared to 35% to 55% of the model dimension (20) for the converging cases of the forced atmosphere system. When using a coupled forecast model the ETKF appears more robust to longer assimilation windows.

### 4.3.2 Comparisons of different coupling approaches using the ETKF

We next investigate the ETKF under different coupling scenarios. We use an ensemble size of 37, which is larger than the degrees of freedom of the coupled model. For the forced model cases, perfect forcing is updated every $t_b = 0.1$ and 12.5 MTU (or about 16 mins and 1.4 days). As previously described, the forced atmosphere ETKF is sensitive to the frequency of the forcing update (Figure 4.3).

The ETKF converges using both SCDA and WCDA, though SCDA consistently produces smaller RMSE in the oceanic component than WCDA.

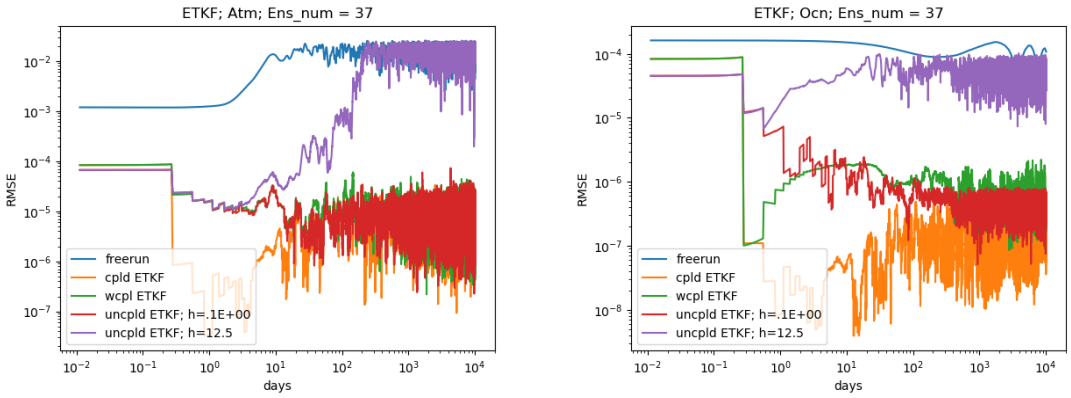


Figure 4.3: RMSEs in atmosphere (left) components and ocean (right) components for SCDA (orange), WCDA (green), UCDA with forcing interval $t_b = 0.1$ MTU (red) and UCDA with forcing interval $t_b = 12.5$ MTU (purple). The DA window is uniformly defined as $\tau = 2.5$ MTU with ensemble size 37. For all the experiments, the full observing network is applied.

The forced atmosphere ETKF diverges if the forcing from the ocean is updated less frequently. For example, when examining the Lyapunov spectrum for a similar case, the forced model with $t_b = 12.5$ MTU ($\sim$ 1.4 days) still has a leading CLE larger than zero (Figure 4.4), indicating filter divergence. The CDA ETKF systems generate atmospheric analyses with RMSE similar to the uncoupled ETKF with frequently updated ($t_b = 0.1$ MTU or 16mins) perfect forcing. Note that for the UCDA with $t_b = 0.1$ MTU, though the RMSEs in both atmosphere and ocean are stablized, the computational leading LE of the whole system in 4.4 is shown as marginally above 0, approximately $3.14 \times 10^{-5}$ (compared to $6.08 \times 10^{-4}$ for the

leading LE of forced ocean model with $t_b = 0.1$). The reason for this can be manifold, such as the machine error of computing the LEs or the testing time period is not long enough to obtain a negative value. Experiments with longer implementation time needs to be conducted for a further verification of this problem.

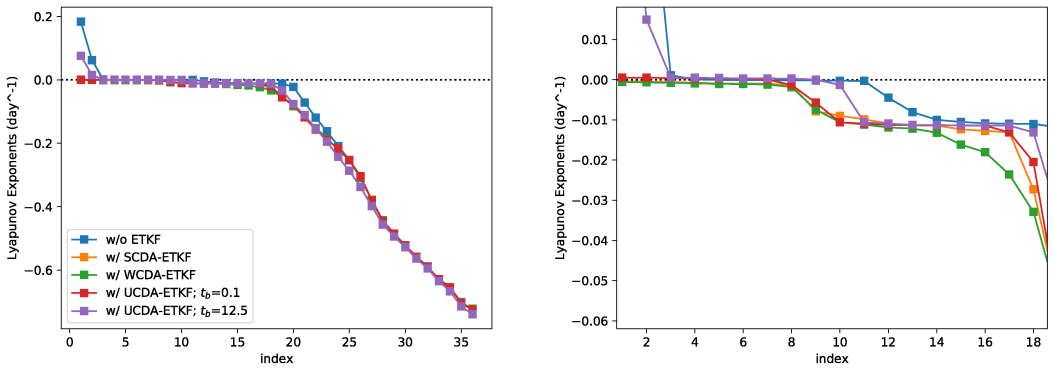


Figure 4.4: Lyapunov spectrum (left) and a magnification of the near-zero LEs (right) for SCDA ETKF, WCDA ETKF and forced atmosphere and ocean ETKF (UCDA) using forcing intervals of $t_b = 0.1$ (around 16 min) and $t_b = 12.5$ (around 1.4 days). Each is implemented with an ensemble size of 37, an analysis cycle window of 2.5 MTU, and no inflation.

### 4.3.3 Comparisons of different observation networks

We next focus on the performance of SCDA and consider the transition of the observation network from one medium to coupled media, including atmosphere only, ocean only and both. In this thesis, we only simplify the discussion by considering the cutting off observing network in terms of the model state variables, which are in fact defined on the spectral space instead of the physical space. We unify the

multiplicative inflation factor as $\rho = 1$ and the ensemble size as 37, in order to demonstrate a clear comparison in the the error reduction. Detailed discussions of observing on the physical network and diversified multiplicative inflation factor are shown in Penny et al. (2019).

We first examine the experiments using the analysis update window as $\tau = 0.1$ MTU ($\sim$ 16 mins), a high-frequency analysis. Figure 4.5 shows the RMSEs in atmosphere and ocean components. The ETKF of observing only the ocean components is comparable to the ETKF of observing only atmosphere states for ocean components (i.e. around $10^{-6}$), while it performs slightly worse than the experiments using the other two types of observing networks in reducing the atmosphere error. Observing both of the domains performs the best in correcting both of the atmosphere and ocean states among all the testing SCDA with $\tau = 0.1$ MTU. In Figure 4.6, the ensemble spread is shown for each experiment case, where experiments of observing only the atmosphere domain and both domains provide less variation in the ensemble spread in all the state variables. The corresponding Lyapunov spectrum is shown in Figure 4.7, where we can observe that the leading LE in all the ETKF experiments is slightly below 0. This implies that ETKFs of the ensemble size 37 with $\tau = 0.1$ MTU are capable of stablizing the system for all observing networks described here.

We then extend the analysis window to $\tau = 2.5$ MTU (see 4.8). Results show that providing only the ocean observation is not sufficient for the ETKF to constrain the error growth in the atmosphere state variables. This coincides with the results from Ballabrera-Poy et al. (2009) and Han et al. (2013), where they show

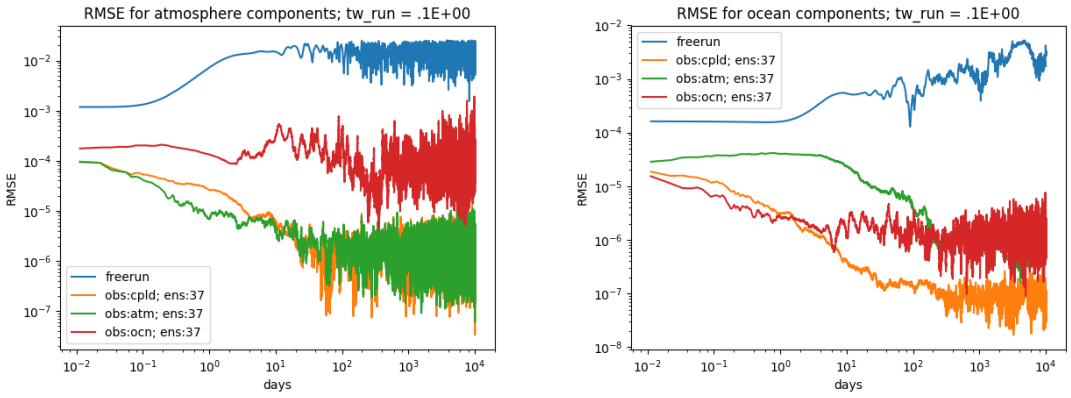


Figure 4.5: RMSEs in atmosphere (left) components and ocean (right) components for SCDA using observations from: 1) both atmosphere and ocean (orange; atmosphere time average: $1.98 \times 10^{-6}$; ocean time average: $8.85 \times 10^{-8}$); 2) atmosphere only (green; atmosphere time average: $2.05 \times 10^{-6}$; ocean time average: $7.00 \times 10^{-7}$) and 3) ocean only (red; atmosphere time average: $1.04 \times 10^{-4}$; ocean time average: $1.35 \times 10^{-6}$), compared with the free run (blue). The DA window is defined as $\tau = 0.1$ MTU ($\sim 16$ mins).

that SCDA has difficulty in using observations from the low-frequency system to accurately update the high frequency system. While the performance of observing only the atmosphere domain is comparable to the case of observing coupled domain in correcting atmosphere state variables, it provides less accuracy in estimating the ocean state variables than observing the coupled domain. Conclusively, SCDA of observing atmosphere and ocean states together is relatively more robust and stable than the ones of observing only single domain.

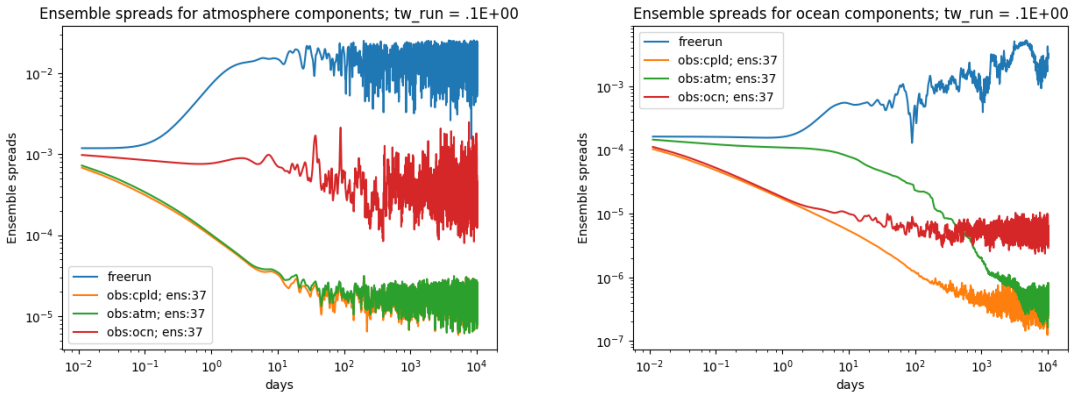


Figure 4.6: Ensemble spreads in atmosphere (left) components and ocean (right) components for SCDA using observations from: 1) both atmosphere and ocean (orange); 2) atmosphere only (green) and 3) ocean only (red), compared with the free run (blue). The DA window is defined as $\tau = 0.1$ MTU ($\sim$ 16 mins).

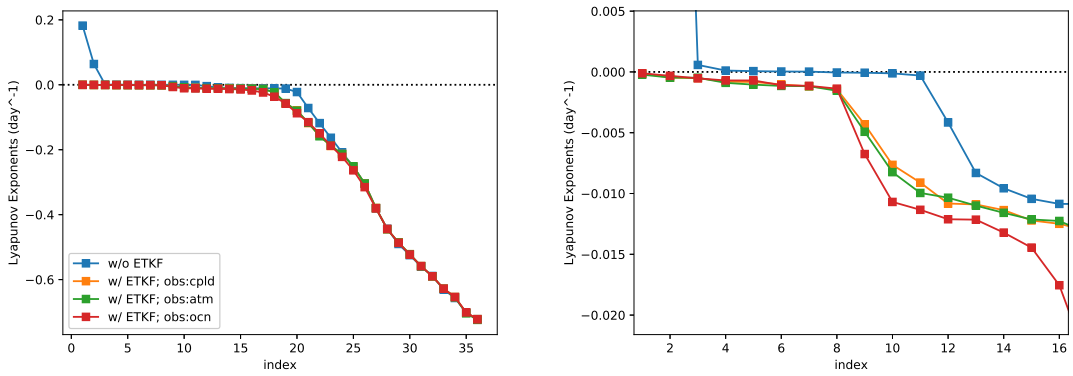


Figure 4.7: The Lyapunov exponents of the coupled model with and without the 37-member ETKF DA, with analysis/observing window $\tau = 0.1$ MTU, observing the atmosphere, the ocean, or both domains (left). Magnified view of the same (right).

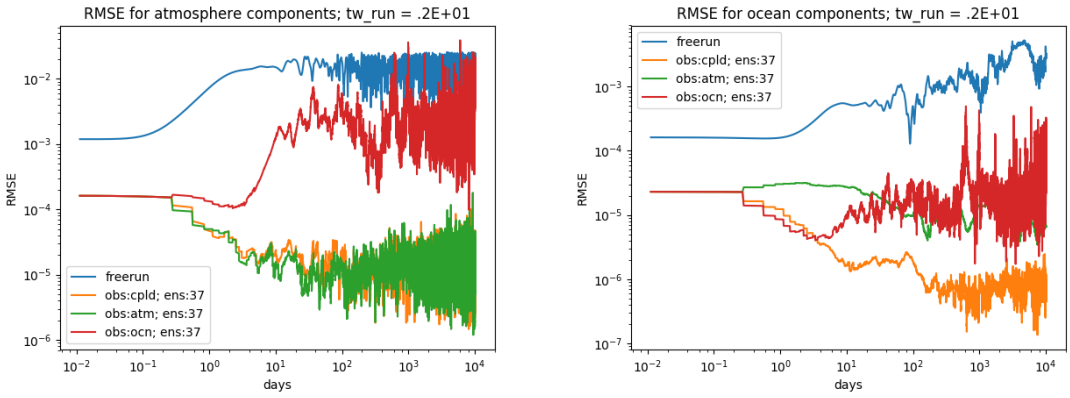


Figure 4.8: RMSEs in atmosphere (left) components and ocean (right) components for SCDS using observations from: 1) both atmosphere and ocena (orange); 2) atmosphere only (green) and 3) ocean only (red), compared with the free run (blue). The DA window is defined as $\tau = 2.6$ MTU ($\sim 6$ hrs).

### 4.3.4 Forecast leading time

Penny (2017) indicated, "For the purpose of seamless prediction, the research community must demonstrate that CDA not only improves initial states but also improves forecasts at all time scales. In this section, we compare the accuracy of forecasts at multiple lead times using DA with different degrees of coupling. For these comparisons we use initial conditions produced from 36,000 DA cycles of a 37-member ETKF using either SCDA, WCDA, or uncoupled DA. At each cycle we initialize forecasts with lead times ranging from 0 to 10 days, compute the RMSEs, and aggregate the reults. For the uncoupled DA, we use a forced model with either perfect forcing or forcing degraded equal to the perfect forcing plus a white noise of 10% of climatological variability. The SCDA forecasts have the highest

accuracy, followed by the WCDA, then the uncoupled DA with perfect forcing. For the oceanic states (Figure 4.9), the forecasts produced by SCDA outperform all the other couplings for the first two days of lead time. For both the SCDA and WCDA, the error grows quickly after day 3 to reach almost the same level by day 10. The ocean forecast errors for the uncoupled DA using perfect forcing are generally stable at all lead times, while at shorter lead times the forcasts generated by degraded forcing result in a relatively larger RMSE than all other methods. A more realistic scenario for degraded forcing might allow forcing error to grow with time until it reaches saturation, for example by relaxing the forcing field to climatology.

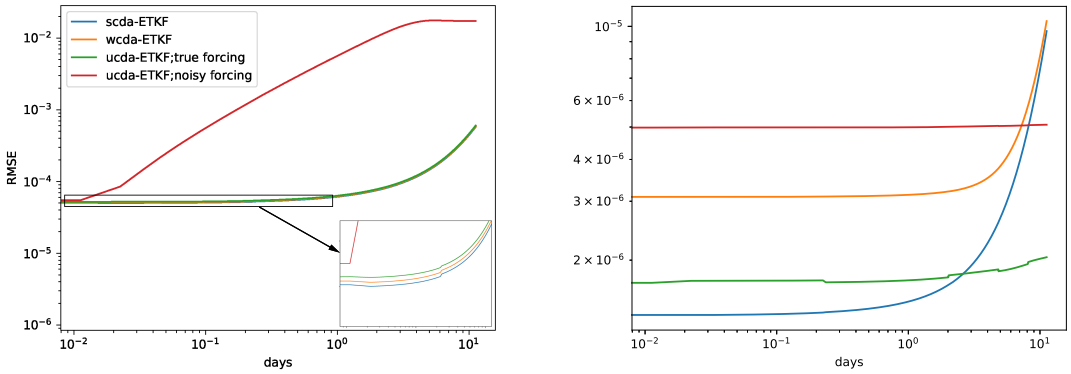


Figure 4.9: The RMSE of forecasts at various lead times in the (a) atmosphere and (b) ocean components for the ETKF using SCDA, WCDA, and the forced ETKF with perfect and noisy forcing. The analysis window is $\tau = 2.5$ MTU (6 hrs) and the forcing window of uncoupled ETKF is $t_b = 0.1$ MTU (16 min).

## 4.4 Conclusion

In this chapter, we take advantage of the Lyapunov spectrum to discuss the impact of coupling appearing in the model and DA system. We first compare the forced ocean/atmosphere model with the coupled model under the framework of drive-response system. Imposed with the prescribed perfect driving information, all the forced ocean/atmosphere models have nearly the same Lyapunov spectrum with different forcing time intervals $t_b$ = 0.01, 100, 100 MTU and $t_b$ =infinity (i.e. constant states). Specifically, among all the near-zero LEs of forced ocean models, a sharply stepwise decay is shown, as a contrast to a smooth decay for the coupled model, which implies a loss of transition phases for the forced ocean model by ignoring the instantaneous bilateral fluxes within the interface layer. More importantly, the non-negative leading LE for the forced ocean/atmosphere models indicates that the responsive forced ocean/atmosphere model cannot generally synchronize with the driving coupled model if no extra information or mechanism is applied. As such, we examine the application ETKF on the forced models with different forcing frequencies and compare it with the coupled forecast model. We discover that using the coupled model as the forecast model is more robust and stable for varied analysis windows than the forced model. In addition, the stability of the analysis window depends highly on the forcing window.

We then compare the UCDA, WCDA and SCDA using ETKF. Smith et al. (2015) found that for a 1D simplified coupled column model using incremental 4D-Var, WCDA was usually comparable to uncoupled assimilation experiments in which

the atmosphere and ocean models were forced using accurate surface fluxes. Here we find a similar result and add that the stability is degraded in the forced system only when the forcing accuracy is reduced (e.g. by extending the forcing time window).

We examine the SCDA assimilating observations from the atmosphere-only, ocean-only, and both medias. We found that when providing only ocean observations, it is difficult to constrain the error growth in all model state variables unless the analysis window is sufficiently short. The inclusion of the atmosphere observation is essential to stablize the SCDA system for larger analysis update windows.

In the last part, we compared the performance of different couplings in the forecast leading time. SCDA provides the most accurate forecast in both the atmosphere and ocean states with a two-day lead time. The forecast errors of the forced ETKF are in general stable for ocean states at various lead times. A small error in the forcing can produce relative large error in all the states even within a short lead time. The perfect forcing can provide a relative small and stable error for ocean state variable in long lead times, however it is impossible to exert such a perfect forcing in any realistic scenario.

Table 4.1: The minimum ensemble size required to prevent filter divergence for various scenarios (without inflation) within the time $9 \times 10^4$ MTU. The quantity $t_b$ indicates the forcing frequency and $\tau$ is the assimilation window

| Experiment type | Forcing/Coupling interval | Analysis cycle window (MTU) | | |
|---|---|---|---|---|
| | | $\tau = 0.1$ ($\sim$ 16 mins) | $\tau = 1.0$ ($\sim$ 2.6 hrs) | $\tau = 10.0$ ($\sim$ 1.1 days) |
| Forced atmosphere | $t_b = 0.01$ ($\sim$ 1.6 mins) | 7 | 7 | 11 |
| | $t_b = 0.1$ ($\sim$ 16 mins) | 7 | 7 | 9 |
| | $t_b = 1.0$ ($\sim$ 2.6 hrs) | Diverges with 21 | 7 | 11 |
| | $t_b = 10.0$ ($\sim$ 1.1 days) | Diverges with 21 | Diverges with 21 | 7 |
| Coupled system, observed atmosphere | | 12 | 9 | 9 |

# Chapter 5: Conclusion and Future Research

In this dissertation, we mainly focus on the developments in Lagrangian Data Assimilation (LaDA) and Coupled Data Assimilation (CDA). Lagrangian floats, drifting along the flow with currents and measuring ocean fields (e.g temperature and salinity), are a promising technology for detecting the ocean flow of small spatial and high-frequency temporal scale (Poje et al., 2014). While assimilating the direct measurement of temperature and salinity is a commonly used approach in many operational prediction centers, the drifter positions are under-exploited in DA for estimating the ocean circulation system because of: (1) its indirect representation of the ocean states in the general forecast ocean models; (2) nonlinear error growth; and (3) the curse of high dimension when applying to a realistic model. In this thesis, we propose a LaDA method LETKF-LaDA using the augmented-state approach (Ide et al., 2002; Kuznetsov et al., 2003), targeting to assimilate the direct drifter positions efficiently and then to readily extend to the realistic system, which has been a challenge in ocean data assimilation for the past 15 years. Using the "identical twin" approach in the OSSEs, we first examine the LETKF-LaDA on the simple 3D double-gyre model of MOM4p1, which shows that the proposed method is efficient in reducing model error in various ocean variables, at all depths down to 1000 m. In

addition we conclude that using a localization radius of 2 to 3 times the baroclinic Rossby radius of deformation gives the best results for this specific system. Adding observations of temperature and salinity to the drifter positions further enhances error reduction.

An application of a more advanced model, i.e. MOM6, using GoM configuration and eddy-resolving resolution has been considered next to further verify the performance of LETKF-LaDA. We first build up a complete online drifter module for the MOM6, successfully solving the untouched problem remaining in the drifter modules of the MOM4p1 and MOM5 about the communication issues between different processors. Different from the previous testing, a more realistic scenario using the time-variant atmosphere forcing form the 20CRv3 database has been attempted to emulate the operational use. Results show that assimilating only the Lagrangian positions improves not only the prediction of ocean flow velocities in different water masses, but also the forecast of temperature and salinity as well. We also verify that LaDA is in general comparable to the results of EuDA but even with higher accuracy in the velocity fields. We also show that variable localization is necessary when assimilating temperature and salinity drifter measurement together with the position observations.

In the follow-on research, we will further upgrade the application of LETKF-LaDA on the Gulf of Mexico configuration using realistic observation and boundary conditions. Different from the previous examination in Chapter 3, the forecast model will be designed using an open boundary condition defined by SODA database instead of the simple "wall" boundary condition. The insitu and Argo observations of

temperature and salinity will be assimilated together with GLAD historical database using LETKF-LaDA. Similar to the work done by Carrier et al. (2014, 2016) and Jacobs et al. (2014), a comparison of the absolute dynamic height (ADH; in meters) from the Archiving, Validation, and Interpretation of Satellite Oceanographic data (AVISO) product will be made to verify the performance of the results. We are expecting that the addition of the LaDA can further improve the forecast of the ocean fields, especially for the mesoscale and sub-mesoscale flow, in order to provide beneficial information about understanding the Loop Current System (LCS) in terms of the variability in strength, location, depth, and the size of the Loop Current (LC). This will serve as a test for later extension to other regional applications and, if successful, eventually tests in global applications.

With real historical data, the drifter observations can be updated with high frequency, i.e. hourly updates for GDP and 15min-updates for GLAD. Meanwhile, as mentioned in chapter 3, the latest version of online drifter module for MOM6 is capable of providing the intermediate drifter trajectory at the end of each forecasting time step. In future research, the LETKF-LaDA can be upgraded to the 4D-LETKF (Harlim and Hunt, 2007), where the minimization of the cost function is based on a observed trajectory within the whole DA window instead of a terminal position,

$$J(\mathbf{w}) = \frac{1}{2}(k-1)\mathbf{w}^T\mathbf{w} + \frac{1}{2}\sum_{l=1}^{n}[\mathbf{y}_l^o - H_l(\overline{\mathbf{x}}_l^b) - \mathbf{Y}_l^b\mathbf{w}]^T\mathbf{R}_l^{-1}[\mathbf{y}_l^o - H_l(\overline{\mathbf{x}}_l^b) - \mathbf{Y}_l^b\mathbf{w}].$$

Moreover, in this thesis, we only focus on assimilating surface drifters, which we have already concluded have a significant impact on improving the surface currents and have a positive influence in estimating the deep ocean. In the next step of

the future research, we will extend the application of the proposed LETKF-LaDA to deep ocean by assimilating the Argo floats positions at 1000m drifting depth and at the surface during transmission phase, of which the temperature and salinity profiles are mainly used as observations in current practice (Scott et al., 2018; Verrier et al., 2017). Advanced studies by Nilsson et al. (2012, 2011) implemented an assimilation of the Argo trajectories to improve the Mediterranean Forecast System by converting the trajectory observations to Eulerian velocity, while no attempt has ever tried to assimilate the direct drifter positions. Figure 5.1 shows the dense distribution of Argo floats within the global ocean. Equipped with the latest version of drifter module of MOM6, it is possible to escalate the current surface drifter simulator to one emulating the Argo float in the deep ocean, providing the forecast states for Argo trajectories. By assimilating the Argo location in addition to the temperature and salinity profiles, we are expecting a better estimation of ocean circulation in the deep ocean.

In Chapter 4, we use the knowledge of synchronization and the Lyapunov spectrum to show that SCDA is more robust than UCDA and WCDA. Using the example of MAOOAM, we conclude that it is necessary to use the coupled model and strong coupling in the DA, taking account of the bi-directional fluxes and interactions in the model and covariances of states from different media. In addition, the motion of the surface drifters locating at the interface layer between the atmosphere and ocean is a consequence of the intertwined reaction from both of the aforementioned media, and therefore the trajectory observations provide valuable information on the coupled dynamics. By combining the discussion in LaDA, we are planning to

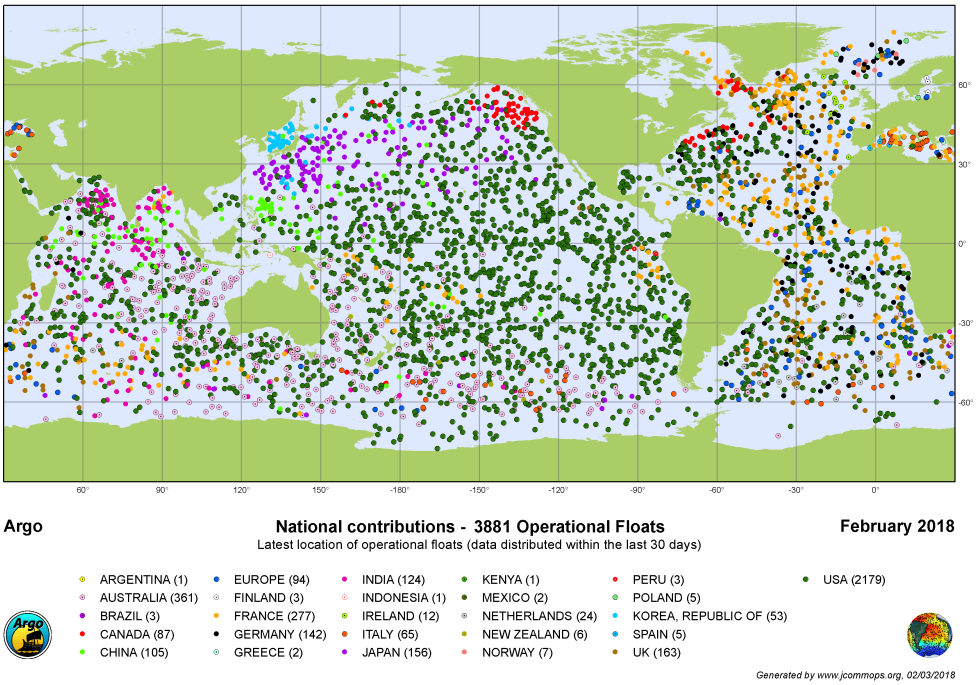


Figure 5.1: The distribution of active floats in the Argo array, colour coded by country that owns the float, as of February 2018, generated by `www.jcommops.org`, 02/03/2018.

utilize LETKF-LaDA within the coupled scenario in order to improve the estimation of the forecast in not only the ocean states, but also the atmosphere states as well.

# Appendix A: Triangular Transformation Theorem for QR Algorithm

In this chapter, we shed light on the details of deriving the QR algorithm based on the triangular transformation. The discussion below is extended by the demonstration shown in Dieci and Vleck (2007). Consider the TLM as a regular N-dimension system,

$$\frac{d}{dt}\delta\mathbf{x}(t) = \mathcal{J}(t)\delta\mathbf{x}(t), \tag{A.1}$$

where $\mathcal{J} : \mathbb{R}^+ \to \mathbb{R}^{N\times N}$ is continuous and bounded uniformly in $t$. We denote $\mathbf{\Psi}(t)$ as the fundamental matrix of the dynamical system (A.1), then it has the property,

$$\frac{d}{dt}\mathbf{\Psi}(t) = \mathcal{J}(t)\mathbf{\Psi}(t) \tag{A.2}$$

Kuptsov and Parlitz (2012) explains that the relation between the fundamental matrix $\mathbf{\Psi}(t)$ and the tangent linear propagator $\mathcal{F}(t, t_0)$ is,

$$\mathcal{F}(t, t_0) = \mathbf{\Psi}(t)\mathbf{\Psi}^{-1}(t_0) \quad \text{for } t > t_0.$$

Then we can also obtain the same property that,

$$\frac{d}{dt}\mathcal{F}(t, t_0) = \mathcal{J}(t)\mathcal{F}(t, t_0). \tag{A.3}$$

Next, we are going to utilize the above properties showing the following theorem.

**Theorem.** *Define the QR factorization of the tangent linear propagator as,* $\mathcal{F}(t, t_0) = \tilde{\boldsymbol{Q}}(t)\tilde{\boldsymbol{R}}(t)$*, where* $\tilde{\boldsymbol{Q}}(t)$ *is an orthonormal matrix and* $\tilde{\boldsymbol{R}}(t)$ *is an upper triangular*

*matrix with all diagonal components positive for all $t > t_0$. If we define $\boldsymbol{y}(t) = \tilde{\boldsymbol{Q}}^T(t)\delta\boldsymbol{x}(t)$, then $\boldsymbol{y}(t)$ satisfies an upper triangular dynamical system,*

$$\frac{d}{dt}\mathbf{y}(t) = \mathbf{B}(t)\mathbf{y}(t),$$

*where $\boldsymbol{B}(t)$ is an upper triangular matrix for all $t > t_0$. In addition,*

$$\ln(\tilde{\mathbf{R}}_{jj}(t)) = \int_{t_0}^{t} \mathbf{B}_{jj}(s)ds, \quad \text{for } j = 1, 2, ..., N.$$

**Proof.** Based on the relation $\mathbf{y}(t) = \tilde{\mathbf{Q}}^T(t)\delta\mathbf{x}(t)$, we consider the derivative of $\mathbf{y}(t)$,

$$\begin{aligned}
\frac{d}{dt}\mathbf{y}(t) &= \frac{d}{dt}(\tilde{\mathbf{Q}}^T(t))\delta\mathbf{x}(t) + \tilde{\mathbf{Q}}^T(t)\frac{d}{dt}\delta\mathbf{x}(t) \\
&= \frac{d}{dt}(\tilde{\mathbf{Q}}^T(t))\delta\mathbf{x}(t) + \tilde{\mathbf{Q}}^T(t)\mathcal{J}(t)\delta\mathbf{x}(t) \\
&= \frac{d}{dt}(\tilde{\mathbf{Q}}^T(t))\tilde{\mathbf{Q}}(t)\mathbf{y}(t) + \tilde{\mathbf{Q}}^T(t)\mathcal{J}(t)\tilde{\mathbf{Q}}(t)\mathbf{y}(t) \\
&= \Big[\frac{d}{dt}(\tilde{\mathbf{Q}}^T(t))\tilde{\mathbf{Q}}(t) + \tilde{\mathbf{Q}}^T(t)\mathcal{J}(t)\tilde{\mathbf{Q}}(t)\Big]\mathbf{y}(t)
\end{aligned}$$

Define $\mathbf{B}(t)$ as,

$$\mathbf{B}(t) := \frac{d}{dt}(\tilde{\mathbf{Q}}^T(t))\tilde{\mathbf{Q}}(t) + \tilde{\mathbf{Q}}^T(t)\mathcal{J}(t)\tilde{\mathbf{Q}}(t). \tag{A.4}$$

Then we are going to show that $\mathbf{B}(t)$ is in fact an upper triangular matrix. Notice that equation (A.4) can be also written as,

$$\frac{d}{dt}(\tilde{\mathbf{Q}}^T(t)) = \mathbf{B}(t)\tilde{\mathbf{Q}}^T(t) - \tilde{\mathbf{Q}}^T(t)\mathcal{J}(t).$$

Using the identity $\tilde{\mathbf{Q}}^T(t)\tilde{\mathbf{Q}}(t) \equiv \mathbf{I}$ for all $t > t_0$, we then can derive the derivative of the matrix $\tilde{\mathbf{Q}}(t)$ as,

$$\frac{d}{dt}\tilde{\mathbf{Q}}(t) = \mathcal{J}(t)\tilde{\mathbf{Q}}(t) - \tilde{\mathbf{Q}}(t)\mathbf{B}(t). \tag{A.5}$$

We then take the derivative of the tangent linear propagator $\mathcal{F}(t, t_0) = \tilde{\mathbf{Q}}(t)\tilde{\mathbf{R}}(t)$

and use (A.5),

$$\begin{aligned}\frac{d}{dt}\mathcal{F}(t,t_0) &= \frac{d}{dt}(\tilde{\mathbf{Q}}(t)\tilde{\mathbf{R}}(t)) \\ &= \Big(\frac{d}{dt}\tilde{\mathbf{Q}}(t)\Big)\tilde{\mathbf{R}}(t) + \tilde{\mathbf{Q}}(t)\Big(\frac{d}{dt}\tilde{\mathbf{R}}(t)\Big) \\ &= \mathcal{J}(t)\tilde{\mathbf{Q}}(t)\tilde{\mathbf{R}}(t) - \tilde{\mathbf{Q}}(t)\mathbf{B}(t)\tilde{\mathbf{R}}(t) + \tilde{\mathbf{Q}}(t)\Big(\frac{d}{dt}\tilde{\mathbf{R}}(t)\Big)\end{aligned} \tag{A.6}$$

Alternative, based on property (A.3), we can also obtain

$$\frac{d}{dt}\mathcal{F}(t,t_0) = \mathcal{J}(t)\mathcal{F}(t,t_0) = \mathcal{J}(t)\tilde{\mathbf{Q}}(t)\tilde{\mathbf{R}}(t) \tag{A.7}$$

Combining (A.6) and (A.7),

$$\frac{d}{dt}\tilde{\mathbf{R}}(t) = \mathbf{B}(t)\tilde{\mathbf{R}}(t) \tag{A.8}$$

Since $\tilde{\mathbf{R}}(t)$ and the derivative of $\tilde{\mathbf{R}}(t)$ are both upper triangular matrices, then we can conclude that $\mathbf{B}(t)$ is an upper triangular matrix. In addition, scrutinizing the diagonal components in (A.8), we also have the following system of the scalar differential equations:

$$\frac{d}{dt}\tilde{\mathbf{R}}_{jj}(t) = \mathbf{B}_{jj}(t)\tilde{\mathbf{R}}_{jj}(t), \quad \text{for } j = 1, 2, \cdots, N.$$

Therefore,

$$\tilde{\mathbf{R}}_{jj}(t) = \exp\Big[\int_{t_0}^{t}\mathbf{B}_{jj}(s)ds\Big]\tilde{\mathbf{R}}_{jj}(0) = \exp\Big[\int_{t_0}^{t}\mathbf{B}_{jj}(s)ds\Big],$$

or,

$$\ln\left(\tilde{\mathbf{R}}_{jj}(t)\right) = \int_{t_0}^{t}\mathbf{B}_{jj}(s)ds.$$

Notice that $\tilde{\mathbf{R}}(t)$ is determined by QR factorizing $\mathcal{F}(t_0,t_0)$, which is an identity matrix $\mathbf{I}$. Thus, $\tilde{\mathbf{R}}(t_0) = \mathbf{I}$, i.e. $\mathbf{R}_{jj}(0) = 1$ $(j = 1, 2, ..., N)$. (Q.E.D.)

# Bibliography


Apel, J.R. (1987), *Principles of Ocean Physics*. International Geophysics, Elsevier Science, URL https://books.google.com/books?id=iuk6nZIzvtkC.

Apte, A., C. K. R. T. Jones, and A. M. Stuart (2008), "A Bayesian approach to Lagrangian data assimilation." *Tellus A*, 60, 336–347, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0870.2007.00295.x.

Ballabrera-Poy, Joaquim, Eugenia Kalnay, and Shu-Chih Yang (2009), "Data assimilation in a system with two scales – combining two initialization techniques." *Tellus A*, 61, 539–549, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0870.2009.00400.x.

Bishop, Craig H., Brian J. Etherton, and Sharanya J. Majumdar (2001), "Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects." *Monthly Weather Review*, 129, 420–436, URL https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Bitterman, David S. and Donald V. Hansen (1993), "Evaluation of Sea Surface Temperature Measurements from Drifting Buoys." *Journal of Atmospheric and Oceanic Technology*, 10, 88–96, URL https://journals.ametsoc.org/doi/abs/10.1175/1520-0426%281993%29010%3C0088%3AEOSSTM%3E2.0.CO%3B2.

Bonjean, Fabrice and Gary S. E. Lagerloef (2002), "Diagnostic Model and Analysis of the Surface Currents in the Tropical Pacific Ocean." *Journal of Physical Oceanography*, 32, 2938–2954.

Carrassi, Alberto, Michael Ghil, Anna Trevisan, and Francesco Uboldi (2008), "Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system." *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 18, 023112, URL https://doi.org/10.1063/1.2909862.

Carrier, Matthew J., Hans Ngodock, Scott Smith, Gregg Jacobs, Philip Muscarella, Tamay Ozgokmen, Brian Haus, and Bruce Lipphardt (2014), "Impact of Assimilating Ocean Velocity Observations Inferred from Lagrangian Drifter Data Using the NCOM-4DVAR." *Monthly Weather Review*, 142, 1509–1524, URL https://doi.org/10.1175/MWR-D-13-00236.1.

Carrier, Matthew J., Hans E. Ngodock, Philip Muscarella, and Scott Smith (2016), "Impact of Assimilating Surface Velocity Observations on the Model Sea Surface Height Using the NCOM-4DVAR." *Monthly Weather Review*, 144, 1051–1068, URL https://doi.org/10.1175/MWR-D-14-00285.1.

Chelton, Dudley B., Roland A. deSzoeke, Michael G. Schlax, Karim El Naggar, and Nicolas Siwertz (1998), "Geographical Variability of the First Baroclinic Rossby Radius of Deformation." *Journal of Physical Oceanography*, 28, 433–460, URL https://doi.org/10.1175/1520-0485(1998)028<0433:GVOTFB>2.0.CO;2.

Coelho, Emanuel F., P. Hogan, G. Jacobs, P. Thoppil, H.S. Huntley, B.K. Haus, B.L. Lipphardt, A.D. Kirwan, E.H. Ryan, J. Olascoaga, F. Beron-Vera, A.C. Poje, A. Griffa, T.M. zgkmen, A.J. Mariano, G. Novelli, A.C. Haza, D. Bogucki, S.S. Chen, M. Curcic, M. Iskandarani, F. Judt, N. Laxague, A.J.H.M. Reniers, A. Valle-Levinson, and M. Wei (2015), "Ocean current estimation using a Multi-Model Ensemble Kalman Filter during the Grand Lagrangian Deployment experiment (GLAD)." *Ocean Modelling*, 87, 86 – 106, URL http://www.sciencedirect.com/science/article/pii/S1463500314001577.

De Cruz, L., J. Demaeyer, and S. Vannitsem (2016), "The Modular Arbitrary-Order Ocean-Atmosphere Model: MAOOAM v1.0." *Geoscientific Model Development*, 9, 2793–2808, URL https://www.geosci-model-dev.net/9/2793/2016/.

Delandmeter, P. and E. van Sebille (2019), "The Parcels v2.0 Lagrangian framework: new field interpolation schemes." *Geoscientific Model Development Discussions*, 2019, 1–24, URL https://www.geosci-model-dev-discuss.net/gmd-2018-339/.

Dieci, Luca, Michael Jolly, and Erik S. Van Vleck (2011), "Numerical Techniques for Approximating Lyapunov Exponents and Their Implementation." *Journal of Computational and Nonlinear Dynamics*, 6.

Dieci, Luca and Erik S. Van Vleck (2007), "Lyapunov and Sacker-Sell Spectral Intervals." *Journal of Dynamics and Differential Equations*, 19, 265–293.

Dong, Shenfu, Denis Volkov, Gustavo Goni, Rick Lumpkin, and Gregory R. Foltz (2017), "Near-surface salinity and temperature structure observed with dual-sensor drifters in the subtropical South Pacific." *Journal of Geophysical Research: Oceans*, 122, 5952–5969, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JC012894.

Evensen, Geir (2004), "Sampling strategies and square root analysis schemes for the EnKF." *Ocean Dynamics*, 54, 539–560, URL https://doi.org/10.1007/s10236-004-0099-2.

Fox-Kemper, B., G. Danabasoglu, R. Ferrari, S.M. Griffies, R.W. Hallberg, M.M. Holland, M.E. Maltrud, S. Peacock, and B.L. Samuels (2011), "Parameterization of mixed layer eddies. III: Implementation and impact in global ocean climate

simulations." *Ocean Modelling*, 39, 61 – 78, URL http://www.sciencedirect.com/science/article/pii/S1463500310001290. Modelling and Understanding the Ocean Mesoscale and Submesoscale.

Fox-Kemper, Baylor, Raffaele Ferrari, and Robert Hallberg (2008), "Parameterization of Mixed Layer Eddies. Part I: Theory and Diagnosis." *Journal of Physical Oceanography*, 38, 1145–1165, URL https://doi.org/10.1175/2007JPO3792.1.

Greybush, Steven J., Eugenia Kalnay, Takemasa Miyoshi, Kayo Ide, and Brian R. Hunt (2011), "Balance and Ensemble Kalman Filter Localization Techniques." *Monthly Weather Review*, 139, 511–522, URL https://doi.org/10.1175/2010MWR3328.1.

Griffies, Stephen M. (2008), "Elements of MOM4p1." URL https://data1.gfdl.noaa.gov/~arl/pubrel/o/old/doc/mom4p1_guide.pdf. Accessed 8 Nov 2018.

Griffies, Stephen M., Ronald C. Pacanowski, and Robert W. Hallberg (2000), "Spurious Diapycnal Mixing Associated with Advection in a z-Coordinate Ocean Model." *Monthly Weather Review*, 128, 538–564, URL https://doi.org/10.1175/1520-0493(2000)128<0538:SDMAWA>2.0.CO;2.

Griffies, Stephen M., Ronald J. Stouffer, Alistair J. Adcroft, Kirk Bryan, Keith W. Dixon, Robert Hallberg, Matthew J. Harrison, Ronald C. Pacanowski, and Anthony Rosati (2015), "A Historical Introduction to MOM." URL https://www.gfdl.noaa.gov/wp-content/uploads/2019/04/mom_history_2017.09.19.pdf. [Online; posted October 2015].

Hallberg, Robert (2013), "Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects." *Ocean Modelling*, 72, 92 – 103, URL http://www.sciencedirect.com/science/article/pii/S1463500313001601.

Halliwell, G. R., A. Srinivasan, V. Kourafalou, H. Yang, D. Willey, M. Le Hnaff, and R. Atlas (2014), "Rigorous Evaluation of a Fraternal Twin Ocean OSSE System for the Open Gulf of Mexico." *Journal of Atmospheric and Oceanic Technology*, 31, 105–130, URL https://doi.org/10.1175/JTECH-D-13-00011.1.

Han, Guijun, Xinrong Wu, Shaoqing Zhang, Zhengyu Liu, and Wei Li (2013), "Error Covariance Estimation for Coupled Data Assimilation Using a Lorenz Atmosphere and a Simple Pycnocline Ocean Model." *Journal of Climate*, 26, 10218–10231, URL https://doi.org/10.1175/JCLI-D-13-00236.1.

Harlim, John and Brian R. Hunt (2007), "Four-dimensional local ensemble transform Kalman filter: numerical experiments with a global circulation model." *Tellus A: Dynamic Meteorology and Oceanography*, 59, 731–748, URL https://doi.org/10.1111/j.1600-0870.2007.00255.x.

Houtekamer, P. L. and Herschel L. Mitchell (1998), "Data Assimilation Using an Ensemble Kalman Filter Technique." *Monthly Weather Review*, 126, 796–811, URL https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

Hunt, Brian R., Eric Kostelich, and Istvan Szunyogh (2007), "Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter." *Physica D: Nonlinear Phenomena*, 230, 112–126.

Ide, Kayo, Leonid Kuznetsov, and Christopher K R T Jone (2002), "Lagrangian data assimilation for point vortex systems." *Journal of Turbulence*, 3, N53, URL https://doi.org/10.1088/1468-5248/3/1/053.

Ilcak, Mehmet, Alistair J. Adcroft, Stephen M. Griffies, and Robert W. Hallberg (2012), "Spurious dianeutral mixing and the role of momentum closure." *Ocean Modelling*, 45-46, 37 – 58, URL http://www.sciencedirect.com/science/article/pii/S1463500311001685.

Jacobs, Gregg A., Brent P. Bartels, Darek J. Bogucki, Francisco J Beron-Vera, Shuyi S Chen, Emanuel F. Coelho, Milan Curcic, Annalisa Griffa, Matthew Gough, Brian K Haus, Angelique C. Haza, Robert W. Helber, Patrick J. Hogan, Helga S. Huntley, Mohamed Iskandarani, Falko Judt, A. D. Kirwan, Nathan Laxague, Arnoldo Valle-Levinson, Bruce L. Lipphardt, Arthur J Mariano, Hans E. Ngodock, Guillaume Novelli, Maria J Olascoaga, Tamay M Ozgokmen, Andrew C. Poje, Ad J H M Reniers, Clark D. Rowley, Edward H. Ryan, Scott R. Smith, Peter L. Spence, Prasad G. Thoppil, and Mozheng Wei (2014), "Data assimilation considerations for improved ocean predictability during the Gulf of Mexico Grand Lagrangian Deployment (GLAD)." *Ocean Modelling*, 83, 98–117.

Jansen, Malte F., Alistair J. Adcroft, Robert Hallberg, and Isaac M. Held (2015), "Parameterization of eddy fluxes based on a mesoscale energy budget." *Ocean Modelling*, 92, 28 – 41, URL http://www.sciencedirect.com/science/article/pii/S1463500315000967.

Kang, Ji-Sun, Eugenia Kalnay, Junjie Liu, Inez Fung, Takemasa Miyoshi, and Kayo Ide (2011), "Variable localization in an ensemble Kalman filter: Application to the carbon cycle data assimilation." *Journal of Geophysical Research: Atmospheres*, 116, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JD014673.

Kuptsov, Pavel V. and Ulrich Parlitz (2012), "Theory and Computation of Covariant Lyapunov Vectors." *Journal of Nonlinear Science*, 22, 727–762.

Kuznetsov, L., K. Ide, and C. K. R. T. Jones (2003), "A Method for Assimilation of Lagrangian Data." *Monthly Weather Review*, 131, 2247–2260, URL https://doi.org/10.1175/1520-0493(2003)131<2247:AMFAOL>2.0.CO;2.

Lange, M. and E. van Sebille (2017), "Parcels v0.9: prototyping a Lagrangian ocean analysis framework for the petascale age." *Geoscientific Model Development*, 10, 4175–4186, URL https://www.geosci-model-dev.net/10/4175/2017/.

Liapounoff, A. (1907), "General problem of the stability of the movement." *Annals of the Faculty of Sciences of Toulouse: Mathematics*, 2nd, 9, 203–474, URL http://www.numdam.org/item/AFST_1907_2_9__203_0.

Lima, L. N., L. P. Pezzi, S. G. Penny, and C. A. S. Tanajura (2019), "An Investigation of Ocean Model Uncertainties Through Ensemble Forecast Experiments in the Southwest Atlantic Ocean." *Journal of Geophysical Research: Oceans*, 124, 432–452, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JC013919.

Lumpkin, Rick, Mayra Pazos, National Oceanographic, and Atmospheric Administration (2007), "Measuring surface currents with Surface Velocity Program drifters: the instrument, its data, and some recent results. Chapter two of Lagrangian Analysis." In *and Prediction of Coastal and Ocean Dynamics*, University Press.

Marshall, David P. and Alistair J. Adcroft (2010), "Parameterization of ocean eddies: Potential vorticity mixing, energetics and Arnolds first stability theorem." *Ocean Modelling*, 32, 188 – 204, URL http://www.sciencedirect.com/science/article/pii/S1463500310000107. The magic of modelling: A special volume commemorating the contributions of Peter D. Killworth Part 2.

Melnichenko, Oleg, Peter Hacker, Nikolai Maximenko, Gary Lagerloef, and James Potemra (2016), "Optimum interpolation analysis of Aquarius sea surface salinity." *Journal of Geophysical Research: Oceans*, 121, 602–616, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JC011343.

Mogensen, Kristian S., Tim Hewson, Sarah Keeley, and Linus Magnusson (2018), "ECMWF Newsletter." URL https://www.ecmwf.int/en/newsletter/156/news/effects-ocean-coupling-weather-forecasts. Last accesed 8 July 2019.

Molcard, Anne, Leonid I. Piterbarg, Annalisa Griffa, Tamay M. zgkmen, and Arthur J. Mariano (2003), "Assimilation of drifter observations for the reconstruction of the Eulerian circulation field." *Journal of Geophysical Research: Oceans*, 108, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001JC001240.

Nilsson, J. A. U., S. Dobricic, N. Pinardi, P.-M. Poulain, and D. Pettenuzzo (2012), "Variational assimilation of Lagrangian trajectories in the Mediterranean ocean Forecasting System." *Ocean Science*, 8, 249–259, URL https://www.ocean-sci.net/8/249/2012/.

Nilsson, Jenny A. U., Srdjan Dobricic, Nadia Pinardi, Vincent Taillandier, and Pierre-Marie Poulain (2011), "On the assessment of Argo float trajectory assimilation in the Mediterranean Forecasting System." *Ocean Dynamics*, 61, 1475–1490, URL https://doi.org/10.1007/s10236-011-0437-0.

Olascoaga, M. J., F. J. Beron-Vera, G. Haller, J. Trianes, M. Iskandarani, E. F. Coelho, B. K. Haus, H. S. Huntley, G. Jacobs, A. D. Kirwan Jr., B. L. Lipphardt Jr., T. M. zgkmen, A. J. H. M. Reniers, and A. Valle-Levinson (2013), "Drifter motion in the Gulf of Mexico constrained by altimetric Lagrangian coherent structures." *Geophysical Research Letters*, 40, 6171–6175, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013GL058624.

Ott, Edward, Brian R. Hunt, Istvan Szunyogh, Aleksey V. Zimin, Eric J. Kostelich, Matteo Corazza, Eugenia Kalnay, D. J. Patil, and James A. Yorke (2004), "A Local Ensemble Kalman Filter for Atmospheric Data Assimilation." *Tellus A*, 56, 415–428, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0870.2004.00076.x.

Pecora, Louis M. and Thomas L. Carroll (1991), "Driving systems with chaotic signals." *Phys. Rev. A*, 44, 2374–2383, URL https://link.aps.org/doi/10.1103/PhysRevA.44.2374.

Pecora, Louis M. and Thomas L. Carroll (2015), "Synchronization of chaotic systems." *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25, 097611, URL https://doi.org/10.1063/1.4917383.

Penny, S. G., E. Bach, K. Bhargava, C.-C. Chang, C. Da, L. Sun, and T. Yoshida (2019), "Strongly Coupled Data Assimilation in Multiscale Media: Experiments Using a Quasi-Geostrophic Coupled Model." *Journal of Advances in Modeling Earth Systems*, 0, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001652.

Penny, S. G. and T. Miyoshi (2016), "A local particle filter for high-dimensional geophysical systems." *Nonlinear Processes in Geophysics*, 23, 391–405, URL https://www.nonlin-processes-geophys.net/23/391/2016/.

Penny, Stephen G. (2017), "Mathematical foundations of hybrid data assimilation from a synchronization perspective." *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27, 126801, URL https://doi.org/10.1063/1.5001819.

Penny, Stephen G., David W. Behringer, James A. Carton, and Eugenia Kalnay (2015), "A Hybrid Global Ocean Data Assimilation System at NCEP." *Monthly Weather Review*, 143, 4660–4677, URL https://doi.org/10.1175/MWR-D-14-00376.1.

Penny, Stephen G. and coauthors (2008), "Coupled Data Assimilation for Integrated Earth System Analysis and Prediction: Goals, Challenges and Recommendations." URL https://www.wmo.int/pages/prog/arep/wwrp/new/documents/Final_WWRP_2017_3_27_July.pdf. Last ccessed 3 April 2018.

Penny, Stephen G. and Thomas M. Hamill (2017), "Coupled Data Assimilation for Integrated Earth System Analysis and Prediction." *Bulletin of the American Meteorological Society*, 98, ES169–ES172, URL https://doi.org/10.1175/BAMS-D-17-0036.1.

Poje, Andrew C., Tamay M. Özgökmen, Bruce L. Lipphardt, Brian K. Haus, Edward H. Ryan, Angelique C. Haza, Gregg A. Jacobs, A. J. H. M. Reniers, Maria Josefina Olascoaga, Guillaume Novelli, Annalisa Griffa, Francisco J. Beron-Vera, Shuyi S. Chen, Emanuel Coelho, Patrick J. Hogan, Albert D. Kirwan, Helga S. Huntley, and Arthur J. Mariano (2014), "Submesoscale dispersion in the vicinity of the Deepwater Horizon spill." *Proceedings of the National Academy of Sciences*, 111, 12693–12698, URL https://www.pnas.org/content/111/35/12693.

Poterjoy, Jonathan (2016), "A Localized Particle Filter for High-Dimensional Nonlinear Systems." *Monthly Weather Review*, 144, 59–76, URL https://doi.org/10.1175/MWR-D-15-0163.1.

Romeiser, R., S. Suchandt, H. Runge, and U. Steinbrecher (2009), "High-resolution current measurements from space with TerraSAR-X along-track InSAR." In *OCEANS 2009-EUROPE*, 1–5.

Saha, S., S. Nadiga, C. Thiaw, J. Wang, W. Wang, Q. Zhang, H. M. Van den Dool, H.-L. Pan, S. Moorthi, D. Behringer, D. Stokes, M. Pea, S. Lord, G. White, W. Ebisuzaki, P. Peng, and P. Xie (2006), "The NCEP Climate Forecast System." *Journal of Climate*, 19, 3483–3517, URL https://doi.org/10.1175/JCLI3812.1.

Salman, H., K. Ide, and C. K. R. T. Jones (2008), "Using flow geometry for drifter deployment in Lagrangian data assimilation." *Tellus A*, 60, 321–335, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0870.2007.00292.x.

Salman, H., L. Kuznetsov, C. K. R. T. Jones, and K. Ide (2006), "A Method for Assimilating Lagrangian Data into a Shallow-Water-Equation Ocean Model." *Monthly Weather Review*, 134, 1081–1101, URL https://doi.org/10.1175/MWR3104.1.

Santitissadeekorn, Naratip, Elaine T. Spiller, Christopher Jones, Regis Rutarindwa, Liyan Liu, and Kayo Ide (2014), "Observing System Simulation Experiments of cross-layer Lagrangian data assimilation." *Dynamics of Atmospheres and Oceans*, 66, 77 – 93, URL http://www.sciencedirect.com/science/article/pii/S0377026514000141.

Scott, K. Andrea, Changheng Chen, and Paul G. Myers (2018), "Assimilation of Argo Temperature and Salinity Profiles Using a Bias-Aware EnOI Scheme for the Labrador Sea." *Journal of Atmospheric and Oceanic Technology*, 35, 1819–1834, URL https://doi.org/10.1175/JTECH-D-17-0222.1.

Slivinski, Laura, Elaine Spiller, Amit Apte, and Björn Sandstede (2015), "A Hybrid Particle-Ensemble Kalman Filter for Lagrangian Data Assimilation." *Monthly Weather Review*, 143, 195–211, URL https://doi.org/10.1175/MWR-D-14-00051.1.

Slivinski, Laura C., Gilbert P. Compo, and Coauthors (2019a), "Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system." To appear.

Slivinski, Laura C., Gilbert P. Compo, and Coauthors (2019b), "Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system: Appendices." To appear.

Sluka, Travis C., Stephen G. Penny, Eugenia Kalnay, and Takemasa Miyoshi (2016), "Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation." *Geophysical Research Letters*, 43, 752–759, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015GL067238.

Smith, Polly J., Alison M. Fowler, and Amos S. Lawless (2015), "Exploring strategies for coupled 4D-Var data assimilation using an idealised atmosphereocean model." *Tellus A: Dynamic Meteorology and Oceanography*, 67, 27025, URL https://doi.org/10.3402/tellusa.v67.27025.

Snyder, Chris, Thomas Bengtsson, Peter Bickel, and Jeff Anderson (2008), "Obstacles to High-Dimensional Particle Filtering." *Monthly Weather Review*, 136, 4629–4640, URL https://doi.org/10.1175/2008MWR2529.1.

Taillandier, Vincent, Annalisa Griffa, and Anne Molcard (2006), "A variational approach for the reconstruction of regional scale Eulerian velocity fields from Lagrangian data." *Ocean Modelling*, 13, 1 – 24, URL http://www.sciencedirect.com/science/article/pii/S1463500305000776.

van Leeuwen, P. J. (2003), "A Variance-Minimizing Filter for Large-Scale Applications." *Monthly Weather Review*, 131, 2071–2084, URL https://doi.org/10.1175/1520-0493(2003)131<2071:AVFFLA>2.0.CO;2.

Vannitsem, Stéphane and Valerio Lucarini (2016), "Statistical and dynamical properties of covariant lyapunov vectors in a coupled atmosphere-ocean model—multiscale effects, geometric degeneracy, and error dynamics." *Journal of Physics A: Mathematical and Theoretical*, 49, 224001, URL https://doi.org/10.1088%2F1751-8113%2F49%2F22%2F224001.

Vernieres, Guillaume, Christopher K.R.T. Jones, and Kayo Ide (2011), "Capturing eddy shedding in the Gulf of Mexico from Lagrangian observations." *Physica D: Nonlinear Phenomena*, 240, 166 – 179, URL http://www.sciencedirect.com/science/article/pii/S0167278910001855. Nonlinear Excursions Symposium and Volume in Physica D to honor Louis N. Howards scientific career.

Verrier, S., P.-Y. Le Traon, and E. Remy (2017), "Assessing the impact of multiple altimeter missions and Argo in a global eddy-permitting data assimilation system." *Ocean Science*, 13, 1077–1092, URL https://www.ocean-sci.net/13/1077/2017/.

Zhang, S., M. J. Harrison, A. Rosati, and A. Wittenberg (2007), "System Design and Evaluation of Coupled Ensemble Data Assimilation for Global Oceanic Climate Studies." *Monthly Weather Review*, 135, 3541–3564, URL https://doi.org/10.1175/MWR3466.1.